Fitting infrared ice spectra with genetic modelling algorithms

Presenting the ENIIGMA fitting tool

W. R. M. Rocha[✉]*, G. Perotti[✉], L. E. Kristensen[✉], and J. K. Jørgensen

Niels Bohr Institute & Centre for Star and Planet Formation, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen K., Denmark
e-mail: rocha@strw.leidenuniv.nl

Received 7 September 2020 / Accepted 15 July 2021

ABSTRACT

Context. A variety of laboratory ice spectra simulating different chemical environments, ice morphologies, and thermal and energetic processing are needed in order to provide an accurate interpretation of the infrared spectra of protostars. To decipher the combination of laboratory data that best fits the observations, an automated, statistics-based computational approach is necessary.

Aims. We aim to introduce a new approach, based on evolutionary algorithms, to searching for molecules in ice mantles via spectral decomposition of infrared observational data with laboratory ice spectra.

Methods. We introduce a publicly available and open-source fitting tool called ENIIGMA (dEcompositionN of Infrared Ice features using Genetic Modelling Algorithms). The tool has dedicated Python functions to carry out continuum determination of the protostellar spectra, silicate extraction, spectral decomposition, and statistical analysis to calculate confidence intervals and quantify degeneracy. We conducted fully blind and non-blind tests with known ice samples and constructed mixtures in order to assess the code. Additionally, we performed a complete analysis of the Elias 29 spectrum and compared our findings with previous results from the literature.

Results. The ENIIGMA fitting tool can identify the correct ice samples and their fractions in all checks with known samples tested in this paper. In the cases where Gaussian noise was added to the experimental data, more robust genetic operators and more iterations became necessary. Concerning the Elias 29 spectrum, the broad spectral range between 2.5 and 20 μm was successfully decomposed after continuum determination and silicate extraction. This analysis allowed the identification of different molecules in the ice mantle, including a tentative detection of $\text{CH}_3\text{CH}_2\text{OH}$.

Conclusions. The ENIIGMA is a toolbox for spectroscopy analysis of infrared spectra that is well-timed with the launch of the *James Webb Space Telescope*. Additionally, it allows different chemical environments and irradiation fields to be explored, allowing the user to correctly interpret astronomical observations.

Key words. ISM: molecules – solid state: volatile – infrared: ISM – stars: protostars – astrochemistry

1. Introduction

Characterising the composition of astrophysical ice mantles in the interstellar medium (ISM) is crucial for understanding the chemical evolution of young stellar objects (YSOs). With this goal, systematic studies in the infrared (IR) with ground- and space-based telescopes (e.g. Chiar et al. 1995; Schutte et al. 1996; Gibb et al. 2000; Pontoppidan et al. 2003; Boogert et al. 2008; Zasowski et al. 2009; Thi et al. 2011; Öberg et al. 2011) aided by laboratory experiments simulating diverse chemical environments and irradiation fields (e.g. Gerakines et al. 1995; Schutte et al. 1993; Palumbo et al. 1998; Watanabe & Kouchi 2002; Fuchs et al. 2009; Pilling et al. 2010; Linnartz et al. 2015; Terwisscha van Scheltinga et al. 2018; Rachid et al. 2020; Ioppolo et al. 2021) have advanced our knowledge of the composition of interstellar ices.

The Infrared Space Observatory (ISO) allowed the first systematic studies of ices in star-forming regions (e.g. Schutte et al. 1996; van Dishoeck et al. 1998). These observations revealed the presence of H_2O , NH_3 , and CH_4 , and allowed the synthesis of new chemical species to be inferred, such as CH_3OH and

NH_4^+ , which are both byproducts of heating and energetic processing by photons and cosmic rays (Gibb et al. 2000, 2004). Tentative assignments of CH_3CHO (acetaldehyde) and HCOOH (formic acid) relying on the absorption features at 7.24 μm and 7.41 μm were made by Schutte et al. (1999). However, the strong bands at 5.80 μm associated with these two molecules have not yet been unequivocally detected, and therefore the presence of these molecules is still a matter of debate.

The *Spitzer Space Telescope*, which has improved sensitivity compared to ISO, allowed ice surveys toward low-mass star-forming regions, background stars, and high-UV environments (e.g. Bergin et al. 2005; Whittet et al. 2009; Reach et al. 2009; Boogert et al. 2013). Systematic analyses of the 5–8 μm absorption complex, the CH_4 band at 7.7 μm , the CO_2 bending mode at 15.2 μm , and NH_3 and CH_3OH absorption features at 9.01 μm and 9.74 μm were performed by Boogert et al. (2008), Öberg et al. (2008), Pontoppidan et al. (2008), and Bottinelli et al. (2010), respectively. All of the above authors used a phenomenological approach and/or analytical function decomposition to estimate environment-dependent variations in order to avoid degenerated solutions in characterising the ice morphology. A statistical analysis of the ice column density and abundance with respect to H_2O in these four studies was done by Öberg et al. (2011), who concluded that, with the exception

* Current address: Laboratory for Astrophysics, Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands.

of CO, CO₂, and CH₄ (the most volatile species), there is no significant statistical abundance variation between low- and high-mass YSOs in the case of OCN⁻, CH₃OH, and the broad residual component C5 described in Boogert et al. (2008). However, in the case of background sources, the abundances of CO and CO₂ are similar to those seen in low-mass YSOs and are higher than those seen in the high-mass YSOs by a factor of around 3.4 (Boogert et al. 2013, 2015).

Considering pathways and morphology, the analysis of *Spitzer* observations demonstrated that CO₂, NH₃, and CH₄ ice abundances do not vary significantly with respect to H₂O ice, which indicates co-formation (Öberg et al. 2011). On the contrary, the same work shows that CH₃OH and CO ices vary by an order of magnitude relative to water ice, indicating different formation pathways. NH₄⁺ is anti-correlated with H₂O, likely because of the high desorption temperature compared to other volatiles (Greenberg & D'Hendecourt 1985). In the case of CO and CO₂, there is evidence in support of cosmic-ray processing –given the abundance ratio CO:CO₂– and of thermal evolution, which is mainly based on the CO₂ bending mode splitting at 15.2 μm (Pontoppidan et al. 2008). Attempts at assigning more complex molecules such as HCOOH, CH₃CHO, and CH₃CH₂OH were also made by Öberg et al. (2011), although they are still relying on the absorption features between 7 and 8 μm.

The main disadvantage of the phenomenological approach is the limited information about ice morphology and the a priori assumption of carriers associated with the absorption bands. In addition, these methods are usually limited to a short wavelength range to address a small set of spectral components (e.g. 3–5; Pontoppidan et al. 2003; Thi et al. 2006; Boogert et al. 2008). Other approaches combining laboratory ice spectra to fit the observed spectra have been adopted in previous works (Merrill et al. 1976; Gibb et al. 2004; Thi et al. 2006; Zasowski et al. 2009; Suutarinen 2015). Although the degeneracies are high in these methods, some issues in the direct comparison of the observations with the experimental data were revealed, as is the case for the absorption peak at 6 μm attributed to H₂O ice –which is found to be deeper than the features at 3 μm and 13 μm (Schutte et al. 1996; Keane et al. 2001)– and the nature of the peaks at 3.55 μm and 6.85 μm, which cannot be solely attributed to CH₃OH ice (Dartois & d'Hendecourt 2001; Schutte & Khanna 2003; Bottinelli et al. 2010).

In the interest of exploiting large data sets of laboratory ice spectra to identify and quantify molecules in ice mantles observed toward YSOs, we present a new computational code for spectral analysis. The code uses a genetic modelling algorithm (Holland 1975) to decompose IR spectra of protostars using a linear combination of laboratory data themselves. Evolutionary algorithms have successfully been used in the literature to derive the dust composition of asymptotic giant branch (AGB) stars (Baier et al. 2010), and to derive the physical properties of protostars (Woitke et al. 2016). This paper is laid out as follows: Sect. 2 lists the public databases containing ice spectra used in this work and Sect. 3 details the code methodology. The accuracy and performance tests of this tool are discussed in Sect. 4, and an application to a real source, the Class I protostar Elias 29, is shown in Sect. 5. A summary of the capabilities of the ENIIGMA fitting tool and the results of the Elias 29 spectral analysis are given in Sect. 6.

2. The ENIIGMA fitting tool

The ENIIGMA (dEcomposition of Infrared Ice features using Genetic Modelling Algorithm) fitting tool is designed for the

analysis of observational spectra containing ice absorption features. As a general overview, this tool is split into three major modules; the first provides Python functions to calculate the continuum spectral energy distribution (SED) of YSO spectra and to remove silicate features. The second focuses on the unbiased spectral decomposition using a linear combination of laboratory ice spectra. The third offers a statistical analysis of the fits such as confidence intervals and degeneracy quantification. The ENIIGMA fitting tool documentation and download instructions are publicly available¹.

In the first module, the ENIIGMA makes use of previous techniques available in the literature for continuum determination (e.g. Boogert et al. 2000) and silicate removal from a template (e.g. Bottinelli et al. 2010). The second and third modules represent the novelty of the code: genetic modelling algorithms are used to search for the optimal solution in a large problem space followed by a robust statistical analysis of the results. Global optimisation is a mathematical method that aims to search for the best overall solution providing the maximum likelihood among the parameters. This method is unbiased in the sense that a large sample of ice spectra are tested without targeting any particular species, and the final statistics-based decision of the best fit is made. In the context of ice observations in space, ENIIGMA allows us to explore a significant amount of IR laboratory data and to decipher the best linear combination of those that fit the observational spectrum. ice components. A flowchart of the entire procedure is shown in Fig. 1, and the different steps are detailed in the following sections.

2.1. Input data, continuum determination, and optical depth

The ENIIGMA fitting tool works with three types of input data. The first possibility is to work with observed spectra representing the flux in units of Jy. The second possibility is to use spectra in an optical depth scale, including the silicate features at 9.8 μm and 18 μm. Finally, it is possible to provide the tool with spectra where the silicate features have been removed. The file format and structure must be ASCII (American Standard Code for Information Interchange), with three columns containing the wavelength in micrometres, flux or optical depth, and the errors, respectively.

The accurate analysis of the chemical composition and column densities of ices in star-forming regions depends on the subtraction of continuum SEDs. However, determining an accurate continuum SED for YSOs is challenging because of the large width of the ice and silicate features. Wavelengths shortwards of 5 μm are better constrained because of the lack of ice bands at the *K*-band and the narrow features of CO₂ and CO ices at 4.27 and 4.67 μm, respectively. Some methods for the continuum determination shortwards of 4 μm have been adopted in the literature such as the near-IR excess removal using the Kurucz stellar atmosphere models (e.g. Pontoppidan et al. 2004) and a linear combination of black bodies to fit near- and mid-IR photometric data (e.g. Ishii et al. 2002; Moutaka et al. 2004; Perotti et al. 2020, 2021). Both methods not only provide the flux continuum, but also physical properties such as near-IR extinction or effective temperature.

The continuum above 5 μm on the other hand is much less constrained, even though short-wavelength data are available. The entire range between 5 and 30 μm is dominated by multiple broad and narrow absorption and emission features associated to ices and polycyclic aromatic hydrocarbons (PAHs),

¹ <https://eniigma-fitting-tool.readthedocs.io>

Fig. 1. Flowchart of the ENIIGMA fitting tool. The light grey dashed box highlights the steps during the spectral decomposition. The light yellow region highlights the Genetic Algorithm (GA) module, whereas the light green region shows the steps performed during the post-processing statistical analysis. The small grey boxes indicate the input data requested by the tool. Small blue boxes are the output data. White boxes indicate the processes performed in each step.

respectively (see reviews by Tielens 2008; Boogert et al. 2015). Hydrogenated amorphous carbon (HAC), or more generally a-C(:H), is another class of material observed toward sources at the Galactic centre with multiple features in this spectral range (Chiar et al. 2000; Jones 2012). Toward YSOs, a-C(:H) has been attributed to the absorption bands at 3.47 μm , 6.85 μm , and 7.25 μm (e.g. Gibb et al. 2004; Alata et al. 2014). Nevertheless, the profile at 3.47 μm shows a good correlation with the water ice band at 3 μm , suggesting that this band is associated with volatile chemical species instead (e.g. Brooke et al. 1999; Dartois et al. 2002; Boogert et al. 2015). The strong feature at 6.85 μm observed towards YSOs has also been excluded as a potential carrier of a-C(:H) –as it is observed in a similar wavelength in the diffuse ISM– because of the absence of the strong feature at 3.40 μm (Boogert et al. 2015). The 6.85 μm band has also been attributed to the CO_3^- ion in minerals, but the lack of absorption bands at short or long wavelengths leads us to discard this possibility (e.g. Tielens & Allamandola 1987; Schutte et al. 1996; Boogert et al. 2008, 2015). Finally, the feature at 7.25 μm can be associated with HCOOH (Schutte et al. 1999), ethanol ($\text{CH}_3\text{CH}_2\text{OH}$; Öberg et al. 2011; Terwisscha van Scheltinga et al. 2018), and aminomethanol ($\text{NH}_2\text{CH}_2\text{OH}$; Bossa et al. 2009). However, Jones (2016) discuss that the other compounds containing C=O bonds (e.g. ketone, organic carbonate) could be present in the grains before or at the onset of ice mantle formation, and therefore the bands at 3.47 μm , 6.85 μm , and 7.25 μm could be consistent with carbonyl functional groups. Because of all of these

uncertainties, a low-order polynomial fit has often been used to trace the continuum between the 5 and 32 μm spectral range (e.g. Gibb et al. 2000, 2004; Pontoppidan et al. 2005; Boogert et al. 2008, 2011; Zasowski et al. 2009; Öberg et al. 2011; Noble et al. 2013, 2017).

Given that context, the low-order polynomial and blackbody combination methods are incorporated in the ENIIGMA tool for the continuum determination of the observational flux. In the polynomial approach, the function below is used:

$$F_{\text{poly}}(\lambda) = \sum_{k=0}^n a_k \lambda^k, \quad (1)$$

where a_k are the constants, λ^k is the wavelength, and k is the polynomial order. The second method adopts a sum of blackbodies, given by

$$F_{\text{BB}}(\lambda, T) = \sum_{i=1}^m f_i \frac{2hc^2}{\lambda^5} \left(\exp \left[\frac{hc}{\lambda k_B T_i} \right] - 1 \right)^{-1}, \quad (2)$$

where f_i is a scale factor, h is Planck's constant, c is the light velocity, k_B the Boltzmann's constant, λ the wavelength, and T_i is the temperature of each blackbody.

Once the continuum SED is determined from Eqs. (1) or (2), the observed spectrum (F_{λ}^{obs}) is converted to an optical depth scale using the following equation:

$$\tau_{\lambda}^{\text{obs}} = -\ln \left(\frac{F_{\lambda}^{\text{obs}}}{F_{\lambda}^{\text{cont}}} \right). \quad (3)$$

Fig. 2. Illustration of the silicate removal with the synthetic silicate method. *Top:* B1-b optical depth is shown by the black solid line, whereas the dashed and solid lines are the scaled GCS 3 silicate profile and the synthetic silicate feature, respectively. *Bottom:* residual optical depth after synthetic silicate removal (black line). The blue and red colours show the residual optical depth after local and template continuum methods, respectively (see text for details; Bottinelli et al. 2010).

2.2. Silicate removal

The spectral range between 8 and 25 μm is dominated by strong and weaker silicate absorption features at 9.7 μm and 18 μm , respectively. However, other functional groups associated with simple or complex molecules are also blended in the silicate band at 9.7 μm . Some examples are ammonia (9.01 μm ; ν_2 umbrella vibrational mode) and methanol (9.74 μm ; ν_4 stretching vibrational mode of C–O bond) as discussed in Boogert et al. (2008) and Bottinelli et al. (2010). Additionally, ethanol ($\text{CH}_3\text{CH}_2\text{OH}$) also shares the C–O stretching mode at 9.51 μm with the silicate broad profile.

Proper removal of the silicate feature is therefore critical, although not straightforward. Different methodologies have been used in the literature, such as the local continuum and silicate spectrum template of sources toward the Galactic centre (e.g. the GCS 3 silicate absorption profile, taken from Kemper et al. 2004). In Boogert et al. (2008), this absorption feature is narrower compared to the observed silicate band of HH 46 IRS. As a consequence, when the GCS 3 absorption band is used for silicate removal, a significant residual is left between 8 and 10 μm . A systematic study by van Breemen et al. (2011) shows that such an IR excess is absent in the sightline of the diffuse ISM, while it is present toward molecular clouds. Although this difference can be associated with ice growth in some cases (e.g. H_2O , NH_3 , and CH_3OH ; Ossenkopf & Henning 1994; Boogert & Ehrenfreund 2004; Boogert et al. 2008; McClure

2009; Bottinelli et al. 2010), it is also observed toward sources with weak or absent absorption of these molecules at other wavelengths (e.g. SStc2d_J182835.8+002616; van Breemen et al. 2011). In this regard, such an IR excess is likely due to the chemical composition of the dust, and most specifically the combination of magnesium-rich pyroxene-type (MgSiO_3) and amorphous olivine-type (Mg_2SiO_4) silicates. Further evidence that the silicate band is not associated with only one composition has been found by Poteet et al. (2015), who fitted the silicate profile with a linear combination of amorphous silicates, in particular Mg_2SiO_4 , MgFeSiO_4 , and MgSiO_3 . This latter study indicates that although small, variations in the silicate band shape cannot be ruled out. In the same framework, the THEMIS (The Heterogeneous dust Evolution Model for Interstellar Solids) code also uses amorphous olivine-type and pyroxene-type silicates with inclusions of iron compounds and a–C(:H) to interpret the observed absorption bands in the diffuse ISM (Jones et al. 2013). Additionally, the effects of dust evolution from the diffuse ISM to denser regions can be addressed with the THEMIS code (e.g. Köhler et al. 2015; Jones et al. 2017).

The debate about the nature (geometry, size, and composition) of the silicate feature toward Galactic centre sources and protostars is still to be clarified by high-quality data provided by the upcoming *James Webb* Space Telescope observations. To avoid the uncertainties underlying the silicate feature, the ENIGMA fitting tool assumes a synthetic silicate profile. This method is similar to the silicate template scaling approach used by Bottinelli et al. (2010). The steps adopted here are the following: (I) the observational GCS 3 silicate profile is scaled to match the bands at 9.7 μm and 18 μm (same as Bottinelli et al. 2010); (II) these two silicate bands are decomposed into six Gaussian components (three for each band); and (III) a small variation of the width and height of the six Gaussian components are performed to match the YSO silicate profile. The synthetic silicate is therefore given by the following set of equations:

$$G(\lambda; A, \lambda_0, \sigma) = \frac{A}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(\lambda - \lambda_0)^2}{2\sigma^2} \right], \quad (4a)$$

$$\tau_{\lambda}^{\text{ss}} = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} G(\lambda^i, \lambda_0^j), \quad (4b)$$

$$\lambda_0(\mu\text{m}) = 8.3, 9.7, 11.2, 16.2, 18.0, 20.8, \quad (4c)$$

where $G(\lambda; A, \lambda_0, \sigma)$ is the Gaussian function, $\tau_{\lambda}^{\text{ss}}$ is the synthetic silicate (ss) optical depth, λ_0 are the fixed peak positions of each component, A is the amplitude and σ is the width of the components. We note that in this approach only A and σ are allowed to vary. This method assumes that the silicate width of protostars is substantially caused by the dust chemical composition as shown by van Breemen et al. (2011) and the ice features are not affected.

Figure 2 shows an example of how the synthetic silicate method is applied to an observational spectrum. As shown in the top panel, the GCS 3 silicate profile is scaled to the peak at 9.7 μm of the B1-b spectrum taken from Boogert et al. (2008) (Step I). After steps (II) and (III), the synthetic silicate feature is calculated to match the observational spectrum. However, the band at 18 μm is less deep than the observed profile. In the bottom panel, the residual spectrum of B1-b after silicate removal is shown and compared with the residual spectra taken from Bottinelli et al. (2010). To extract the 9.7 μm silicate feature, these latter authors used the local and template continuum methods. Briefly, the local continuum mimics the silicate feature

by a short-range polynomial fit, and in particular the intervals of 8.25–8.75, 9.23–9.37, and 9.98–10.4 μm . Conversely, the template continuum uses empirical profiles separated into three general categories, namely (1) sources with a straight 8 μm wing, (2) sources with a curved 8 μm wing, and (3) sources with a rising 8 μm wing (‘emission’ sources). In the case of the source B1-b, the curved 8 μm wing has been used. As mentioned by Bottinelli et al. (2010), the template method is not suitable to fit the 9.7 μm silicate feature toward all YSOs in their sample, and the local continuum approach was adopted instead. However, the latter method provides narrower bands for the feature at 9 μm – often associated with the NH_3 ice – than in the method adopting the silicate template as pointed out by Öberg et al. (2011). From a simple comparison with the naked eye, the synthetic silicate method introduced in this paper can be seen to agree very well with the methods adopted by Bottinelli et al. (2010). In this way, it can be an alternative to the local continuum method when the template profile is not available. It is worth noting that an advantage of the synthetic silicate method is that it provides the silicate extraction of both 9.7 μm and 18 μm bands simultaneously, which allows a better analysis of the H_2O libration band at around 13 μm . Despite the challenges, new laboratory data combining ice and dust could provide a better understanding of the link between ice and dust in star-forming regions. To this end, Potapov et al. (2018, 2021) measured spectrum and optical constants of silicate (MgSiO_3) and H_2O ice, and showed that this experimental data can partially fit the 3 μm of some astronomical objects.

2.3. Laboratory ice spectra and databases

In order to build an internal database of experimental data for the ENIIGMA to decompose the spectra of YSOs, a data set of around 100 laboratory ice spectra was compiled from the public online databases listed below:

- NASA Ames: The Astrophysics & Astrochemistry Lab at NASA Ames Research Center²;
- Leiden DB: Leiden Ice Database³;
- UNIVAP: Laboratorio de Astroquímica & Astrobiología da UNIVAP⁴.

The ice samples compiled in this paper were recorded in transmittance (T) or absorbance (A) mode using the Fourier Transform Infrared Spectrometer (FTIR). All spectra in transmittance scale were converted to absorbance using the equation $A = -\log_{10}(T)$. In the absorbance scale, the ice spectra can be converted to optical depth using the following equation (D’Hendecourt & Allamandola 1986):

$$\tau_{\lambda}^{\text{lab}} = 2.3 \cdot A_{\lambda}. \quad (5)$$

In addition to the compiled data, two IR spectra containing organic residue material (henceforth: residue) at high temperature (>100 K) were taken from the literature. Residue is produced with or without UV irradiation of the ice mixture samples and subsequent warming up to room temperature (300 K). In particular, the spectra taken from Schutte & Khanna (2003) are HNCO:NH_3 heated to 120 K without energetic processing and $\text{H}_2\text{O:CO}_2:\text{NH}_3:\text{O}_2$ irradiated with UV and heated to 200 K. Both of these experiments identified the ammonium ion (NH_4^+) formation, a likely carrier of the absorption band at 6.85 μm in YSO spectra (Schutte et al. 1996).

² <http://www.astrochem.org>

³ <https://icedb.strw.leidenuniv.nl>

⁴ <https://www1.univap.br/gaa/nkabs-database/data.htm>

Accurate baseline correction of the laboratory data is required to avoid misinterpretations of the carriers. As an example, the band at 5.85 μm that was attributed to HCOOH by Schutte et al. (1996, 1999) was put into question by Boogert et al. (2008) and Öberg et al. (2011) because of unreliable laboratory baselines. Instead, ethanol ($\text{CH}_3\text{CH}_2\text{OH}$) was proposed as the carrier by those authors. To avoid this issue, the baselines of the data compiled in this paper were checked and corrected when necessary. Appendix A lists the data sets used in this paper and the reference from which the data were taken. Briefly, these ice data are grouped according to common characteristics, such as (i) pure, (ii) pure and heated, (iii) heated mixtures, (iv) irradiated mixtures, and (v) residue.

2.4. Genetic algorithm module

Inspired by the theory of natural selection and biological evolution, the genetic algorithms (GA) aim to provide a global optimisation process for a proposed problem. For clarification, the GA nomenclature is briefly explained: (i) *gene* corresponds to the value of a parameter (e.g. scaling factor of the ice spectrum to match the observation), (ii) *chromosome* is a set of values proposed as a solution of the parameters, (iii) *fitness function* is the criteria to evaluate the fit (e.g. squared residual, reduced χ^2), *selection* is the election of the best parameters from the entire set of potential solutions (chromosome) used to generate improved values, (iv) *crossover* is the mixing of values from two distinct solutions in order to create a new solution, (v) *mutation* is a small variation of one or more parameters (gene), (vi) *global minimum* is the optimal solution of the parameters, and (vii) *generation* corresponds to an interaction where potential solutions are tested.

The GA module implemented in the ENIIGMA fitting tool is built on Pyevolve⁵ (Perone 2009), an open-source and extendible library dedicated to performing evolutionary computation in Python programming language. The GA computation is composed of three main characteristics, namely generation of a random population of probable solutions, fitness-oriented to evaluate the population, and variation-driven to improve the next population (Holland 1975; Koza 1992). The population follows the chromosome-like structure, in which the vector $w_{ij} \in \mathbb{R}^{m \times n}$ is given by:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{bmatrix}, \quad (6)$$

where each variable w is called gene, and the rows are the combination of genes that provide a solution. The quality of the population is evaluated by a fitness function (F) – the squared residual as shown below:

$$F = \sum_{i=0}^{n-1} \left(\tau_{\lambda,i}^{\text{obs}} - \sum_{j=0}^{m-1} w_j \tau_{\lambda,j}^{\text{lab}} \right)^2, \quad (7)$$

where the optical depth of the observational spectrum ($\tau_{\lambda,i}^{\text{obs}}$) was defined in Eq. (3), $\tau_{\lambda,j}^{\text{lab}}$ is the optical depth of the laboratory data, w_j is the scale factor (the genes in the chromosome; Eq. (6)), and σ_i^2 is the error of optical depth propagated from the error in the flux.

⁵ http://pyevolve.sourceforge.net/0_6rc1

After checking the random population with the fitness function in the first iteration, the genetic operators are used to improve the population in the subsequent generations. Selection is the first operator and aims to guarantee the survival of the best genes (coefficients) from one generation to another. One of the selection methods adopted in this paper is called the ‘roulette wheel’ and elects the best individuals based on their probability of providing good solutions, which is defined as:

$$P_i = \frac{F_i}{\sum_{j=1}^N F_j}, \quad (8)$$

where F_i is the fitness of individual i in the population, and N is the number of individuals in the population. The second selection method is called ‘tournament’, and randomly selects a fixed number of individuals from the entire population. The ‘winner’ of the tournament, that is, the coefficient with the highest fitness, is passed to the next generations. In comparison with the roulette wheel method, the tournament selection has a lower computational cost. Moreover, the roulette wheel method suffers from the problem of premature convergence if there are individuals with a high probability to be selected, whereas the selection pressure in the tournament selection is constant and maintains the diversity of the population (Blickle & Thiele 1996).

The best individuals passed from one generation to another are elected to the crossover operation. Formally, if Eq. (6) is given for only two chromosomes of two genes, that is, $\mathbf{W}^1 = [w_{11}, w_{12}]$ and $\mathbf{W}^2 = [w_{21}, w_{22}]$, the single-point crossover operator is given by H_{W^1, W^2} , which will allow the elected genes to exchange positions with one another. Depending on the complexity of the problem, for example, a linear combination of more than four components, two-point or multi-point crossovers can also be used in the ENIIGMA. The crossover rate is defined as 80% in Pyevolve and provides good solutions for many problems in the literature (e.g. Peng et al. 2003; Baier et al. 2010). However, a crossover rate of 90% provided better results in the present study when the tournament selection method was adopted. The mutation operator is used next, and consists of applying a small perturbation ζ in one or more genes (scale factor), formally expressed as $w'_{mn} = w_{mn} + \zeta$. In particular, a Gaussian mutation is adopted with a mutation rate of 10% in this paper. Compared with other mutation methods, a Gaussian distribution of the genes was found to be superior to other approaches (Hinterding 1995). Once the genetic operators are applied, the fitness function evaluates how close the new genes are to the solution. This process is repeated a number of times until the convergence criteria are reached, which in this paper is the lowest values of the fitness function after all generations.

Although genetic algorithms are one of the simplest random-based evolutionary algorithms (EAs), it is not less robust than other EAs or even other conventional optimisation methods. As such, it has been used to solve complex problems in astrophysics such as the huge degeneracy behind the physical parameters of protoplanetary discs and the silicate composition in the mid-IR (e.g. Charbonneau 1995; Hetem & Gregorio-Hetem 2007; Baier et al. 2010; Woitke et al. 2019).

2.5. Searching for the best solution

The ENIIGMA fitting tool requires an initial guess (G) composed of four laboratory spectra to start searching for the best combination that minimises the residual in Eq. (7). At this initial stage, the GA module is applied to fit G to the observational optical depth $\tau_{\lambda}^{\text{obs}}$, and the residual F_1 is calculated. Next, each laboratory data sample is added to the initial guess one at a time in

a numerical loop, and another combination is used to fit $\tau_{\lambda}^{\text{obs}}$, which leads to another residual calculation, F_2 . All the laboratory data providing $F_2 < F_1$ are stored, and after this stage, the GA module is applied to combine them into groups of eight or more genes, which results in the global minimum. It is worth noting that during the search for the best solution, the information from the initial guess might be lost, because other ice spectra can provide better solutions. In this sense, the ENIIGMA fitting tool does not target a specific molecule or ice sample, but rather it searches for the best linear combination from a large database performing an unbiased search for the global minimum solution.

The population (W) and the generation numbers are two of the most important parameters in GAs. If these two numbers are too small, the GA module cannot find a proper solution, because the variability of the genes is essential in the context of EAs. High numbers on the other hand will provide good solutions, but the computation limit must also be taken into account. Following previous works employing GAs (e.g. Szkody et al. 2010; Harrison & Hamilton 2015), the ENIIGMA fitting tool uses the population-to-generation ratio of 90/100 or 150/200, which allows it to search the global minimum parameters from a large population whilst avoiding premature convergence.

2.6. Ice column density

The optical depths obtained from the GA optimisation are used to calculate the ice column densities following the equation:

$$N_{\text{ice}} = \frac{1}{\mathcal{A}} \int_{\nu_1}^{\nu_2} (w \cdot \tau_{\nu}) d\nu, \quad (9)$$

where $\mathcal{A}$ is the band strength of a specific vibrational mode. If the ice is composed of only one molecule (pure ice), the absorption features themselves are used to calculate the column density. However, in the opposite case, blended features are avoided, and the isolated vibrational modes are used instead. If blended profiles cannot be avoided, a Gaussian or Lorentzian decomposition is applied to separate the components. Table 1 lists the band strengths of molecules present in the ENIIGMA database used to calculate the ice column densities.

2.7. Statistical analysis

The heuristic nature of GA optimisation makes the final result unpredictable, although one expects that the global minimum is always found. In addition, the decomposition of the IR ice features is a source of huge degeneracy itself on top of the fitting procedure. Given these two aspects, a statistical analysis becomes necessary to evaluate how good the solutions are after an enormous number of combinations of laboratory ice data. For example, if the GA module is applied to combine 13 ice spectra in groups of 8 and without repetition, there are 1287 different solutions for the same problem. The aim of the post-processing statistical analysis is to check how close the other combinations of the optimal solution are. In this way, both the heuristic aspect and degeneracy itself can be evaluated as a whole.

The statistical analysis is based on the minimum $\Delta\chi^2$ method (Avni & Bahcall 1980), given by the following equations:

$$\chi^2 = \frac{F}{\sigma^2}, \quad (10a)$$

$$\Delta\chi^2(\nu, \alpha) = \chi^2 - \chi_{\text{min}}^2, \quad (10b)$$

where F is the squared residual in Eq. (7) and σ the standard deviation taken from the observational optical depth; ν and α are

Table 1. Ice features and band strengths of the identified molecules during the spectral decomposition.

Molecule	λ [μm]	ν [cm^{-1}]	Identification	$\mathcal{A}$ [cm mol^{-1}]	References
H ₂ O	3.01	3322	O–H stretch	2.2×10^{-16}	Bouilloud et al. (2015)
H ₂ O	6.00	1666	H ₂ O bend	1.1×10^{-17}	Bouilloud et al. (2015)
NH ₃	9.01	1109	NH ₃ umbrella	2.1×10^{-17}	Bouilloud et al. (2015)
CH ₄	7.67	1303	CH ₄ deformation	8.4×10^{-18}	Bouilloud et al. (2015)
CH ₃ OH	9.74	1128	C–O stretch	1.78×10^{-17}	Bouilloud et al. (2015)
CH ₃ CN	4.44	2252	CN stretch	2.2×10^{-18} ^(a)	Hudson & Moore (2004)
HCOOH	8.22	1216	C–O stretch	8.22×10^{-17}	Bouilloud et al. (2015)
H ₂ CO	5.83	1715	C–O stretch	9.60×10^{-18}	Schutte et al. (1993)
CH ₃ CH ₂ OH	9.17	1090	CH ₃ rock	7.35×10^{-18}	Terwisscha van Scheltinga et al. (2018)
CH ₃ OCH ₃	8.59	1163	COC stretch. + CH ₃ rock	3.00×10^{-17}	Terwisscha van Scheltinga et al. (2018)
CH ₃ CHO	5.80	1723	C–O stretch	1.3×10^{-17}	Terwisscha van Scheltinga et al. (2018)
COO [−]	6.30	1586	COO [−] stretch	6.0×10^{-17}	Muñoz Caro & Schutte (2003)
NH ₄ ⁺	6.85	1460	NH ₄ ⁺ bend	3.0×10^{-17}	Schutte & Khanna (2003)
CO ₂	4.27	2341	CO stretch	1.3×10^{-16}	Bouilloud et al. (2015)
CO	4.67	2141	CO stretch	1.4×10^{-17}	Bouilloud et al. (2015)

Notes. ^(a)The band at 4.44 μm (C $\equiv$ N) was used because we were unable to find the band strength for CH₃CN at 9.61 μm (CH₃ rock).

the number of free parameters and the statistical significance, respectively. χ^2_{min} corresponds to the goodness-of-fit in the global minimum solution. One can observe, that confidence regions do not depend on the accuracy of the fit, but the number of free parameters. Equations (10a),b are calculated from a normal variation around the optimal scale factors, w , by using a probability distribution $p(x)$ available in the `numpy.random.normal` numpy (Harris et al. 2020) routine:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right], \quad (11)$$

where the mean, μ , is the optimal scale factor itself, and σ is the standard deviation taken from the observational optical depth. From this analysis, the confidence intervals of the linear combination coefficients are calculated with a significance of 3σ .

Column density error bars. The uncertainty in the column densities is calculated from the confidence intervals derived with the minimum χ^2 method. In this regard, the lower and upper scale parameters are used to calculate the minimum and maximum optical depths. The ice column density calculated from these limits provides the bounds for the error bars of each molecule found during the spectral decomposition. It must be stressed that in this method the error bar depends on the number of parameters in the fit instead of the covariance matrix in non-heuristic methods (e.g. non-linear least squares problems). For example, for a fit with two components, the 1–3 σ confidence intervals are obtained for $\Delta\chi^2$ equal to 2.41, 4.61, and 9.21, respectively. On the other hand, if eight laboratory data samples are used, the confidence intervals become $\Delta\chi^2$ equal to 9.52, 13.36, and 20.09, respectively.

Recurrence plots. The degeneracy of the spectral decomposition is quantified with a recurrence plot analysis. All solutions found by the ENIIGMA are sorted into increasing order by the χ^2 value of the fit. Next, Eq. (10a) is used to calculate $\Delta\chi^2$ and define the bounds for the 3 σ confidence interval. All solutions outside this level of significance are excluded from the analysis. Finally, the frequency of each ice sample is calculated, and the recurrence is derived using the following equation:

$$R = \frac{f_i}{S}, \quad (12)$$

where f_i is the absolute frequency of sample i and S is the sum of all solutions. The recurrence of the ice data is shown as a percentage, and 100% means that a given sample was found in all solutions, indicating that such a component is required for the spectral decomposition. Lower percentage values mean that the corresponding ice sample can be replaced by another data sample and still provide a solution with high statistical significance.

Histograms. Because of the degeneracy involved in the fit of observational YSO spectra, ice column densities derived from local minimum solutions might deviate from values calculated from the optimal combination. In this regard, the column density variation of the ice components shown in the Recurrence plot is estimated from a histograms analysis. The bin sizes are proportional to the column density variances and were calculated by the Freedman Diaconis Estimator taken from the `numpy.histogram_bin_edges` Python package. Additionally, the histograms are log-transformed and normalised by the median ice column density. With this analysis, both the mean and 3 σ confidence interval for the column densities are calculated (see also Appendix B).

3. Accuracy tests

In this section, we address the capability of the ENIIGMA fitting tool to provide accurate solutions. For these tests, known ice samples present in the database (see Appendix A) are used as input data to check the accuracy of the tool to search for the correct sample and fractionation.

Depending on the complexity of the problem, the method to find the global minimum solution with genetic algorithms may require more robust genetic operators. In this section, three methods were adopted to tackle the proposed problems, and their parameters are shown in Table 2. The population size and generation number are important input values. If lower values are adopted for these two parameters, the solution space might not be entirely covered. In addition, two selection methods and crossover rates were adopted depending on the population size.

Table 2. Laboratory data of pure and thermally processed ices used in this paper.

Method	Mutation rate	Selection method	Crossover rate	Crossover method	Gen.	Pop. size
1	0.1%	Roulette wheel	80%	Single-point	100	90
2	0.1%	Tournament	90%	Single-point	200	150
3	0.1%	Tournament	90%	Two-point	200	150

3.1. Identification of known samples

The ice spectra used to test the ENIIGMA tool with known ice samples were taken from the public databases shown in Sect. 2.3, and are listed in Tables A.1 and A.4.

Three tests were performed in this section, namely (i) non-blind: the solution is indicated in the initial guess and is also present in the database; (ii) fully blind: the solution is not included in the initial guess but it is available in the database, and (iii) solution not available: the correct sample is deliberately removed from the database and not chosen as initial guess. In all cases, the three searching methods in Table 2 provided the expected global minimum solution.

Pure ice. The test with the pure ice sample is performed with H₂O (Hudgins et al. 1993) and ethanol (CH₃CH₂OH; Terwisscha van Scheltinga et al. 2018) ice at 15 K. Water ice is selected because it is the most abundant molecule in the solid phase observed towards YSOs. On the other hand, ethanol was selected because of the multiple narrow features in the IR spectrum, and because it has some common absorption features with methanol (CH₃OH) ice, which is also observed in star-forming regions. Figures 3 and 4 show the results of the three tests. The non-blind and blind tests in the left and middle panels indicate that the ENIIGMA tool can identify the correct ice spectra associated with pure water and ethanol. The residuals shown in the bottom boxes are below 5% and are the result of the intrinsic randomness behind the genetic optimisation. It is worth noting that the difference in the residuals in tests 1 and 2 are not associated to the presence or absence of the correct ice sample in the initial guess, but rather are related to the randomness of the method itself. The right panel shows the fit when the correct solution is not available in the database. In that case, the ENIIGMA tool fits the input spectra with the most similar data sample available in the database, that is, H₂O at 40 K and CH₃CH₂OH at 30 K, respectively. As a result, the residuals in some wavelengths are higher than in tests 1 and 2.

Binary ice mixture. A mixture dominated by H₂O (95%) ice containing a fraction of ethanol (5%) at 15 K is used in the test with the binary ice sample. Despite the small fraction, ethanol clearly shows features at around 9 μ m and 11 μ m. The fitting results are shown in Fig. 5. The left and middle panels showing the fully sighted and fully blinded tests indicate that the correct ice spectra were assigned with residuals lower than 5%. In the case where the correct solution is not available (right panel), the H₂O:CH₃CH₂OH at 30 K was selected by the fitting routine. As in the pure ice case, this fit indicates that the ENIIGMA tool is able to select the most similar ice sample to fit the input spectrum. Because of the small differences of this ice mixture at 15 K and 30 K, the residual is around 5%.

Ternary ice mixture. The third test is performed with a tertiary ice mixture composed by CO:CH₃OH:CH₃CH₂OH (20:20:1). As seen in Fig. 6, the fits were also accurate for the fully sighted and blinded tests (left and right panels). The residual in the middle panel is higher than in the left panel, clearly

showing the random nature of the method. The fit in the right panel was performed by combining two ice samples, namely CO:CH₃CH₂OH at 30 K and CH₃CH₂OH at 30 K. Because of the prominent CO feature in the input spectrum not present in the ethanol ice, the code combined the two samples in order to provide a good fit. The pure CO ice was not used in this fit because its full width at half maximum is 30% narrower than in the mixture with ethanol.

Quaternary ice mixture. The final test for the identification of known samples is performed with H₂O:CH₃OH:CO:NH₃ (100:50:1:1) at 10 K, where both water and methanol features are prominent. Left and middle panels of Fig. 7 show the results from the non-blind and fully blind tests, respectively. Again, the fits result in residuals of lower than 5%. Slightly higher residuals are seen in the right panel, where the solution is not available in the database. In that case, the ice sample composed by the sample molecules at 40 K was provided with the fit. As in the case of the tertiary mixture, this result indicates that a combination of the pure molecules does not result in a good fit. For example, Dawes et al. (2016) show that the peak position of the O–H vibrational at around 3.0 μ m is redshifted in a mixture of H₂O:CH₃OH compared to the pure H₂O. Nevertheless, the temperature difference of this mixture at 10 and 40 K is evident in the residuals. At this temperature, most of the CO ice has been thermally desorbed (Collings et al. 2004; Acharyya et al. 2007), whereas a small fraction trapped in the porous ice has enough energy to migrate over active sites.

3.2. Fractionation

As water represents the major component of astrophysical ices observed toward YSOs, in this section we aim to check the capability of the ENIIGMA tool in finding the correct coefficients in a linear combination of H₂O ice and another component. In the tests shown in Sects. 3.2.1–3.2.3, two IR ice spectra were summed with different coefficients, and Gaussian noise of standard deviation equal to 0.01 has been added to the data to resemble the noise present in the astronomical observations. Although the tests in Sects. 3.2.1 and 3.2.2 use water ice spectra at two different temperatures, the aim is to discuss the fractionation of constructed mixtures with similar spectral bands, rather than thermal processing. The searching Methods 1–3 in Table 2 provided the expected global minimum solutions. In the test shown in Sect. 3.2.5, eight IR ice spectra were considered to mimic the ice absorption features observed toward protostars. In this case, only Methods 2 and 3 provided the best fits.

3.2.1. H₂O (15 K) and H₂O (40 K)

Infrared spectra of pure water at 15 K and 40 K were summed with proportions of 50:50 and 90:10, respectively. These two ice data samples were selected because they have small differences in the band shapes and thus represent a challenging task for fitting methods. Figure 8 shows the test results for these two H₂O

Fig. 3. ENIIGMA fit (dashed green line) of the pure H₂O ice sample at 15 K (black line). *Left and middle panels:* fit in non-blind and blind tests, respectively. *Right panel:* fit when the correct solution does not exist. The residuals of the fit are shown in the boxes below the fits.

Fig. 4. Same as Fig. 3, but for pure CH₃CH₂OH at 15 K.

Fig. 5. Same as Fig. 3, but for H₂O:CH₃CH₂OH (20:1) at 15 K.

ice spectra. The top left panel shows the good match between the ENIIGMA fit and combined H₂O ice data with proportions of 50:50. The residual plot in the bottom part of this figure shows that only the Gaussian residual noise is left after subtracting the model from the data. The top right panel shows a $\Delta\chi^2$ map, where the confidence intervals ($1-3\sigma$) are indicated by the green, yellow, and red contours, respectively. As we note from this degeneracy analysis, the increasing H₂O proportion at 15 K with simultaneous decreasing H₂O proportion at (40 K) also provides statistically significant results inside the confidence intervals.

The two bottom panels of Fig. 8 show the same test for the proportion of 90:10 at 15 K and 40 K, respectively. The spectral decomposition shown in the left panel also shows a good match between the ENIIGMA fit and the combined H₂O ice data, with residuals of lower than 5%. In this test, the $\Delta\chi^2$ map shown in the right panel indicates that a solution considering only H₂O (15 K) is also statistically possible when the fraction of H₂O (40 K) is lower than 10%. The same solution is still found by the ENIIGMA tool with the σ_{noise} equal to 0.02. Above that limit, the ENIIGMA is unable to find the H₂O (40 K) component.

Fig. 6. Same as Fig. 3, but for CO:CH₃OH:CH₃CH₂OH (20:20:1) at 15 K.

Fig. 7. Same as Fig. 3, but for H₂O:CO:NH₃:CH₃OH at 10 K.

For the case where the H₂O ice proportions are 90:10, a population evolution analysis is shown in Fig. 9. At the initial generations, most of the population values result in a high fitness score (white region) and therefore the optimal coefficients have not yet been found. As the optimisation evolves, the population is improved and better coefficients are generated because of the survival of the best solution. Close to generation 85, most of the population results in a low fitness score, which increases the probability of finding the global minimum solution of the problem. The fitness score evolution over the generations is shown in the bottom panel of Fig. 9. As noted in the top panel, the fitness score at the initial generations is high for any value in the population. With the improvement of the population, the fit is improved until the optimal solution is found between generations 80 and 85, with $\chi^2_v = 1.01$. As the population has not converged in this problem, the fitness score starts increasing again until generation 100, which is the stopping criteria in this problem.

3.2.2. H₂O (15 K) and H₂O (75 K)

A second check with water ice samples at (15 K) and (75 K) was also made. At high temperatures the water band shapes are significantly different from those in the cases of 15 K and 40 K because of the transition from amorphous to cubic crystalline ice (Hudgins et al. 1993). In particular, the O–H band at 3 μm is gradually sharpened until the full crystalline ice is formed between 100 and 140 K (Sack & Baragiola 1993). As shown in the left panels of Fig. 10, the ENIIGMA tool is able to identify the correct proportions of the two samples in the cases of 50:50 and

10:90, respectively. This is also indicated by the low fit residuals of around 5% in both situations.

The confidence intervals of the optimal solution are shown in the right panels of Fig. 10. In these plots, the reduced area of the contours compared to Fig. 8 indicates that the solution is less degenerated. The reason is the differences of the band shapes of the cold and warm H₂O ice components. In the case of 50:50 proportion shown in the top right panel, the 3σ confidence interval of both coefficients is around 20%. On the other hand, in the bottom right panel, the degeneracy of the 10:90 proportion is 50% and 10%, respectively, which indicates that the H₂O (75 K) component is better constrained than the H₂O (15 K) component.

After the global minimum solution is found, the degeneracy analysis of the ice composition is performed for the case with proportions of 10:90. Figures 11a–c show the IR spectra of H₂O ice at 15, 40, and 75 K. We note here that the spectral shape of the bands is similar between the samples at 15 and 40 K, and significantly different at 75 K because of the crystallisation of the ice matrix via warm-up. An initial test has been done with the sum of the two H₂O ice samples at 15 and 75 K without Gaussian noise and the recurrence plot was created. The recurrence plot shown in panel d indicates that no degeneracy exists inside the 3σ confidence interval when the spectrum is free of noise. On the other hand, if Gaussian noise is added to the summed spectrum, the degeneracy becomes evident. The dominant ice component, H₂O at 75 K, is still present in all solutions inside the same confidence interval, whereas the recurrence of the water component at 15 K is reduced to 50% because of its lower contribution. The presence of other samples in this recurrence analysis indicates

Fig. 8. Spectral decomposition and confidence interval analysis of the ENIIGMA results. *Left panels:* spectral decomposition of the combined water ice spectra (black solid line) with proportions of 50:50 and 90:10. The components of the decomposition are given by the red and blue dotted lines, whereas the ENIIGMA fit is given by the green dashed line. The residual of the fit is shown in the bottom parts of these graphs. *Right panels:* confidence interval analysis of the coefficients in the linear combination. The grey-scale colours are the difference between chi-square and minimum chi-square values; the olive, yellow, and red contours are 1σ , 2σ , and 3σ confidence intervals, respectively.

that the H₂O (15 K) can be replaced by one of those data samples and a good solution inside 3σ confidence interval is still provided.

3.2.3. H₂O (15 K) and CO:CH₃OH:CH₃CH₂OH (15 K)

In this test, the combination of H₂O and the mixture CO:CH₃OH:CH₃CH₂OH (20:20:1) at 15 K is used. In general, these two ice data samples have different band peaks and shapes in the IR spectrum, although the O–H stretching modes of H₂O and CH₃OH overlaps around 3 μm. Another difference is that the samples do not have the same ice thickness. Figure 12 shows the fits of the proportion 50:50. Because of the thickness difference, the water ice has a major contribution to the band at 3 μm compared to methanol, whereas at 4.67 μm and 8.86 μm the presence of the mixture sample becomes evident. The residuals of the fit are lower than 5% and indicate a good match between fit and combined data. The confidence interval analysis of the coefficients shown in the top right panel indicates that the water and mixture components vary by about 20% and 40%, respectively, inside 3σ .

The bottom left panel shows the fit of the proportions 90:10 for the water and mixture samples, respectively. In this case, the contribution of methanol at 3 μm is minimal, although it is still

visible around 9 μm. As seen from the residuals of the fit, the decomposition process was also able to identify the correct coefficients of the two samples. However, because of the small contribution of the mixture, the confidence interval analysis (bottom right panel) indicates that solutions considering only pure water are statistically significant as well.

3.2.4. Spurious features

To test the effect of spurious bands that could be present in a real observational spectrum, we added emission and absorption features of non-ice compounds to the constructed mixture shown in Sect. 3.2.3. The emission features mimic the hydrogen lines at 3.04 μm (HI Pfε 5–10), 3.30 μm (HI Pfδ 5–9), and 3.74 μm (HI Pfγ 5–8). These lines are seen in spectra observed with ground-based telescopes when the standard star is not properly modelled (e.g. Pontoppidan et al. 2004; Thi et al. 2006). These emission profiles were created with the Python package `pysynphot`⁶ (STScI Development Team 2013) by adopting a full width at half maximum of 50 Å. For the absorption feature, we use the naphthalene (C₁₀H₈) spectrum taken from Hudgins & Sandford (1998), which is the simplest polycyclic aromatic hydrocarbon

⁶ <https://pysynphot.readthedocs.io>

Fig. 9. *Top:* heat map of the fitness score for each population over the generations. *Bottom:* fitness score over the generations. The blue and red colours indicate the best and worst fitness based on the population of each generation. The grey shaded area between the two lines shows the difference between the worst and best fitness.

containing two rings. Naphthalene has multiple bands between 3.20 μm and 3.44 μm , where features of methanol and ethanol are also present.

Figure 13 displays the spectral decomposition of a constructed mixture containing ice bands (H_2O and $\text{CO}:\text{CH}_3\text{OH}:\text{CH}_3\text{CH}_2\text{OH}$) and spurious features. Once these emission and absorption features are located shortwards of 4 μm , the ENIIGMA fit is performed between 2.5 μm and 4.0 μm . This test shows that the ENIIGMA fitting tool is able to correctly identify the ice bands and can also fit the narrow features associated with spurious absorption and emission lines. In this example, the spectral profile between 3.2 μm and 3.5 μm is the most affected with spurious features, but those features do not change the shape of the ice bands. We highlight that in real observations, other spurious features might occur in the spectrum due to poor continuum subtraction, as well as the effect of spectral glitches and baseline jumps. In these cases, a broad band fit of the spectrum covering other vibrational modes of the same molecule could minimise the effect of spurious features in the spectral decomposition with the ENIIGMA package.

3.2.5. Synthetic ice spectrum

Multiple ice features have been observed in the IR spectra of protostars (e.g. Gibb et al. 2004; Zasowski et al. 2009). For example, a linear combination of five empirical components associated to more than one carrier have been adopted

to decompose the short spectral range between 5 and 8 μm (Boogert et al. 2008). In this test, a synthetic ice spectrum is created based on the median ice composition derived by Öberg et al. (2011) towards low-mass protostars, namely $\text{H}_2\text{O}:\text{CO}:\text{CO}_2:\text{CH}_3\text{OH}:\text{NH}_3:\text{CH}_4:\text{XCN}$ (1:0.29:0.29:0.03:0.05:0.05:0.003). However, the chemical species XCN is replaced by 2% CH_3CHO and 1% $\text{CH}_3\text{CH}_2\text{OH}$ with respect to the water ice to check the feasibility of identifying these molecules with the ENIIGMA tool. The synthetic spectrum shown in Fig. 14a is created with the linear combination of IR spectra of eight pure ices (temperatures: 10–15 K) listed in Table A.1 and the silicate spectrum from GCS 3 taken from Kemper et al. (2004). Also, a Gaussian noise of σ equal to 0.01 is added to the spectrum. This highlights that a real spectrum might have such strong silicate features, and that they must be subtracted before spectral decomposition with the ENIIGMA tool. The procedure of silicate removal in an observed spectrum is shown in Sects. 2.2 and 4.1.

The initial guess adopted in this test is made of H_2O (40 K), NH_3 (10 K), and CH_3OH (75 K). Figures 14b–e show the result of the spectral decomposition. Although the expected global minimum solution was found by the three methods in Table 2, Method 1 provided less accurate coefficients by about 4–7%. In the top panel, the fit is shown for the range between 2.5 and 20 μm . At this scale, only the major features are seen, such as H_2O (3 μm , 6 μm , 13 μm), CO_2 (4.27 μm , 15.53 μm), CO (4.67 μm), and CH_4 (7.673 μm). The residual plot shows that only the Gaussian noise is left after subtracting the data from the model, which indicates that both the samples and the coefficients were found by the ENIIGMA fitting tool. The bottom panels show a zoom onto the regions where the weak features are detected. The left panel highlights the C–H stretching mode at 3.53 μm . The middle panel shows the contribution of CH_3CHO at 5.8 μm and 7.4 μm , the contribution of NH_3 at 6.2 μm , and the contribution of CH_3OH at 6.8 μm . The right panel shows the dominant absorption features of NH_3 at 9.35 μm and CH_3OH at 9.8 μm , with small contributions of CH_3CHO at 8.9 μm and the $\text{CH}_3\text{CH}_2\text{OH}$ double peaks at 9.2 μm and 9.5 μm .

In summary, these tests show that ENIIGMA can find the correct components and linear combination coefficients in the combined spectral data. However, Methods 2 and 3 provide better results than Method 1 when the number of components is large.

4. Application to infrared spectra of the Class I protostar Elias 29

To test the ENIIGMA fitting tool with a real source, a spectral analysis of the well-known Class I YSO Elias 29 was carried out. This object is located in the Ophiuchus cloud at an estimated distance of 135 pc (Ortiz-León et al. 2018). Elias 29 is the brightest source in the ρ Oph E core and was observed in broad IR spectral range by the Infrared Space Observatory. Multiple absorption bands were securely and tentatively detected in the IR spectrum of this protostar (Gibb et al. 2004; Boogert et al. 2000; Rocha & Pilling 2015). Millimeter observations toward Elias 29 show high abundances of simple molecules (e.g. CO , H_2CO , HCO^+ , SO , SO_2 ; Boogert et al. 2002; Oya et al. 2019) and a deficiency in complex molecules (e.g. CH_3OH , HCOOCH_3 , CH_3OCH_3 ; Artur de la Villarmois et al. 2019; Oya et al. 2019). In particular, Oya et al. (2019) attribute the lack of organic molecules and richness in SO and SO_2 to the evolved nature of the source or the enhancement of the dust temperature above the CO freeze-out regime (>20 K) as suggested by Rocha & Pilling (2018) based on radiative transfer models. Artur de la Villarmois et al. (2019)

Fig. 10. Same as Fig. 8, but for H_2O at 15 K and 75 K.

Fig. 11. Degeneracy analysis for the combination of H_2O (15 K) + H_2O (75 K): 10:90. Panels a–c: stretching, bending, and librational modes of H_2O ice, respectively, at 15, 40, and 75 K. These panels highlight the temperature effect on the band shapes in the water ice IR spectra. Panel d: recurrence pie chart for the fit of the combined data H_2O (15 K): H_2O (75 K) with proportions 10:90 and without adding Gaussian noise to the spectrum. Panel e: same as panel d, but after adding Gaussian noise to the spectrum ($\sigma = 0.01$).

Fig. 12. Same as Fig. 8, but for H₂O and CO:CH₃CH₂OH at 15 K.

suggest that the non-detection of warm CH₃OH toward Elias 29 and other sources in the Ophiuchus molecular cloud is due to the absence of a hot-core-like region in the inner envelope close to the protostar or the due to higher temperatures in the precursor environments, as also discussed in the case of Corona Australis (Lindberg et al. 2014).

4.1. Continuum and optical depth calculation, and silicate feature extraction

In the present paper, the continuum SED of Elias 29 was determined with a blackbody function shortwards of 4.2 μm and a low-order polynomial function between 4.1 μm and 30 μm. In the former case, the spectrum between the ranges 2.5–2.8 μm and 3.9–4.2 μm is free of ice features and was therefore used to constrain the fitting. In this case, a temperature of 717 K provided the best match with the observations. Longwards of 4.1 μm, a third-order polynomial fitting passing through spectral ranges without ice features (4.1–4.25 μm, 4.3–4.5 μm, 5.3–5.5 μm and 25–30 μm) has been adopted. Pontoppidan et al. (2005) mention that although this method has some uncertainty involved, the shapes of the bands are not affected, even though the total optical depth is uncertain by up to 20%. The top panel of Fig. 15 shows the result of this fit and other calculations in the literature. The fits from Boogert et al. (2000) and Rocha & Pilling (2015) agree better with the continuum SED calculated in the

present paper, whereas the fit from Boogert et al. (2002) does not match the observations.

Once the continuum SED is determined, the observational optical depth is calculated with Eq. (3), and the result is shown in the middle panel of Fig. 15. The continuum methods used in this paper and Boogert et al. (2000) result in similar optical depths. The results from the two methods are shown in the bottom panel of Fig. 15 and differ by less than 10% in the range 2.5–30 μm. A similar variation is reported by van Breemen et al. (2011) for the band at 9.8 μm, when different continuum models are assumed for the objects StRs 164 and SSTc2d_J163346.2–242753. Most importantly, these latter authors concluded that the shape of the 9.8 μm bands remains virtually unaffected, which is in agreement with the result shown in this work. On the other hand, the effects of alterations in the shape of the silicate band at 18 μm are much less constrained.

The silicate absorption in the spectrum of Elias 29 was removed following the procedure described in Sect. 2.2. From the top panel of Fig. 16, we see that the GCS 3 silicate feature closely resembles the absorption band seen toward Elias 29, suggesting they could have similar mineralogy. In this particular case, the modification of the 9.7 μm band shape is minimal as can be noted in the result of the silicate extraction in the bottom panel of Fig. 16. The major difference between the two methods is only the absorption excess at 8.8 μm, which is discussed in Sect. 2.2. With the synthetic silicate band, this leftover residual is subtracted. Figure 16a also shows an absorption excess in

Fig. 13. Spectral decomposition of the H_2O and $\text{CO}:\text{CH}_3\text{OH}:\text{CH}_3\text{CH}_2\text{OH}$ constructed mixture containing spurious features mimicked with hydrogen lines and naphthalene bands.

the range 11–18 μm between the broad silicate bands at 9.8 μm and 18 μm that is clearly seen in the spectrum where the silicate feature is subtracted (Fig. 16b). Because of the uncertainties in the continuum and in the silicate band profile at 18 μm , some effect could be added to this band with silicate subtraction. However, Boogert et al. (2008) and Zasowski et al. (2009) convincingly show that this entire feature is associated with the H_2O ice libration mode.

4.2. Spectral decomposition

The Methods 1–3 in Table 2 were used to decompose the IR spectrum of Elias 29. The initial guess adopted in this study is composed of the pure ice samples H_2O , CO , CO_2 , and NH_3 . After selecting the best candidates for the solution, the entire sample is combined into groups of seven components. The best global minimum was found in Method 3, and the result is shown in Figs. 17a–f. In panel a, the spectral decomposition in the range 2.5–20 μm highlights the fitting of the most prominent bands, whereas the panels b–f show a zoom into small spectral ranges and the contributions of shallow features. The spectral features of a-C(:H) taken from Alata et al. (2014) are also shown in these panels to illustrate potential contributions of carbonaceous grains as discussed by Jones (2012, 2016).

The global minimum solution in the Elias 29 spectral analysis is given by the combination of pure and mixed ice samples. The dominant component in this broadband analysis is $\text{H}_2\text{O}:\text{NH}_3:\text{CO}_2:\text{CH}_4$ (10:1:1:1) processed by heavy ions at 35 K, that simulates the effect of cosmic rays in the ice mantle. This sample has significant contributions at 3 μm , 6 μm , 11–18 μm , and 15 μm . The second prominent component is CO_2 with vibrational modes at 4.27 μm and 15 μm . Pure and mixed H_2O components also contribute about 20% of the remaining absorption at 3 μm . The shallow features observed in this spectral decomposition are CO , NH_3 , CH_4 , and $\text{CH}_3\text{CH}_2\text{OH}$.

An old-standing problem in the analysis of mid-IR spectra of YSOs and background sources is that the band at 6 μm is deeper than what is expected from the 3 μm and 13 μm H_2O ice bands (e.g. Cox 1989; Gibb et al. 2000; Boogert et al. 2008, 2011; Bottinelli et al. 2010). A potential explanation for the absorption excess at 6 μm is the presence of organic refractory material (Gibb & Whittet 2002), although this hypothesis lacks independent spectroscopic confirmation (Boogert et al. 2015).

Another open question in this context is the correct carrier of the band at 6.85 μm . Although the most likely candidate is NH_4^+ (Schutte & Khanna 2003), a convincing profile fit is lacking, as pointed out by Boogert et al. (2015).

In the present paper, the global optimisation with the ENIIGMA tool provided a good fit of the absorption bands observed in the Elias 29 spectrum in the range 2.5–20 μm . The feature at 3 μm associated with the O–H vibrational mode is fitted by multiple ice samples containing water. However, there is an absorption excess in the Elias 29 spectrum that is not fitted with the ice samples in this case (Fig. 17b). One of the explanations for this excess is the absorption caused by ammonia hydrates (e.g. Knacke et al. 1982; Dartois & d’Hendecourt 2001). Another cause is attributed to the light scattering due to large grains (Smith et al. 1989). Jones (2016) also suggest that compounds containing the carbonyl functional group ($\text{C}=\text{O}$) present in the dust before or during the formation of the ice mantle could be a potential explanation for this absorption excess. Figure 17b shows that a-C(:H) could contribute to the large absorption excess around 3.5 μm , although the absorption features at 6.85 μm and 7.24 μm lack a visual match with the observations (Fig. 17e). Because of the low spectral resolution, this does not rule out the presence of a small fraction of a-C(:H) in the dust grains toward Elias 29. However, Boogert et al. (2000) argue that for the case of Elias 29, the scattered light due to large grains is the most likely explanation because of the close match of this red wing obtained with large icy grains.

The CO_2 stretching component at 4.27 μm (Fig. 17c) is found to be a combination of pure and mixed CO_2 . However, the peak optical depth in the fit does not match with the observations because it is also constrained by the CO_2 bending mode at 15.1 μm . Adding more CO_2 to the feature at 4.27 μm would lead to absorption excess at 15.1 μm . On the other hand, the CO ice band at 4.67 μm (Fig. 17d) is reproduced in this study by three components, namely pure CO ice, $\text{CO}:\text{CH}_3\text{OH}$, and CO formed with the cosmic-ray processing of $\text{H}_2\text{O}:\text{NH}_3:\text{CO}_2:\text{CH}_4$ (10:1:1:1) at 35 K, which is in agreement with the literature. Previous studies of the structure of the CO ice band (e.g. Pontoppidan et al. 2003; Thi et al. 2006; Perotti et al. 2020) showed that this band is well fitted by three Gaussian components or a combination of laboratory data and analytical functions. Such components are associated with pure CO , and CO mixed in polar ice matrices (e.g. H_2O and CH_3OH ; Cuppen et al. 2011).

The broadband in the range 5.5–8.0 μm (Fig. 17e) is well fitted by combining multiple components. In particular, two ice samples processed by UV ($\text{H}_2\text{O}:\text{NH}_3:\text{CH}_3\text{OH}:\text{CO}:\text{CO}_2$; Muñoz Caro & Schutte 2003) and cosmic rays ($\text{H}_2\text{O}:\text{NH}_3:\text{CO}_2:\text{CH}_4$ Pilling et al. 2010) significantly contribute to fit the Elias 29 spectrum. The ammonium ion (NH_4^+) is a common byproduct of the ice processing in these two cases, and is one of the proposed carriers of the band at 6.85 μm . The significant contribution of these two ice samples to the spectrum of Elias 29 might indicate that the chemical evolution of this YSO is mostly induced by energetic processing of the ice mantles.

In the presented spectral analysis, $\text{CH}_3\text{CH}_2\text{OH}$ mixed in H_2O ice is a likely carrier of absorption bands at 7.5 μm and 9.6 μm (Figs. 17e,f). Although the ethanol bands cannot be resolved at 7.5 μm with ISO observations, the band at 9.6 μm could indicate a tentative detection of this complex organic molecule toward Elias 29. This result also shows the contribution of methanol ice at 9.75 μm . Most importantly, this result shows that the band between 9.5 and 10 μm might not

Fig. 14. Spectral decomposition of the synthetic ice spectrum composed of eight components of pure ice samples (black line) and Gaussian noise ($\sigma = 0.01$). *Panel a*: synthetic spectrum along with the silicate band from 2.5–20 μm . *Panel b*: synthetic spectrum without the silicate feature and residual plot. The ENIIGMA fit is shown by the dashed green line and the ice components by the other colours. *Panels c–e*: small spectral ranges highlighting the contribution of shallow bands.

only be associated with methanol as suggested previously (e.g. Boogert et al. 2008; Bottinelli et al. 2010), but might be due to the presence of other alcohols in ices as discussed by Jones (2016). Consequently, the reduced amount of methanol in ices would contribute to partially solve the conundrum of the low gas-to-ice ratio observed toward star-forming regions ($\sim 10^{-4}$; Öberg et al. 2009; Perotti et al. 2020, 2021), as also conjectured by Jones (2016). However, we stress that high-resolution spectral data are required to unambiguously detect complex molecules in ices beyond methanol.

4.3. Recurrence analysis of the spectral decomposition

Because the ice spectral features vary with the physical and chemical environment, that is parameters such as composition, temperature, and radiation field, the fits of observational IR spectra of YSOs are often degenerate. The recurrence pie chart (Sect. 2.7) is used to verify the uniqueness of the solution found with the ENIIGMA and the result is displayed in Fig. 18.

This analysis shows that samples 1–7 have recurrences above 75%, and are also the components providing the global minimum solution as seen in Fig. 17. Samples 1–3 have a recurrence of 100%, and thus are solutions that cannot be replaced; additionally, they provide a significant contribution to the strong absorption features in the Elias 29 spectrum. The CO sample has a recurrence of 93.8% because the C–O stretching mode at 4.67 μm can be combined with other solutions containing CO, such as sample 1 (CO formed after processing; see Sect. 4.2) and samples 7 and 9. The recurrence of pure H₂O (87.5%) indicates that this component is necessary for the global minimum solution, but is not crucial. Other samples containing H₂O (e.g. 6, 9 and 10) can also be adopted to provide a good fit of the water absorption bands. Sample 6 is the only component containing CH₃CH₂OH. However, because this sample is dominated by water ice, which can be replaced by another component, its recurrence is 78.1%. Sample 7 has a recurrence of 75% and is important to fit the red wing of the absorption band at 4.67 μm .

Fig. 15. Continuum determination and optical depth calculation of the Elias 29 spectrum. *Top*: blue and red dashed lines show the black-body and polynomial continuum calculation used in this paper, respectively. The grey dotted line shows the continuum determination from Boogert et al. (2000), which is similar to the continuum in the present paper. The dotted magenta line is the continuum calculation with radiative transfer modelling from Boogert et al. (2002). The green solid line is the continuum calculation with radiative transfer modelling from Rocha & Pilling (2015). In both cases, the silicate feature is redshifted in the models compared with the observations. *Middle*: grey line shows the optical depth scale of Elias 29 spectrum adopting the continuum from Boogert et al. (2000), whereas the blue and red lines show the same by respectively adopting blackbody and polynomial continuum derived in this paper. *Bottom*: difference between optical depth estimates from this paper and those in Boogert et al. (2000).

Samples 8–11 are less recurrent (37.7–68.8%), although still inside the 3σ confidence interval. In particular, sample 8 contains a strong NH_4^+ feature that could replace component 2 in the fits. However, it is not part of the global minimum solution because the bandwidth at $6.85\ \mu\text{m}$ of component 8 is lower than in component 2, which provided a good match with the observational spectrum. Samples 9 and 10 provide a less accurate fit than components 1, 5, and 6. The less recurrent component, HCOOH, has been tentatively detected towards Elias 29. Despite formic acid not being part of the global minimum solution, it is not excluded as a potential carrier to fit the Elias 29 spectrum.

4.4. Ice column density

Ice bands can be blended in the IR spectrum, and therefore the selection of clear absorption features is essential to providing accurate quantification of the ice column densities. With the spectral decomposition used in this paper, the bands associated with different molecules were isolated, and their ice column densities calculated using Eq. (9) and the band strengths listed in Table 1. The derived values are shown in Table 3 and are compared with previous calculations by Boogert et al. (2000). The lower and upper bounds are calculated from the 3σ confidence interval analysis shown in Fig. 19. Table 3 also shows the median ice column densities calculated from the histograms shown in Appendix B. These values take into account all solutions found by the ENIIGMA tool inside 3σ confidence interval.

Fig. 16. Silicate extraction procedure in the Elias 29 spectrum. *Top*: black line shows the optical depth scale of the Elias 29 spectrum calculated in this paper. Red and blue lines show, respectively, the scaled CGS 3 and synthetic silicate bands at $9.7\ \mu\text{m}$ and $18\ \mu\text{m}$. *Bottom*: red and blue lines show the Elias 29 mid-IR optical depth after silicate band extractions with scaled CGS 3 and synthetic silicate profiles, respectively.

Water ice has multiple bands detected in the Elias 29 spectrum. In this paper, the water libration mode at $14\ \mu\text{m}$ is used in the calculation as it is easily separated from the CO_2 vibrational mode at $15\ \mu\text{m}$. The bands at $3\ \mu\text{m}$ and $6\ \mu\text{m}$ are blended with strong features of some molecules such as CH_3OH , HCOOH and processed ice, which makes the correct extraction of the water band difficult. In the cases of CO_2 and CO , the respective bands at $4.27\ \mu\text{m}$ and $4.67\ \mu\text{m}$ were used to derive the column densities. However, the pure and mixed CO_2 components found in the decomposition were not sufficient to reproduce the observed absorption band, and the N_{CO_2} ice column density in this paper is 28% lower than that found by Boogert et al. (2000). On the other hand, the N_{CO} values match well within the uncertainties. The ice column densities of calculated in this paper for H_2CO ($5.83\ \mu\text{m}$), CH_3OH ($6.83\ \mu\text{m}$), CH_4 ($7.67\ \mu\text{m}$), and NH_4^+ ($6.85\ \mu\text{m}$) are also in agreement with the findings of Boogert et al. (2000).

From the ENIIGMA fits, $\text{CH}_3\text{CH}_2\text{OH}$ contributes to the absorption band at $9.6\ \mu\text{m}$, and has abundances with respect to H_2O and CH_3OH ices of 0.25% and 10%, respectively. Other molecules such as HCOOH , OCS and OCN^- were not found in this work.

5. Limitations

The methodology adopted in the ENIIGMA fitting tool seems to work well to decompose ice bands in the spectra of YSOs. However, we mention in this section the limitations of this tool, and list opportunities for further development.

First. A poorly subtracted continuum or silicate band would add spurious features to the spectrum that could bias the ENIIGMA spectral decomposition. For example, due to inherent uncertainties in the broad $18\ \mu\text{m}$ amorphous silicate band subtraction, the shape of the H_2O libration and CO_2 bands around $13\ \mu\text{m}$ could be affected. We therefore recommend fitting other water and carbon dioxide absorption bands when possible to increase the reliability of the spectral decomposition method. We highlight that the approaches included in the

Fig. 17. Spectral decomposition of the mid-IR spectrum of Elias 29 protostar. *a*: optical depth scale of the observed spectrum between 2.5 and 20 μm , and the seven ice components used to fit the observed spectrum. The best fit is shown by the dashed green line. The residual spectrum is shown below this panel. *b–f*: zoom onto short spectral ranges highlighting the contribution of small features. In these plots, the spectral features of amorphous carbon, a-C(:H) , that were not used in the spectral decomposition are shown by the dashed grey lines above the Elias 29 spectrum and the ice data components. This illustrates a potential contribution of features other than vibrational modes of ice species. *Panel b*: *L*-band red wing absorption excess that is not entirely reproduced by the ice components (see Sect. 4.2). *Panels c and d*: CO_2 and CO features. *Panel e*: important contributions of UV and cosmic-ray processed samples. Finally, *panel f* highlights the contribution of the broad NH_3 ice feature at 9 μm , $\text{CH}_3\text{CH}_2\text{OH}$ at 9.6 μm , and CH_3OH at 9.8 μm .

ENIIGMA fitting tool are simplified techniques that have been successful while dealing with IR observations containing ice bands (e.g. Boogert et al. 2008; Bottinelli et al. 2010). Nevertheless, more accurate approaches than the ENIIGMA package could be adopted, instead. For example, a separate subtraction of silicate and other dust features could be performed with other computational tools (e.g. THEMIS code; Jones et al. 2017), and the

silicate-subtracted spectrum decomposed with the ENIIGMA fitting tool.

Second. The solutions found with the ENIIGMA fitting tool rely on the amount and diversity of data stored in the ENIIGMA database, which is flexible and can be enlarged or reduced. While adding data to the database, we recommend caution, and

1. $\text{H}_2\text{O}:\text{NH}_3:\text{CO}_2:\text{CH}_4$ (10:1:1:1) at 35K [CR]: 100 %
2. $\text{H}_2\text{O}:\text{NH}_3:\text{CH}_3\text{OH}:\text{CO}:\text{CO}_2$ (2:1:1:1:1) at 12K [UV]: 100 %
3. CO_2 at 12K: 100 %
4. CO: 93.8 %
5. H_2O at 15K: 87.5 %
6. $\text{H}_2\text{O}:\text{CH}_3\text{CH}_2\text{OH}$ (20:1) at 15K: 78.1 %
7. $\text{CO}:\text{CH}_3\text{OH}$ (1:1) at 10K: 75 %
8. Residue (SK03: high T): 68.8 %
9. $\text{H}_2\text{O}:\text{NH}_3:\text{CO}$ (10:0.6:0.4) at 13K: 56.2 %
10. $\text{H}_2\text{O}:\text{NH}_3$ (1:0.5) at 13K: 53.1 %
11. HCOOH at 14K: 37.8 %

Fig. 18. Recurrence pie chart of the IR ice data contributing to solutions inside 3σ confidence interval ($\Delta\chi^2 \leq 18.5$). The size of each piece is proportional to the contribution of each sample.

encourage the addition of properly baselined laboratory data, as inaccurate baselines would bias the results as well.

Third. Experimental works show that both the temperature and chemical environment change the shape and peak position of ice bands (e.g. Schutte & Khanna 2003; Dawes et al. 2016). This imposes an intrinsic caveat to the methodology adopted in the ENIIGMA fitting tool, namely the effects of intermolecular interactions within the ice are limited to the mixtures present in the ENIIGMA database. It is worth noting, however, that to the best of our knowledge, there is no analytical method that systematically explains such a change in the ice bands with the chemical environment. Works by Bonfim & Pilling (2018) started addressing this issue using quantum chemistry calculations, where an average dielectric constant simulates a generic mixture. One way to overcome this limitation is to add reliable experimental data of icy mixtures to the database.

Fourth. The current version of the ENIIGMA fitting tool does not account for changes in the shapes of the spectral bands caused by the geometry and properties of the grain where the ice is formed in interstellar conditions.

6. Summary

In this paper, the ENIIGMA fitting tool is introduced as a new approach to perform spectral decomposition of ice features in the IR using a genetic modelling algorithm. Python functions to perform continuum calculation and silicate removal in YSO

Fig. 19. Two-dimensional $\Delta\chi^2$ maps in grey scale showing the correlation among the linear combination coefficients of seven ice laboratory data used to decompose the Elias 29 spectrum. The green, yellow, and red contours indicate 1– 3σ confidence intervals, respectively. For the seven components, they are defined for $\Delta\chi^2$ equal to 8.38, 12.02, and 18.48.

spectra before the decomposition are also available in the tool. We also added a post-processing module dedicated to statistical analysis of the solutions to the capabilities of the ENIIGMA tool. Compiling 103 laboratory ice spectra split into pure and mixed, as well as thermally or energetically processed ices, we created an internal database for use in the spectral analysis.

In order to test all capabilities of the ENIIGMA fitting tool, fully blind and non-blind tests were performed on known species. We applied the statistical module to derive the confidence intervals of coefficients in the global minimum solution. Moreover, the code was able to successfully decompose a synthetic ice spectrum containing different fractions of chemical species with respect to water, as well as a constructed mixture containing spurious absorption and emission features.

A final test on a real source showed that the ENIIGMA successfully decomposed the Elias 29 spectrum with seven components, which correspond to pure and mixed ice samples. From this analysis, multiple molecules were identified, including a tentative detection of $\text{CH}_3\text{CH}_2\text{OH}$ at $9.6\mu\text{m}$. Moreover, the ice column densities and their lower and upper limits derived with the ENIIGMA tool from the global minimum solution and histogram analysis are in agreement with the previous estimations in the literature.

The tests performed in the present study with synthetic and observational ice spectra show that the ENIIGMA fitting tool can identify both strong and weak absorption features associated

Table 3. Ice column densities towards Elias 29.

Specie	N_{ice} [10^{17} cm^{-2}] Boogert et al. (2000)	This paper: N_{ice} [10^{17} cm^{-2}]	
		Global minimum	Median (Histogram ^(b))
H ₂ O	34 (6)	33.11 ^{41.13} _{21.09}	35.42 ^{41.87} _{25.23}
NH ₃	<3.5	1.01 ^{1.47} _{0.75}	2.42 ^{3.88} _{0.43}
CO	1.7 (0.3)	1.55 ^{1.81} _{0.73}	1.42 ^{1.87} _{1.23}
CO ₂	6.7 (0.5)	5.22 ^{7.13} _{3.4}	4.42 ^{6.87} _{2.23}
CH ₄	<0.5	0.38 ^{0.51} _{0.17}	0.42 ^{1.12} _{0.02}
H ₂ CO	<0.6	0.38 ^{0.42} _{0.13}	0.27 ^{0.52} _{0.03}
CH ₃ OH	<1.5	0.86 ^{1.16} _{0.06}	1.70 ^{2.38} _{0.21}
CH ₃ CH ₂ OH	NA	0.08 ^{0.11} _{0.03}	0.12 ^{0.40} _{0.02}
NH ₄ ⁺	1.01 (0.3) ^(a)	1.15 ^{0.72} _{0.3}	1.02 ^{3.12} _{0.23}
HCOOH	<0.3	NA	0.55 ^{0.91} _{0.13}
OCS	<0.015	NA	NA
OCN ⁻	<0.067	NA	NA

Notes. Tentative detections in the literature are shown as 3σ upper limits and non-detections as “NA”. ^(a)Boogert et al. (2008). ^(b)From the histograms shown in Appendix B.

with different molecules. In this regard, it will be a useful toolbox for the analysis of data provided by future near- and mid-IR observations of YSOs, such as the data that will be provided by the *James Webb* Space Telescope.

Acknowledgements. We thank the referee, Anthony Jones, for his constructive comments and suggestions that significantly improved this paper. This work benefited from support by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program through ERC Consolidator Grant “S4F” (grant agreement No 646908) to KJ. The research of LEK is supported by a research grant (No 19127) from VILLUM FONDEN. WRMR also thanks Dr. Ahmed Gad for the fruitful discussions about the genetic algorithm principles.

References

- Acharyya, K., Fuchs, G. W., Fraser, H. J., van Dishoeck, E. F., & Linnartz, H. 2007, *A&A*, **466**, 1005
- Alata, L., Cruz-Díaz, G. A., Muñoz Caro, G. M., & Dartois, E. 2014, *A&A*, **569**, A119
- Artur de la Villarmois, E., Jørgensen, J. K., Kristensen, L. E., et al. 2019, *A&A*, **626**, A71
- Avni, Y., & Bahcall, J. N. 1980, *ApJ*, **235**, 694
- Baier, A., Kerschbaum, F., & Lebzelter, T. 2010, *A&A*, **516**, A45
- Bergin, E. A., Melnick, G. J., Gerakines, P. A., Neufeld, D. A., & Whittet, D. C. B. 2005, *ApJ*, **627**, L33
- Bisschop, S. E., Fuchs, G. W., Boogert, A. C. A., van Dishoeck, E. F., & Linnartz, H. 2007, *A&A*, **470**, 749
- Blickle, T., & Thiele, L. 1996, *Evol. Comput.*, **4**, 361
- Bonfim, V. S., & Pilling, S. 2018, in *IAU Symposium*, eds. M. Cunningham, T. Millar, & Y. Aikawa, 332, 346
- Boogert, A. C. A., & Ehrenfreund, P. 2004, in *Astrophysics of Dust*, eds. A. N. Witt, G. C. Clayton, & B. T. Draine, *ASP Conf. Ser.*, **309**, 547
- Boogert, A. C. A., Tielens, A. G. G. M., Ceccarelli, C., et al. 2000, *A&A*, **360**, 683
- Boogert, A. C. A., Hogerheijde, M. R., Ceccarelli, C., et al. 2002, *ApJ*, **570**, 708
- Boogert, A. C. A., Pontoppidan, K. M., Knez, C., et al. 2008, *ApJ*, **678**, 985
- Boogert, A. C. A., Huard, T. L., Cook, A. M., et al. 2011, *ApJ*, **729**, 92
- Boogert, A. C. A., Chiar, J. E., Knez, C., et al. 2013, *ApJ*, **777**, 73
- Boogert, A. C. A., Gerakines, P. A., & Whittet, D. C. B. 2015, *ARA&A*, **53**, 541
- Bossa, J. B., Theule, P., Duvernay, F., & Chiavassa, T. 2009, *ApJ*, **707**, 1524
- Bottinelli, S., Boogert, A. C. A., Bouwman, J., et al. 2010, *ApJ*, **718**, 1100
- Bouilloud, M., Fray, N., Bénilan, Y., et al. 2015, *MNRAS*, **451**, 2145
- Brooke, T. Y., Sellgren, K., & Geballe, T. R. 1999, *ApJ*, **517**, 883
- Charbonneau, P. 1995, *ApJS*, **101**, 309
- Chiar, J. E., Adamson, A. J., Kerr, T. H., & Whittet, D. C. B. 1995, *ApJ*, **455**, 234
- Chiar, J. E., Tielens, A. G. G. M., Whittet, D. C. B., et al. 2000, *ApJ*, **537**, 749
- Collings, M. P., Anderson, M. A., Chen, R., et al. 2004, *MNRAS*, **354**, 1133
- Cox, P. 1989, *A&A*, **225**, L1
- Cuppen, H. M., Penteado, E. M., Isokoski, K., van der Marel, N., & Linnartz, H. 2011, *MNRAS*, **417**, 2809
- Dartois, E., & d’Hendecourt, L. 2001, *A&A*, **365**, 144
- Dartois, E., d’Hendecourt, L., Thi, W., Pontoppidan, K. M., & van Dishoeck, E. F. 2002, *A&A*, **394**, 1057
- Dawes, A., Mason, N. J., & Fraser, H. J. 2016, *Phys. Chem. Chem. Phys. (Inc. Faraday Trans.)*, **18**, 1245
- D’Hendecourt, L. B., & Allamandola, L. J. 1986, *A&AS*, **64**, 453
- Fuchs, G. W., Cuppen, H. M., Ioppolo, S., et al. 2009, *A&A*, **505**, 629
- Gerakines, P. A., Schutte, W. A., Greenberg, J. M., & van Dishoeck, E. F. 1995, *A&A*, **296**, 810
- Gerakines, P. A., Schutte, W. A., & Ehrenfreund, P. 1996, *A&A*, **312**, 289
- Gibb, E. L., & Whittet, D. C. B. 2002, *ApJ*, **566**, L113
- Gibb, E. L., Whittet, D. C. B., Schutte, W. A., et al. 2000, *ApJ*, **536**, 347
- Gibb, E. L., Whittet, D. C. B., Boogert, A. C. A., & Tielens, A. G. G. M. 2004, *ApJS*, **151**, 35
- Greenberg, J. M., & D’Hendecourt, L. B. 1985, in *NATO Advanced Science Institutes (ASI) Series C*, eds. J. Klinger, D. Benest, A. Dollfus, & R. Smoluchowski, 156, 185
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Harrison, T. E., & Hamilton, R. T. 2015, *AJ*, **150**, 142
- Hetem, A., & Gregorio-Hetem, J. 2007, *MNRAS*, **382**, 1707
- Hinterding, R. 1995, *Proceedings of 1995 IEEE International Conference on Evolutionary Computation*, 1, 384
- Holland, J. H. 1975, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence* (Ann Arbor: University of Michigan Press)
- Hudgins, D. M., & Sandford, S. A. 1998, *J. Phys. Chem. A*, **102**, 329
- Hudgins, D. M., Sandford, S. A., Allamandola, L. J., & Tielens, A. G. G. M. 1993, *ApJS*, **86**, 713
- Hudson, R. L., & Moore, M. H. 2004, *Icarus*, **172**, 466
- Ioppolo, S., Fedoseev, G., Chuang, K. J., et al. 2021, *Nat. Astron.*, **5**, 197
- Ishii, M., Nagata, T., Chrysostomou, A., & Hough, J. H. 2002, *AJ*, **124**, 2790
- Jones, A. P. 2012, *A&A*, **540**, A1
- Jones, A. P. 2016, *R. Soc. Open Sci.*, **3**, 160224
- Jones, A. P., Fanciullo, L., Köhler, M., et al. 2013, *A&A*, **558**, A62
- Jones, A. P., Köhler, M., Ysard, N., Bocchio, M., & Verstraete, L. 2017, *A&A*, **602**, A46
- Keane, J. V., Tielens, A. G. G. M., Boogert, A. C. A., Schutte, W. A., & Whittet, D. C. B. 2001, *A&A*, **376**, 254

- Kemper, F., Vriend, W. J., & Tielens, A. G. G. M. 2004, *ApJ*, **609**, 826
- Knacke, R. F., McCorkle, S., Puetter, R. C., Erickson, E. F., & Kraetschmer, W. 1982, *ApJ*, **260**, 141
- Köhler, M., Ysard, N., & Jones, A. P. 2015, *A&A*, **579**, A15
- Koza, J. R. 1992, *Genetic programming: On the Programming of Computers by Means of Natural Selection* (Cambridge: The MIT Press)
- Lindberg, J. E., Jørgensen, J. K., Brinch, C., et al. 2014, *A&A*, **566**, A74
- Linnartz, H., Ioppolo, S., & Fedoseev, G. 2015, *Int. Rev. Phys. Chem.*, **34**, 205
- McClure, M. 2009, *ApJ*, **693**, L81
- Merrill, K. M., Russell, R. W., & Soifer, B. T. 1976, *ApJ*, **207**, 763
- Moultaka, J., Eckart, A., Viehmann, T., et al. 2004, in *The Dense Interstellar Medium in Galaxies*, eds. S. Pfalzner, C. Kramer, C. Staubmeier, & A. Heithausen, 91, 295
- Muñoz Caro, G. M., & Schutte, W. A. 2003, *A&A*, **412**, 121
- Noble, J. A., Fraser, H. J., Aikawa, Y., Pontoppidan, K. M., & Sakon, I. 2013, *ApJ*, **775**, 85
- Noble, J. A., Fraser, H. J., Pontoppidan, K. M., & Craigmiles, A. M. 2017, *MNRAS*, **467**, 4753
- Öberg, K. I., Boogert, A. C. A., Pontoppidan, K. M., et al. 2008, *ApJ*, **678**, 1032
- Öberg, K. I., Bottinelli, S., & van Dishoeck, E. F. 2009, *A&A*, **494**, L13
- Öberg, K. I., Boogert, A. C. A., Pontoppidan, K. M., et al. 2011, *ApJ*, **740**, 109
- Ortiz-León, G. N., Loinard, L., Dzib, S. A., et al. 2018, *ApJ*, **869**, L33
- Ossenkopf, V., & Henning, T. 1994, *A&A*, **291**, 943
- Oya, Y., López-Sepulcre, A., Sakai, N., et al. 2019, *ApJ*, **881**, 112
- Palumbo, M. E., Baratta, G. A., Brucato, J. R., et al. 1998, *A&A*, **334**, 247
- Peng, S., Xu, Q., Ling, X. B., et al. 2003, *FEBS Lett.*, **555**, 358
- Perone, C. S. 2009, *Pyevolve: A Python Open-Source Framework for Genetic Algorithms* (New York: Association for Computing Machinery)
- Perotti, G., Rocha, W. R. M., Jørgensen, J. K., et al. 2020, *A&A*, **643**, A48
- Perotti, G., Jørgensen, J. K., Fraser, H. J., et al. 2021, *A&A*, **650**, A168
- Pilling, S., Seperuelo Duarte, E., da Silveira, E. F., et al. 2010, *A&A*, **509**, A87
- Pontoppidan, K. M., Fraser, H. J., Dartois, E., et al. 2003, *A&A*, **408**, 981
- Pontoppidan, K. M., van Dishoeck, E. F., & Dartois, E. 2004, *A&A*, **426**, 925
- Pontoppidan, K. M., Dullemond, C. P., van Dishoeck, E. F., et al. 2005, *ApJ*, **622**, 463
- Pontoppidan, K. M., Boogert, A. C. A., Fraser, H. J., et al. 2008, *ApJ*, **678**, 1005
- Potapov, A., Mutschke, H., Seiber, P., Henning, T., & Jäger, C. 2018, *ApJ*, **861**, 84
- Potapov, A., Bouwman, J., Jäger, C., & Henning, T. 2021, *Nat. Astron.*, **5**, 78
- Poteet, C. A., Whittet, D. C. B., & Draine, B. T. 2015, *ApJ*, **801**, 110
- Rachid, M. G., Pilling, S., Rocha, W. R. M., et al. 2020, *MNRAS*, **494**, 2396
- Reach, W. T., Faied, D., Rho, J., et al. 2009, *ApJ*, **690**, 683
- Rocha, W., & Pilling, S. 2014, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.*, **123**, 436
- Rocha, W. R. M., & Pilling, S. 2015, *ApJ*, **803**, 18
- Rocha, W. R. M., & Pilling, S. 2018, *MNRAS*, **478**, 5190
- Rocha, W. R. M., Pilling, S., de Barros, A. L. F., et al. 2017, *MNRAS*, **464**, 754
- Rocha, W., Pilling, S., Domaracka, A., Rothard, H., & Boduch, P. 2019, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.*, **228**, 117826
- Sack, N. J., & Baragiola, R. A. 1993, *Phys. Rev. B*, **48**, 9973
- Schutte, W. A., & Khanna, R. K. 2003, *A&A*, **398**, 1049
- Schutte, W. A., Allamandola, L. J., & Sandford, S. A. 1993, *Icarus*, **104**, 118
- Schutte, W. A., Tielens, A. G. G. M., Whittet, D. C. B., et al. 1996, *A&A*, **315**, L333
- Schutte, W. A., Boogert, A. C. A., Tielens, A. G. G. M., et al. 1999, *A&A*, **343**, 966
- Smith, R. G., Sellgren, K., & Tokunaga, A. T. 1989, *ApJ*, **344**, 413
- STScI Development Team 2013, Astrophysics Source Code Library [record ascl:**1303.023**]
- Suutarinen, A. 2015, PhD Thesis, The Open University, UK
- Szkody, P., Mukadam, A., Gänsicke, B. T., et al. 2010, *ApJ*, **716**, 1531
- Terwisscha van Scheltinga, J., Ligterink, N. F. W., Boogert, A. C. A., van Dishoeck, E. F., & Linnartz, H. 2018, *A&A*, **611**, A35
- Thi, W.-F., van Dishoeck, E. F., Dartois, E., et al. 2006, *A&A*, **449**, 251
- Thi, W.-F., Woitke, P., & Kamp, I. 2011, *MNRAS*, **412**, 711
- Tielens, A. G. G. M. 2008, *ARA&A*, **46**, 289
- Tielens, A. G. G. M., & Allamandola, L. J. 1987, in Physical Processes in Interstellar Clouds, eds. G. E. Morfill, & M. Scholer, *NATO Adv. Study Inst. (ASI) Ser. C*, **210**, 333
- van Breemen, J. M., Min, M., Chiar, J. E., et al. 2011, *A&A*, **526**, A152
- van Dishoeck, E. F., Wright, C. M., Cernicharo, J., et al. 1998, *ApJ*, **502**, L173
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Meth.*, **17**, 261
- Watanabe, N., & Kouchi, A. 2002, *ApJ*, **571**, L173
- Whittet, D. C. B., Cook, A. M., Chiar, J. E., et al. 2009, *ApJ*, **695**, 94
- Woitke, P., Min, M., Pinte, C., et al. 2016, *A&A*, **586**, A103
- Woitke, P., Kamp, I., Antonellini, S., et al. 2019, *PASP*, **131**, 064301
- Zasowski, G., Kemper, F., Watson, D. M., et al. 2009, *ApJ*, **694**, 459

Appendix A: List of ice laboratory data

We compiled a data set containing 102 laboratory data for astrophysical ices from different public online databases and divided them into four categories, namely (i) pure and non-processed ices – Table A.1, (ii) pure and thermally processed ices – Table A.2, (iii) organic residue samples – Table A.3, and (iv) mixed processed ices – Table A.4. In the last category, the mixtures are grouped into three subcategories, namely non-processed, thermally processed, and energetically processed.

These ice samples were formed inside a high-vacuum chamber ($P \sim 10^{-7} - 10^{-10}$ mbar) by the condensation of liquids or gases onto an IR transparent substrate (e.g. ZnSe, CsI) cooled to temperatures of around 10 K. In the cases where the ice samples were irradiated by UV, a hydrogen discharge lamp dominated by Ly- α emission with a flux of 10^{14} photons $\text{cm}^{-2} \text{s}^{-1}$ was used (Schutte & Khanna 2003; Muñoz Caro & Schutte 2003). Some samples were irradiated by heavy ions simulating the effect of cosmic rays in the ISM. In this case, different ions were used as a projectile with energies in the range 0.2–632 MeV.

Table A.1. Laboratory data of pure and non-processed ices used in this paper.

Label/Temp.	Database	Reference
H ₂ O (15 K)	NASA Ames	[1]
NH ₃ (10 K)	...	[2]
CH ₄ (10 K)	NASA Ames	[1]
CO (12 K)	UNIVAP	[3]
CO ₂ (12 K)	UNIVAP	[3]
H ₂ CO (10 K)	...	[2]
CH ₃ OH (10 K)	NASA Ames	[1]
HCOOH (15 K)	Leiden DB	[4]
CH ₃ CN (12 K)	UNIVAP	[3]
CH ₃ COOH (12 K)	UNIVAP	[3]
CH ₃ CHO (15 K)	Leiden DB	[5]
CH ₃ OCH ₃ (15 K)	Leiden DB	[5]
CH ₃ CH ₂ OH (15 K)	Leiden DB	[5]

Notes. [1] [Hudgins et al. \(1993\)](#), [2] [Gerakines et al. \(1996\)](#), [3] [Rocha & Pilling \(2014\)](#), [4] [Bisschop et al. \(2007\)](#), [5] [Terwisscha van Scheltinga et al. \(2018\)](#).

Table A.2. Laboratory data of pure and thermally processed ices used in this paper.

Label	Temperature (K)	Database	Reference
H ₂ O	40, 75, 100, 120, 140	NASA Ames	[1]
CH ₄	20, 30	NASA Ames	[1]
H ₂ CO	30, 70	...	[2]
CH ₃ OH	50, 75, 100, 120	NASA Ames	[1]
HCOOH	30, 60, 75, 90, 105	Leiden DB	[3]
CH ₃ CHO	30, 70, 90, 110, 120	Leiden DB	[4]
CH ₃ OCH ₃	30, 70, 90, 100	Leiden DB	[4]
CH ₃ CH ₂ OH	30, 70, 100, 120, 130, 140, 150	Leiden DB	[4]

Notes. [1] [Hudgins et al. \(1993\)](#), [2] [Gerakines et al. \(1996\)](#), [3] [Bisschop et al. \(2007\)](#), [4] [Terwisscha van Scheltinga et al. \(2018\)](#).

Table A.3. Laboratory data of organic residue.

Label/Fraction	ID	Temperature (K)	Database	Reference
HNCO:NH ₃ (1:1.2)	NH ₄ ⁺ heating	120	...	[1]
H ₂ O:CO ₂ :NH ₃ :O ₂ ^a (10:2:1.1:1)	NH ₄ ⁺ heating + UV	200	...	[1]

Notes. ^(a)The ice sample was firstly photolysed by UV and subsequently warmed to 200 K. [1] [Schutte & Khanna \(2003\)](#).

Table A.4. Laboratory data of mixed and processed ices (thermal or energetic) used in this paper.

Non-processed mixtures					
Label/Fraction	Temperature (K)	Ionizing agent	$\log_{10}(\text{Fluence})$	Database	Reference
H ₂ O:CH ₃ CH ₂ OH (20:1)	15	...	...	Leiden DB	[1]
H ₂ O:CH ₃ CHO (20:1)	16	...	...	Leiden DB	[1]
CO:CH ₃ OH (1:1)	15	...	...	...	[2]
Thermally-processed mixtures					
Label/Fraction	Temperature (K)	Ionizing agent	$\log_{10}(\text{Fluence})$	Database	Reference
H ₂ O:CH ₃ OH:CO:NH ₃ (100:50:1:1)	10–140 ^a	...	...	NASA Ames	[3]
CO:CH ₃ CH ₂ OH (20:1)	15, 30	...	...	Leiden DB	[1]
CO:CH ₃ CHO (20:1)	15, 30	...	...	Leiden DB	[1]
CO:CH ₃ OCH ₃ (20:1)	15, 30	...	...	Leiden DB	[1]
CO:CH ₃ OH:CH ₃ OCH ₃ (20:1)	15, 30	...	...	Leiden DB	[1]
CO:CH ₃ OH:CH ₃ CH ₂ OH (20:20:1)	15, 30	...	...	Leiden DB	[1]
Energetically processed mixtures					
Label/Fraction	Temperature (K)	Ionizing agent	$\log_{10}(\text{Fluence})^b$	Database	Reference
H ₂ O:NH ₃ (1:0.5)	13	40 MeV ⁵⁸ Ni ¹³⁺	0, 12, 13.2	UNIVAP	[4]
H ₂ O:NH ₃ :CO (1:0.6:0.4)	13	46 MeV ⁵⁸ Ni ¹³⁺	0, 12, 13.3	UNIVAP	[4]
H ₂ O:CH ₄ (1:0.6)	16	40 MeV ⁵⁸ Ni ¹¹⁺	0, 12, 13	UNIVAP	[4]
H ₂ O:CH ₄ (10:1)	16	40 MeV ⁵⁸ Ni ¹¹⁺	0, 12, 13	UNIVAP	[4]
H ₂ O:H ₂ CO:CH ₃ OH (100:0.2:0.8)	15	220 MeV ¹⁶ O ⁷⁺	0, 12, 13	UNIVAP	[4]
H ₂ O:HCOOH (1:1)	15	46 MeV ⁵⁸ Ni ¹¹⁺	0, 12, 13	UNIVAP	[4]
H ₂ O:NH ₃ :c-C ₆ H ₆ (1:0.3:0.7)	13	632 MeV ⁵⁸ Ni ²⁴⁺	0, 12.3, 13.5	UNIVAP	[4]
H ₂ O:CH ₃ OH (1:1)	15	40 MeV ⁵⁸ Ni ¹¹⁺	0, 12, 13	UNIVAP	[4]
H ₂ O:NH ₃ :CO ₂ :CH ₄ (10:1:1:1)	35	15.7 MeV ¹⁶ O ⁵⁺	0, 12, 13	UNIVAP	[4]
H ₂ O:NH ₃ :CO ₂ :CH ₄ (10:1:1:1)	72	15.7 MeV ¹⁶ O ⁵⁺	0, 12, 13	UNIVAP	[4]
CO:NH ₃ (1:1)	14	5.8 MeV ¹⁶ O ²⁺	0, 12.1, 13	UNIVAP	[5]
NH ₃ :CH ₃ OH (1:1)	14	0.2 MeV ⁵⁸ Ni ²⁴⁺	0, 12, 13	UNIVAP	[5]
H ₂ O:NH ₃ :CH ₃ OH:CO:CO ₂ (2:1:1:1:1)	12	UV (7.3–10.5 eV)	14	...	[6]

Notes. ^(a)Heating ramp: 10, 40, 80, 100, 120, 140 K. ^(b)Fluence unit: ions cm⁻². [1] Terwisscha van Scheltinga et al. (2018), [2] Cuppen et al. (2011), [3] Hudgins et al. (1993), [4] Rocha et al. (2017), [5] Rocha et al. (2019), [6] Muñoz Caro & Schutte (2003).

Appendix B: Histograms

All solutions inside a 3σ confidence interval out of 1100 combinations (11 components) are analysed with histograms (Figure B.1). The mean, lower, and upper limits of the column densities are derived in this analysis (see Table 3). The bin size is proportional to the column density variation and is calculated by the Freedman Diaconis Estimator, which is given by:

$$h = 2 \frac{IQR}{n^{1/3}}, \quad (\text{B.1})$$

where IQR is the interquartile range, and is robust to the outliers. In order to reduce the skewness in the distribution, the histograms are log-transformed, although some data still show tails in the distribution. To take into account that skewness factor, the mean and upper limits are calculated by `scipy.stats.skewnorm` available in the SciPy library (Virtanen et al. 2020). The skewness parameter is calculated, and the statistical quantities derived. The histograms

Fig. B.1. Log-transformed and median-centred histograms of the ice column densities derived from all solutions inside a 3σ confidence interval in the case of Elias 29.