UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS FÍSICAS



TESIS DOCTORAL

El origen del material orgánico complejo en precursores de sistemas planetarios

The origin of the complex organic material in precursors of planetary systems

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Andrés Megías Toledano

DIRECTORA

Izaskun Mayte Jiménez Serra

© Andrés Megías Toledano, 2025

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS

PROGRAMA DE DOCTORADO EN ASTROFÍSICA





TESIS DOCTORAL

EL ORIGEN DEL MATERIAL ORGÁNICO COMPLEJO EN PRECURSORES DE SISTEMAS PLANETARIOS

THE ORIGIN OF THE COMPLEX ORGANIC MATERIAL IN PRECURSORS OF PLANETARY SYSTEMS

AUTOR

ANDRÉS MEGÍAS TOLEDANO

DIRECTORA

IZASKUN MAYTE JIMÉNEZ SERRA

CENTRO DE ASTROBIOLOGÍA

FEBRERO DE 2025

Contents

1 Introduction **1**

1.1 Interstellar medium . . . 2

1.2 Low-mass star formation . . . 5

1.3 Chemistry in starless and prestellar cores . . . 7

1.3.1 Gas phase . . . 8

1.3.2 Solid phase . . . 9

1.4 Astrochemical observations of starless and prestellar cores . . . 10

1.5 Objectives of this thesis . . . 12

2 Methodology **15**

2.1 Molecular spectroscopy . . . 15

2.1.1 Rotational transitions . . . 16

2.1.2 Vibrational transitions . . . 19

2.2 Radiative transfer . . . 20

2.2.1 Local thermodynamical equilibrium . . . 22

2.2.2 Black body radiation . . . 24

2.2.3 Einstein coefficients . . . 25

2.2.4 Column densities . . . 28

2.2.5 MADCUBA and SLIM . . . 30

2.2.6 Non-local thermodynamical equilibrium . . . 32

2.2.7 RADEX Line Fitter . . . 34

2.2.8 Column densities for ice absorption spectra . . . 35

2.3 Astronomical observations . . . 37

2.3.1 Radio astronomical observations . . . 37

2.3.1.1 Antenna temperature . . . 37
2.3.1.2 Source size . . . 38
2.3.1.3 Main beam efficiency . . . 39
2.3.1.4 Radial velocity . . . 42
2.3.1.5 Signal noise . . . 42
2.3.1.6 Frequency switching . . . 43
2.3.2 Automated GILDAS/CLASS Pipeline . . . 44
2.3.3 Infrared absorption spectra . . . 47
2.4 Telescopes used in this thesis . . . 48
2.4.1 IRAM 30 m telescope . . . 48
2.4.2 James Webb Space Telescope . . . 49
2.5 Astrochemical models . . . 51
2.5.1 Reaction rate astrochemical codes . . . 51
2.5.2 UCLCHEM . . . 52

3 Chemical complexity in starless and prestellar cores 57
3.1 Introduction . . . 58
3.2 Observations . . . 60
3.3 Results . . . 63
3.3.1 Detected transitions for L1517B . . . 63
3.3.2 Molecular column densities and excitation temperatures . . . 65
3.3.2.1 LTE analysis . . . 65
3.3.2.2 Non-LTE analysis . . . 66
3.3.3 Molecular abundances . . . 68
3.4 Comparison with other cores . . . 69
3.4.1 Comparison with L1498 and L1544 . . . 69
3.4.2 Comparison with L1521E . . . 72
3.5 Modelling the formation of COMs in L1517B . . . 73
3.5.1 The model . . . 73
3.5.2 Comparing simulations and observations . . . 76
3.6 HC_3N / CH_3CN ratio across low-mass star formation . . . 77
3.7 Conclusions . . . 78

4 Predicting the composition of interstellar ices with machine learning 81
4.1 Introduction . . . 82
4.2 Methodology: the AICE tool . . . 84
4.2.1 Artificial neural networks . . . 85
4.2.2 Architecture used for AICE . . . 85
4.2.3 Training dataset . . . 87
4.2.3.1 Infrared molecular spectra . . . 88
4.2.3.2 Spectra processing and baseline removal . . . 90
4.2.3.3 Resampling and normalization . . . 90
4.2.3.4 Data augmentation using pure ice spectra . . . 91
4.2.3.5 Training and validation subsets . . . 92
4.3 Choosing the model hyperparameters . . . 93
4.4 Results . . . 94
4.4.1 Training and validation . . . 94
4.4.2 Testing with astronomical ice . . . 96
4.4.3 Comparison of the AICE predicted column densities . . . 98
4.4.4 Effects of band saturation . . . 99
4.4.5 Effects of changes in the spectral range . . . 101
4.5 Discussion . . . 102
4.6 Conclusions . . . 104

5 Astrochemical simulations of molecular clouds for the Ice Age project 105
5.1 Introduction . . . 106
5.2 The models . . . 107
5.2.1 UCLCHEM . . . 108
5.2.2 Physical parameters . . . 108
5.3 Results . . . 111
5.3.1 Static phase . . . 111
5.3.2 Collapse phase . . . 112
5.3.3 Best fit models . . . 113
5.3.4 Models with a lower temperature . . . 116
5.4 Discussion . . . 118

5.4.1 Formation of CH_3OH ice . . . 118
5.4.2 Conversion of CO into CO_2 . . . 118
5.4.3 Formation of hydrides in the ice: H_2O, NH_3 and CH_4 . . . 119
5.4.4 Uncertainties in the extinction toward NIR38 and J110621 . . . 120
5.5 Conclusions . . . 120

6 Discussion **123**

7 Conclusions and perspectives **129**
7.1 Conclusions . . . 129
7.1.1 Chapter 3: Chemical complexity in starless and prestellar cores . . . 129
7.1.2 Chapter 4: Composition of astronomical ices and machine learning . . . 130
7.1.3 Chapter 5: Astrochemical simulations of molecular clouds . . . 130
7.1.4 Additional conclusions . . . 131
7.2 Wrap-up . . . 131
7.3 Perspectives . . . 132
7.3.1 Maps of COM emission in the prestellar core L1544 . . . 132
7.3.2 Larger sample of starless cores . . . 133
7.3.3 Statistical sample of ice chemical composition with JWST . . . 133

Bibliography **134**

A Chapter 2 Appendix **157**
A.1 Automated GILDAS/CLASS Pipeline . . . 157
A.1.1 Configuration file . . . 157
A.1.2 Selection mode . . . 158
A.1.3 Line search and reduction . . . 158
A.2 Example chemical network . . . 160

B Chapter 3 Appendix **163**
B.1 Derivation of the methanol column density and its uncertainty with RADEX . 163
B.2 Propagation of uncertainties and upper limits . . . 166
B.3 Abundance ratio of cianoacetylene and acetonitrile . . . 169

C Chapter 4 Appendix **171**

C.1 Astronomical ice spectra 171

C.2 Artificial neural networks 173

C.3 Baseline removal 176

C.4 List of training dataset 179

C.5 Evolution of the training 184

C.6 Our augmented dataset 184

C.7 Propagation of uncertainties 186

C.8 Effect of band saturation 190

C.9 Retraining with a smaller spectral range 191

C.10 Automatic band integrator 193

Abstract

Starless and prestellar cores, located inside giant molecular clouds, are the very first stages in the formation of low-mass stars and of their subsequent planetary systems, like our Solar system. In this thesis, we explore the level of chemical complexity found in these objects – from their earlier stages in molecular clouds to starless and prestellar cores –, focusing on interstellar complex organic molecules (COMs). These molecules are relevant for the formation of prebiotic molecules, believed to be the precursors of life as we know it on Earth. We try to understand the relevance of the chemistry at these stages in the chemical inheritance of planetesimals and cometesimals to the final planetary system, given that complex organics such as amino acids have been found in bodies of our Solar system like meteorites, asteroid Ryugu and comet 67P/Churyumov-Gerasimenko. For this, we use radio astronomical observations, to detect gas-phase chemical species in the interstellar medium (ISM), and infrared absorption spectra, to observe solid-phase species in the icy layers of interstellar dust grains. To obtain a comprehensive view of how these COMs are formed under interstellar conditions, we also carry out the astrochemical modelling of these regions using two models, MONACO and UCLCHEM.

In this work, we first present new radio astronomical observations of the emission of COMs toward the starless core L1517B, and compare our results with other observations of the starless core L1498, the prestellar core L1544, and the starless core L1521E. We find that L1517B, believed to be very young, shows the lowest level of chemical complexity, suggesting a chemical evolution along the process of gravitational collapse of these cores, where chemical complexity increases with time. In this hypothetical scenario, simple nitrogen- and oxygen-bearing molecules would form first – such as acetonitrile (CH_3CN) or ketene (H_2CCO), followed by more complex molecules, especially containing oxygen, once the catastrophic depletion of carbon monoxide (CO) takes place, due to adsorption on dust grains. Finally, in the prestellar core phase, more complex compounds such as dimethyl ether (CH_3OCH_3) would be formed. However, the starless core L1521E presents relatively large COMs – such as dimethyl ether – despite its youth, suggesting that deeper observations of this type of objects must be performed to better characterize its level of chemical complexity and its possible evolutionary path.

The chemical modelling of COMs under the coldest interstellar conditions in starless/prestellar cores suggest that these species are mainly formed on the surface of dust grains to be subsequently injected into the gas phase by non-thermal chemical processes (as, e.g., chemical

reactive desorption). In order to confirm this scenario is therefore essential to constrain the presence and abundance of COMs in interstellar ices. With this goal in mind, we present a tool based in artificial neural networks called AICE (*Automatic Ice Composition Estimato*r), coded in Python programming language and trained on hundreds of laboratory experiments, coming from different experimental groups around the world and retrieved mainly through the Leiden Ice Database for Astrochemistry (LIDA). This tool allows the user to quickly predict the composition and temperature of astronomical ices from its infrared absorption spectra. The error in the validation step for the neural networks is relatively low (~3 % for the fractional composition and ~11 K for temperature). When testing the tool with data from James Webb space telescope (JWST) we find similar results to those previously obtained by other authors with other intense computational methods, but AICE's performance provides a much faster execution time (~0.1 s in a regular laptop), which constitutes an advantage with respect to other methods. Additionally, this model intrinsically learns the changes in the molecular band strengths due to the changes in chemical composition and temperature. AICE therefore represents a powerful tool to analyze new data from JWST, which is expected to deliver hundreds of ice absorption spectra in the near future.

The detail provided by recent ice absorption spectra measured with the JWST, requires a better understanding of the interstellar ice formation processes. Within the context of the chemical modelling group of the Ice Age collaboration, we have recently contributed with the modelling of the chemistry of the ice toward two background stars in the Chamaeleon I molecular cloud recently observed with JWST. To do this, we have used the UCLCHEM astrochemical software and we have compared the simulations with previous observations of ices from the Ice Age project. We conclude that methanol (CH_3OH) is mainly formed on the surface of dust grains via hydrogenation of carbon monoxide (CO). In addition, we also find that the models that best reproduce the observations are those with intermediate collapse time scales (~10^5 yr) and intermediate final densities (~$2 \cdot 10^5$ cm^{-3}).

From our work, we conclude that ice observations are crucial in order to better understand and constrain the chemistry of COMs in starless and prestellar cores. Because of this, new observations by JWST will play a key role, and our new AICE analysis tool could therefore be extremely useful for the fast and efficient analysis of such complex JWST datasets. Additionally, more gas-phase observations of starless/prestellar cores with deeper sensitivity are also needed to better characterize their chemical complexity in a deeper level, like it has been done for L1517, L1498, L1544 and, in a smaller extent, L1521E.

Resumen

Los núcleos sin estrella y preestelares, ubicados en el interior de grandes nubes moleculares, constituyen las primeras etapas en la formación de estrellas de baja masa y de sus posteriores sistemas planetarios, como nuestro sistema solar. En esta tesis exploramos el nivel de complejidad química encontrado en estos objetos –desde sus primeras etapas en nubes moleculares hasta núcleos sin estrella y preestelares–, centrándonos en moléculas orgánicas complejas interestelares (COMs). Estas son relevantes para la formación de moléculas prebióticas, que se cree que son los precursores de la vida tal como la conocemos en la Tierra. Nuestro objetivo es comprender la relevancia de la química en estas etapas iniciales de la herencia química desde los planetesimales y cometesimales hasta el sistema planetario final, dado que se han encontrado compuestos orgánicos complejos como los aminoácidos en cuerpos de nuestro sistema solar como meteoritos, el asteroide Ryugu y el cometa 67P/Churyumov-Gerasimenko. Para ello utilizamos observaciones radioastronómicas, que permiten detectar especies químicas en fase gaseosa en el medio interestelar (ISM), y espectros de absorción en el infrarrojo, para observar especies en fase sólida en las capas heladas de los granos de polvo interestelares. Para obtener una visión completa de cómo se forman estas COMs en condiciones interestelares, también hemos llevado a cabo el modelado astroquímico de estas regiones utilizando dos modelos, MONACO y UCLCHEM.

En este trabajo, presentamos por primera vez nuevas observaciones radioastronómicas de la emisión de COMs en el núcleo sin estrella L1517B, comparando nuestros resultados con otras observaciones del núcleo sin estrellas L1498, el núcleo preestelar L1544, y el núcleo sin estrella L1521E. Encontramos que L1517B, que se cree que es muy joven, posee el nivel más bajo de complejidad química, lo que sugiere una evolución química a lo largo del proceso de colapso gravitacional de estos núcleos, donde la complejidad química aumenta con el tiempo. En este escenario hipotético, primero se formarían moléculas simples que contienen nitrógeno y oxígeno, como el acetonitrilo (CH_3CN) o la cetena (H_2CCO), seguidas de moléculas más complejas, especialmente con átomos de oxígeno, una vez que se produce la depleción catastrófica del monóxido de carbono (CO), debido a la adsorción en los granos de polvo. De esta manera, en la fase del núcleo preestelar se formarían compuestos más complejos como el dimetil éter (CH_3OCH_3). Sin embargo, el núcleo sin estrellas L1521E presenta COM relativamente grandes –como el éter dimetílico– a pesar de su juventud, lo que sugiere que se deben realizar observaciones más profundas de este tipo de objetos para caracterizar mejor su rango de complejidad química y su posible camino evolutivo.

El modelado químico de COMs en las frías condiciones interestelares de los núcleos sin estrella / preeselares sugiere que estas especies se forman principalmente en la superficie de los granos de polvo para ser posteriormente inyectados en la fase gaseosa mediante procesos químicos no térmicos (como, por ejemplo, desorción reactiva química). Para confirmar este escenario es, por tanto, esencial determinar la presencia y abundancia de COMs en los

hielos interestelares. Con este objetivo en mente, presentamos una herramienta basada en redes neuronales artificiales llamada AICE (*Automatic Ice Composition Estimator*; *Estimador automático de composición de hielos*), programada en el lenguaje de programación Python y entrenada con cientos de experimentos de laboratorio, provenientes de diferentes grupos experimentales alrededor del mundo y obtenidos principalmente a través de la *Leiden Ice Database for Astrochemistry* (LIDA; *Base de datos de hielos para astroquímica de Leiden*). Esta herramienta, basada en redes neuronales artificiales, permite predecir rápidamente la composición y temperatura de hielos astronómicos a partir de sus espectros de absorción infrarroja. El error en la fase de de validación de las redes neuronales es relativamente bajo (~3 % para la composición fraccionaria y ~11 K para la temperatura). Al probar la herramienta con datos del telescopio espacial James Webb (JWST), obtenemos resultados similares a los ya estimados previamente por otros autores con métodos computacionales más intensos, pero AICE requiere un tiempo de ejecución muy rápido (~0.1 s en un ordenador portátil normal), lo que constituye una ventaja con respecto a otros métodos. Además, este modelo aprende intrínsecamente los cambios en la intensidad de las bandas moleculares debido a los cambios en la composición química y la temperatura. Por lo tanto, AICE podría ser una herramienta clave en el análisis de nuevos datos de JWST, del que se esperan cientos de espectros en un futuro próximo.

El nivel de detalle proporcionado por los recientes espectros de absorción de hielo obtenidos con el JWST hace necesario un estudio detallado de los procesos de formación de hielos interestelares. En el contexto del grupo de modelado químico de la colaboración Ice Age, hemos llevado a cabo el modelado de la química del hielo en dos estrellas de fondo en la nube molecular Chamaeleon I observada recientemente con el JWST. Para ello, hemos utilizado el programa de astroquímica UCLCHEM y hemos comparado las simulaciones con observaciones previas de hielos del proyecto Ice Age. Concluimos que el metanol (CH_3OH) se forma principalmente en la superficie de los granos de polvo mediante la hidrogenación de monóxido de carbono (CO). Además, también encontramos que los modelos que mejor reproducen las observaciones son aquellos con escalas de tiempo de colapso intermedias (~10^5 años) y densidades finales intermedias (~$2 \cdot 10^5\,\mathrm{cm}^{-3}$).

De todo esto, una de las principales conclusiones es que las observaciones de hielos interestelares son cruciales para comprender mejor y determinar la química de COMs en los núcleos sin estrella y preestelares. Debido a esto, las nuevas observaciones del JWST desempeñarán un papel clave, para cuyo análisis AICE podría ser una herramienta muy útil para el análisis rápido, preciso y exhaustivo de estos datos. Además, en el futuro se planea realizar observaciones adicionales en fase gaseosa de núcleos sin estrella / preestelares con una mejor sensibilidad para caracterizar mejor su complejidad química en un nivel más profundo, como se ha hecho para L1517, L1498, L1544 y, en menor medida, L1521E.

Chapter 1

Introduction

The formation of a planetary system is a process that spans millions of years divided in several stages, linked to the formation of the parent star. To properly understand the origin of the complex organic material that is observed in our Solar System –not only on Earth but also in other minor bodies like meteorites or the asteroid Ryugu– it is necessary to study all the phases involved in the formation of a low-mass star like our Sun. Since our galaxy, the Milky Way, hosts thousands of millions of stars, we can observe many stars in different evolutionary stages, and even the phases previous to the formation of the star itself, when matter is concentrated in relatively dense molecular clouds. Thanks to astronomical observations and the help of laboratory experiments and theoretical simulations, these different objects can be chemically characterized. By comparing the chemical inventory of multiple objects at consecutive evolutionary stages, we can then infer how chemical complexity evolves with the formation of a Solar-like star and a planetary system like our own.

This thesis focuses on the chemistry of the initial conditions of low-mass star formations, in particular in molecular clouds, starless cores and prestellar cores. In this first chapter, these objects are described in detail from an astrophysical and astrochemical point of view, reviewing the chemistry involved in these environments. Chapter 2 goes on more details on the physics and chemistry involved in the methodology used in this thesis to study those astronomical objects. Chapters 3, 4 and 5 present the results of the main research studies carried out as a part of this thesis. Chapter 6 presents a general discussion of the results of this thesis. Finally, Chaper 7 summarizes these results and presents the perspectives for possible future work.

Within this chapter, to give a bit of context we will briefly review in Section 1.1 the astronomical landscape in which these objects are inmersed, the interstellar medium. Then, Section 1.2 describes the phases of the formation of a low-mass star like our Sun, focusing on the molecular cloud and starless phase. Finally, Section 1.3 explains the basics of the chemical processes that take place in these early stages, differenciating between the gas-pase and solid-phase chemical reactions.

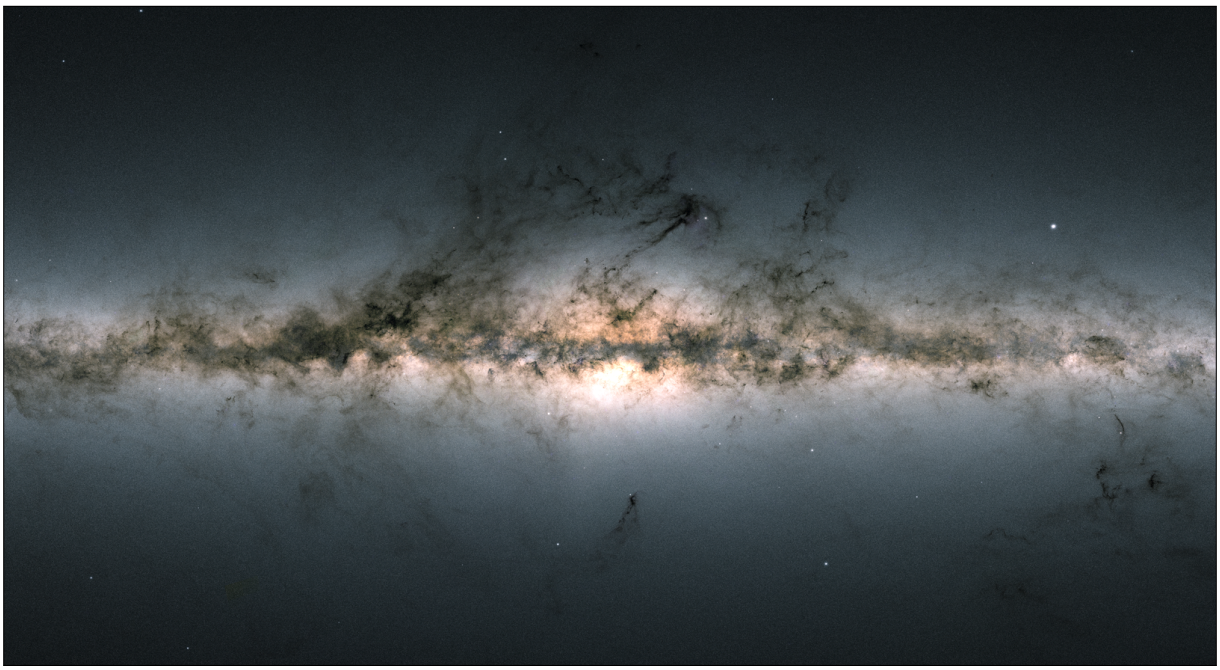


Figure 1.1.1. View of the Milky Way taken by the spacecraft Gaia, from the European Space Agency (ESA). The light emitted by the almost-indistinguishable stars, more abundant towards the galactic center, is absorbed by the dust present in the interstellar medium. Credit: ESA / Gaia / DPAC.
`https://sci.esa.int/s/ApPJaGA`

1.1 Interstellar medium

Stars are the main component of galaxies, but besides them –and apart from dark matter– there are other components that occupy the space in between and play a key role in the evolution of a galaxy: the interstellar medium (ISM). The ISM is composed mainly of gas, dust particles (Figure 1.1.1), cosmic rays and electromagnetic radiation.

Gas represents around 99 % of the mass of the interstellar medium, while the remaining 1 % corresponds to dust particles (Tielens, 2005). Within the gas, the majority (74 %) is atomic and molecular hydrogen (H, H_2), followed by helium (He, 24 %). The remaining 2 % includes simple atomic and molecular species like oxygen (O), carbon (C), nitrogen (N) and carbon monoxide (CO), but also less abundant and more complex species, like methanol (CH_3OH) or even larger molecules**.** However, the percentage corresponding to these species larger than CO is quite low. The typical abundance (with respect to hydrogen) of CO in cold clouds (relatively dense regions) is already as low as ~10^{-5} (Shalabiea and Greenberg, 1995; Wakelam and Herbst, 2008); for CH_3OH, the abundances are at maximum ~10^{-6} (e.g., Jiménez-Serra et al., 2005), while larger organic molecules ar at least a factor 10 less abundant than CH_3OH (e.g., Jin and Garrod, 2020). Regarding the dust, it is composed by irregular particles, normally micron-sized and submicron-sized, composed mainly by a silicate or carbonaceous core (Draine, 2003) and an outer icy layer –if the temperatures are cold enough (Boogert et al., 2015). This outer layer can contain simple species like water (H_2O), carbon monoxide (CO), carbon dioxide (CO_2), methanol (CH_3OH), ammonia (NH_3) and methane (CH_4), but probably also larger and much less abundant species.

Depending on the physical conditions, the interstellar medium (ISM) can be divided in different phases. Broadly speaking, we can distinguish between ionized and neutral regions, where the ionized ones have high temperatures (~ 10^4–10^6 K) and usually low densities ($\lesssim$ 10 particles per cm^3) and the neutral ones can be colder (as low as ~10 K) and denser ($\gtrsim$ 200 cm^{-3} in molecular clouds). In more detail, six phases can be distinguished (Tielens, 2005; Draine, 2011):

- **Hot ionized medium**. Large regions shocked by supernova explosions and stellar winds of early-type stars (O/B, the largest and most luminous), that ionize the gas by collisions. They present very high temperatures ($\gtrsim 10^6$ K) and very low densities (~ 10^{-3} cm^{-3}), with cooling times of the order of millions of years (Myr). As a consequence, a large fraction of the ISM (~50 %) is filled with this ionized gas, occupying most of the volume above and below the galactic disc.
- **Warm ionized medium**. Regions ionized by radiation emitted by early-type stars (O/B) or other sources, but not as hot and diffuse as the hot ionized medium (~10^4 K, ~0.1 cm^{-3}). They can occupy a significant fraction of the ISM (~10 %).
- **H II regions**. Small regions located in the surroundings of early-type stars (O/B), ionized due to their ultraviolet (UV) radiation, with similar temperatures as the warm ionized medium (~10^4 K). They are mainly observable due to atomic ionized hydrogen (H^+ or H II) recombination transitions, that is, emission lines after the ion captures a free electron. The densities range from low (~10 cm^{-3}) to relatively high values (~10^4 cm^{-3}) for the ultra-compact H II regions. They occupy quite a small fraction of the ISM (<1 %).
- **Warm neutral medium**. Mainly composed of atomic gas, slightly colder than the warm ionized medium (~5000 K) and slightly denser (~0.5 cm^{-3}). Similarly, it can occupy a large fraction of the interstellar medium (~40 %).
- **Cold neutral medium**. Mainly composed of atomic gas but at relatively low temperatures (~100 K) and medium densities (~30 cm^{-3}). It fills a small fraction of the ISM (~1 %).
- **Molecular clouds**. Relatively small structures composed mainly by molecular hydrogen (H_2) and dust grains, with cold temperatures ($\lesssim$100 K). Despite representing very small regions of the whole ISM (<1 %), they correspond to around 50 % in mass of the ISM, due to their higher densities. Depending on the density and temperature we can distiguish three subclasses:
  - *Diffuse molecular clouds*. They still contain atomic hydrogen (H) as well as H_2, and have ionized carbon (C^+). They have relatively high temperatures (~30–100 K) and densities of ~300 cm^{-3}.
  - *Translucent clouds*. With lower temperatures and higher densities ($\lesssim$30 K, ~10^3 cm^{-3}), hydrogen is mainly found as H_2, and carbon is predominantly in atomic form (C) or as carbon monoxide (CO).
  - *Dense molecular clouds*. The higher densities ($\gtrsim 10^3$ cm^{-3}) protect their interiors from external radiation (except for cosmic rays), allowing a decrease in tempera-

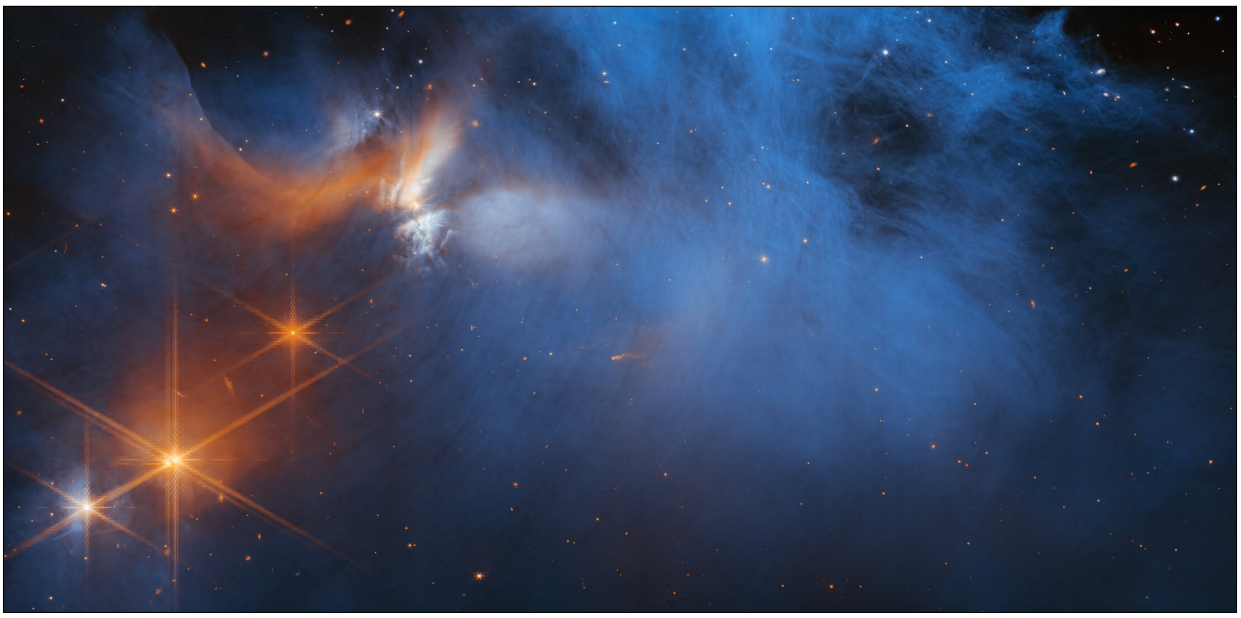


Figure 1.1.2. Image of the molecular cloud Chamaleon I, 630 light years away, taken by James Webb space telescope, in the near infrared. The young and outflowing protostar on the upper left illuminates the cold cloud material (in blue). Several background stars and galaxies can be seen across the cloud. Credit: NASA / ESA / CSA, M. Zamani (ESA/Webb); Science: F. Sun (Stewart Observatory), Z. Smith (Open University), and the Ice Age ERS Team. `https://esawebb.org/images/weic2303a/`

ture ($\lesssim$10 K). Carbon is mainly found in form of CO, and other small molecules can be formed. Within these clouds, concentrations of matter form filaments and cores, producing starless and prestellar cores, with densities as high as ~10^6–10^7 cm^{-3} (Caselli and Ceccarelli, 2012). Eventually, prestellar cores continue collapsing until they form a protostar, as explained in next section.

Molecular clouds

For the scope of this thesis, molecular clouds (e.g., Fig. 1.1.2) are of special relevance. They are asymmetric structures with extensions of tens to hundreds of parsecs [1] and can contain hundreds of thousands of solar masses (Herbst and van Dishoeck, 2009). Internally, molecular clouds can have substructures, filaments of matter which originate as a consequence of the cloud turbulent activity compressing the gas (Mac Low and Klessen, 2004; Myers, 2009; André et al., 2014). This was one of the main results obtained by ESA's Herschel space observatory (e.g., Palmeirim et al., 2013; Fig. 1.1.3). Within these filaments, there can be clumps or concentrations of matter without any young stellar object (YSO) formed inside yet, called starless cores, with extension of ~0.1 pc, temperatures of ~10 K and densities of $\gtrsim 10^4$ cm^{-3} (Bergin and Tafalla, 2007). In the work published by André (2017), the distribution of masses found in the cores inside these filaments was compared to the initial mass function of stars, linking the dark cores embedded in filaments to the future stars formed in molecular clouds. Starless cores can have ages of less than a million years (Myr) or even more, although it is

[1] One parsec is approximately 3.26 light years: 1 pc $\simeq$ 3.26 ly $\cong 3.09 \cdot 10^{16}$ m.

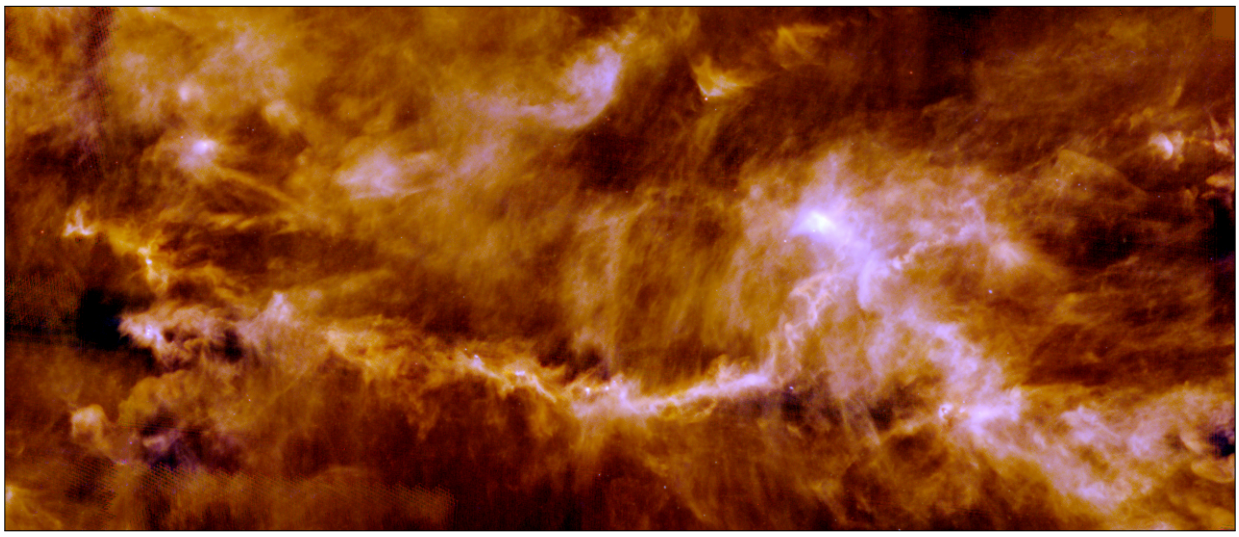


Figure 1.1.3. Image of the B211/B213 filament in the Taurus molecular cloud (TMC), obtained by Herschel Space Observatory in 2013. Credit: ESA / Herschel / PACS, SPIRE / Gould Belt survey key programme / Palmeirim et al. (2013).

not straight-forward to derive their age (e.g.: Jiménez-Serra et al., 2021; Scibelli et al., 2021).

1.2 Low-mass star formation

The formation and evolution of stars depend on their final mass. In this thesis we focus on low-mass stars (similar to the solar mass, $M_\odot \simeq 1.98 \cdot 10^{30}$ kg, or less), since their formation process can be applied to our Solar system.

The formation of low-mass stars begins in giant molecular clouds. Some of the starless cores present in molecular clouds can be dense enough that they can start collapsing, overcoming thermal and magnetic pressure and turbulence (Öberg and Bergin, 2021). In this case they are called prestellar cores, since eventually a protostar will be formed in the center. During this phase, matter slowly accumulates toward the center, increasing the density (which can reach ~10^7 cm^{-3} and slightly reducing the temperature due to shielding (Keto and Caselli, 2010; Caselli and Ceccarelli, 2012), reaching values of less than 7 K (Bergin and Tafalla, 2007). Due to this increase of density, atoms and molecules present in the gas phase start to get absorbed onto dust grains, in whose surfaces several chemical reactions can take place. Prestellar cores have typical ages under 1 Myr (e.g.: Jiménez-Serra et al., 2016; Vasyunin et al., 2017).

As density continues increasing, pressure increases in the centre, which slowly raises the temperature, while an extense envelope forms in the surroundings (Caselli et al., 2022) and a broad disc is formed, following the angular momentum of the core. At this moment, we can speak of a protostar (Öberg and Bergin, 2021). With the increase in temperature in the centre, it starts emitting radiation as a black body, but it is still embedded in a thick

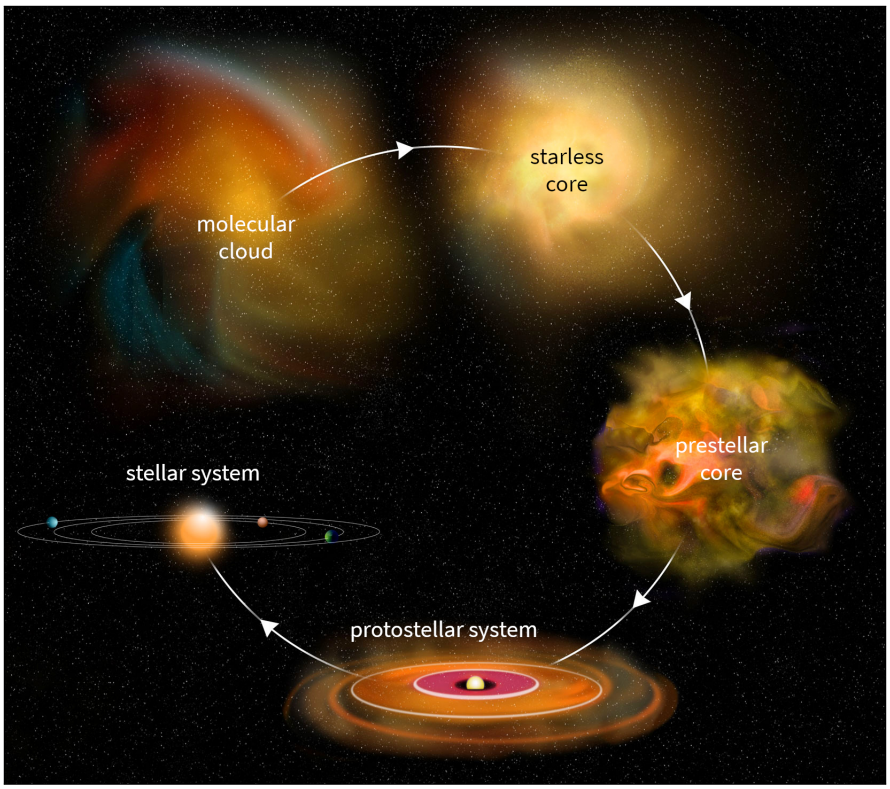


Figure 1.2.1. Scheme of the process of low-mass star formation. Inside molecular molecular clouds, starless cores can form, which can experience gravitational collapse (prestellar cores), eventually forming a protostar and a planetary system. Adapted image; Credit: Bill Saxton / NRAO / AUI / NSF. `https://supernova.eso.org/exhibition/images/0408_F_Clump-core_med-CCfinal/`

envelope and a broad disc –with typical sizes of 0.05 pc or ~10^4 astronomical units [2]–, which prevents the radiation to escape outside. Additionallly, the protostar ejects matter in the form of jets perpendicular to the plane of rotation. This very first stage of young stellar objects is classified as Class 0, and it is estimated to last around 50 000 yr (50 kyr; Kristensen and Dunham, 2018). Some of these objects are found to be chemical rich, probably due to the increase of temperatures around the central protostar, and are known as *hot corinos* (De Simone et al., 2020).

As time passes, the envelope and disc reduce their size, reaching extensions of ~500 au (Williams and Cieza, 2011), and part of the radiation from the central protostar starts to escape, mainly in the infrared; this would be a Class I protostar, which can last several 100 kyr, and can also present a hot corino (De Simone et al., 2020). When the envelope is mainly consumed and there is only a disc (~100 au in extension) with outflows, we then refer to Class II objects, with typical ages of ~1 Myr or more (André, 2002). Finally, Class III objects have a very thin disc (with extensions of $\lesssim$100 au) and no outflows, having ages up to ~10 Myr (Williams and Cieza, 2011). These circumstellar discs are also called protoplanetary discs, since they can eventually form planets inside them. In practice, these objects are classified in the mentioned classes according to their spectral energy distribution (SED). Along this evolution, the protostar continues increasing its temperature. When the centre achieves

[2] An astronomical unit (au) is the mean distance between the Earth and the Sun: 1 au $\simeq 1.50 \cdot 10^{11}$ m

temperatures high enough (~10^6 K), deuterium (^{2}H or D) fusion reactions start. Eventually, if it reaches a high enough value (~10^7 K), fusion reactions of hydrogen take place, yielding a new pre-main-sequence star (Öberg and Bergin, 2021); otherwise, it would be a brown dwarf (Luhman, 2012). Within all these phases **(**schematically represented in Fig. 1.2.1), the increasing radiation emitted by the protostar increases the temperature of its surroundings and ionizes the gas, enhancing the chemistry both in the gas and the solid phase due tu thermal and energetic processing (van Dishoeck et al., 2006; Herbst and Garrod, 2022).

Interestingly, in the last decade it has been shown that there is a good correlation between the molecular abundances found in Solar System comets (like 67P/Churyumov-Gerasimenko, or 67P/C-G) and those measured in protostellar systems (Öberg et al., 2015; Drozdovskaya et al., 2019; López-Gallifa et al., 2024), which suggests that there is a chemical inheritance between the phases of the formation of a Solar-like system and the final stage with a fully developed planetary system (see Fig. 1.2.2). In fact, one of the main results from ESA's Rosetta mission on comet 67P/C-G is that the chemical composition of this comet is likely prestellar, which triggers the question about how much of that chemical complexity is built during the initial conditions of Solar-system formation. Furthermore, large COMs such as amino acids –some of them being prebiotic– have been found in carbonaceous meteorites (e.g.: Cronin et al., 1979; Cronin and Pizzarello, 1983, 1999; Glavin et al., 1999; Koga et al., 2021, 2024), in the asteroid Ryugu (Oba et al., 2023) and in the coma of the comet 67P/C-G (Altwegg et al., 2016). These facts altogether stress the importance of studying the origin of COMs in the different stages of low-mass star formation to understand the origin of prebiotic organic chemistry on Earth, and its possible occurence in other planetary systems. However, the chemistry of starless/prestellar has not been studied as much as protostellar systems, even though they represent the earliest stages in which COMs can be formed. This is the main motivation of this thesis, in which we study the chemistry of these young objects, with special interest in COMs.

1.3 Chemistry in starless and prestellar cores

Chemistry is a useful tool to constrain the physical structure of dense cores in molecular clouds. The chemistry in these cold objects is an interplay between gas phase processes and the chemical reactions occurring on dust grains. In the following we describe the basics of the gas phase and grain surface reactions occurring in starless and pre-stellar cores.

The physical conditions of molecular clouds and starless/prestellar cores are quite different to those found on Earth, since the densities are much lower (even in prestellar cores) and temperatures are extremely cold. As a reference, typical temperatures and pressures on Earth are, respectively, ~15 ºC (~288 K) and ~1 atm (~1 bar $\simeq 10^5$ Pa), while on starless/prestellar cores we have typical values of ~10 K (Caselli and Ceccarelli, 2012) and <10^{-10} mbar (Congiu et al., 2020). The low densities prevent the occurence of three-body reactions, limiting the variety of bond formation (Herbst and Klemperer, 1973). Therefore,

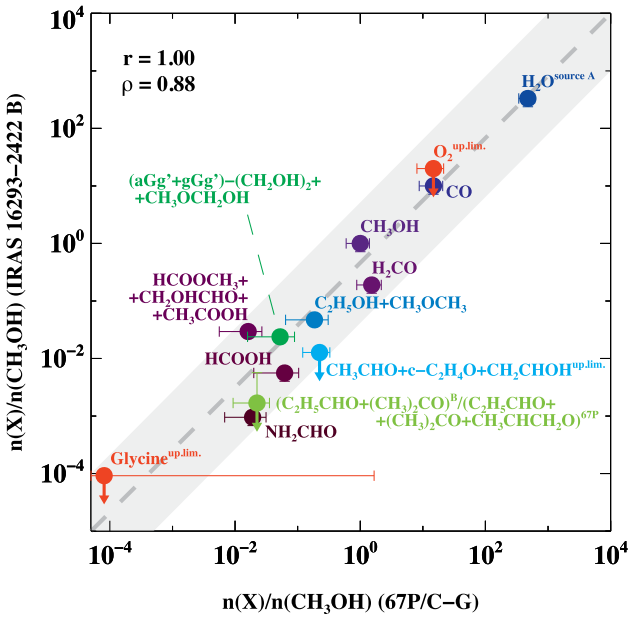


Figure 1.2.2. Plot of the molecular abundances with respect to methanol (CH_3OH) measured toward the protostellar source IRAS 16293-2422 B versus those measured in the comet 67P/Churyumov-Gerasimenko (67P/C-G), as presented by Drozdovskaya et al. (2019). The good correlation found (in logarithmic scale) suggests that there is a chemical inheritance between the two types of sources.

two-body reactions with one unit of each reactant are the common type of reactions. Due to the very low temperatures and the consequently low kinetic energy of the reactants, barrierless exotermic reactions or reactions with very low barrier gain importance, although quantum tunneling can mitigate this in some cases (Cazaux et al., 2011; Hama and Watanabe, 2013). The possible chemical reactions occuring both in gas-phase and solid-phase can be experimentally simulated and studied in laboratory, as well as reproduced by quantum chemical calculations, helping to understand the observed abundances in the interstellar medium.

1.3.1 Gas phase

Figure 1.3.1 summarises the most common type of reactions occuring in interstellar environments for gas-phase species. A typical bond formation reaction is the radiative association, where two reactants A and B collide and form a product AB, releasing a photon ($h\nu$) that carries away the released energy: $A + B \longrightarrow AB + h\nu$ (Öberg and Bergin, 2021). Due to the need of collisions, it is common that one of the reactants is either atomic or molecular hydrogen (Herbst and Klemperer, 1973). Another possible reaction is the associative detachment, where an anion (A^-) collides with a neutral reactant (B) to form the product AB, this time releasing an electron (e^-) that carries the produced energy: $A^- + B \longrightarrow AB + e^-$. This last type of reactions should be less common that radiative associative, since it requires that one of the reactants is an anion (Millar et al., 2007), although several anions have been recently detected in the Taurus Molecular Cloud (Agúndez et al., 2023c).

Similarly but in an inverse way, molecular bonds are typically destroyed by absorbing a

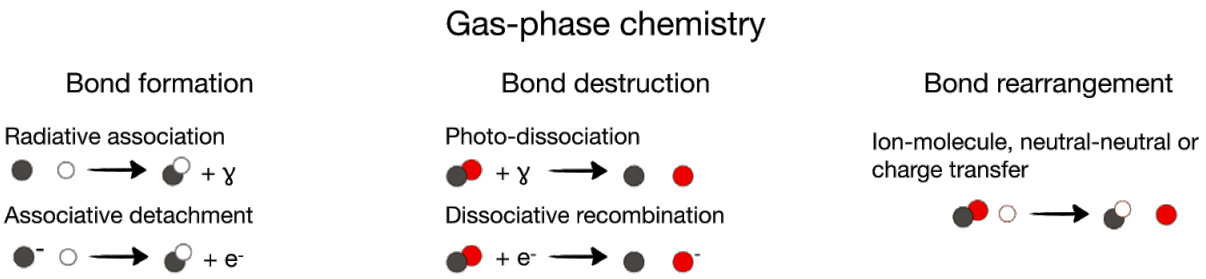


Figure 1.3.1. Scheme of the most common type of gas-phase chemical reactions in the interstellar medium. Photons are denoted as γ, but they are not necessarily gamma rays. Credit: Öberg and Bergin (2021).

photon (photo-dissociation) or absorbing an electron (dissociative recombination). In photodissociation reactions, the species AB absorbs a photon ($h\nu$) and gets dissociated into A and B: $AB + h\nu \longrightarrow A + B$ (Hollenbach and Tielens, 1999). Usually, ultraviolet (UV) photons are required to destroy molecular bonds (Öberg and Bergin, 2021). In dissociative recombination, collisions with electrons can break molecular bonds, producing an anion and a neutral species. Additionally, in shocked regions molecular bonds can be broken by high-velocity collisions between the reactants (Öberg and Bergin, 2021). Lastly, a two-body reaction between two reactants in which at least one of them has a molecular bond (AB, C) can led to a bond rearrangement: $AB + C \longrightarrow AC + B$ (Herbst and Klemperer, 1973). These reactions usually involve radicals (neutral species with unpaired electrons), ions or both. For example, ion-molecule reactions are typically exotherminc.

1.3.2 Solid phase

Even if dust particles represent only 1 % of the matter in the interstellar medium (compared to the 99 % of gas), they play a key role in the chemistry of the interstellar medium. This is mainly because atoms and molecules can get absorbed to their surfaces forming ices (*freeze-out*) and then different surface reactions can take place, producing molecules that may eventually get desorbed and released to the gas phase (Ceccarelli et al., 2007). Figure 1.3.2 summarises the most important processes involving solid phase chemistry in dust grains.

Once atoms and molecules are in the surface of the grain, chemistry is initially dominated by hydrogenation, that is, the addition of atomic hydrogen (Hasegawa et al., 1992), since it has a high mobility through the surface due to its low mass. Actually, molecular hydrogen (H_2) is originally formed due to the adsorption of two hydrogen atoms into dust grains and their further arrangement (Hollenbach and Tielens, 1999; Wakelam et al., 2017). In this way, many simple species can be formed, such as water (H_2O) when hydrogenating oxygen atoms (Dulieu et al., 2010); ammonia (NH_3) when hydrogenating nitrogen atoms (Redaelli et al., 2023); and methanol (CH_3OH) when hydrogenating carbon monoxide (Watanabe and Kouchi, 2002). Additionally to hydrogenation, radicals that are close enough can react and form new species, a process that is enhanced at higher temperatures (~25 K), since diffussion increases.

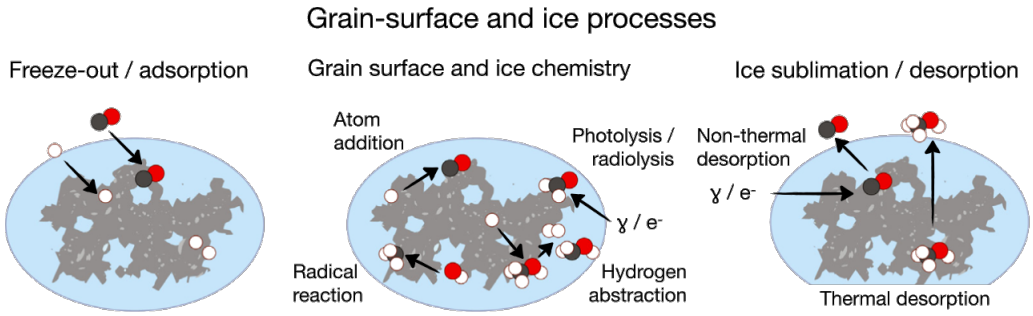


Figure 1.3.2. Scheme of the most common type of solid-phase chemical reactions in the interstellar medium. Photons are denoted with as γ, but they are not necessarily gamma rays. Credit: Öberg and Bergin (2021).

As for bond destruction, it can occur through collisions with radiation or electrons (or even cosmic rays), but also by hydrogen abstraction, in which a single hydrogen atom can *steal* another hydrogen atom from a molecule, forming H_2 (Öberg and Bergin, 2021). Then, molecules formed in the icy surface can get desorbed to gas phase, either by thermal desorption (by an increase of temperatures, e.g. in the protostellar phase) or by non-thermal desorption (by UV radiation, cosmic rays, or even by H_2 formation). These molecules released to the gas-phase can then react with other gaseous species and produce new species, as it happens with methanol (Balucani et al., 2015; Vasyunin et al., 2017).

1.4 Astrochemical observations of starless and prestellar cores

One characteristic of starless core is the depletion in gaseous carbon monoxide (CO). As density increases, more and more CO gets adsorbed onto dust grains, therefore producing a depletion in the observed gas-phase CO (Crapsi et al., 2005). Therefore, CO depletion is higher in prestellar cores, and it can be used as a proxy to identify starless/prestellar cores and their evolutionary stage. A consequence of this CO depletion is a high level of deuteration, that is, a high level of deuterium (^{2}H or D) either in atomic form or contained within molecules (Crapsi et al., 2005). This is due to the fact that CO is one of the first agents that destroy the deuterated trihydrogen cation (H_2D^+), which initiates the chemistry of deuterium (Bergin and Langer, 1997; Bergin and Tafalla, 2007; Caselli et al., 2022). For example, this can be measured with the deuteration fraction in the diazenylium cation (N_2H^+), that is, with the ratio in column density of N_2D^+/N_2H^+, having typical values of $\lesssim 0.05$. For the more evolved prestellar cores, this deuterium fraction increases, reaching higher values of ~0.25. Therefore, this can be used as a proxy to identify both starless and prestellar cores and their level of evolution.

In addition, a physical direct proxy used to distinguish prestellar cores from static starless cores is the sign of infall motions towards the center, either only from the more external

regions (the envelope) or from the whole structure. This can be seen as a characteristic double-peaked shape in some molecular lines that are tracers of the moving material, such as carbon monosulfide (CS) or isotopologues of the formic cathion (HCO^+), as observed by, e.g., Redaelli et al. (2022a,b). A higher level of infall would correspond to a more evolved prestellar core, while a slight sign of infall only in the envelope could not be enough to classify a starless core as prestellar.

Lastly, observations of the density profile of these objects can also be performed with bolometric observations of dust continuum (Tafalla et al., 2004, 2006). It has been observed that they have a small flat region in the center and then a profile that usually decreses being inversely proportional to the distance to the power of ~1.5–2.0. The determination of a high enough central density ($\gtrsim 10^4$ cm^{-3}; Bergin and Tafalla, 2007) determines the status of starless core and can even indicate the level of evolution, since prestellar cores present higher densities (up to ~10^7 cm^{-3}; Caselli et al., 2017).

Interstellar complex organic molecules

In astrochemistry, interstellar complex organic molecules (COMs or iCOMs) are defined as molecules containing at least one carbon atom and having a minimum of 6 atoms in total (Herbst and van Dishoeck, 2009; Ceccarelli et al., 2017). It was believed that in the extreme and cold conditions of cold molecular clouds and starless/prestellar cores, the chemistry would be relatively simple. However, in 2007 and 2012, higher sensitivity observations (Marcelino et al., 2007; Bacmann et al., 2012) revealed the presence of interstellar complex organic molecules (COMs) even under these conditions in which CO is fully frozen onto grains. This has triggered a lot of interest in the astrochemistry community to understand how those COMs are present in these cold objects.

The first detections of COMs in the interstellar medium (ISM) were reported in the decade of 2000 in relatively hot sources ($\gtrsim$100 K), such as massive hot cores (Hollis et al., 2000, 2006; Belloche et al., 2008, 2013) or low-mass hot corinos (Bottinelli et al., 2004; Jørgensen et al., 2013). In these years, it was believed that COMs needed such relatively high temperatures ($\gtrsim$100 K) to be formed, so that radical-radical diffusion reactions can occur. However, in the following decade and until now, COMs have been reported as well in colder starless and prestellar cores (e.g.: Öberg et al., 2010; Bacmann et al., 2012; Vastel et al., 2014; Jiménez-Serra et al., 2016, 2021; Scibelli and Shirley, 2020; Scibelli et al., 2021, 2024; Fig. 1.4.1) and in molecular clouds (e.g.: McGuire et al., 2017, 2018, 2020; Cernicharo et al., 2021a,b, 2024; Agúndez et al., 2021, 2023a,b, 2024). This indicates that COMs can indeed be formed at these low temperatures (as low as ~10 K or even less). At the same time, detections of COMs on more evolved protostellar systems have been reported, including in the later protoplanetary disc stage (e.g.: Walsh et al., 2016; Bergner et al., 2017; Ilee et al., 2021). Therefore, COMs are present in all the phases of the formation of low-mass stars, being formed already in the early stages of starless/prestellar cores and even in molecular clouds (e.g., Garrod et al.,

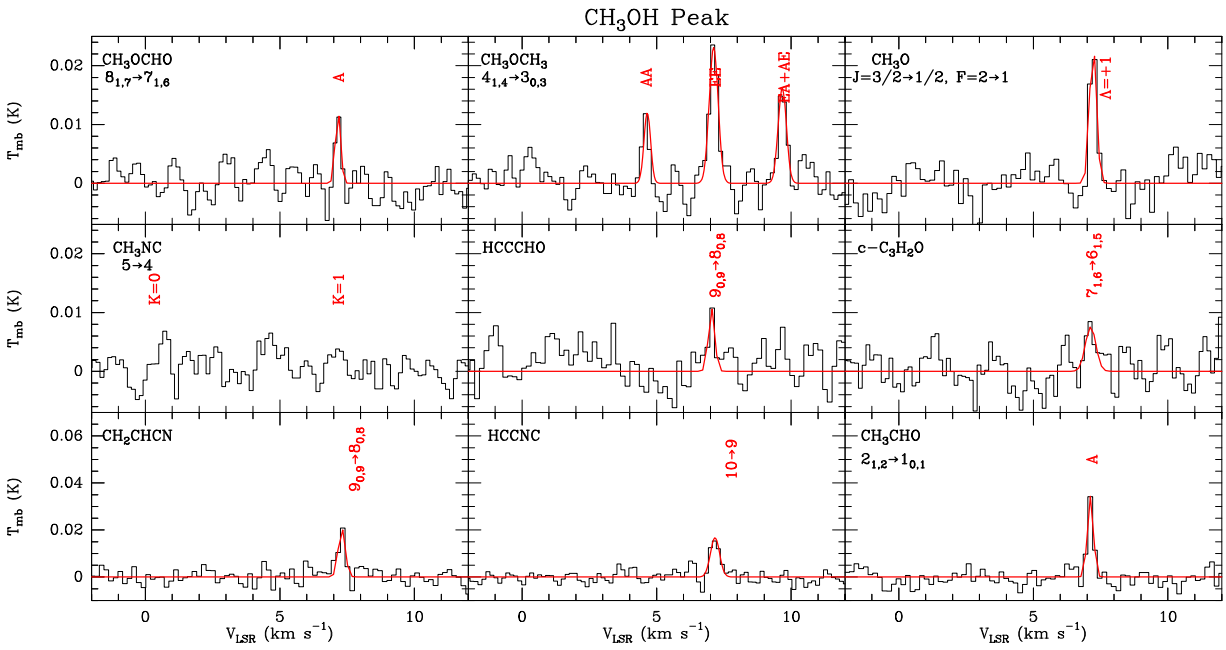


Figure 1.4.1. Some molecular transitions found toward the methanol peak of the prestellar core L1544, as presented by Jiménez-Serra et al. (2016). Intensities are presented as main beam temperature (as explained in Section 2.3.1) versus radial velocity (in local standard of rest, LSR).

2022).

The references mentioned in the last paragraph about observations of molecular clouds correspond to the Taurus Molecular Cloud (TMC), in particular to the zone TMC-1. Two large projects have been observing this source in the recent years detecting numerous COMs and PAHs (policyclic aromatic hydrocarbons): GOTHAM (***G**BT **O**bservations of **T**MC-1: **H**unting **A**romatic **M**olecules*), using the 100 m Green Bank Telescope (GBT) in West Virginia (United States of America); and QUIJOTE (***Q**-band **U**ltrasensitive **I**nspection **J**ourney to the **O**bscure **T**MC-1 **E**nvironment*), using the Yebes 40 m telescope in Guadalajara (Spain). This last project has provided the first detection in space of a PAH, indene (C_6H_8) as well as other PAHs, PAH precursors and many hydrocarbon chains, among others.

In this thesis we focus on the formation of COMs and in studying the composition of the gas and dust (ices) that yields the relatively high level of chemical complexity observed in these cold objects. We focus on the very first stages of the formation of a Solar-like system, i.e., molecular clouds and starless/prestellar cores. We aim at expanding the chemical characterization of such sources, both in gas and solid phase, by the use of radio astronomical observations and the development of software tools that facilitate the analysis of astronomical observations, both for radio and infrared observations.

1.5 Objectives of this thesis

In the following we present the main scientific objectives of this thesis:

1. To constrain the chemistry of starless and prestellar cores in gas phase and its evolu-

tion, with radio astronomical observations, focusing on interstellar complex organic molecules (COMs) and its precursors.

2. To study the chemistry of starless and prestellar cores in solid phase (i.e., ices in dust grains), with infrared astronomical observations, focusing on COMs and their precursors. For this, a new analysis tool should be developed to fully exploit the large database of interstellar ices obtained by James Webb space telescope.
3. To use astrochemical modelling to study the formation processes of COMs and of its precursors in starless and prestellar cores, and compare it with astronomical observations to derive or constrain information about the chemistry of these objects.

Chapter 2

Methodology

This chapter describes tools that have been used in this thesis to derive the physical properties and chemical inventory of precursors of planetary systems –that is, prestellar cores. We start reviewing how molecular spectroscopy allows to identify molecules in the interstellar medium studying the emission and absorption of radiation at certain frequencies, and how the analysis of the astronomical observations allows to derive the composition and the properties of the observed material. Then, the telescopes used in this work are listed and described. Finally, we describe the basics of the chemical modeling of regions of the interstellar medium by using astrochemical codes. The chapter ends with a list of bibliography used as a reference for each section.

2.1 Molecular spectroscopy

Molecules can absorb and emit radiation at specific frequencies or wavelengths of the electromagnetic spectrum (Figure 2.1.1). For molecules, the transitions can be divided into three types: electronic, vibrational and rotational (Wilson et al., 2013). Electronic transitions are associated with changes in the energy state of one of the electrons of the molecule, whose atomic nuclei remain at the same relative positions;[1] molecular electronic transitions take place in the optical and ultraviolet range. Vibrational transitions correspond to oscillating movements of the atomic nuclei around its equilibrium positions, where the frequency of the emitted or absorbed radiation corresponds to the physical frequency of vibration of the atoms; the energies associated with them are lower than those of electronic transitions, having frequencies in the range of the near and mid infrared. Lastly, rotational transitions correspond to the different ways of rotation of atomic nuclei taking into account the relative chemical bonds and their spatial distribution, where the physical frequency of rotation corresponds to the frequency of the emitted or absorbed radiation; the energies are even lower

[1] In reality, once the reaction has taken place, the molecule can get relaxed to a stable state with a different geometry than initially; hence the relative positions of the atoms can actually change.

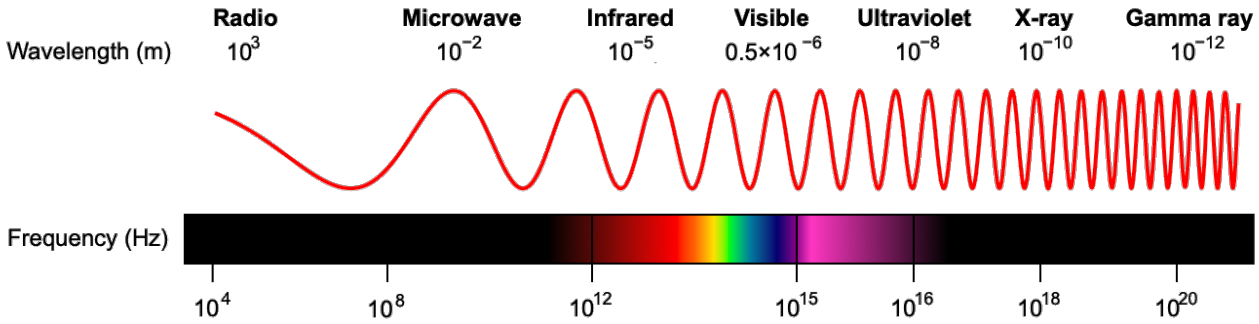


Figure 2.1.1. Scheme representing the range of the electromagnetic spectrum. The frequency (ν) and the wavelength (λ) are related with each other and the speed of light (c) by the equation $c = \lambda\nu$, where $c \simeq 2.998 \times 10^{8}$ m/s. Adapted image, original by Inductiveload and NASA: `https://commons.wikimedia.org/wiki/File:EM_Spectrum_Properties_edit.svg`.

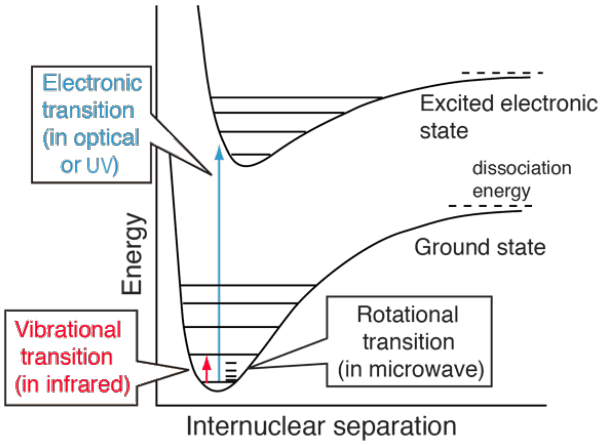


Figure 2.1.2. Scheme of the different types of molecular transitions, considering a diatomic molecule. The curves show the energy of the system as a function of the distance between the two atoms. Image by R. Nave: `http://hyperphysics.phy-astr.gsu.edu/hbase/molecule/molec.html`.

than vibrational transitions, being the frequencies in the range of microwaves and radio waves.

To ilustrate this energy classification, Figure 2.1.2 represents schematically the different types of molecular transitions and the corresponding energy levels, using a diatomic molecule as an example. In this case, the energy of the system can be represented against the distance between the two atomic nuclei, and it will present a minimum around an equilibrium distance in the ground state. In the following, the two last type of transitions (vibrational and rotational) will be described with a bit more detail, since they are the type of transitions studied in this thesis.

2.1.1 Rotational transitions

This type of transitions require a property on the molecule to be present: electric dipole moment (Wilson et al., 2013). If a molecule has a dipole moment, incident electromagnetic waves can induce a torque on it and excite its rotational levels. Similarly, if a molecule is already in an excited rotational level, it can emit radiation to go back to a lower energy

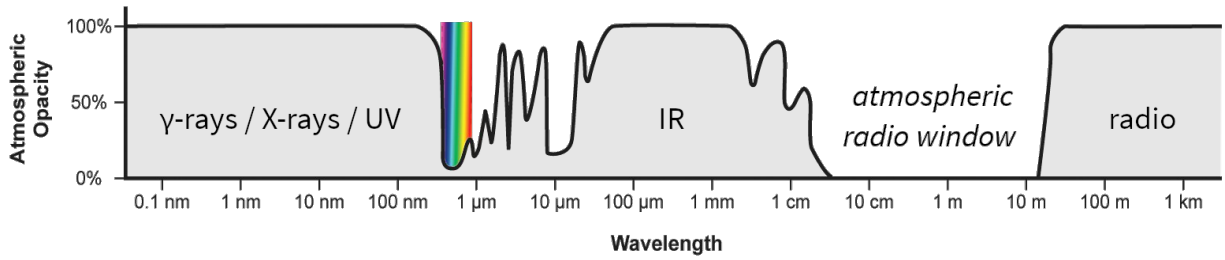


Figure 2.1.3. Plot of the atmospheric opacity with respect to the wavelength of the incident radiation. Adapted image, original by NASA:
`https://commons.wikimedia.org/wiki/File:Atmospheric_electromagnetic_transmittance_or_opacity.jpg`.

rotational level.

The typical energies associated with these transitions are of the order of 0.1-1 meV, which corresponds to wavenumbers of mm and cm (that is, at wavelengths of ranges of microwaves and radio waves). It happens that Earth's atmosphere is transparent to radio waves in the range of, approximately, 0.2 mm to 20 m (15 MHz to 1.5 THz), as shown in Figure 2.1.3 (Wilson et al., 2013). Therefore, rotational transitions can be detected from ground based observatories. Usually they are detected in emission, although they can also be detected in absorption. The temperatures needed to excite the molecules in their rotational states are of the order of several to tens of Kelvins, which are temperatures easily reached in the interstellar medium.

The values of the energy levels depend on at least one quantum number (an integer) that is usually designed as J, which corresponds to the projection of the angular momentum on the axis of the molecule. For linear molecules, the formula for the rotational levels is (Mangum and Shirley, 2015):

$$E = \frac{\hbar^2}{2cI} J(J+1) \ , \tag{2.1}$$

where I is the momentum of inertia of the molecule, c is the speed of light ($c \simeq 2.998 \times 10^8$ m/s) and $\hbar$ is the rationalized Planck's constant ($\hbar \equiv h/(2\pi)$, $h \simeq 6.63 \cdot 10^{-34}$ J·s $\cong 4.14 \cdot 10^{-15}$ eV·s). For linear molecules, this number is sufficient to describe the rotational states. However, for molecules with other spatial symmetries, like symmetric tops and asymmetric tops, there are additional quantum numbers that must be used to describe the energy states (Wilson et al., 2013). For example, for symmetric tops the number K is associated with the projection of the angular momentum on the symmetry axis. The formula for the energy levels would be (Wilson et al., 2013):

$$E = \frac{\hbar^2}{2cI_\perp} J(J+1) + \frac{\hbar^2}{2c}\left(\frac{1}{I_\parallel} - \frac{1}{I_\perp}\right) K^2 \ , \tag{2.2}$$

where $I_\parallel$ and $I_\perp$ are the molecular momenta of intertia on the axis of symmetry and on a perpendicular axis, respectively. For asymmetric top molecules, we would have two quantum numbers K_a and K_c, corresponding to the projection of the angular momentum on the first and last symmetry axes (a and c).

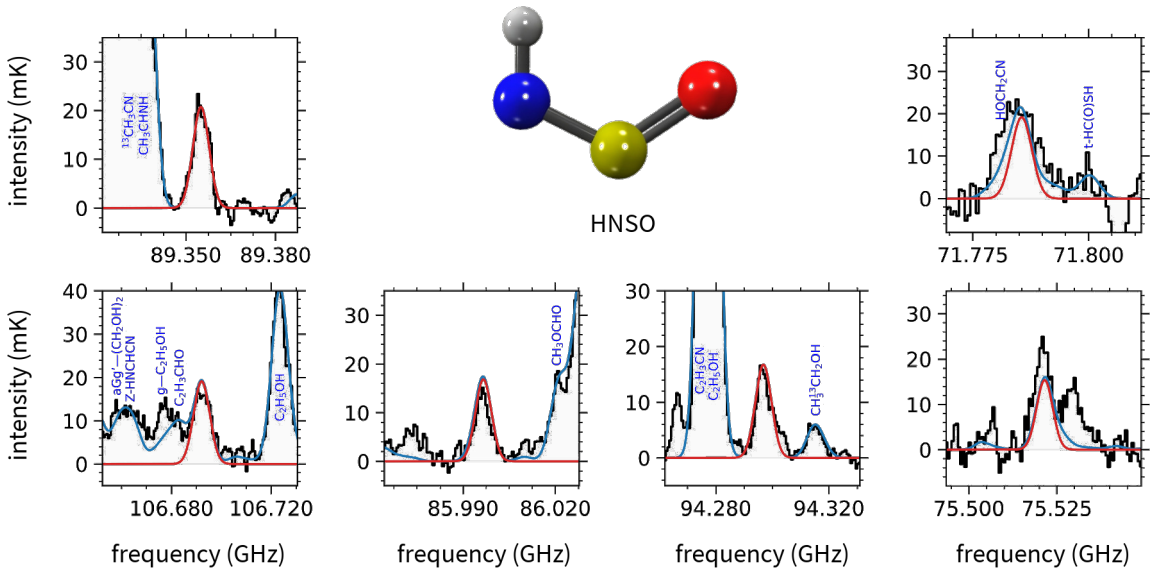


Figure 2.1.4. Some rotational transitions of the molecule HNSO (thionylimide) observed by Sanz-Novo et al. (2024b) in the spectrum of a molecular cloud called G0693, located in the Galactic centre. Intensity is expressed as antenna temperature (in mK, see Sections 2.2.2 and 2.3.1.1). The modelled lines of HNSO are marked in red, while in blue the contribution of the lines from all molecular species identified toward the cloud is shown. In the top you can see an scheme of the molecular structure of HNSO.

Additionally, if the corresponding electronic state of the molecule presents an orbital angular momentum, it can interact with the rotation of the molecule, producing a splitting in the rotational levels. In this case, there is another quantum number involved, usually denoted as Λ, which corresponds to the projection of the electronic orbital angular momentum on the molecular axis (Wilson et al., 2013). Lastly, the rotational energy levels can be splitted in more sub-levels in an effect that is called hyperfine structure. It is caused when at least one of the atoms forming the molecule has a nuclear spin, which creates a magnetic moment that interacts with the electrons and the rest of nuclei. In this case, another quantum number (again, an integer) enters in scene, usually denoted as F and corresponding to the total angular momentum, including the electronic spin (Wilson et al., 2013).

Normally, the frequencies of rotational transitions of a certain molecule can be characterized in the laboratory with spectroscopy experiments in which the molecule is synthetized and excited to observe its emission lines. Additionally, quantum chemical calculations can help to extrapolate some of the measured transitions to a broader spectral range (Ozaki et al., 2021). This is what is recorded in molecular rotational specroscopy catalogues such as the Jet Propulsion Laboratory (JPL) catalog (Pickett et al., 1998)[2] and the Cologne Database for Molecular Spectroscopy (CDMS; Endres et al., 2016).[3]

As an example of rotational transitions, Figure 2.1.4 shows several emission lines of the molecule HNSO (thionylimide) observed in an astronomical spectrum of a molecular cloud in the Galactic centre (G0693), from a work by Sanz-Novo et al. (2024b). Emission lines of HNSO are marked in red, while other emission lines from other molecules are marked in

[2] https://spec.jpl.nasa.gov/ftp/pub/catalog/catform.html
[3] https://cdms.astro.uni-koeln.de/

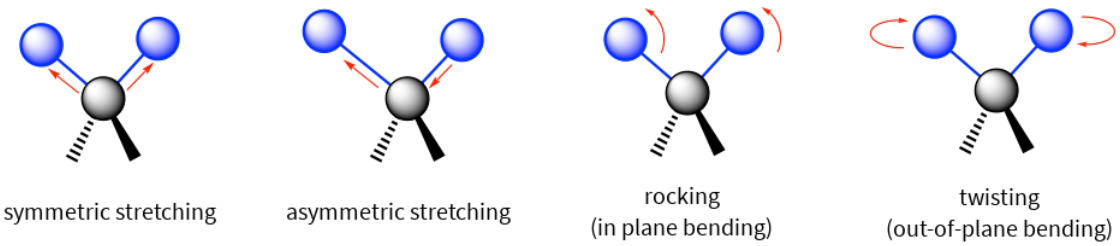


Figure 2.1.5. Schematic view of examples of molecular vibrations, for a group of three atoms inside a molecule. Adapted image, original by Liu (2021): `https://kpu.pressbooks.pub/organicchemistry/chapter/6-2-infrared-ir-spectroscopy/`.

blue. Rotational transitions from different molecules can be blended, depending on the central frequencies of the lines and the width. The shape of rotational lines is usually Gaussian, with a width that depends on the temperature and the velocity dispersion of the source material.

2.1.2 Vibrational transitions

This type of transitions corresponds to movements of the atomic nuclei inside the molecule around its equilibrium position, in a way that the center of mass of the molecule is preserved. For each group of atoms, there are different movements that can be peformed depending on the direction of the movement of each atom. Figure 2.1.5 shows some examples of molecular vibrations for a group of three atoms, being two of them connected to the third one.

The typical energies associated with these transitions are of the order of 0.1–0.4 eV (Wilson et al., 2013), which corresponds to wavenumbers of the order of μm and tens of μm (that is, in the near and mid infrared). The corresponding excitation temperatures, calculated by dividing the energy by Boltzmann's constant, are of the order of thousands of Kelvins, which is why they are typically measured in star-forming regions due to the presence of hot gas or illuminated by ultravioltet (UV) radiation. Alternatively, these transitions can also be detected in absorption when stellar radiation propagates through the molecular clouds and interact with interstellar ices, in whose surface there can be ices of different molecules, such as water (H_2O), carbon monoxide (CO) and carbon dioxide (CO_2) can be found.

As an example, Figure 2.1.6 displays some observations of vibrational transitions in a molecular cloud, from a work by McClure et al. (2023) that used data from James Webb space telescope. A background star behind the molecular cloud, for the two positions observed, emits light that transmits through all the dust and gets absorbed at certain wavelength regions corresponding to the vibrational transitions of the molecules present in the ices in solid form. The observed absorption features, or molecular infrared (IR) bands, can have a very large width. This is a difficulty for identifying molecules with this method, since it increases the probability of overlapping bands.

An additional difficulty for identifying these spectral features is that the vibrational tran-

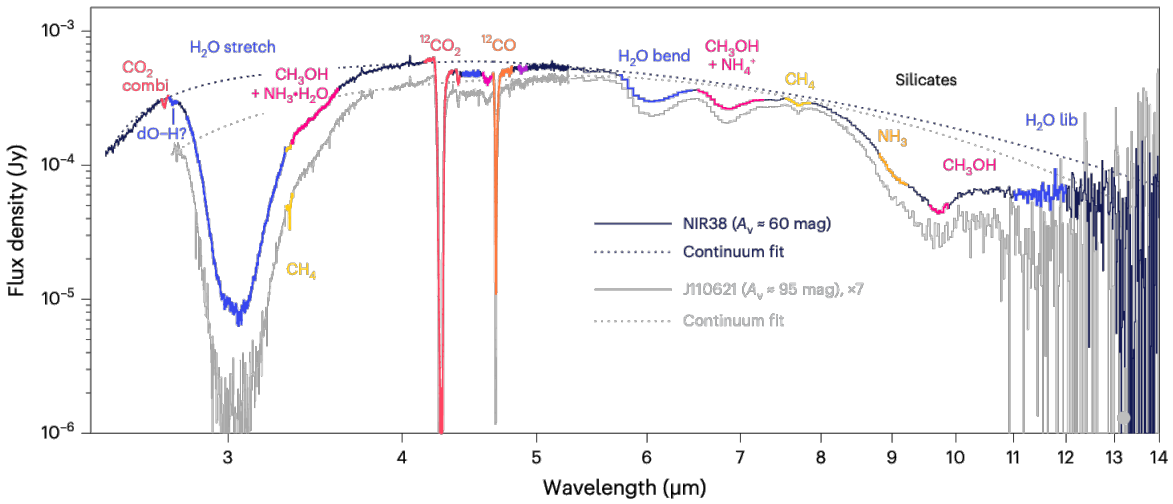


Figure 2.1.6. Spectra observed by McClure et al. (2023) using James Webb space telescope towards two positions of the Chamaleon I molecular cloud, with a background star each (NIR38, spectrum in black; and J110621, spectrum in grey). Dots indicate the fitted continuum of the background stars. Marked with colours are different molecular absorption bands, produced by the ice layers covering interstellar dust grains.

sitions are associated to chemical bonds between groups of atoms. Therefore, different molecules that share functional groups can share absorption bands, with only slight differences due to the rest of the atoms in each molecule. Figure 2.1.7 ilustrates this effect, showing the absorbance spectra measured in the laboratory of pure water ice (H_2O), pure methanol ice (CH_3OH) and pure ethanol ice (CH_3CH_2OH). Absorbance indicates how much light is absorbed at certain frequencies. As shown in Figure 2.1.7, there are several features that are shared between two or three of the molecules, like the wide OH stretching band. Vibrational bands are typically characterized in the laboratory with absorption spectroscopy, although quantum chemical calculations can also be performed to try to estimate the frequencies.

2.2 Radiative transfer

Radiative transfer studies how radiation propagates in a medium with matter. Let us consider a ray of light that is traveling through space, with a specific intensity I_ν (power per area per frequency per solid angle). This intensity will only change if radiation is absorbed or emitted due to the matter in the path (Figure 2.2.1).

The loss of intensity due to absorption is proportional to the intensity, so that the intensity change produced in an infinitesimal space $\mathrm{d}x$ is $-\kappa\, I_\nu \mathrm{d}x$; here κ is the absorption coefficient (also denoted α), that is an intrinsic property of the material and deppends on the frequency. If there is also emission, there will be a gain of intensity dependent on the emissivity, ε_ν (also denoted j_ν), which is also an intrinsic property of the material, so that the intensity change is $\varepsilon_\nu\, \mathrm{d}x$. Combining the two contributions, we obtain the equation of transfer (Condon and

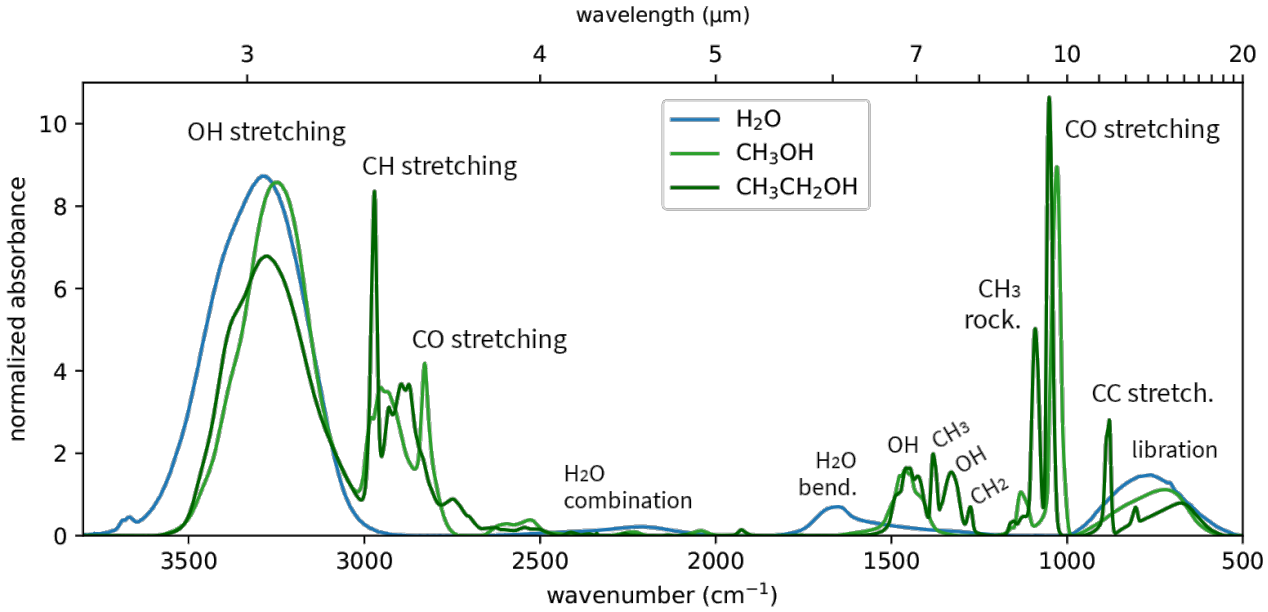


Figure 2.1.7. Infrared absorption spectra of water (H_2O), methanol (CH_3OH) and ethanol (CH_3CH_2OH) measured on laboratory, at a temperature of 15 K. The absorbance spectra were baseline-corrected and normalized to its mean absorbance value. The first experiment was performed by Öberg et al. (2007), while the other two were performed by Terwisscha van Scheltinga et al. (2018). Data retrieved from LIDA (Leiden Ice Database for Astrochemistry): `https://icedb.strw.leidenuniv.nl/`.

Ransom, 2016):

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}x} = \varepsilon_\nu - \kappa\, I_\nu \ . \tag{2.3}$$

It is convenient to define a new adimensional magnitude called *optical depth*, τ, also called *opacity*. It indicates the level of absorption produced by matter in a certain path, being defined as the integral of the absorption coefficient along the line of sight: $\tau \equiv \int_{x_0}^{x} \kappa\, \mathrm{d}x$, where x_0 is the location where the radiation first contacts with the matter along its path (Wilson et al., 2013). Therefore, τ is 0 initially, where there is no matter, but it will increase along the path of the radiation; in particular, if κ were constant along the path, the optical depth would be just $\tau = \kappa\, l$, with l being the distance traveled ($l = x - x_0$). Appart from the optical depth, we also define the source function as the ratio between the emissivity and the absorption coefficient: $S_\nu \equiv \varepsilon_\nu / \kappa$. Therefore, the equation of transfer can be written as:

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau} = S_\nu - I_\nu \ . \tag{2.4}$$

In an homogeneus medium, both the emissivity and the absorption coefficient are spatially invariant, hence the source function is also spatially invariant. In this case, the solution to the radiative transfer equation along the extension of a region along which radiation is traveling is (Condon and Ransom, 2016):

$$I_\nu(\tau) = I_{\nu,0}\, \mathrm{e}^{-\tau} + S_\nu \left(1 - \mathrm{e}^{-\tau}\right) \ , \tag{2.5}$$

where $I_{v,0}$ is the intensity at the start of the region, where the optical depth is 0: $I_{v,0} = I_v(0)$.

From an observational point of view, radiation is traveling towards the observer along

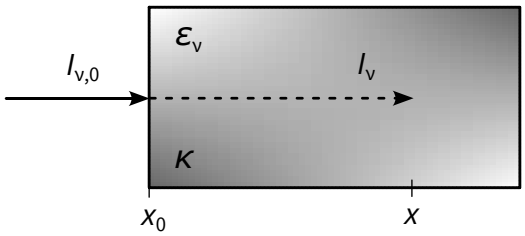


Figure 2.2.1. Scheme of the propagation of radiation through a material. Radiation incides through the material with an specific intensity $I_{\nu.0}$. Given to the absorption coefficient (κ) and the emissivity (ε_ν) of the material, the intensity of the radiation is modified through the path across the material, changing from the initial $I_{\nu.0}$ at the start of the material (x_0) to I_ν in the position x.

the line of sight. Therefore, $I_{v,0}$ corresponds to the background intensity *behind* the bulk of matter located along the line of sight. Therefore, the observed intensity will correspond to the both the background intensity $I_\nu(0)$ and the emission produced by the source function (S_ν) emitted by the matterial, as well as the absorption coefficient (κ, which affects τ) along the path of the radiation while it travels across it. In many cases, $I_\nu(0)$ is just the intensity of the cosmic microwave background.

In the optically thin limit ($\tau \to 0$), the solution of the radiative transfer equation can be written as:

$$I_\nu(\tau) \cong I_{\nu,0} + (S_\nu - I_{v,0})\,\tau \ ; \tag{2.6}$$

that is, intensity would increase linearly with optical depth (if $S_\nu > I_{\nu,0}$). In the optically thick limit ($\tau \to \infty$), the intensity would be equal to the source function: $I_\nu \cong S_\nu$. In this case, intensity does not change along the path of propagation, having a complete equilibrium of radiation with its surroundings. This situation of equilibrium is achieved, for instance, in stellar interiors or in the innermost regions of the molecular dusty envelopes around massive protostars or hot cores.

2.2.1 Local thermodynamical equilibrium

In many cases, the calculations related with radiative transfer can be simplified if there is local thermodynamical equilibrium (LTE). This means that one can consider a unique temperature that locally describes the properties of the material, in a way that there is an equilibrium with the surroundings (Condon and Ransom, 2016). A special case of this would be the full thermodynamical equilibrium (TE), where the unique temperature is spatially invariant, so the whole system is at the same temperature. If there is not local thermodynamical equilibrium (non-LTE), multiple temperatures can be used to describe the behaviour of the system in each local region.

First of all, the *kinetic temperature* (T_k) determines the distribution of velocities (v) of the particles that form the system (atoms, molecules), which can collide with each other. If we consider only one type of particles with mass m, they would be described by the Maxwell-

Boltzmann probability density function (PDF) (Condon and Ransom, 2016):

$$f(v) \;=\; \frac{4/\sqrt{2\pi}}{\sqrt{k_{\rm B}T_{\rm k}\,/m}}\,\frac{\frac{1}{2}mv^2}{k_{\rm B}T_{\rm k}}\exp\left(-\frac{\frac{1}{2}mv^2}{k_{\rm B}T_{\rm k}}\right) \quad . \tag{2.7}$$

This function relates the kinetic energy of the particles ($\frac{1}{2}mv^2$) with the thermal energy ($k_{\rm B}T_{\rm k}$) of the system.[4] With this distribution of velocities, that is isotropic, the most probable velocity is $v_{\rm p} = \sqrt{2\,k_{\rm B}T_{\rm k}/m}$. Similarly, the average velocity is $\langle v\rangle = \sqrt{8/(2\pi)}\,v_{\rm p} \simeq 1.13\,v_{\rm p}$, while the average relative velocity between two particles is $\langle v_{\rm rel}\rangle = \sqrt{2}\,\langle v\rangle = \left(4/\sqrt{2\pi}\right)\,v_{\rm p} \simeq 1.60\,v_{\rm p}$. As a reference, for molecular hydrogen (H_2), the most probable velocities at 10 and 100 K woul be, respectively, $v_{\rm p} \simeq 9.1$ m/s and $v_{\rm p} \simeq 28.7$ m/s.

Now, consider two energy levels in the molecules/atoms that form our system, a lower level with energy $E_{\rm l}$ and an upper level with energy $E_{\rm u}$. Both energy levels are populated with number densities $n_{\rm l}$ and $n_{\rm u}$, respectively. The *excitation temperature* ($T_{\rm ex}$) determines the population ratios between the energy levels, according to the Boltzmann equation (Wilson et al., 2013):

$$\frac{n_{\rm u}}{n_{\rm l}} \;=\; \frac{g_{\rm u}}{g_{\rm l}}\,\exp\left(-\frac{E_{\rm u}-E_{\rm l}}{k_{\rm B}T_{\rm ex}}\right) \quad , \tag{2.8}$$

where $g_{\rm l}$ and $g_{\rm u}$ are the statistical weights of the two states. This means that at higher temperatures, the upper energy level will be more populated. In general, the population of any energy level i with energy E_i can be expressed in terms of the degeneracy of that level, g_i, the total population, n, and the partition function, $Q(T_{\rm ex})$, which depends on the excitation temperature of the molecule ($T_{\rm ex}$), having the following expression (Goldsmith and Langer, 1999):

$$n_i \;=\; \frac{n}{Q(T_{\rm ex})}\,g_i\,\exp\left(\frac{-E_i}{k_{\rm B}T_{\rm ex}}\right) \quad . \tag{2.9}$$

Lastly, if we focus only on the radiation emitted and absorbed by the atoms/molecules present in the system, ignoring the effect of collisions, and if the radiation is in equilibrium with its surroundings, the system should behave as a black body. Here, the *radiation temperature* ($T_{\rm r}$) defines the spectral energy distribution of the system, where the specific intensity (I_ν) emitted towards any direction should follow the Planck's law (Mangum and Shirley, 2015):

$$I_\nu \;=\; \frac{2hc\,(\nu^3/c^3)}{\exp\left(\frac{h\nu}{k_{\rm B}T_{\rm r}}\right)-1} \quad ; \tag{2.10}$$

where h is Planck's constant, $k_{\rm B}$ is Boltzmann's constant, and c is the speed of light.[5]

In local thermodynamical equilibrium (LTE), all these three temperatures are equal ($T_{\rm k} = T_{\rm ex} = T_{\rm r}$), therefore a single temperature (T) can be used to describe locally the thermodynamical properties of the system. This allows for a simplification of the calculations and

[4] Remember that $k_{\rm B}$ is the Boltzmann constant: $k_{\rm B} \simeq 1.38 \times 10^{-23}$ J/K $\cong 8.72 \times 10^{-5}$ eV/K.
[5] $hc \simeq 1240$ eV·nm.

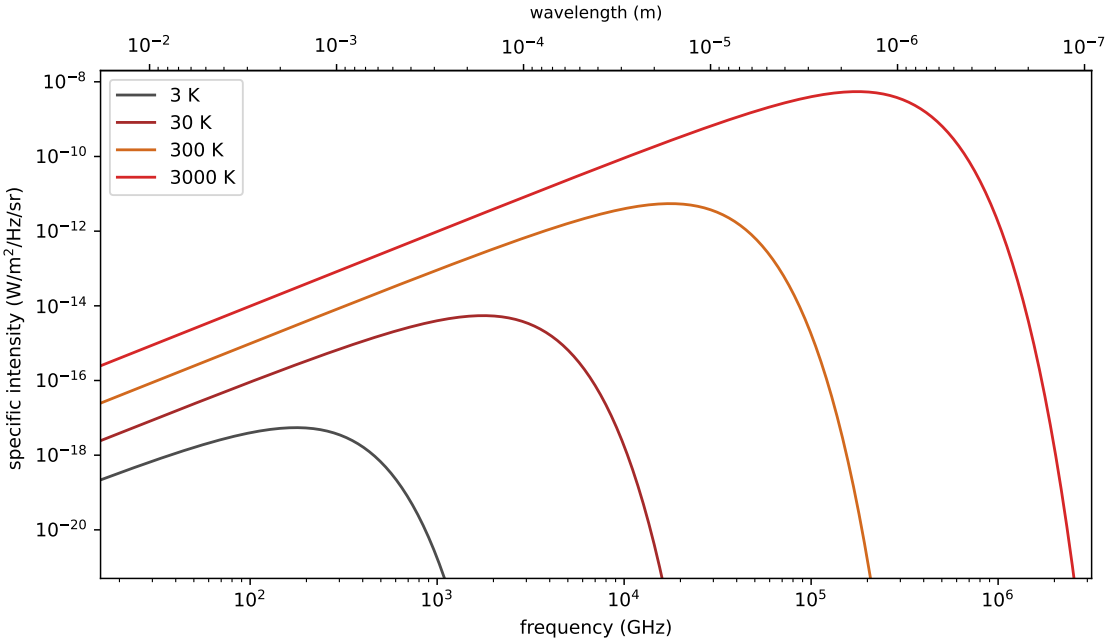


Figure 2.2.2. Specific intensity emitted by a black body at several temperatures, according to Planck's law. In the limit of low frequencies, we can see the behaviour described by the Rayleigh-Jeans formula.

formulas related with radiative transfer, as it is shown in the following subsections.

2.2.2 Black body radiation

Going back of the radiative transfer equation, if the medium is in local thermodynamic equilibrium (LTE), Kirchhoff's law states that the source function (S_ν) is that of a black body (that is, it follows the Planck's law, equation 2.10), which depends on the temperature (T) of the material (Mangum and Shirley, 2015):

$$S_\nu = \frac{2h\nu\,\nu^2/c^2}{\exp\left(\frac{h\nu}{k_B T}\right) - 1} \quad . \tag{2.11}$$

In this case, the majority of the radiation will be emited at a certain spectral range. In particular, the wavenumber of the peak emission (λ_{max}) follows Wien's law: $\lambda_{max} T = 2.898$ mm·K. If the spectral range that we are considering is relatively low in energy compared to the source temperature, $h\nu \ll k_B T$, the black body radiation can be approximed as:

$$S_\nu \cong \frac{2\nu^2}{c^2} k_B T \quad , \tag{2.12}$$

which is the Rayleigh-Jeans law (Fig. 2.2.2). In this regime, the intensity is proportional to temperature. In fact, this relation is used to define the brightness temperature for any intensity I_ν, just inverting the relation (Wilson et al., 2013): $T_b \equiv I_\nu c^2/(2\nu^2 k_B)$. In radioastronomy, intensity is usually measured as a temperature by using this relation (as in Fig. 2.1.4).

In the Rayleigh-Jeans regime ($h\nu \ll k_B T$) and in the optically thin limit ($\tau \to 0$), the solu-

tion to the transfer equation can be written in a simple form:

$$I_\nu(\tau) \;\simeq\; I_{\nu,0} + \left(\frac{2\nu^2}{c^2} k_\mathrm{B} T - I_{\nu,0}\right) \tau \;\; . \tag{2.13}$$

If we consider that the background radiation is the radiation of the cosmic microwave background, then $I_{\nu,0}$ also has the form of a black body: $I_\mathrm{bg} \simeq (2\nu_\mathrm{ul}^2/c^2)\, k_\mathrm{B} T_\mathrm{bg}$, with $T_\mathrm{bg} = 2.73$ K. In terms of brightness temperature, the equation would be very simple:

$$T_\mathrm{b}(\tau) \;\equiv\; \frac{I_\nu(\tau) c^2}{2\nu^2 k_\mathrm{B}} \;\simeq\; T_\mathrm{bg} + (T - T_\mathrm{bg})\, \tau \;\; , \tag{2.14}$$

This means that the brightness temperature increases or decreases linearly with the optical depth, and the rate of change is the difference between the thermodynamical temperature of the material (T) and the temperature of the cosmic microwave background (T_bg). Normally, $T > T_\mathrm{bg}$, so the intensity of radiation increases while it travels across the material (emission is produced along the way; e.g., Fig. 2.1.4); but if $T < T_\mathrm{bg}$, the intensity would decrease along the path (absorption is produced across the path; e.g., Fig. 2.1.6).

2.2.3 Einstein coefficients

We consider again a molecule that has two energy levels with energies E_l and E_u. For now, we neglect the effect of collisions between molecules, hence the density must be low enough, as well as the kinetic temperature.

A molecule in the upper energy level will have a probability of transitioning to the lower energy level, by emiting a photon with a frequency ν_ul so that $E_\mathrm{u} - E_\mathrm{l} = h\nu_\mathrm{ul}$ (Condon and Ransom, 2016). This probability is described by the Einstein coefficient A_ul, the spontaneous emission rate. Additionally, there can be other emissions and absorptions of photons, with the corresponding change in the energy level, due just by the presence of more radiation and its interaction with the molecule. This is controlled by the Einstein coefficients B_lu and B_ul, the stimulated absorption and emission coefficients, respectively. If the intensity of the radiation is I_ν , the stimulated emission and absorption rates will be $B_\mathrm{ul} I_\nu$ and $B_\mathrm{lu} I_\nu$, respectively.[6] The radiation would be emitted and absorbed from any direction isotropically, but here we focus on a specific line of sight.

We assume that the system is in an equilibrium state. Therefore, the number of upward transitions must be equal to the number of downward transitions, which yields the principle of detailed balance (Wilson et al., 2013):

$$A_\mathrm{ul} n_\mathrm{u} + B_\mathrm{ul} n_\mathrm{u} I_\nu \;=\; B_\mathrm{lu} n_\mathrm{l} I_\nu \;\; . \tag{2.15}$$

[6] In many books, the Einstein coefficients are defined so that the stimulated emission and absorption rates depend on the specific energy density (u_ν) of the radiation field instead of the specific intensity (I_ν). For an isotropic medium, both quantities are proportional: $u_\nu = (4\pi/c) I_\nu$. Furthermore, when considering a real system, there would be a spread in energies, and consequently, a line profile centered at ν_ul. In this more realistic case, it is usual to consider the average absorption and emission rates, that depend on the profile-averaged specific intensity ($\bar{I}$) or the profile-averaged specific energy density ($\bar{u}$).

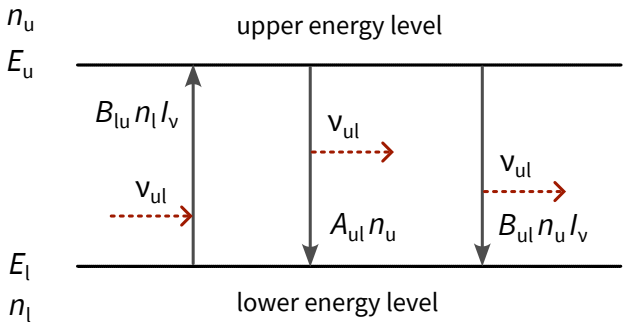


Figure 2.2.3. Scheme of the absorption and emission processes in a two-level system dominated by radiation. The system (e.g., a molecule) has two states with energies E_l and E_u, populated by densities n_l and n_u. The system in the lower level can transition to the upper level absorbing a photon of energy equal to the difference between the levels ($E_\mathrm{u} - E_\mathrm{l} = h\nu_\mathrm{ul}$), with a rate depending on the stimulated absorption coefficient (B_lu) and the intensity of the incident radiation (I_ν). From the upper level, the system can transition to the lower level emmiting a photon of frequency ν_ul, either spontaneously (according to the spontaneous emission coefficient, A_ul) or induced by the presence of other radiation (according to thestimulated emission coefficient, B_ul).

In other words, there is a radiation equilibrium (Fig. 2.2.3). The simpler case would be that collisions are negligible and therefore changes in the energy levels are mainly dominated by radiation transitions (absorptions and emissions). However, this equilibrium in radiation is also fulfilled in local thermodynamical equilibrium, where despite of having collisions between the molecules, following the Boltzmann equation for the occupation of the popuation levels (equation 2.8), the radiation is also in equilibrium.

Additionally, since we have this radiation equilibrium, I_ν must follow the black body emission (equation 2.10). Therefore, combining equations 2.11, 2.8 and 2.15, we obtain the following relations (Condon and Ransom, 2016) between the Einstein coefficients:

$$g_\mathrm{l} B_\mathrm{lu} = g_\mathrm{u} B_\mathrm{ul} \ . \tag{2.16}$$

$$B_\mathrm{ul} = \frac{c^2}{2h\nu_\mathrm{ul}^3} A_\mathrm{ul} \ . \tag{2.17}$$

All in all, this means that there is only one independent Einstein coefficient. The Einstein coefficients are intrinsic properties of each molecule and each pair of energy levels involved in the transition. Thus, even if we assumed a radiation equilibrium to derive these relations, they are fully valid whether or not there is termodynamical equilibrium (TE) or even local thermodynamic equilibrium (LTE).

Rotational transitions

For rotational transitions, the coefficient A_ul can be expressed in terms of the value of the electric dipole moment, μ (Condon and Ransom, 2016):

$$A_\mathrm{ul} = \frac{2}{3} \frac{(2\pi\nu_\mathrm{ul})^3}{\epsilon_0\, hc^3} \mu^2 \ , \tag{2.18}$$

where ϵ_0 is the vacuum permitivity ($\epsilon_0 \simeq 8.85 \cdot 10^{-12}$ C/(V·m)). Here we see that if the molecule has no electric dipole moment, it cannot emit or absorb radiation to transition between rotational energy levels. The corresponding $B_{\rm ul}$ coefficient, using relation 2.17, would be:

$$B_{\rm ul} = \frac{1}{3}\frac{2\pi}{\epsilon_0 \hbar^2 c}\mu^2 \ , \tag{2.19}$$

where $\hbar \equiv h/(2\pi)$. Note that in this case $B_{\rm ul}$ does not deppend on the frequency of the transition (neither does $B_{\rm lu}$, using relation 2.16).

Equation of transfer

The equation of radiative transfer can be written in terms of the Einstein coefficients. The terms in equation 2.15 refer to the number of photons emitted and absorbed between the lower and upper energy levels of the molecule per unit volume (since we use number densities, n). If we convert this to the intensity associated with those photons, we can then recover the radiative transfer equation.

First of all, although we have considered two exact energy levels ($E_{\rm u}$, $E_{\rm l}$), that correspond to a unique photon frequency ($\nu_{\rm ul}$), in reality we have a finite spread in energies (due to thermal variability and micro-turbulent motions, and by Doppler effect). Therefore, we will have a distribution of photon frequencies given by a probability function $\phi(\nu)$, with a peak around $\nu_{\rm ul}$ and normalized to unity when integrating over frequencies.[7] Secondly, each photon with frequency ν carries an energy equal to $h\nu$, which will be distributed over the full solid angle of 4π. Then, we can compute the changes in the energy emitted and absorbed due to these transitions in a particular line of sight, that will define a perpendicular area through which to compute the line intensity. Since the specific intensity I_ν is defined as the change in energy over time (power) per area, per solid angle and per frequency, multiplying equation 2.15 in both sides by $\phi(\nu)h\nu/(4\pi)$ and regrouping, we obtain the changes in intensity along the line of sight ($\mathrm{d}I_\nu/\mathrm{d}x$) due to spontaneous emission and stimulated absorption and emission, that must be equal:

$$\frac{h\nu}{4\pi}A_{\rm ul}n_{\rm u}\phi(\nu) = \frac{h\nu}{4\pi}\left(B_{\rm lu}n_{\rm l} - B_{\rm ul}n_u\right) I_\nu \, \phi(\nu) \ . \tag{2.20}$$

This can be compared to the radiative transfer equation (equation 2.3 for the case of $\mathrm{d}I_\nu/\mathrm{d}x$ = 0 (radiation equilibrium), in a way that we can obtain expressions for the emission and absorption coefficients (Mangum and Shirley, 2015):

$$\varepsilon_\nu = \frac{h\nu}{4\pi}A_{\rm ul}n_{\rm u}\phi(\nu) \ , \tag{2.21}$$

[7] The shape of $\phi(\nu)$ comes from the Maxwell-Boltzmann distribution, performing a variable change from velocity (v) to frequency (ν) through the non-relativistic expression for Doppler effect: $\nu \cong \nu_0(1 - v/c)$. In reality, there is also an intrinsic Gaussian line profile independent of temperature or kinematics, determined by Heisenberg's uncertainty principle, but it is very thin compared to the one produced by Doppler effect.

$$\kappa \;=\; \frac{h\nu}{4\pi}\left(B_{\rm lu}n_{\rm l} - B_{\rm ul}n_u\right)\phi(\nu) \;=\; \frac{1/2}{4\pi}\frac{c^2}{\nu^2}\frac{g_{\rm u}}{g_{\rm l}}A_{\rm ul}n_{\rm l}\left(1-\exp\left(-\frac{h\nu}{k_{\rm B}T}\right)\right)\phi(\nu)\;. \tag{2.22}$$

In the last equation, the relations between the Einstein coefficients (equations 2.16 and 2.17) and the upper and lower populations (equation 2.8) were used to express κ in terms of $A_{\rm ul}$ and $n_{\rm l}$.

2.2.4 Column densities

Consider that we are observing towards a certain position in the sky, where there is a bulk of material that can emit and absorb radiation (thanks to transitions between energy levels), whose furthest and closest points to the observer along the line of sight are x_1 and x_2. The column density (N) quantifies the amount of material along the line of sight toward the observed position in the sky. It is defined as the integral along that line of sight of the number density (Wilson et al., 2013): $N \equiv \int_{x_1}^{x_2} n\,{\rm d}x$, where the integral is performed along the extension of the material (from x_1 to x_2).

Similarly, the optical depth along that line of sight will be the integral of the absorption coefficient from x_1 to x_2: $\tau \equiv \int_{x_1}^{x_2}\kappa\,{\rm d}x$. If the system is in local thermodynamical equilibrium (LTE) and we focus on a certain transition, we can use equation 2.22 as the expression for the absorption coefficient. If we integrate along the line of sight in both sides of the formula, we obtain a relation between the optical depth and the column density of the lower energy level ($N_{\rm l}$, coming from $n_{\rm l}$), for the case of local thermodynamic equilibrium (Condon and Ransom, 2016):

$$\tau \;=\; \frac{1/2}{4\pi}\frac{c^2}{\nu^2}\frac{g_{\rm u}}{g_{\rm l}}A_{\rm ul}N_{\rm l}\left(1-\exp\left(-\frac{h\nu}{k_{\rm B}T}\right)\right)\phi(\nu)\;. \tag{2.23}$$

At this point, it is convenient to do a variable change and shift from frequency (ν) to radial velocity (v). Assuming that the bulk of the material is already moving at a rather constant radial velocity v_0 with respect to the observer, the conversion would be $(v-v_0)/c \cong -(\nu-\nu_{\rm ul})/\nu_{\rm ul}$, where c is the speed of light (the slight approximation comes from the fact that speeds are not relativistic). Remember that $\phi(\nu)$ is a probability density function, so we must change its form to preserve probability. The probability in terms of frequency is $\phi(\nu){\rm d}\nu$, which must be equal to the corresponding probability in terms of velocity, $-\phi(v){\rm d}v$ (the minus sign arises because when frequency increases, radial velocity decreases); therefore, the transformation between the probability densities is $\phi(\nu) = -\phi(v)\,{\rm d}v/{\rm d}\nu$. Given the variable change, we have that ${\rm d}v/c \cong -{\rm d}\nu/\nu_{\rm ul}$, therefore the mentioned derivative is ${\rm d}v/{\rm d}\nu \cong -c/\nu_{\rm ul}$.

At the same time, we should express the frequencies ν into velocities: $\nu = \nu_{\rm ul}\,(1-(v-v_0)/c)$. However, the thermal and micro-turbulent movements of the material emitting or absorbing radiation are far from being relativistic ($v - v_0 \ll c$), and also we know that the line profile $\phi(\nu)$, which is multiplying in the right side of equation 2.23, was centered around

$\nu_{\rm ul}$, so we can do the aproximation of $\nu \cong \nu_{\rm ul}$. Then, equation 2.23 transforms into:

$$\tau \;\cong\; \frac{1/2}{4\pi}\,\frac{c^3}{\nu_{\rm ul}^3}\,\frac{g_{\rm u}}{g_{\rm l}}A_{\rm ul}N_{\rm l}\left(1-\exp\left(-\frac{h\nu_{\rm ul}}{k_{\rm B}T}\right)\right)\phi(v)\;\;. \tag{2.24}$$

Note that $A_{\rm ul}$ can depend on the frequency ν, but we can use the previous reasoning to also change ν by $\nu_{\rm ul}$.

At this point, since we know that $\phi(v)$ is normalized to unity, we can integrate both sides of the equation in radial velocity, to make $\phi(v)$ dissapear. In the left side of the equation, we will have the integral of τ (with respect to velocity, v). In local thermodynamic equilibrium and with a low optical depth ($\tau \ll 1$) the shape of the optical depth will be a Gaussian curve centered around v_0 (Martín et al., 2019):

$$\tau \;\simeq\; \tau_0 \exp\left(-4\,\ln 2\,\left(\frac{v-v_0}{\Delta v}\right)^2\right)\;\;, \tag{2.25}$$

where τ_0 is the peak opacity ($\tau' \ll 1$) and Δv is the full-width at half-maximum (FWHM). Since the curve is Gaussian, it would be more convenient to work with the standard deviation (σ_v), but it in spectroscopy it is common to work with the FWHM, since it is easier to measure; in any case, the relation between them is $\Delta v = 2\sqrt{2\ln 2}\,\sigma_v \simeq 2.35\,\sigma_v$.

Going back to the integral in velocity, we know that the area of a Gaussian of these parameters will be $\tau'\sigma_v\sqrt{2\pi}$. Therefore, integrating in both sides of equation 2.24 in velocity we obtain the following relation (Martín et al., 2019):

$$\frac{1}{2}\sqrt{\frac{2\pi}{2\ln 2}}\,\tau'\Delta v \;\simeq\; \frac{1/2}{4\pi}\,\frac{c^3}{\nu_{\rm ul}^3}\,\frac{g_{\rm u}}{g_{\rm l}}A_{\rm ul}N_{\rm l}\left(1-\exp\left(-\frac{h\nu_{\rm ul}}{k_{\rm B}T}\right)\right)\;\;. \tag{2.26}$$

This expression relates the width and height of the line, in optical depth, with the physical parameters of the transition. Since what one observes in reality is intensity (I_ν) –or, equivalently, brightness temperature ($T_{\rm b}$)–, we will rescue equation 2.13, that relates the intensity with the optical depth in the Rayleigh-Jeans approximation. We apply it for the peak values of optical depth (τ_0) and intensity (I_0), obtaining that: $\tau_0 \cong (I_0 - I_{\rm bg})c^2/(2\nu_{\rm ul}^2 k_{\rm B}(T - T_{\rm bg}))$, where $I_{\rm bg}$ is the background intensity, which is normally the cosmic microwave background, $I_{\rm bg} \simeq (2\nu_{\rm ul}^2/c^2)\,k_{\rm B}T_{\rm bg}$, with $T_{\rm bg}$ = 2.73 K (Wilson et al., 2013).[8] It is worth noticing that if $T > T_{\rm bg}$, the line will appear in emission ($I_0 > I_{\rm bg}$), while if $T < T_{\rm bg}$ the line would appear in absorption ($I_0 < I_{\rm bg}$); in both cases, the optical depth is positive ($\tau_0 > 0$). If $T = T_{\rm bg}$, there would be no line.

Finally, since we are interested in the total column density of the molecule (N), rather of the column density of a certain energy level, we will use equation 2.9, but applied to N instead of n, to obtain an expression including N. Additionally, we use the fact that $h\nu_{\rm ul} = E_{\rm u} - E_{\rm l}$. Then, substituting all the terms mentioned in the last paragraph in equation 2.26

[8] In these equations, we use the same reasoning as before to substitute ν for $\nu_{\rm ul}$.

and reagrouping terms, we obtain the following relation (Martín et al., 2019):

$$(I_0 - I_{\rm bg})\,\Delta v \;\simeq\; \frac{\sqrt{\frac{2\ln 2}{2\pi}}}{2\pi}\,\frac{k_{\rm B}(T-T_{\rm bg})}{Q(T)}\,\frac{c}{\nu_{\rm ul}}A_{\rm ul}g_{\rm u}N\left(\exp\left(\frac{-E_{\rm l}}{k_{\rm B}T}\right)-\exp\left(\frac{-E_{\rm u}}{k_{\rm B}T}\right)\right)\quad . \tag{2.27}$$

This enormous formula is in the end a relation between the observed parameters of the line (peak intensity, I_0; width, Δv; and central frequency, $\nu_{\rm ul}$), the physical parameters associated with the molecule (the column density, N; and the temperature, T), and with the physical properties of the molecular transition (the Einstein coefficient $A_{\rm ul}$, the partition function $Q(T)$, the degeneracy $g_{\rm u}$, and the energies $E_{\rm u}$, $E_{\rm l}$. We should not forget either about the background intensity, usually also in the Rayleigh-Jeans regime: $I_{\rm bg} \simeq (2\nu_{\rm ul}^2/c^2)\,k_{\rm B}T_{\rm bg}$ (with $T_{\rm bg}$ = 2.73 K in the case of the cosmic microwave background).

The last mentioned parameters of the previous formula are usually characterized and tabulated thanks to spectroscopical experiments and quantum chemical simulations. Actually, for the case of rotational transitions, we can use the computed expression of the Einstein coefficient $A_{\rm ul}$ with respect to the molecular dipolar moment, μ (equation 2.18). Therefore, the above formula would be transformed into:

$$(I_0 - I_{\rm bg})\,\Delta v \;\simeq\; \frac{2}{3}(2\pi)^2\sqrt{\frac{2\ln 2}{2\pi}}\,\frac{k_{\rm B}(T-T_{\rm bg})}{\epsilon_0\,h\,Q(T)}\,\frac{\nu_{\rm ul}^2}{c^2}\,\mu^2 g_{\rm u}N\left(\exp\left(\frac{-E_{\rm l}}{k_{\rm B}T}\right)-\exp\left(\frac{-E_{\rm u}}{k_{\rm B}T}\right)\right)\,. \tag{2.28}$$

This formula can be applied to several transitions of the same species, adjusting the spectroscopic parameters accordingly. Then, observing several lines (normally in emission), one can characterize the column density and also the temperature of the molecule. The column density modulates the intensity of all the lines, while the temperature modulates the relative intensities between the different observed transitions; note that in all the lines the velocity width should be the same. If there is only one transition available, then one should fix the temperature of the molecule to obtain an estimate of the column density.

Lastly, we should mention that there are generalizations of the previous formulas for the case of larger values of optical depth and for the cases outside the Rayleigh-Jeans regime, that can be checked in the works by Goldsmith and Langer (1999) and by Martín et al. (2019). For higher optical depths, the shape of the lines start to broaden and flatten at the peak, therefore deviating from a Gaussian profile.

2.2.5 MADCUBA and SLIM

Based on the explained theoretical background, there is a software used in this thesis that allows to analyse rotational lines observed in radioastronomical observations: MADCUBA (***Ma**drid **D**ata**cub**e **A**nalysis*). It is a software developed at the Centre for Astrobiology (Madrid, Spain) to analyse astronomical datacubes and spectra from the main astronomical facilities.

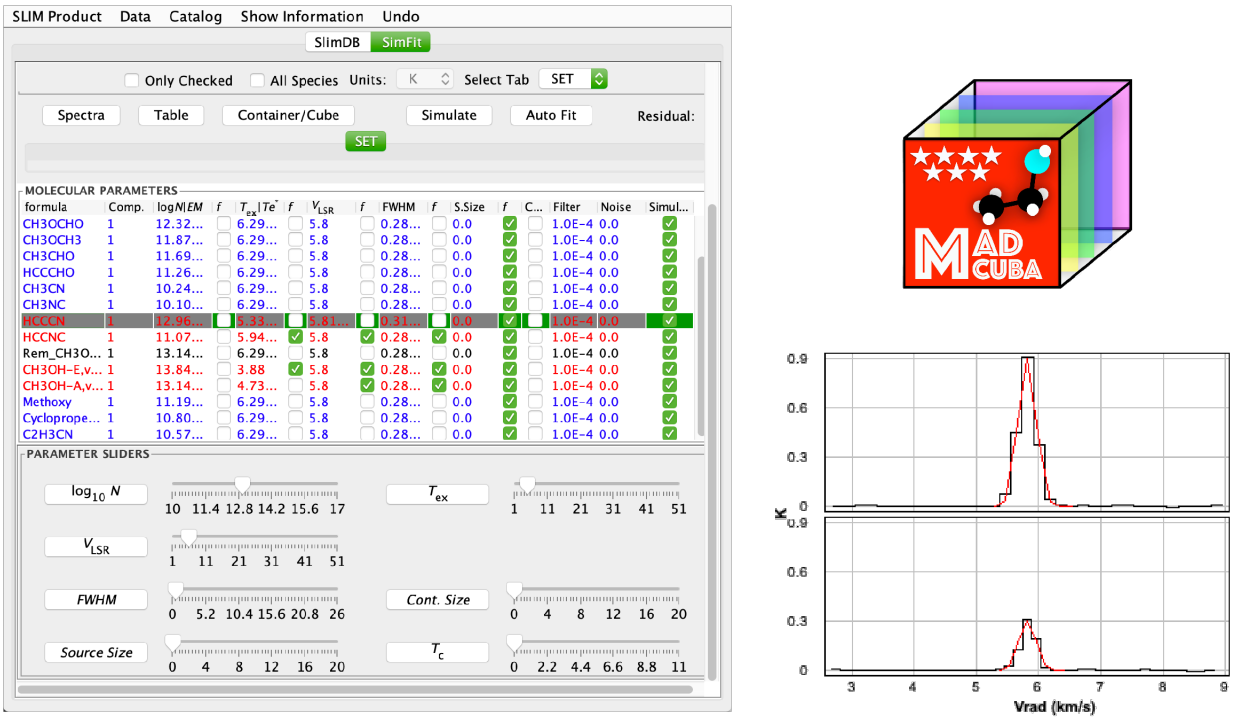


Figure 2.2.4. Screenshots and logo of the software MADCUBA. In the left we show the principal graphical user interface of SLIM, showing the list of targeted species. For the selected species in the list (HCCCN, cyanoacetilene), the panels on the right show the transitions found on the observed spectra (in black) and the synthetic profile (in red); the horizontal axis shows the radial velocity (in km/s), while the vertical axis depicts the brightness temperature (in K).

It is open source, written in Java and it has a user-friendly graphical interface.[9] Within this software, there is a tool called SLIM (*Spectral Line Identification and Modelling*), designed specifically to analyse astronomical spectra, and described in detail in the paper by Martín et al. (2019).

First of all, MADCUBA allows to import astronomical spectra (such as in FITS format), process them, and visualize them. Then, the tool SLIM allows to search molecular transitions in the loaded spectra thanks to several spectroscopic catalogs that are incorporated within the software: mainly, the Cologne Database for Molecular Spectroscopy (CDMS; Endres et al., 2016) and the Jet Propulsion Laboratory (JPL) molecular catalogue (Pickett et al., 1998). After manually selecting the molecules of interest, SLIM automatically searches the transitions of the selected species searched for and covered in the frequency range of the observed spectra. Finally, SLIM includes the possibility of modelling the observed lines in local thermodynamical equilibrium (LTE) conditions, generating Gaussian profiles to fit the observations, that can be fitted to the observed lines thanks to its *autofit* option.

Within the autofit, there are several tunable parameters: column density of the molecule (N), excitation temperature (T), local sidereal rest (LSR) central velocity of the line ($v_{\rm LSR}$) and full-width at half maximum (FWHW) of the line (Δv).[10] Usually, these parameters are first

[9] `https://cab.inta-csic.es/madcuba/`

[10] There are some additional parameters (source size, continuum size and continuum temperature), but

manually initialized at some values that produce a reasonable match to the observed spectrum, and then they can be automatically fitted with the autofit option. The line profiles are generated using an equation similar to equation 2.28 presented here. Then, SLIM uses an optimization method to minimize the difference between the observed spectra and the model, for all the covered transitions for each molecule. As a result, we obtain the fitted values for each parameter with its uncertainty. Figure 2.2.4 shows some screenshots of MADCUBA and SLIM.

2.2.6 Non-local thermodynamical equilibrium

In previous subsections we have assumed local thermodynamical equilibrium (LTE), which leads to the energy population ratios given by the Boltzmann distribution (equation 2.8), which is defined by a unique temperature T.

However, this is not always the case. If densities are not high enough, inelastic collisions with abundant species –normally, molecular hydrogen, H_2– might need to be taken into account. In these cases, kinetic energy is exchanged between the molecule and the collision partner (H_2), hence the molecule can be excited and de-excited. The *kinetic* temperature, T_k, defines the distribution of relative velocities between the colliding particles, v_r, following the Maxwell-Boltzmann probability density function (PDF) (Condon and Ransom, 2016):

$$f(v_r) \;=\; \frac{4/\sqrt{2\pi}}{\sqrt{k_B T\,/\,m_r}}\,\frac{\frac{1}{2}m_r v_r^2}{k_B T}\exp\left(-\frac{\frac{1}{2}m_r v_r^2}{k_B T_k}\right) \quad . \tag{2.29}$$

where this time m_r is the reduced mass of the colliding particles (that have masses m_1 and m_2), defined as $m_r = m_1 m_2\,/\,(m_1 + m_2)$.

The probability of an individual collision that excites a molecule from a lower energy level E_l to an upper level E_u is the expected value of the product of the collision cross section (σ_{ul}) and the relative velocity (v_r):

$$\gamma_{ul} \;=\; \langle \sigma_{ul} v_r \rangle \;=\; \int_0^\infty \sigma_{ul}(v_r)\, v_r\, f(v_r)\, \mathrm{d}v_r \quad , \tag{2.30}$$

and it is called collision rate coefficient (Condon and Ransom, 2016); as a rough estimate, the formula can be simplified as $\gamma_{ul} \approx \langle \sigma_{ul} \rangle \langle v_r \rangle = \langle \sigma_{ul} \rangle \left(4/\sqrt{2\pi}\right) \sqrt{k_B T\,/\,m_r}$. Then, the collision rate per molecule of interest will be $C_{ul} = n_c \gamma_{ul}$ (Wilson et al., 2013), where n_c is the number density of the collisional partner (usually, H_2).

We can now use the Einstein coefficients (A_{ul}, B_{ul}, B_{lu}) to write the number of transitions from one level to another, and use the principle of detailed balance to equal the number of transitions upwards and downwards. The number of downwards transitions due to spontaneous emission is $A_{ul} n_u$, the number due to stimulated emission is $B_{ul} n_u I_\nu$, and the number due to collisions (the system transfers energy to the collisional partner) is $C_{ul} n_u$; similarly, the number of upwards transitions due to stimulated absorption is $B_{lu} n_l I_\nu$, and the number

they are not used in this thesis (they are all set to the default paramters, i.e., zero).

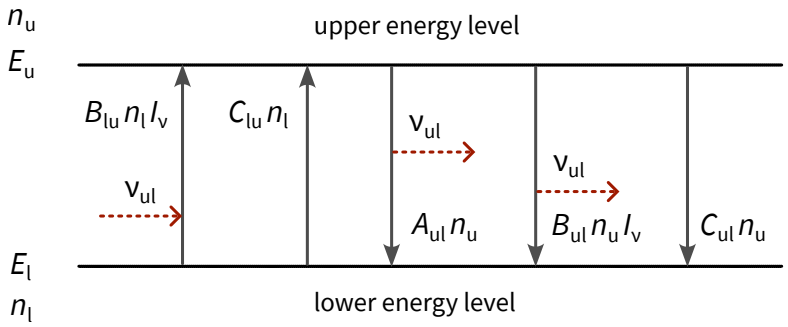


Figure 2.2.5. Scheme of the absorption and emission processes in a two-level system. The system (e.g., a molecule) has two states with energies E_l and E_u, populated by densities n_l and n_u. The system can transition from one level to another either by radiation processes (following the Einstein coefficients B_{lu}, A_{ul}, B_{ul}, and on the external radiation I_ν) or by collisions with other abundant species –normally, H_2– (following the collision coefficients C_{lu} and C_{ul}, which are proportional to the density of the collisional partner).

due to collisions (the collisional partner transfers energy to the system) is $C_{lu}n_l$. Then, in an equilibrium state, the number of upwards transitions must be equal to the number of downwards transitions (detailed balance; Condon and Ransom, 2016):

$$A_{ul}n_u + B_{ul}n_u I_\nu + C_{ul}n_u \;=\; B_{lu}n_l I_\nu + C_{lu}n_l \;, \tag{2.31}$$

represented schematically in Fig. 2.2.5. In local thermodynamical equilibrium (LTE), the number of updwards transitions due to collisions is equal to the number of downwards collisional transitions: $C_{ul}n_u = C_{lu}n_l$. Similarly, the number of emissions due to radiation would be equal to the number of absorptions. In other words, in LTE both radiation and collisions are in statistical equilibrium.

If collisions dominate the system, the population levels of our system will follow the Boltzmann equation with the kinetic temperature T_k: $n_u/n_l = (g_u/g_l)\exp(-h\nu_{ul}/(k_B T_k))$, where g_u and g_l are the statistical weights of the upper and lower energy levels, and where $h\nu_{ul} = E_u - E_l$. At the same time, from the detailed balance we would get that $C_{ul}n_u = C_{lu}n_l$. Then, it just follows that (Condon and Ransom, 2016):

$$\frac{C_{lu}}{C_{ul}} \;=\; \frac{g_u}{g_l}\,\exp\left(-\frac{E_u - E_l}{k_B T_k}\right) \;. \tag{2.32}$$

This relation is always fulfilled, since the collisional rates do not change if the system is under local thermodynamical equilibrium or not, or if collisions dominate or not.

On the other hand, we assume that the expression of the specific intensity I_ν is given by a black body with a radiation temperature T_r. Similarly as before, we can now assume that radiation transitions dominate (collisions are negligible), to derive the relations between the Einstein coefficients already derived in Section 2.2.3 (equations 2.16 and 2.17). These relations must hold even if radiation transitions do not dominate, as the Einstein coefficients remain the same (like the collisional rates).

Combining all of these relations, we can derive the excitation temperature of the system (T_{ex}), which defines the population ratios through Boltzmann equation, as a function of the

radiation and kinetic temperatures Wilson et al., 2013:

$$T_{\rm ex} \;=\; T_{\rm k}\,\frac{T_{\rm r}A_{\rm ul}+T_0C_{\rm ul}}{T_{\rm k}A_{\rm ul}+T_0C_{\rm ul}}\;, \tag{2.33}$$

where $T_0 \equiv h\nu_{\rm ul}/k_{\rm B}$. From this expression, we find that if collisions dominate ($C_{\rm ul} \gg A_{\rm ul}$), the excitation temperature approaches the kinetic temperature ($T_{\rm ex} \cong T_{\rm k}$), while if radiation dominates ($A_{\rm ul} \gg C_{\rm ul}$), the excitation temperature approaches the radiation temperature ($T_{\rm ex} \cong T_{\rm r}$). In practice, since the collision rate is proportional to the density of the collisional partner ($C_{\rm ul} = n_{\rm c}\gamma_{\rm ul}$), collisions become important when the density is high enough. A reference value for this is the critical density, defined as the density in which the collision rate equals the spontaneous emission rate ($C_{\rm ul} = A_{\rm ul}$), therefore having $n_{\rm crit} = A_{\rm ul}/\gamma_{\rm ul}$. If the density is much greater than this critical value, the molecular emission/absorption will be dominated by collisions, being in local thermodynamical equilibrium (LTE). The critical density depends on the spectroscopic properties of the molecular line, therefore some lines for certain molecules can be in a state of LTE while others may not in the same region.

Appart from this derivation, there are other approaches that allow to obtain expressions that relate the optical depth and peak intensities of the lines with the column density of the molecule, similarly as equation 2.28 for local thermodynamical equilibrium (LTE). Therefore, for non-LTE the column density and the excitation temperature of the molecule can be derived by observing several transition lines. For this, the kinetic temperature of the system, the H_2 number density, and the collisional rate coefficients for the transitions ($\gamma_{\rm ul}$) between the molecule and H_2 must be known beforehand (it can be estimated via quantum mechanical calculations or measured by experiments). For more details, see the work by van der Tak et al. (2007).

2.2.7 RADEX Line Fitter

RADEX is an open-source software, written in Fortran and published by van der Tak et al. (2007),[11] that calculates the intensities of atomic and molecular transition lines produced in a uniform medium, assuming a non-local thermodinamical equilibrium (non-LTE), as explained in the previous subsection. It makes use of molecular data files, including collisional coefficients, available in the **L**eiden **A**tomic and **M**olecular **Da**tabase (LAMDA).[12]

RADEX calculates the predicted peak intensities of the lines given certain values of the H_2 number density, gas kinetic temperature, background temperature, molecular column density, and linewidth of the observed emission.

In order to use RADEX in an integrated way with Python and to allow for a fit of the observed lines to derive the physical parameters (mainly the column density, N), including an uncertainty estimation, a Python script has been written as a part of this thesis that allows

[11] `https://personal.sron.nl/~vdtak/radex/index.shtml`
[12] `https://home.strw.leidenuniv.nl/~moldata/`

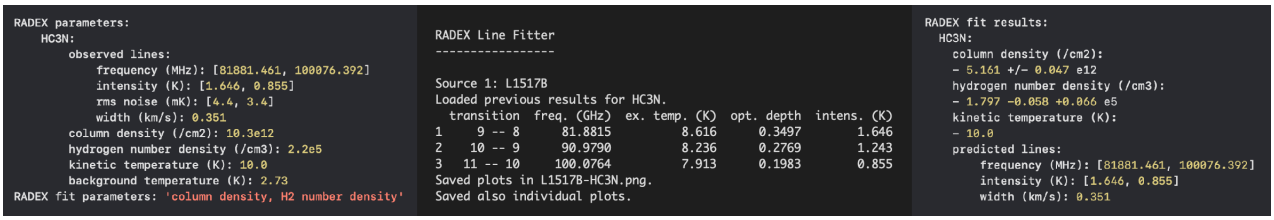


Figure 2.2.6. Some screenshots of RADEX Line Fitter, as a simple use example. *Left*: Input configuration file in YAML formatting style, for two observed transition of cyanoacetilene (HC_3N) in the starless core L1517B. *Center*: Results of the fit performed using RADEX. *Right*: Output file produced containing the results of the fit, mainly the HC_3N column density and the H_2 number density.

for such task: it is called RADEX Line Fitter. It has been published in GitHub,[13] including some examples and a complete user guide, as well as in the Astrophysics Source Code Library (ASCL).[14] It requires RADEX to be installed locally in order to use it

RADEX Line Fitter needs a configuration file written in the YAML formatting style (plain text) specifying the input parameters for RADEX as well as the information regarding the observed transitions –mainly, the observed intensities with the corresponding noise, and the observed line width in radial velocity. The scripts performs an automatic fit where three posible parameters are optimized: the molecular column density only; the molecular column density as well as the H_2 number density; and the molecular column density as well as the kinetic temperature. In any of these three cases, the one-dimensional or bi-dimensional parameter space is explored to search for the optimal solution and to estimate the uncertainty in the fitted parameters. In principle, the three-dimensional parameter space of the molecular column density, the kinetic temperature and the H_2 number density could also be explored, obtaining three values as a solution. However, this is not supported within RADEX Line Fitter, since it would require quite a lot of computational time, and furthermore it would require at least three transitions to avoid degeneracies. Figure 2.2.6 shows some screenshots of a single fit done with RADEX Line Fitter, for two observed transition of cyanoacetilene (HC_3N) in the starless core L1517B, corresponding to the observations presented in Chapter 3. Additionally, Appendix B.1 presents a more detailed explanation of the fitting procedure, including plots of the fit.

2.2.8 Column densities for ice absorption spectra

In the previous subsections, we have seen how to obtain the column densities by observing several rotational transition lines (normally in emission), either assuming local thermodynamical equilibrium (LTE) or not in molecular clouds.

In this thesis, however, we also analyze the infrared (IR) absorption features produced by radiation propagating through a molecular cloud and that gets absorbed by interstellar ices. Given a sufficiently bright background source of light, with specific intensity $I_{\nu,0}$, the

[13] https://github.com/andresmegias/radex-python/
[14] https://ascl.net/2302.022/

interstellar ices located in our line of sight will absorb part of the light, yielding a molecular absorption feature characterized by an absorbed intensity I_ν. At the same time, ices do not emit radiation and therefore the source function (S_ν) of the material will be zero.

In this case, the solution of the equation of transfer would be simply:

$$I_v = I_{v,0}\, \mathrm{e}^{-\tau} \ . \tag{2.34}$$

This is indeed the Lambert-Beer's law (Swinehart, 1962; Mayerhöfer et al., 2020). In particular, the optical depth (τ) will be centered around the central frequency of the line or molecular band, absorbing light in its surroundings. From this, the optical depth can thus be calculated as:

$$\tau = \ln(I_{\nu,0}/I_\nu) \ . \tag{2.35}$$

This relation also holds if we use spectral flux ($\mathcal{F}_\nu$) instead of specific intensity (I_ν), since they are proportional. In particular, $\mathrm{d}\mathcal{F}_\nu = I_\nu\, \mathrm{d}\Omega$, with Ω being the solid angle. Therefore, the emitted spectral flux for a source emitting isotropically in all directions would be $\mathcal{F}_\nu = 4\pi\, I_\nu$.

In the scale of optical depth, the contributions of several layers of materials is additive, that is, the optical depth is proportional to the column density of the molecule that is absorbing the light. This constant of proportionality can be measured in laboratory experiments by integrating the observed absorbance spectrum in optical depth and dividing by the column density (which has to be obtained experimentally by other methods), yielding what is called *band strength*, S (Muñoz Caro and Escribano, 2018).

Therefore, to obtain the column density of a molecule in an ice absorption spectrum, one has to integrate a certain band of the molecule (from ν_1 to ν_2, covering all the band) and then divide by the corresponding experimentally-derived band strength (S_{ν_1,ν_2}), obtaining (Muñoz Caro and Escribano, 2018):

$$N = \frac{\int_{\nu_1}^{\nu_2} \tau(\nu)\, \mathrm{d}\nu}{S_{\nu_1,\nu_2}} \ . \tag{2.36}$$

In practice, the band strengths and the optical depth are integrated in wavenumber instead of frequency, which is the inverse of the wavelength: $\tilde{\nu} = 1/\lambda$ (the wavenumber is proportional to the frequency, that is why it is notated as $\tilde{\nu}$). Sometimes, a Gaussian line is fitted to the band and integrated, instead of integrating the spectrum itself.

This method is used to determine column densities of interstellar ices. However, this method can involve uncertainties up to 20 %, since the shape of the bands can change with temperature and with the total composition of the ice mixture, and the band strengths can change accordingly (Muñoz Caro and Escribano, 2018).

2.3 Astronomical observations

This section describes the basics of the observations used in this thesis. In summary, they are spectra taken toward single positions in different molecular clouds, using different spectral ranges: radio for rotational transitions and infrared for vibrational transitions.

2.3.1 Radio astronomical observations

Radio telescopes are made of big parabolas that reflect and concentrate radiation into the secondary mirror and then to the receiver, inside the external structure. More technical aspects of radiotelescopes can be seen in Section 2.4.1, which talks in particular about the IRAM 30 m radio telescope.

2.3.1.1 Antenna temperature

Instead of working with intensities or power, usually temperature is used as a magnitude to express the received flux by the radio telescope, since it can be related to the incident brightness temperature (T_b, which itself is related to the incident intensity I_ν). In particular, the *antenna temperature* is defined as the specific power ($\mathcal{P}_\nu$) generated in the antenna receiver due to the incident radiation (I_ν) divided by Boltzmann's constant (Condon and Ransom, 2016):

$$T_\mathrm{A} \equiv \mathcal{P}_\nu / k_\mathrm{B} \ . \tag{2.37}$$

This would be the temperature corresponding to an imaginary resistor that would consume the same amount of specific power ($\mathcal{P}_\nu$) in the Rayleigh-Jeans regime ($h\nu \ll k_\mathrm{B}T$), since in this regime that power would be $P_\nu \simeq k_\mathrm{B}T$ (Condon and Ransom, 2016). In reality, the received power has being partially absorbed by the atmosphere (although in the radio window this is not usually high), and it could also be contaminated by atmospheric thermal emission and other minor effects related to the telescope. In practice, these effects can be corrected, and usually the corrected antenna temperature (that is, removing the mentioned effects) is denoted as T_A^* (Wilson et al., 2013). Sometimes this T_A^* is actually the *forward beam* brightness temperature, in which case one should multiply it by the forward efficiency of the telescope to obtain the usual antenna temperature. In any case, we will continue using T_A for the sake of simplicity.

On the other hand, the received specific power is related to the specific flux ($\mathcal{F}_\nu$) received in the effective aperture area of the telescope (A_eff): $\mathcal{P}_\nu = \mathcal{F}_\nu A_\mathrm{eff}$. At the same time, the received spectral flux is related to the specific intensity (I_ν) received from a solid angle (Ω) in the sky: $\mathrm{d}\mathcal{F}_\nu = I_\nu \, \mathrm{d}\Omega$. This gives $\mathrm{d}\mathcal{P}_\nu = A_\mathrm{eff} \, I_\nu \, \mathrm{d}\Omega$. The effective area would be more or less the area of a circle with the telescope radius, including the effect of curvature, optical distortions and construction imperfections.

However, the antenna does not collect the radiation from all the sky. In contrast, it has a spatial sensitivity, measured as the *antenna pattern*, $P(\Omega)$. It usually has a symmetric main beam (usually Gaussian), centered at the position of the sky where the telescope is pointing, with some additional side lobes whose sensitivity decreases as one moves away from the pointing position (Wilson et al., 2013). The value of the antenna pattern ranges from 0 to 1, being 1 in the pointing direction. Additionally, in practice radio detectors only detect one polarization in the radiation, so there is a factor of $1/2$ between the received and the measured intensity. All in all, the spectral power received by the telescope is (Condon and Ransom, 2016):

$$\mathcal{P}_\nu \;=\; \frac{1}{2} A_{\text{eff}} \int_{\bigcirc} I_\nu(\Omega)\, P(\Omega)\, \mathrm{d}\Omega \;\;, \tag{2.38}$$

where the integral is done over all the sky hemisphere ($\bigcirc$).[15] We can define the *beam solid angle* of the antenna as the integral of the antenna pattern over all the hemisphere: $\Omega_{\text{A}} = \int_{\bigcirc} P(\Omega)\, \mathrm{d}\Omega$. It can be interpreted as the size (in solid angle) of an imaginary antenna pattern consisting of a conic beam that would have a sensitivity of 1 inside it and 0 outside it, so that the integral of the pattern would be the same as with the real pattern, $P(\Omega)$. Note that the beam solid angle Ω_{A} is related to the effective area of the telescope (related to the telescope radius) and the wavelength (λ) of the incident radiation, such as $A_{\text{eff}}\Omega_{\text{A}} = \lambda^2$. [16] Since $c = \lambda\nu$, we obtain that $\frac{1}{2}A_{\text{eff}} = \frac{1}{2}(c^2/\nu^2)/\Omega_{\text{A}}$. We can now move the factor $\frac{1}{2}\, c^2/\nu^2$ into the integral and, remembering the definition of brightness temperature, $T_{\text{b}} \equiv I_\nu c^2/(2k_{\text{B}}\nu^2)$, we obtain the following expression:

$$\mathcal{P}_\nu \;=\; \frac{1}{\Omega_{\text{A}}} \int_{\bigcirc} k_{\text{B}} T_{\text{b}}(\Omega)\, P(\Omega)\, \mathrm{d}\Omega \;\;. \tag{2.39}$$

Equivalently, the antenna temperature ($T_{\text{A}} \equiv \mathcal{P}_\nu/k_{\text{B}}$) would be the average of the brightness temperature along the antenna pattern:

$$T_{\text{A}} \;=\; \frac{1}{\Omega_{\text{A}}} \int_{\bigcirc} T_{\text{b}}(\Omega)\, P(\Omega)\, \mathrm{d}\Omega \;\;. \tag{2.40}$$

In other words, the measured antenna temperature is the average of the brightness temperature of the source convolved with the antenna pattern.

2.3.1.2 Source size

There are some special cases in which the above integral can be simplified. We define the source size, Ω_{s}, as the solid angle of the region of the sky from which the received intensity (I_ν) is coming. If the source is much larger than the beam solid angle (extended source,

[15] The solid angle Ω can be decomposed into two angles, the azimuth φ and the polar angle θ. Then, the differential solid angle would be $\mathrm{d}\Omega = \sin\theta \mathrm{d}\varphi\, \mathrm{d}\theta$. Normally, the antenna pattern, $P(\theta, \varphi)$, is symmetric with respect to the pointing position (θ_0, φ_0).

[16] This can be demonstrated with an imaginary experiment in which a cavity in thermodynamic equilibrium containing a resistor is coupled to an antenna, both elements being at the same temperature, and with a filter blocking electromagneting radiation but allowing the flow of currents having frequencies in the range of ν and $\mathrm{d}\nu$, by using the conservation of energy (Condon and Ransom, 2016).

$\Omega_s \gg \Omega_A$), and under the assumption of an isotropic intensity distribution, the antenna temperature is just the brightness temperature:

$$T_A \cong \frac{1}{\Omega_A} T_b \int_{\bigcirc} P(\Omega)\, d\Omega = T_b \ . \tag{2.41}$$

In the contrary case in which the intensity I_ν, isotropic, is emitted from a region in the sky much smaller than the beam ($\Omega_s \ll \Omega_A$), the integral in equation 2.38 would be applied over the source size:

$$T_A \cong \frac{1}{\Omega_A} T_b \int_s P(\Omega)\, d\Omega \cong \frac{1}{\Omega_A} T_b \int_s d\Omega = T_b \frac{\Omega_s}{\Omega_A} \ , \tag{2.42}$$

where we have used that $P(\Omega) \cong 1$ in the region of the source size, since $\Omega_s \ll \Omega_A$. The antenna temperature would be then a fraction of the brightness temperature. It is said that the brightness temperature is being diluted in the antenna pattern, obtaining the measured antenna temperature; the ratio Ω_s/Ω_A is called *filling factor*, which relates to the brightness temperature with the measured antenna temperature (Condon and Ransom, 2016). Then, the brightness temperature would be higher than the measured antenna temperature: $T_b \cong T_A\, \Omega_A/\Omega_s$.

2.3.1.3 Main beam efficiency

The antenna pattern does not only have a main beam, but also some residual side lobes, which allow radiation from outside the main beam to come into the measured signal (T_A) and increase it. We define the main beam efficiency, η_{mb}, as the ratio between the integral of the antenna pattern along the main beam and the integral around all the hemisphere (Wilson et al., 2013; Condon and Ransom, 2016). The last integral would be the beam solid angle (Ω_A), and the first one can be interpreted as the *main beam* solid angle (Ω_{mb}). Again, we could imagine a conic antenna pattern with solid angle Ω_{mb} inside which the sensitivity is 1 and outside of which is 0. Then, the ratio between these solid angles would be the beam efficiency, $\eta_{mb} \equiv \Omega_{mb}/\Omega_A$, which is less than 1 ($\Omega_{mb} < \Omega_A$). It is more common to use the main beam solid angle (Ω_{mb}) rather than the full beam solid angle (Ω_A) of the antenna, since it is easy to estimate –as explained in next paragraph–, that is why the main beam efficiency is commonly used.

The shape of the main beam is usually well approximed by a bidimensional Gaussian (Condon and Ransom, 2016; see Figure 2.3.1). Furthermore, the side lobes are normally located far enough from the main beam so that there is a clear minimum between the main beam and the next side lobes, that we call effective beam amplitude, ϑ^*_{mb}. Then, Ω_{mb} can be approximed as the volume under the Gaussian (that is, the integral of a one-dimensional Gaussiann squared); if $\sigma_{\mathbf{mb}}$ is its standard deviation, this volume will be $2\pi\sigma^2_{mb}$. Thus, the main beam solid angle can be approximed as $\Omega_{mb} \cong 2\pi\sigma^2_{mb}$, being σ_{mb} the standard-deviation width of the main beam (Condon and Ransom, 2016). In practice, it is more common to express the width of the main beam as the half-power beam width (HPBW), which is indeed the same as the full-width at half maximum (FWHM). The relation between

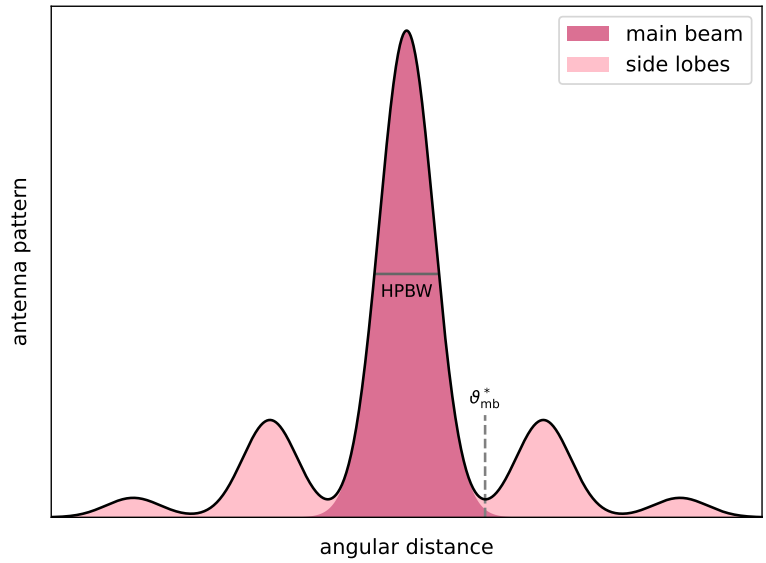


Figure 2.3.1. Scheme of a prototypical antenna pattern, with a Gaussian main beam and several side lobes. The half-power beam width (HPBW) and the effective main beam amplitude (ϑ^*_{mb}) are indicated with gray lines.

the HPBW (θ_{mb}) and the standard deviation is just $\theta_{mb} = 2\sqrt{2\ln 2}\,\sigma_{mb} \simeq 2.35\,\sigma_{mb}$. Then, knowing the HPBW we can easily estimate the main beam solid angle, $\Omega_{mb} \cong 2\pi\sigma^2_{mb} = (2\pi/(8\ln 2))\,\theta^2_{mb} \simeq 1.13\,\theta^2_{mb}$ (Wilson et al., 2013). Actually, the HPBWvaries with the frequency of the incident radiation, due to the diffraction limit. It is related with the diameter (d) of the antenna and the wavelength (λ) of the radiation, given as $\theta_{mb} \approx 1.2\,\lambda/d$ (Condon and Ransom, 2016).

Main beam temperature

Now we will assume a simple case in which the spatial distribution of the source intensity (I_ν) is isotropic with an extension in the sky equal to the main beam ($\Omega_s = \Omega_{mb}$). In this case, the integral will be calculated just around the main beam (Ω_{mb}), and since the intensity is constant it will be just $I_\nu\Omega_{mb}$. Then, we can convert the antenna solid angle Ω_A in Ω_{mb}/η_{mb}. Therefore:

$$T_A = \frac{1}{\Omega_A}\int_{\bigcirc} T_b(\Omega)\,P(\Omega)\,d\Omega = \frac{\eta_{mb}}{\Omega_{mb}}T_b\int_{mb} P(\Omega)\,d\Omega = \eta_{mb}T_b \ . \tag{2.43}$$

Thus, the brightness temperature (T_b) will be just T_A/η_{mb}. We define this ratio in general as the *main beam temperature*, $T_{mb} \equiv T_A/\eta_{mb}$ (Condon and Ransom, 2016). It corresponds to the brightness temperature of a source emitting uniformly in the main beam size that would produce the same antenna temperature (T_A) as the actual source. In this particular case, $T_{mb} = T_b$.

Going back to the simple case of a uniform source much smaller than the antenna pattern ($\Omega_s \ll \Omega_A$, and also $\Omega_s \ll \Omega_{mb}$), we can obtain the relation between the measured main

beam temperature and the brightness temperature:

$$T_{\rm mb} \equiv \frac{T_{\rm A}}{\eta_{\rm mb}} \cong T_{\rm b} \frac{\Omega_{\rm s}}{\eta_{\rm mb}\Omega_{\rm A}} = T_{\rm b} \frac{\Omega_{\rm s}}{\Omega_{\rm mb}} \ . \tag{2.44}$$

The *main beam* filling factor would be the ratio $\Omega_{\rm s}/\Omega_{\rm mb}$. It is said that the brightness temperature ($T_{\rm b}$) gets diluted in the main beam, obtaining the measured main beam temperature ($T_{\rm mb}$).

A more complicated but realistic assumption would be that the spatial distribution of the source intensity (I_ν) is a bidimensional Gaussian centered in the source position, with a peak intensity I'_ν and a standard deviation of $\sigma_{\rm s}$; the corresponding peak brightness temperature would be $T'_{\rm b} \equiv I'_\nu c^2/(2k_{\rm B}\nu^2)$, and its solid angle, $\Omega_{\rm s} = 2\pi\sigma_{\rm s}$. We also consider that the source size is not significantly emitting in the region of the side lobes (we could say as an approximate condition that $2\sigma_{\rm s} \lesssim \vartheta^*_{\rm mb}$). In this simple case, the integral in equation 2.38 would involve the integral of the product of two Gaussians (neglecting the side lobes of the antenna pattern). This product will be a new Gaussian with a standard deviation equal to the harmonic mean of the two original ones: $\sigma = 1/(1/\sigma_{\rm s}^2 + 1/\sigma_{\rm mb}^2)^{1/2} = \sigma_{\rm s}\sigma_{\rm mb}/(\sigma_{\rm s}^2 + \sigma_{\rm mb}^2)^{1/2}$. The integral of this new gaussian will then be $2\pi\sigma^2 = 2\pi\sigma_{\rm s}^2\sigma_{\rm mb}^2/(\sigma_{\rm s}^2 + \sigma_{\rm mb}^2)$. Now, we can use the main beam efficiency ($\eta_{\rm mb}$) to substitute $\Omega_{\rm A}$ by $\Omega_{\rm mb}/\eta_{\rm mb}$. Also, $\Omega_{\rm mb} \cong 2\pi\sigma_{\rm mb}^2$. After all this, we get:

$$T_{\rm A} = \frac{1}{\Omega_{\rm A}} \int_{\bigcirc} T_{\rm b}(\Omega)\, P(\Omega)\, {\rm d}\Omega \simeq \eta_{\rm mb} T'_{\rm b} \frac{\sigma_{\rm s}^2}{\sigma_{\rm s}^2 + \sigma_{\rm mb}^2} \cong \eta_{\rm mb} T'_{\rm b} \frac{\Omega_{\rm s}}{\Omega_{\rm s} + \Omega_{\rm mb}} \ . \tag{2.45}$$

The main beam temperature ($T_{\rm mb}$) is thus related to with the peak brightness temperature ($T'_{\rm b}$) as:

$$T_{\rm mb} \equiv \frac{T_{\rm A}}{\eta_{\rm mb}} \simeq T'_{\rm b} \frac{\Omega_{\rm s}}{\Omega_{\rm mb} + \Omega_{\rm s}} \ . \tag{2.46}$$

Here the main beam filling fraction would be $\Omega_{\rm s}/(\Omega_{\rm mb} + \Omega_{\rm s})$. As already mentioned, these relations would be accurate only if the source is not significantly emitting in the side lobes ($2\sigma_{\rm s} \lesssim \vartheta^*_{\rm mb}$), and this accuracy increases as the source size gets smaller.

However, in general we can have an arbitrary spatial distribution for the received intensity I_ν, with a source size that can be larger than the main beam. The antenna pattern, with its main beam, is in practice limiting our spatial resolution, since we cannot distinguish spatial variations smaller than the main beam. We can substitute $\Omega_{\rm A}$ by $\Omega_{\rm mb}/\eta_{\rm mb}$ in the expression of the antenna temperature:

$$T_{\rm A} = \frac{1}{\Omega_{\rm A}} \int_{\bigcirc} T_{\rm b}(\Omega)\, P(\Omega)\, {\rm d}\Omega = \frac{\eta_{\rm mb}}{\Omega_{\rm mb}} \int_{\bigcirc} T_{\rm b}(\Omega)\, P(\Omega)\, {\rm d}\Omega \ . \tag{2.47}$$

Now, since we defined the main beam temperature as $T_{\rm mb} \equiv T_{\rm A}/\eta_{\rm mb}$, we would get the following expression:

$$T_{\rm mb} = \frac{1}{\Omega_{\rm mb}} \int_{\bigcirc} T_{\rm b}(\Omega)\, P(\Omega)\, {\rm d}\Omega \ . \tag{2.48}$$

Therefore, if the source is not very much extended in the side lobes, the main beam tem-

perature can be viewed as an average of the brightness temperature over the main beam. However, for a source extended over the side lobes, the main beam temperature would differ from this beam-averaged brightness temperature.

2.3.1.4 Radial velocity

Instead of measuring the antenna / main beam temperature (or the intensity I_ν) with respect to frequency (ν), obtaining a frequency spectrum, it is also common in radio astronomy to use the radial velocity (v) as the spectral variable.

The transformation is performed thanks to the Doppler effect, which relates the difference in frequency (with respect to the central frequency of a transition, ν_0) to the radial velocity between the source of the emitted/absorbed radiation and the observer. For a non-relativistic moving source, the formula is as follow:

$$\frac{\nu - \nu_0}{\nu_0} \cong -\frac{v - v_0}{c} \ , \tag{2.49}$$

where v_0 is the velocity of the source as a whole and c is the speed of light (Wilson et al., 2013). Here the velocity would be the *local standard of rest* (LSR) velocity ($v_{\rm LSR}$), which refers to the frame of the mean motion of the Solar neighbourhood around the Milky Way centre.

The use of radial velocity instead of frequency is useful in the sense that we obtain the relative radial velocities of the material emitting or absorbing the radiation, which can give us information about the kinematic structure of the observed object. In particular, many molecules will have the same line profile in terms of radial velocity, so the comparison of the line shape can give information about the chemical and kinetic structure of the source (as in, e.g., Figure 1.4.1). Note that in starless/prestellar cores, the linewidths (or FWHM) of the associated molecular emission can be as narrow as <1 km/s, and therefore high-spectral resolution observations are typically needed to fully resolve the line profile and to obtain detailed information about the dynamics of the gas (e.g., Redaelli et al., 2019).

2.3.1.5 Signal noise

The measured antenna temperature when taking a radio spectrum will always have a component of random noise, due to the receiver electronics, variations of the atmospheric water vapor, atmospheric emission, imperfections in the antenna..., and even the proper signal of the source (Condon and Ransom, 2016). These effects are collected in the system noise temperature, $T_{\rm sys}$. The detectors are usually cooled to reduce the contribution of the receiver itself, although it is not the only source of noise for the system temperature.

The *radiometer equation* (Condon and Ransom, 2016) relates the root mean squared (RMS) value of the noise in the measured antenna temperature ($T_{\rm A}$) to the system temperature

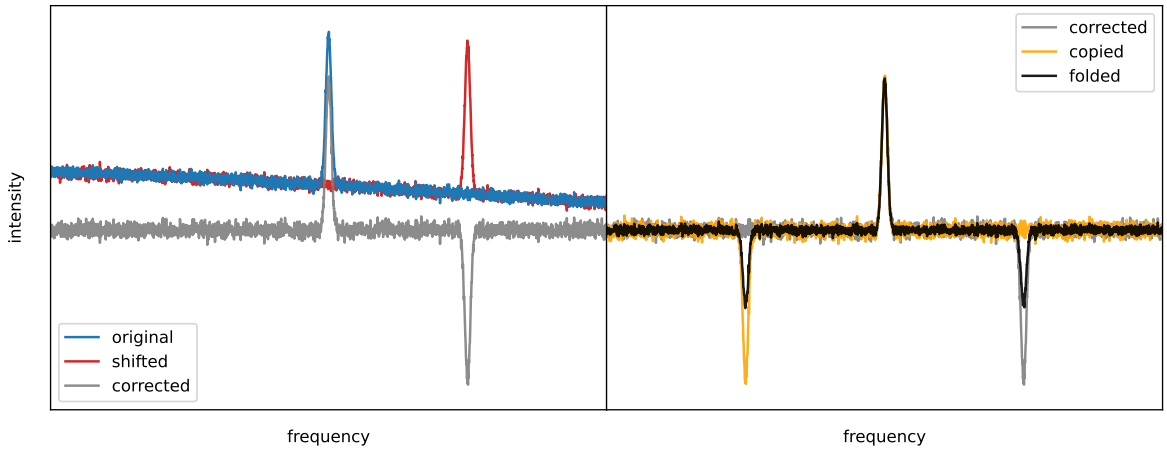


Figure 2.3.2. Scheme of the frequency switching method used to substract the background contribution from the spectrum. The frequency of the receiver is slightly shifted several times per second, obtaining two different signals, the original (in blue) and the shifted one (in red). The registered signal is their difference (in gray). Then, one can fold it –that is, make a copy, invert it, shift it (obtaining the orange curve)– and combine it with the first signal (in gray), to obtain the final curve (in black) after dividing it by 2.

($T_{\rm sys}$), the integration time ($t_{\rm int}$) and the bandwidth of the spectrum ($\Delta\nu_{\rm band}$):

$$\sigma_T \simeq \frac{T_{\rm sys}}{\sqrt{t_{\rm int}\,\Delta\nu_{\rm band}}} \ . \tag{2.50}$$

The obtained noise is inversely proportional to $\sqrt{t_{\rm int}}$. Similarly, if several observations of the same integration time are combined together (averaged), the noise σ_T will be divided by a factor equal to the square root of the number of observations.

2.3.1.6 Frequency switching

In order to remove the contributions to the signal that do not belong to the source, there are several possible observing methods to process the radio signal with this aim.

The most typical ones are called *position switching* and *frequency switching* (Wilson et al., 2013). The first consists of taking observations toward the source and also toward an offset position that should be relatively empty or free of significant radiation, so that it can be substracted from the on-source signal. The second method (frequency switching) is the method used in this thesis and therefore it is explained in more detail in the following.

It mainly consists of regularly shifting the frequency at which the receiver is tuned by a few MHz along the observations, so that these shifts covers several times the typical width of the source lines –which should be relatively narrow, as observed in cold sources like prestellar cores (Wilson et al., 2013). This shifted signal is then substracted to the normal signal. Since the noise contribution of the receiver and the atmosphere usually does not change in the time period in which the shifts are performed (several times per second), it is removed from the signal when doing the substraction. However, the lines would have a shifted twin counterpart in absorption, as a result of that substraction. To increase the

signal-to-noise ratio, one can *fold* the observations, that is, make a copy of the spectrum, invert it and shift it so that the originally absorbed line twin coincides now with the original line, and thus we can average the two spectra. As a result, we reduce the noise level in a factor of $\sqrt{2}$, but now there will be *two* twin lines at both sides in absorption. This must be taken into account to not confuse them with real absorption features. Figure 2.3.2 illustrates this process. This method is much more efficient than position switching, since it does not require observations of an offset position, and therefore all the observing time is spent on the source. However, as a requirement the source should emit narrow and isolated lines, to not mix the artifact lines with real features.

However, even if this method allows to remove most of the background signal, there is still a remaining signal (*ripple*) which constitutes the *baseline* of the final spectrum, although the frequency amplitude of the variations is larger than the line widths. Indeed, the baselines obtained with frequency switching are usually worse (with more ripples) that those obtained with position switching. These baselines must be fitted and removed from the spectrum in order to obtain a final spectrum with flat baselines, whose molecular lines can be analyzed –for example, with the software MADCUBA or RADEX.

In this thesis, we have used frequency switching with data obtained with the IRAM 30 m telescope. As a part of this thesis, I have developed a data reduction pipeline to perform the removal of the ripples produced by this observing technique in an automatic way. This pipeline has been used not only for my work (Megías et al., 2023), but also in other works (Scibelli et al., 2024) and even for observations performed in position switching but for which significant ripples where present due to the use of a far-away offset position (see the works of Sanz-Novo et al., 2023, and Rivilla et al., 2023). This pipeline is described in detail in the following section.

2.3.2 Automated GILDAS/CLASS Pipeline

In order to process and reduce the spectra taken by radio telescopes, one can use the public software GILDAS (*Grenoble Image and Line Data Analysis System*), coded in Fortran and C++ and developed by IRAM (*Institut de RadioAstronomie Millimétrique*).[17] It is a collection of tools to visualyze, analyze and reduce radioastronomical observations –spectra and maps–, compatible with the data files from different observatories, like the IRAM 30 m itself and the Yebes 40 m.

Within GILDAS, there is the tool CLASS (*Continuum and Line Analysis Single-dish Software*). As the name indicates, it allows to analyze spectra taken with single-dish radio telescopes. It is opened directly from the terminal, from which one can open an observation file –usually corresponding to one observation file– and explore the spectra stored within it. Individual spectra can also be visualized in an additional window and explored interactively, zooming,

[17] https://www.iram.fr/IRAMFR/GILDAS/

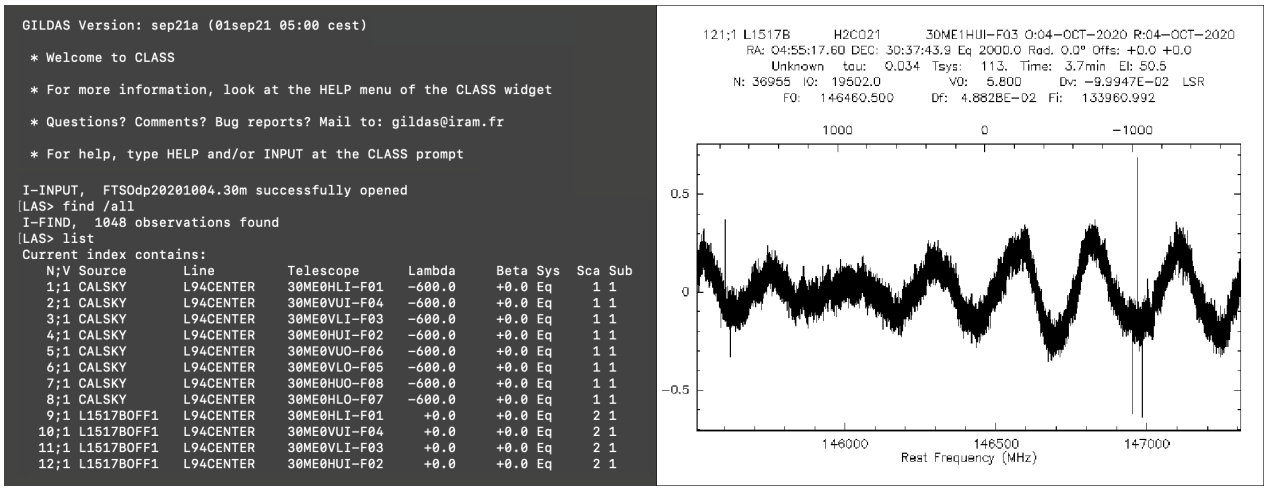


Figure 2.3.3. Screenshots of the software GILDAS and CLASS. *Left*: Start of the tool CLASS, opening an observation file and listing of spectra contained in the observation file. *Right*: Visualization of one of the contained spectra; the vertical axis shows the measured antenna temperature in K, and on top of the plot there is additional information of the spectrum. The brightest detected molecular lines appear as bright spikes around 145 600 MHz and 147 000 MHz, where two twin artificial absorption features can also be seen, produced by the frequency switching method.

changing the units to velocity, and more. Figure 2.3.3 shows some screenshots of CLASS when opening an observation file.

As can be seen on the right, the selected spectrum presents a very noticeable baseline (ripples). Fortunately, the spectral amplitude of the oscillations is much larger than the extension of the transition lines (see brught spikes in the spectrum at around 145 600 MHz and 147 000 MHz). In order to eventually analyze the lines and derive molecular column densities (which will be done with MADCUBA), those ripples must be removed from every single spectrum, in order to obtain a flat baseline. Appart from this baseline removal, one should also average the spectra coming from different observation days, which can be done before or after the baseline removal. Additionally, there is the possibility that spectra retrieved with different spectral band widths (different receiver configurations) overlap in some regions. In this case, they should also be averaged in the overlapping ranges. All of these operations can be performed within CLASS, although it requires a manual inspection of each individual spectrum, taking quite a lot of time in the end.

In order to perform all of those opperations in a more automated and faster way, a set of Python scripts have been written as a part of this thesis that run GILDAS and CLASS following the instructions given in a configuration file. This software has been published with the name of Automated GILDAS/CLASS Pipeline both in GitHub [18] and in the Astrophysics Source Code Library (ASCL).[19] It contains example observation and configuration files and a complete user guide explaining the use of each script.

The Automated GILDAS/CLASS Pipeline offers seven different working modes, each with a different purpose. The most important ones are:

[18] https://github.com/andresmegias/gildas-class-python/
[19] https://ascl.net/2302.018/

- **Selection**. Performs all the merging and averaging of the spectra indicated in the configuration file, creating several final CLASS data files containing the averaged spectra. The averages can be done over all the observing days or just day by day.
- **Line search**. Searchs for significant emission or absorption lines and save their frequency ranges, in order to mask them to perform a fit of the baseline.
- **Reduction**. Performs a baseline fit for each spectrum and subtracts it to obtain reduced spectra, using the provided file with the regions to be masked (which can be the output of the line search mode mentioned before). The fits can be visualized and the masks can be interactively modifed in order to improve the baseline fit. The results are saved in new CLASS data files.
- **Averaging**. In case the averaging in the selection mode was done only day by day, this mode allows to combine and average the reduced spectra obtained with the reduction mode. In also includes an algorithm that automatically detects artifacts and inconsistencies between the combined spectra, which are then removed from the final spectra. New CLASS data files are generated.
- **Merging**. In case there are several individual spectra that overlap in some spectral ranges, this mode allows to average those spectra, improving the final noise. It creates new individual spectra corresponding to the overlapping regions, and writes the results in a new CLASS data file.

As already mentioned, this pipeline was used not only for reducing observations presented in this thesis, but also for reducing observations of a broadband spectroscopic survey carried out within the astrochemistry group at the Centre for Astrobiology, which has yielded in several papers (Rivilla et al., 2022a,b, 2023; Sanz-Novo et al., 2023, 2024a,b). The description on how to use this pipeline can be found in Appendix A.1. Figure A.1.1 shows some screenshots of the use of this tool, while Figure 2.3.4, we show an example spectrum before and after applying the pipeline. As seen in this Figure (top panel), the spectrum presents a clear ripple, which is perfectly removed after applying the pipeline (bottom panel).

In these synthetic spectra, the number of lines present per each single spectrum is very low. However, in real data there are usually tens of lines per spectrum, and there can be tens of spectra to reduce. Here is where the Automated GILDAS/CLASS Pipeline can save a lot of time to the user with respect to using the standalone CLASS. If so, one should manually identify the significant lines and mask them to perform several baseline fittings, and then export the results in new CLASS files, a process that can take hours for analyzing real observations. When used in real data, the automatic line search can take from tens of seconds to several minutes, depending on the number of spectra to be analyzed –the reduction part is very fast, as well as the selection mode.

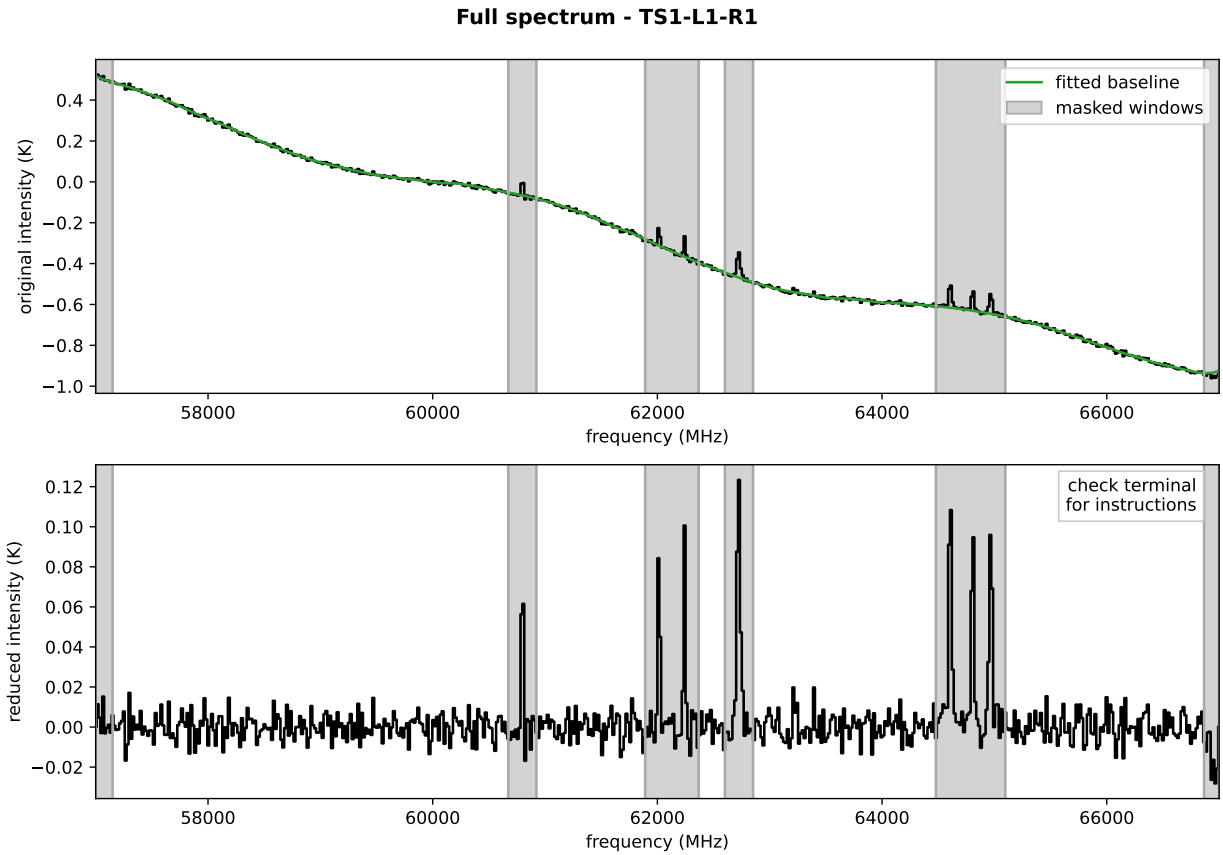


Figure 2.3.4. Result of the automatic line search and the reduction for one of the spectra contained in the files created after the selection mode, for the example shown in Fig. A.1.1. *Top*: Original spectrum showing the identified regions to mask (in gray) and the derived baseline (in green). *Bottom*: Reduced spectrum after the substraction of the fitted baseline, showing also the masked regions (in gray).

2.3.3 Infrared absorption spectra

In contrast to radio observations, infrared spectra can be taken with the same type of telescopes used for optical observations, with similar mirrors, solid states detectors, etc. In this case we do not have any antenna pattern, although there are other effects that should be taken into account for calibrating the observations (like the flat field correction).

Fortunately, the infrared spectra used in this thesis were already calibrated, consisting in specific flux ($\mathcal{F}_\nu$) versus wavelength (λ). They correspond to absorption spectra taken with James Webb Space Telescope (JWST) towards two background stars in the Chamaleon I molecular cloud, as part of the Ice Age project (McClure et al., 2023). They are intended to observe the absorption features produced by the icy layers covering interstellar dust grains in the mentioned molecular cloud. Therefore, there are still some reduction steps that must be done to convert the observed spectrum to absorption spectrum in optical depth (or absorbance).

First of all, one must perform an estimate of the background continuum, that is, the spectrum emitted by the background star. If stellar atmospheric lines are not relevant, one can just make a polynomial fit in certain spectral regions where it is known that there are no significant absorptions due to ice from the grains. Once the continuum is estimated ($\mathcal{F}_{\nu,0}$), we

can convert the observed spectra ($\mathcal{F}_\nu$) to optical depth scale, taking the logarithm (Muñoz Caro and Escribano, 2018):

$$\tau = \ln(\mathcal{F}_{\nu,0}/\mathcal{F}_\nu) \ . \tag{2.51}$$

After this, the contribution due to the silicate core of dust grains must be removed, since we are interested in the outer icy layers. This contribution is seen as two wide bands in the region of 8–12 μm (centered at 9.7 μm) and another contribution centered at 18 μm (Boogert et al., 2015). The shape of this contribution can be estimated by synthetic models, like the ones used by the software OpTool (Dominik et al., 2021). Similarly as with the continuum estimation, one can select several regions where there should not be any ice absorption feature, to make a fit of the synthetic model to the observed optical depth. Once the silicate contribution is estimated, it can be substracted directly from the spectrum in optical depth, obtaining the final absorption ice spectrum.

This procedure is explained in more detail in Appendix C.1, corresponding to Chapter 4, where a real case with observations taken by James Webb Space Telescope (JWST) from the Ice Age project (McClure et al., 2023) is presented.

2.4 Telescopes used in this thesis

This section briefly describes the characteristics of the telescopes used in this thesis. While I have performed observing runs at the IRAM 30 m telescope (both in person and in remote mode), I have not participated in the design of the JWST observations since they were planned before I joined the group at the Centre for Astrobiology.

2.4.1 IRAM 30 m telescope

IRAM (*Instituto de RadioAstronimía Milimétrica*) is a Spanish-French-German consortium funded by the Spanish IGN (*Instituto Geográfico Nacional*), the French CNRS (*Centre National de la Recherche Scientifique*) and the German MPG (*Max-Planck-Gesellschaft*). It operates two radio facilities: NOEMA (**No**rthern **E**xtended **M**illimetre **A**rray), located in the French Alps, and the IRAM 30 m observatory, located in Sierra Nevada (Granada, Spain).[20]

This last telescope, built in 1984, is located at 2850 m of altitude in Sierra Nevada, close to Pico Veleta (Fig. 2.4.1). With its antenna of 30 m in diameter, it is one of the largest and most sensitive radio telescopes observing millimetric wavelengths. It operates in the spectral range of 80 to 370 GHz (3 to 0.8 mm). The half-power beam width (HPBW) of the antenna varies from ~27″ at 89 GHz and ~7″ at 335 GHz (Greve et al., 1998). The telescope is currently equipped with two instruments: EMIR (**E**ight **Mi**xer **R**eceiver) and NIKA2 (**N**ew

[20] https://www.iram.fr/GENERAL/calls/w23/30mCapabilities.pdf



Figure 2.4.1. Photograph of the IRAM 30 m telescope in Sierra Nevada (Spain). Credit: IRAM / K. Zacher. `https://iram-institute.org/science-portal/30-meter-telescope/`.

IRAM **K**ID **A**rray 2). For this thesis we used observations obtained with EMIR, which is explained in more detail below.

EMIR

The **E**ight **Mi**xer **R**eceiver (EMIR) is an heterodyne receiver (frontend) that has four different bands covering the four main atmospheric windows in the millimetre range: E090, E150, E230 and E330, with the numbers refering to the central frequency in GHz. Each of the four bands is subdivided in four blocks: LO (lower-outer), LI (lower-innter), UI (upper-inner) and UO (upper-outer), with two polarizations each. At the same time, three spectral line backends are possible to use: FTS (Fourier Transform Spectrometer), with either 50 or 200 kHz of spectral resolution; VESPA (**Ve**rsatile **S**pectrometric and **P**olarimetric **A**rray), with a resolution ranging from 3.3 kHz to 1.25 MHz; and WILMA (**Wi**deband **L**ine **M**ultiple **A**utocorrelator). For this thesis we used the FTS backend in the 50 kHz spectral resolution mode, since it provides a good enough velocity resolution to resolve emission lines in starless and prestellar cores (with linewidths as narrow as <1 km/s).

2.4.2 James Webb Space Telescope

The James Webb Space Telescope (JWST) was developed by USA's National Aeronautics and Space Administration (NASA), in cooperation with the European Space Agency (ESA) and the Canadian Space Agency (CSA). It was launched in December 2021, arriving at its

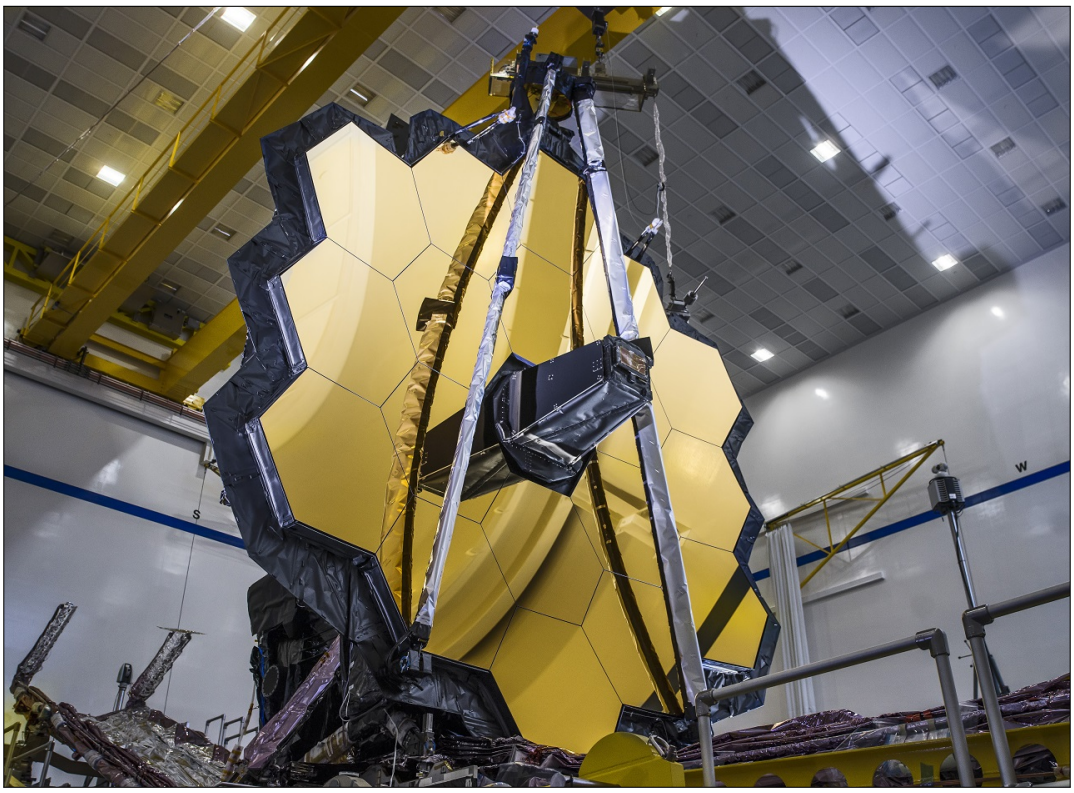


Figure 2.4.2. Photography of James Webb Space Telescope after a succesful deployment test of its primary mirror. Credit: NASA / C. Gunn. https://www.esa.int/ESA_Multimedia/Images/2020/04/James_Webb_Space_Telescope_s_primary_mirror_unfolded.

destination in January 2022, close to the Sun-Earth L_2 Lagrange point, around $1.5 \cdot 10^6$ km from Earth (McElwain et al., 2023).

It has a 6.5 m wide mirror, consisting of a honeycomb-like pattern of 18 hexagonal, gold-coated mirror segments (Fig. 2.4.2). It operates in the visible and infrarred range, being able to obtain single-position, images also spectral image cubes. It hosts four instruments: MIRI (**M**id-**I**nfra**r**ed **I**nstrument), NIRCam (**N**ear-**I**nfra**r**ed **Cam**era), NIRSpec (**N**ear-**I**nfra**r**ed **S**pectrograph) and NIRISS/FGS (**N**ear-**I**nfra**r**ed **I**mager and **S**litless **S**pectrograph / **F**ine **G**uidance **S**ensor).[21] Only data from the three first instruments is used in this thesis, since they provide a good enough spectral resolution and the proper spectral range coverage to identify the desired molecular absorption bands (widths of tens of cm^{-1} at minimum and a spectral range of ~ 2–20 μm).

NIRCam

The Near Infrared Camera (NIRCam) offers imaging, coronagraphy, wide field slitless spectroscopy, and time-series monitoring both in imaging and spectroscopy. It can observe wavelengths from 0.6 to 5.0 μm, with a resolving power of ~1600 and a total field of view of 2.2′ × 2.2′ (the resolving power, $\mathcal{R}$, is defined as the ratio between the wavelength, λ,

[21] https://jwst-docs.stsci.edu/

and the spectral resolution, $\Delta\lambda$: $\mathcal{R} \equiv \lambda/\Delta\lambda$). The spatial resolution is either 0.031 or 0.063 ″ per pixel. In this thesis we used spectra that were extracted from NIRCam datacubes over a single-position (around a background star)

NIRSpec

The Near Infrared Spectrograph (NIRSpec) provides near-infrared spectroscopy from 0.6 to 5.3 μm within a 3.4′ × 3.6′ arcmin field of view, with resolving powers of ~100, ~1000 and ~2700 in four observing modes. The spatial resolution is either 0.1 ″ per pixel. It can use a micro-shutter assembly (MSA), an integral field unit (IFU), and fixed slits (FSs). For this thesis, we used NIRSpec observations with fixed slits, hence having single-position spectra.

MIRI

The Mid-Infrared Instrument (MIRI) provides imaging and spectroscopic observing modes from 4.9 to 27.9 μm. For imaging, the resolving power ranges from 3.5 to 17.2, with a field of vew of 75″ × 113″. For spectroscopy, it has two resolution modes: LRS (Low-Resolution Spectroscopy), with a resolving power of ~100; and MRS (Medium-Resolution Spectroscopy), with a resolving power ranging from ~1550 to ~3250. The spatial resolution is 0.11 ″ per pixel, except for MRS, where it ranges from 0.20 to 0.27 ″ per pixel. In this thesis we used single-position spectra with LRS.

2.5 Astrochemical models

The chemistry of a certain astrophysical region can be modeled to compare the observed abundances or column densities with simulations, allowing to constrain physical conditions of the region. There are different types of astrochemical models that simulate the interaction between atoms, molecules and other elements like cosmic rays or electromagnetic radiation. Among them, the most simple ones are the ones based on reaction rates. In this thesis, we have used the UCLCHEM astrochemical software to simulate the chemistry of the Chamaeleon I molecular cloud toward the position of two background stars observed with the JWST, in the context of the work carried out within the chemical modelling group of the Ice Age collaboration.

2.5.1 Reaction rate astrochemical codes

In reaction rate astrochemistry codes, every single reaction can be modelled with a kinetic reaction rate, that determines the rate at which the reactants get combined to create the products (Holdship et al., 2017). If we have a set of chemical reactions involving several

atoms and molecules and the reaction rates are known, one can easily solve numerically the system of differential equations that determines the time evolution of every atomic or molecular abundance. Normally, reaction rates can be characterized by experiments and/or with quantum chemical calculations. However, these rates depend on conditions such as the density and the temperature. The conditions present in the interstellar medium (high vacuum, very low temperatures) are quite different from those achieved in Earth's surface, therefore specific experiments with those conditions must be performed in order to properly characterize these rate equations.

The amount of chemical reactions involved in an astrophysical region, like a molecular cloud, can be very large –hundreds and even thousands of reactions. Additionally, many reaction rates cannot be succesfully determined in many cases, because of the lack of experiments or simulations. This constitutes one of the largest uncertainties associated with astrochemical models based on reaction rates. However, despite this limitation, reaction rate astrochemical codes have proved to be a useful tool to understand the physical and chemical processes behind the chemistry observed in a wide range of environments in the interstellar medium (e.g.: (Viti et al., 2004; Viti et al., 2011; Jiménez-Serra et al., 2012, 2018)).

As an example, Figure A.2.1 in Appendix A.2 shows an scheme of a simple chemical network as a graph (see Appendix A.2 for a list of all the involved reactions). In the following, we describe the astrochemical code UCLCHEM, which is the one we have used in this thesis for the Ice Age collaboration (Chapter 5); the MONACO code, which is also used in this thesis, is described in detail in Chapter 3, where its predictions are compared to the observational results obtained towards the L1517B starless core.

2.5.2 UCLCHEM

UCLCHEM is an astrochemical software based on reaction rates, written in Fortran with a Python interface, and developed originally at University College London (UCL), hence its name (Holdship et al., 2017). It is open-source and publicly available to download,[22] with documentation and tutorials (see Figure 2.5.1).

UCLCHEM considers gas-phase species but also solid-phase species in interstellar dust grains. For dust grains, it differenciates between the surface and the bulk of the ices (the refractory part of the grain core is not treated in this core). Therefore, it is a three-phase chemical model. Note that this is the usual approach used in astrochemical codes (e.g.: Garrod, 2008; Wakelam et al., 2010). Furthermore, these three phases can interact; for example, gas-phase species can get adsorbed on the grain's surface, and solid species in the surface can get desorbed to the gas. In this way, the evolution of the abundance of multiple chemical species can be simulated in a specific point of a molecular cloud, or a set of points within a certain direction. UCLCHEM was first presented by (Holdship et al., 2017) then extended

[22] https://uclchem.github.io/

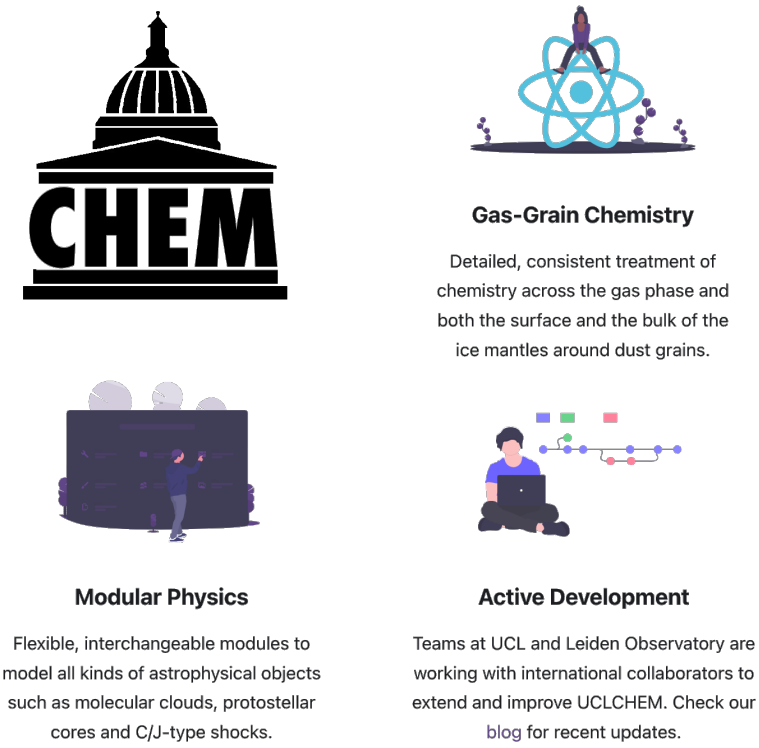


Figure 2.5.1. Screenshots from UCLCHEM's website, including its logo (upper left corner).

by Quénard et al. (2018), and it is actively mantained by researchers from the University of Leiden. It contains default chemical networks for gas and grain chemistry, but they can be manually modified to include custom reactions. The default gas chemical network comes from the UMIST database (McElroy et al., 2013), while the grain one is explained by Quénard et al. (2018). Note, however, that the gas-phase chemical network can be updated by using the latest release of the UMIST database done by Millar et al. (2024).

This astrochemical software can be run through Python, specifying a type of astrophysical environment (static or collapsing molecular cloud, hot core...) and defining a set of input parameters such as the initial elemental abundances, the initial density and temperature, or the ultraviolet (UV) photoionization rate. In the following we list the most relevant numeric parameters in UCLCHEM:

- **Temperature**. Temperature of the gas and the dust, which will be constant in space and time, in K.
- **Initial density**. Initial gas density, refering to the number of hydrogen nuclei (H) per volume, in atoms per cm^3.
- **Elapsed time**. Range of time simulated by the model, in years (yr).
- **Base visual extinction**. Visual extinction (A_V) outside the molecular cloud, which will be summed to the extinction produced inside the cloud, in magnitudes.
- **Cloud radius**. Size of the molecular cloud, in parsecs (pc). It determines the visual extinction (A_V) in the simulated point of the molecular cloud. In particular, if $A_{V,b}$ is the base visual extinction, n_H is the gas density (in H nuclei), and r_{cl} is the cloud

radius, the visual extinction in the center can be estimated as: $A_V \simeq A_{V,b} + r_{cl}\, n_H / (1.6 \cdot 10^{21}\ /\mathrm{cm}^3)$.

- **Cosmic ray ionization rate**. Number of ions produced by cosmic rays per time, as a multiple of $1.3 \cdot 10^{-17}/\mathrm{s}$.
- **Interstellar radiation field**. External intensity in the far ultraviolet (FUV), in units of Habing ($1\,\mathrm{H} = 1.2 \cdot 10^{-10}\,\mathrm{W/cm^2/sr}$).

Apart from these numerical parameters, there are several options that can be activated or deactivated:

- **Free fall**. Determines if the molecular cloud evolves in time collapsing according to free fall.
- **Thermal desorption**. Activates the desorption of species from dust grains to gas phase due to temperature.
- **Non-thermal desorption**. Activates the non-thermal desorption of species from dust grains to gas phase. Different types of non-thermal desorption can be modified individually: desorption due to molecular hydrogen (H_2) formation, due toimpact of cosmic rays, and due to impact of ultraviolet (UV) radiation.

Additionally, there are specific parameters in case that shock events are simulated in the molecular cloud, although they have not been used in this thesis. In Chapter 5, we present and discuss the simulations performed with UCLCHEM, in the context of a molecular cloud that collapses over time, to explain the ice molecular abundances derived from JWST toward the two background stars observed within the Ice Age Early Release Science (ERS) project (McClure et al., 2023).

Bibliography of this chapter

The following list contains the different materials used as a reference for each section:

- Section 2.1, Atomic and molecular spectroscopy:
  - Book *Tools of radioastronomy* (Wilson et al., 2013), chapters 1 and 15.
  - Paper by Mangum and Shirley (2015) about the calculation of column densities with radiative transfer and statistical mechanics.
- Section 2.2, Radiative transfer:
  - Book *Tools of radioastronomy* (Wilson et al., 2013), chapters 1 and 12.
  - Book *Essential radioastronomy* (Condon and Ransom, 2016), chapters 2 and 7.
  - Paper by Mangum and Shirley (2015) about the calculation of column densities with radiative transfer and statistical mechanics.
  - Paper by Goldsmith and Langer (1999) about molecular line emission and radiative transfer.
  - Paper by Martín et al. (2019) about the software MADCUBA.
  - Paper by van der Tak et al. (2007) about the software RADEX.

  - Review by Boogert et al. (2015) about astronomical ices.
  - Book *Laboratory astrophysics*, edited by Muñoz Caro and Escribano (2018), chapters 3 (written by R. Escribano and G. Muñoz Caro) and 5 (written by B. Maté).
- Section 2.3, Astronomical observations:
  - Book *Tools of radioastronomy* (Wilson et al., 2013), chapter 8.
  - Book *Essential radioastronomy* (Condon and Ransom, 2016), chapter 3
- Section 2.4, Telescopes:
  - Website of IRAM (`https://iram-institute.org/`) and of the documentation of the IRAM 30 m telescope (`https://publicwiki.iram.es/`).
  - Paper by Greve et al. (1998) about the IRAM 30 m beam pattern.
  - Article by Kramer & Sánchez Portal about the IRAM 30 m capabilities: `https://www.iram.fr/GENERAL/calls/w23/30mCapabilities.pdf`.
  - Website of documentation of the James Webb Space Telescope (JWST) (`https://jwst-docs.stsci.edu/`).
  - Paper by McElwain et al. (2023) about the JWST.
- Section 2.5, Astrochemical models:
  - Websites of the astrochemical software KROME (`https://www.kromepackage.org/`) and UCLCHEM (`https://uclchem.github.io/`).
  - Papers by Holdship et al. (2017) and Quénard et al. (2018) about the software UCLCHEM.

Chapter 3

Chemical complexity in starless and prestellar cores in gas phase

This section has been adapted from the paper titled *The complex organic molecular content in the L1517B starless core*, published in the journal Monthly Notices of the Royal Astronomical Society (Megías et al., 2023).[1] The complete list of authors is: Andrés Megías Toledano, Izaskun Jiménez Serra, Jesús Martín-Pintado, Anton I. Vasyunin, Silvia Spezzano, Paola Caselli, Giuliana Cosentino, Serena Viti.

Abstract

Recent observations of the pre-stellar core L1544 and the younger starless core L1498 have revealed that complex organic molecules (COMs) are enhanced in the gas phase toward their outer and intermediate-density shells. Our goal is to determine the level of chemical complexity toward the starless core L1517B, which seems younger than L1498, and compare it with the other two previously studied cores to see if there is a chemical evolution within the cores. We have carried out 3 mm high-sensitivity observations toward two positions in the L1517B starless core: the core's centre and the position where the methanol emission peaks (at a distance of $\sim$5000 au from the core's centre). Our observations reveal that a lower number of COMs and COM precursors are detected in L1517B with respect to L1498 and L1544, and also show lower abundances. Besides methanol, we only detected CH_3O, H_2CCO, CH_3CHO, CH_3CN, CH_3NC, HCCCN, and HCCNC. Their measured abundances are $\sim$3 times larger toward the methanol peak than toward the core's centre, mimicking the behaviour found toward the more evolved cores L1544 and L1498. We propose that the differences in the chemical complexity observed between the three studied starless cores are a consequence of their evolution, with L1517B being the less evolved one, followed by

[1] https://doi.org/10.1093/mnras/stac3449

L1498 and L1544. Chemical complexity in these cores seems to increase over time, with N-bearing molecules forming first and O-bearing COMs forming at a later stage as a result of the catastrophic depletion of CO.

3.1 Introduction

In Astrochemistry, complex organic molecules (or COMs) are defined as carbon-bearing compounds with at least 6 atoms in their structure (Herbst and van Dishoeck, 2009). It was initially believed that COMs could only form on dust grains in the presence of a source of heat via hydrogenation, atom addition and radical-radical reactions (Watanabe and Kouchi, 2002; Garrod et al., 2008). Indeed, COMs were firstly detected in relatively hot sources with $T \gtrsim 100$ K, such as massive hot cores (Hollis et al., 2000, 2006; Belloche et al., 2008, 2013) or low-mass warm cores (hot corinos; Bottinelli et al., 2004; Jørgensen et al., 2013). However, the detection of several COMs in the gas phase in starless/pre-stellar cores and dark cloud cores (e.g., Marcelino et al., 2007; Öberg et al., 2010; Bacmann et al., 2012; Cernicharo et al., 2012; Vastel et al., 2014; Jiménez-Serra et al., 2016, 2021; Scibelli and Shirley, 2020) indicates that there must be another chemical pathway that allows the presence of COMs in the gas phase at temperatures as low as 10 K. In the last years, there have been different proposals to explain the formation of these COMs under these conditions. However, the chemical pathways giving rise to COMs in cold cores is not well understood yet, and more observations are needed to constrain the proposed models (Rawlings et al., 2013; Vasyunin and Herbst, 2013; Balucani et al., 2015; Vasyunin et al., 2017; Holdship et al., 2019; Jin and Garrod, 2020; Punanova et al., 2000). For example, Scibelli et al. (2021) studied the starless core L1521E, claiming that COMs are not only formed by gas-phase reactions, bus also by surface reactions on dust grains. Other authors suggest that cosmic rays induce desorption from icy mantles on dust grains (Sipilä et al., 2021; Redaelli et al., 2021). It is remarkable that methanol (CH_3OH), which plays a central role in the formation for larger COMs, is detected systematically in starless cores (Scibelli and Shirley, 2020), and thought to form on the surface of dust grains, partially returning into the gas phase upon reactive desorption (Garrod et al., 2006, 2007; Vasyunin et al., 2017).

In order to understand how COMs form under the cold conditions of pre-stellar/starless cores, Jiménez-Serra et al. (2016, 2021) investigated the radial distribution of large COMs as a function of radius in the L1544 and L1498 starless cores. Methanol tends to show a ring-like morphology circumventing the dust continuum emission (see Tafalla et al., 2006; Bizzocchi et al., 2014; Spezzano et al., 2016; Punanova et al., 2000). Since mapping the emission of larger COMs in starless cores requires large amounts of telescope time, Jiménez-Serra et al. (2016, 2021) observed two positions within these cores: the centre, defined by the position of the dust peak, and the location where methanol peaks. The latter position is representative of an outer, intermediate-density shell located at radii between $\sim$4000 and $\sim$11000 au from the core centre in L1544 and L1498, respectively. Several COM precursors

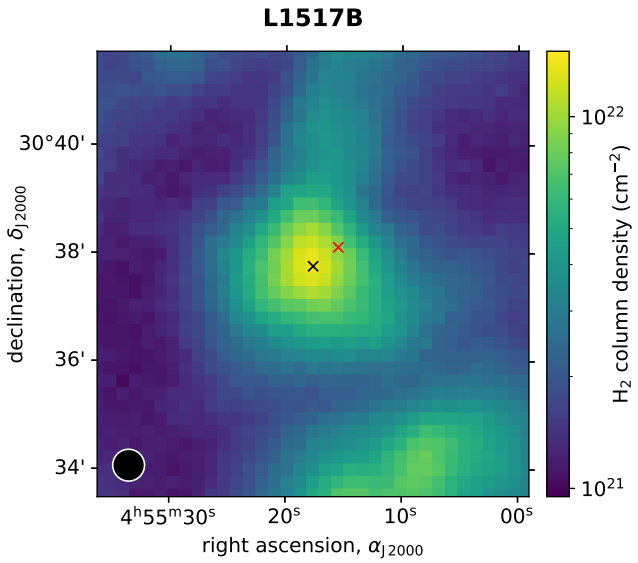


Figure 3.1.1. H_2 column density map obtained from *Herschel*/SPIRE data at 0.25, 0.35 and 0.50 mm, using the same procedure as the one used for L1544 in Spezzano et al. (2016). Crosses indicate the positions observed in the core: the dust peak (in black) and the methanol peak (in red). The beam size for each filter is 17.9, 24.2 and 35.4 arcseconds (0.25, 0.35 and 0.50 mm, respectively), although the images of the two first filters were smoothed to the resolution of 0.50 mm, which is marked with a circle at the bottom-left corner of the image. The positions of the dust and methanol peaks are obtained from Tafalla et al. (2004, 2006). The beam sizes were retrieved from the SPIRE Handbook: `https://www.cosmos.esa.int/web/herschel/spire-overview`.

have been found toward both cores such as tricarbon monoxide (CCCO) and cyanoacetylene (HCCCN), but more complex molecules like acetaldehyde (CH_3CHO), methyl formate (CH_3OCHO) or dimethylether (CH_3OCH_3) have only been detected toward L1544, which is at a more advanced stage of evolution than L1498 (see Jiménez-Serra et al., 2016, 2021). However, N-bearing COMs are detected toward L1498 with abundances close to those measured toward L1544.

In this work, we present a similar study carried out toward another starless core, L1517B, believed to be at an even earlier stage of evolution than L1498 based on the deuterium fractionation and the absence of infall motions (Crapsi et al., 2005). The goal is to compare chemical complexity measured toward L1517B with that observed in L1544 and L1498, and to establish if the observed trend of increasing chemical complexity is due to evolution (see Jiménez-Serra et al., 2016, 2021). These cores, located in the Taurus molecular cloud complex, show different observational signatures that provide information about their stage of gravitational collapse. L1544 is classified as a pre-stellar core because it shows clear evidence of gravitational collapse toward its innermost regions as probed by the emission of N_2H^+ (Caselli et al., 2002; Redaelli et al., 2019). Its high deuterium fractionation ($N_{N_2D^+}/N_{N_2H^+} = 0.23 \pm 0.04$), CO depletion factor ($f_{CO} = 14 \pm 3$; Crapsi et al., 2005; Redaelli et al., 2019), and central H_2 density ($n_{H_2} \sim 10^7$ cm^{-3}; Caselli et al., 2022), are also consistent with this idea. L1498 presents signatures of infall motions in the outer envelope of the core as revealed by the asymmetries observed in the line profiles of CS (Tafalla et al., 2004) but its deuterium fractionation, CO depletion factor and central H_2 density are

Table 3.1.1. Frequency ranges, velocity resolution and RMS noise level of our observations.

Frequency	**Resolution**	**RMS noise** (mK)	
(GHz)	(km s^{-1})	Dust peak	Methanol peak
78.2–80.0	0.18	4.8	3.9
81.4–83.3	0.18	4.2	3.5
83.3–85.2	0.17	2.8	3.3
86.7–88.5	0.17	3.0	3.5
93.5–95.7	0.15	4.1	3.4
95.8–96.8	0.15	10.8	10.9
97.1–99.0	0.15	4.1	3.6
99.1–100.9	0.15	3.0	3.5
102.4–104.2	0.14	4.1	4.7
109.2–111.0	0.13	10.1	11.0

lower than those measured toward L1544 ($N_{N_2D^+}/N_{N_2H^+} = 0.04 \pm 0.01$, $f_{CO} = 7.5 \pm 2.5$, and $n_{H_2} \simeq 9.4 \times 10^4$ cm^{-3}; Crapsi et al., 2005; Tafalla et al., 2004). For L1517B, the CO depletion factor and deuterium fractionation are similar to those of L1498, $f_{CO} = 9.5 \pm 2.8$ and $N_{N_2D^+}/N_{N_2H^+}$ = 0.06 ± 0.01 (Crapsi et al., 2005). However, although L1517B shows a slightly higher central H_2 density ($n_{H_2} \simeq 2.2 \times 10^5$ cm^{-3}), the absence of infall motions either toward the innermost regions or toward the envelope (Tafalla et al., 2004), suggests that L1517B is at a younger evolutionary stage than L1498 and L1544. Hence, L1517B, located at a distance of 159 pc (Galli et al., 2019), would be the dynamically youngest of the three cores.

The paper is organized as follows. In Section 3.2, we describe the observations carried out toward L1517B. In Section 3.3 we present the results of the analysis of the COMs and COM precursor emission and report the values of the derived excitation temperatures, column densities and molecular abundances. Section 3.4 compares the abundances obtained toward L1517B with those measured toward L1498 and L1544. We also compare our results with those reported by Nagy et al. (2019) and Scibelli et al. (2021) toward the young L1521E starless core. This core has been found to be rich in COM emission despite its youth. In Section 3.5, we present the modelling of the COMs and COM precursor chemistry of the L1517B core, and compare the model predictions with the observations. In Section 3.6, we compare the column densities ratios between the N-bearing species HC_3N and CH_3CN derived toward L1517B, L1498, L1544, and L1521E, with those obtained in protostellar systems, protoplanetary discs, and comets, and discuss the observed discrepancies. Finally, in Section 3.7, we summarise our conclusions.

3.2 Observations

The observations of the L1517B starless core were carried out from September 30th to October 4th of 2020, with the Instituto de Radioastronomía Milimétrica (IRAM) 30 m telescope (Granada, Spain). As for L1544 and L1498, we observed two positions within the L1517B core: the location of the dust peak and the position where the emission of methanol peaks.

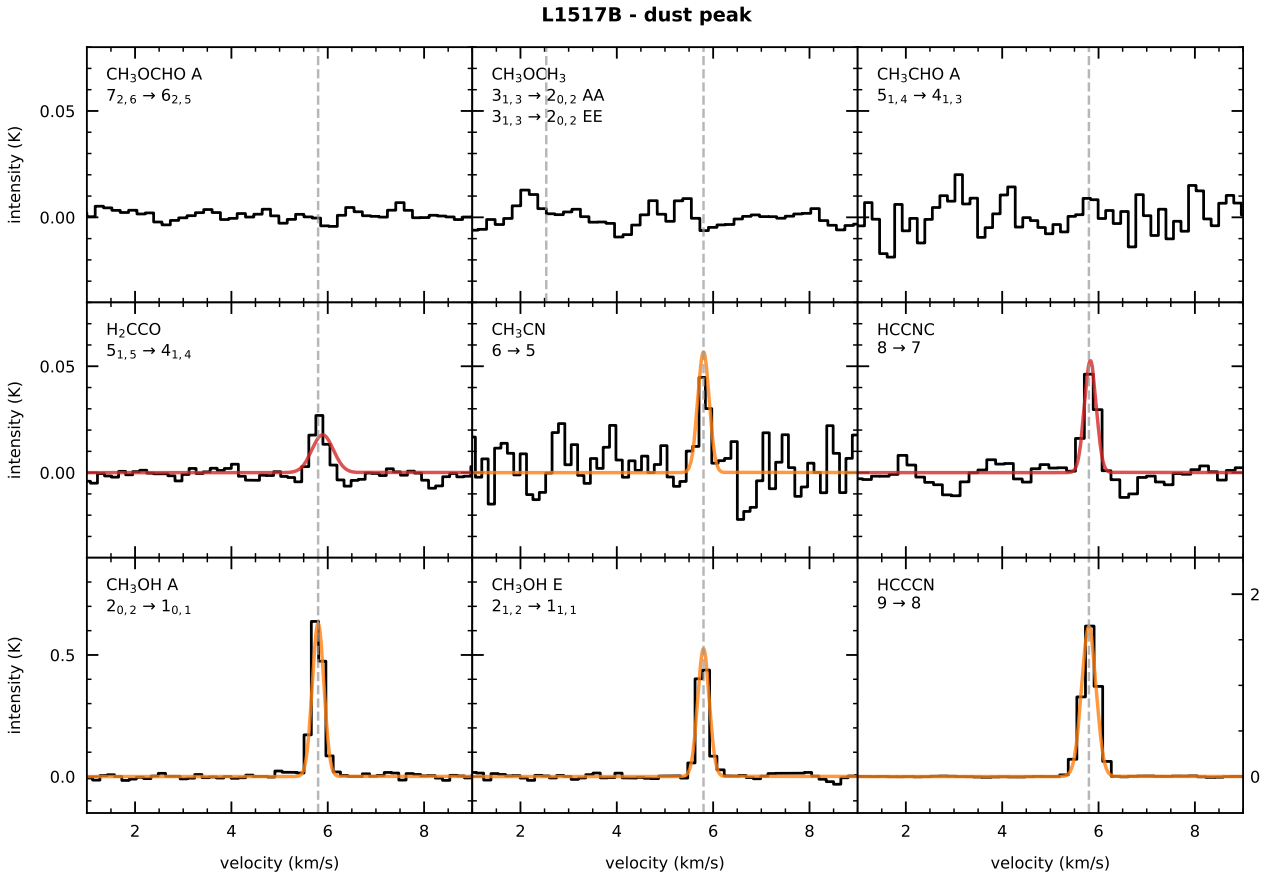


Figure 3.1.2. Sample of the COM and COM precursor lines observed toward L1517B's dust continuum peak and their corresponding fits obtained with MADCUBA (LTE, in red) and RADEX (non-LTE, in orange). See Sections 3.2 and 3.3.1 for more details.

These two positions have the equatorial coordinates (in J2000 system) $\alpha = 4^{h}55^{m}17.6^{s}$, $\delta = 30°37'44''$ for the dust continuum peak, and $\alpha = 4^{h}55^{m}15.7^{s}$, $\delta = 30°38'04''$ for the position of the methanol peak (see Tafalla et al., 2004, 2006). The latter is located $\sim$32″ away from the dust peak (see Fig. 1), which corresponds to $\sim$5000 au at a distance of 159 pc.

The high-sensitivity 3 mm spectra were obtained in frequency-switching mode using a frequency throw of 7.14 MHz. The EMIR E090 receivers were tuned at 84.37 GHz and 94.82 GHz with rejections of $\geq$10 dB. The observed frequency ranges are shown in Table 3.1.1. To identify possible weak spurious features in the observed spectra, we carried out part of the observations by shifting slightly the central frequencies by $\pm$20 MHz (see also Jiménez-Serra et al., 2016). We used the narrow mode of the FTS spectrometer that provided a spectral resolution of 49 kHz, equivalent to 0.13–0.18 km s^{-1} at 3 mm. Typical system temperatures ranged between 75–110 K and the telescope beam size was 22–31 ″ between 78 and 111 GHz; as the beams are almost gaussian and the dust and methanol peaks of L1517B are located $\sim$32″ away ($>$ 2 σ), the contamination between both positions should be negligible (less than 5 percent). The spectra were calibrated in units of antenna temperature, T_{A}^{*}, and converted into main beam temperature, T_{mb}, by using beam efficiencies of 0.81 at 79–101 GHz and of

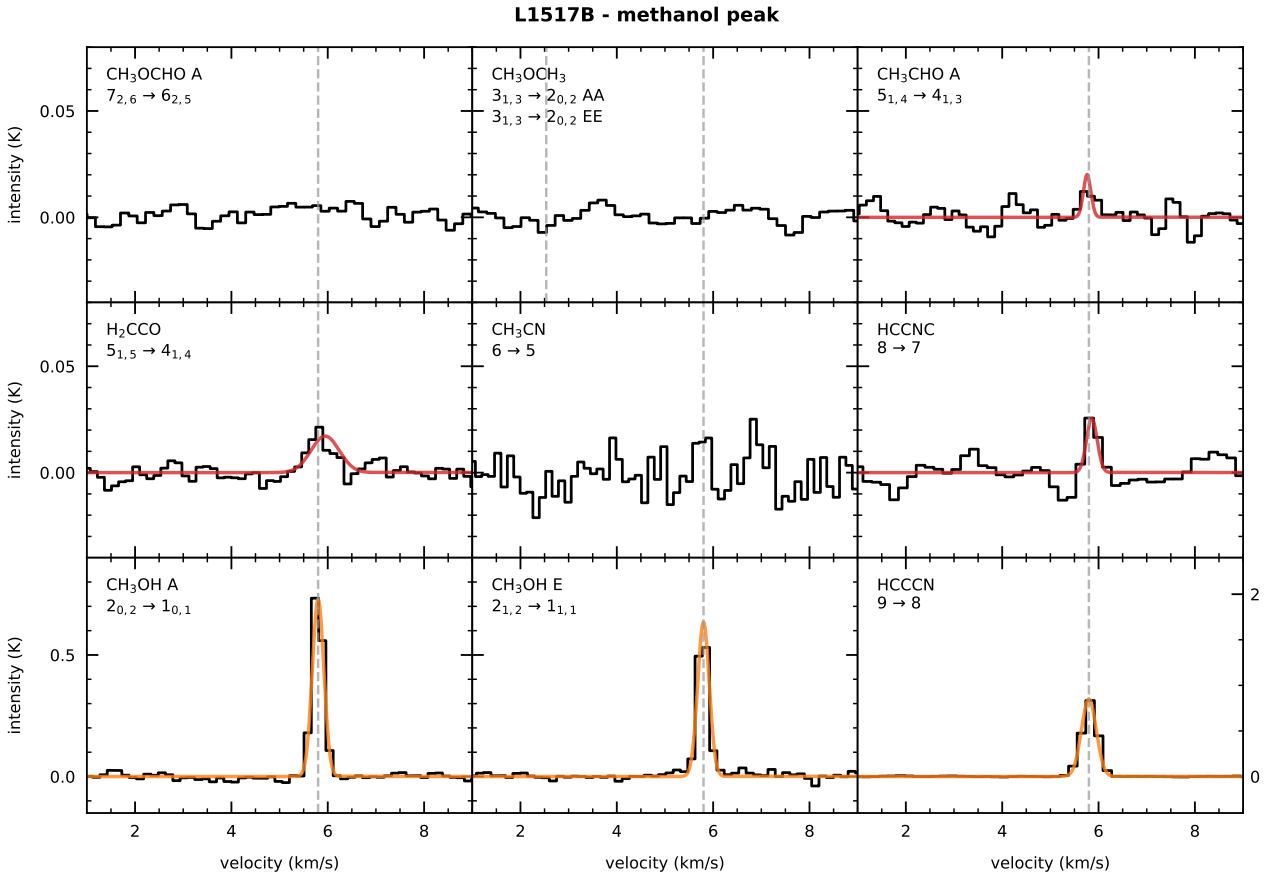


Figure 3.2.1. Sample of the COM lines observed toward L1517B's methanol peak and their corresponding fits obtained with MADCUBA (LTE, in red) and RADEX (non-LTE, in orange). See Sections 3.2 and 3.3.1 for more details. Acetaldehyde (CH_3CHO) transitions are quite noisy, but we could actually detect it because we observed several ones. It seems that acetonitrile (CH_3CN) may show a transition, but we could not properly fit the two covered transitions neither with MADCUBA nor RADEX; plus, the signal to noise is quite low (below 2 for the shown transition and below 1 for the other one).

0.78 at 102–111 GHz.[2] The root mean square (RMS) noise level of the original observations ranged between 3 and 11 mK for both observing positions (see Table 3.1.1), having similar values for each frequency range than the ones obtained by Jiménez-Serra et al. (2016, 2021) for L1544 and L1498.

Our observations have covered the transitions of both O-bearing and N-bearing COMs and COM precursors, summarised in Table 2. For O-bearing species, we have targeted methanol (CH_3OH), methoxy (CH_3O), tricarbon monoxide (CCCO), ketene (H_2CCO), formic acid (t-HCOOH), acetaldehyde (CH_3CHO), methyl formate (CH_3OCHO), dimethylether (CH_3OCH_3), cyclopropenone (c-C_3H_2O), and propynal (HCCCHO). As N-bearing COMs and precursors, we have observed cyanoacetylene (HCCCN), isocyanoacetilene (HCCNC), vinyl cyanide (CH_2CHCN), acetonitrile (CH_3CN), and methyl isocyanide (CH_3NC).

The raw spectra have been analysed and reduced with CLASS, from the package GILDAS,[3] as well as with a Python pipeline written specifically for this purpose.[4] This pipeline fits

[2] https://publicwiki.iram.es/Iram30mEfficiencies/
[3] https://www.iram.fr/IRAMFR/GILDAS/
[4] https://github.com/andresmegias/gildas-class-python/

baselines to the spectra using an iterative method that first masks the strongest lines using sigma-clips and then applies rolling medians and rolling averages, interpolating the masked regions with third order splines. Finally, we used the software MADCUBA (Martín et al., 2019) to search for the molecular transitions of all targeted species, and to carry out the fitting of the molecular line profiles under the assumption of local thermodynamic equilibrium (LTE). The physical parameters derived in the fitting are the molecular column density ($N_{\rm obs}$), excitation temperature ($T_{\rm ex}$), linewidth (Δv) and LSR radial velocity ($v_{\rm LSR}$). For this, we used the tool SLIM (Spectral Line Identification and Modelling) of MADCUBA, employing the Cologne Database for Molecular Spectroscopy (CDMS; Endres et al., 2016) and the Jet Propulsion Laboratory (JPL) molecular catalogue (Pickett et al., 1998). In addition, for the cases of CH_3OH, CH_3CN, and HCCCN, we used the non-LTE code RADEX (van der Tak et al., 2007) to do the fitting of the lines.

3.3 Results

3.3.1 Detected transitions

Figs. 3.1.2 and 3.2.1 show the observed spectra of some representative transitions of the COM and COM precursors detected toward L1517B, while Table 2 lists all the transitions covered in our observations with their derived line parameters. The targeted transitions are the same as in Jiménez-Serra et al. (2016, 2021), and correspond to those expected to be the brightest for an excitation temperature of $\sim$10 K.

Besides methanol, our spectra reveal the detection of other COMs and COM precursor species in L1517B: CH_3O, H_2CCO, CH_3CHO, HCCCN, HCCNC, CH_3CN, and CH_3NC, although acetaldehyde (CH_3CHO) was only detected toward the methanol peak and acetonitrile (CH_3CN) and methyl isocyanide (CH_3NC) were only detected toward the dust peak. More complex molecules such as CH_3OCH_3 or CH_3OCHO were not detected within our noise levels. All detected lines lie above the 3 σ level in integrated intensity (area), where 1 σ is calculated as $\Delta T\sqrt{\Delta v\,\delta v}$, where ΔT is the RMS noise level, Δv is the line width, and δv is the velocity resolution of the spectrum (see Table 1). We are confident about the detection of the transitions since their derived radial velocities correspond to the $v_{\rm LSR}$ of the source ($\sim$5.8 km s^{-1}). Moreover, except for CH_3OH A, CH_3NC and HCCNC, we have measured at least two transitions above the 3 σ level in integrated intensity for the detected species, stressing the identification of the species since the linewidths are narrow ($\sim$ 0.3–0.5 km s^{-1}). The level of line confusion at the targeted RMS noise level is also very low, as commonly found in starless/pre-stellar cores. For the non-detections, we provide upper limits to the integrated intensities by using 3 $\Delta T\sqrt{\Delta v\,\delta v}$.

Table 3.3.1. Most of the COMs and COM precursors transitions covered in our L1517B observations and their derived line parameters.

			Dust peak				Methanol peak			
Species	**Line**	**Frequency** (MHz)	**Area** [a] (mK km s^{-1})	**Linewidth** (km s^{-1})	**LSR velocity** [b] (km s^{-1})	**S/N** [c]	**Area** [a] (mK km s^{-1})	**Linewidth** (km s^{-1})	**LSR velocity** [b] (km s^{-1})	**S/N** [c]
CH_3OH	$2_{0,2} \rightarrow 1_{0,1}$ A	96741.371	206.1 ± 1.9	0.278 ± 0.004	5.785 ± 0.002	109	241.1 ± 2.0	0.277 ± 0.005	5.788 ± 0.002	121
	$2_{1,2} \rightarrow 1_{1,1}$ A	95914.310	< 7	...	...	...	< 7	...	...	...
	$2_{1,1} \rightarrow 1_{1,0}$ A	97582.798	< 2.4	...	...	...	< 2.4	...	...	...
	$2_{1,2} \rightarrow 1_{1,1}$ E	96739.358	150.4 ± 1.9	0.262 ± 0.008	5.786 ± 0.003	80	181.6 ± 2.0	0.262 ± 0.003	5.785 ± 0.003	92
	$2_{0,2} \rightarrow 1_{0,1}$ E	96744.545	9.7 ± 1.9	0.244 ± 0.012	5.84 ± 0.04	5.1	11.2 ± 2.0	0.23 ± 0.08	5.78 ± 0.03	5.6
	$2_{1,1} \rightarrow 1_{1,0}$ E	96755.501	< 7	...	...	...	< 7	...	...	...
CH_3O [d]	$F = 1 \rightarrow 0$, $\Lambda = -1$	82455.980	< 4	...	...	...	< 3	...	...	...
	$F = 2 \rightarrow 1$, $\Lambda = -1$	82458.252	5.4 ± 1.2	0.41 ± 0.08	5.81 ± 0.03	4.5	4.2 ± 0.9	0.39 ± 0.08	5.82 ± 0.03	4.7
	$F = 2 \rightarrow 1$, $\Lambda = +1$	82471.825	5.4 ± 1.2	0.41 ± 0.08	5.81 ± 0.03	4.5	4.2 ± 0.9	0.39 ± 0.08	5.82 ± 0.03	4.7
	$F = 1 \rightarrow 0$, $\Lambda = +1$	82524.180	< 4	...	...	...	< 3	...	...	...
CCCO	$10 \rightarrow 9$	96214.813	< 7	...	...	...	< 7	...	...	...
t-HCOOH	$1_{1,1} \rightarrow 0_{0,0}$	87926.863	< 1.9	...	...	...	< 2.2	...	...	...
H_2CCO	$4_{1,3} \rightarrow 3_{1,2}$ o-	81586.299	11.1 ± 1.3	0.53 ± 0.10	5.89 ± 0.04	8.5	12.6 ± 1.2	0.70 ± 0.25	5.95 ± 0.11	10
	$5_{1,5} \rightarrow 4_{1,4}$ o-	100094.510	9.9 ± 0.9	0.53 ± 0.10	5.89 ± 0.04	11	12.6 ± 1.2	0.70 ± 0.25	5.95 ± 0.11	10
	$4_{0,4} \rightarrow 3_{0,3}$ p-	80832.189	21 ± 4	0.53 ± 0.10	5.89 ± 0.04	5.2	17 ± 4	0.70 ± 0.25	5.95 ± 0.11	4.2
CH_3OCHO	$7_{2,6} \rightarrow 6_{2,5}$ A	84454.754	< 2.0	...	...	...	< 2.5	...	...	...
CH_3OCH_3	$3_{1,3} \rightarrow 2_{0,2}$ EE	82650.180	< 3	...	...	...	< 2.4	...	...	...
	$4_{1,4} \rightarrow 3_{0,3}$ EE	99326.000	< 2.0	...	...	...	< 2.2	...	...	...
CH_3CHO	$5_{1,4} \rightarrow 4_{1,3}$ A	98900.944	< 3	...	...	...	4.0 ± 0.8	0.21 ± 0.05	5.765 ± 0.018	5.0
	$5_{0,5} \rightarrow 4_{0,4}$ A	95963.459	< 7	...	...	...	< 6	...	...	...
	$4_{1,3} \rightarrow 3_{1,2}$ A	79150.166	< 3	...	...	...	3.9 ± 0.7	0.21 ± 0.05	5.765 ± 0.018	5.6
	$4_{1,3} \rightarrow 3_{1,2}$ E	79099.313	< 3	...	...	...	3.9 ± 0.7	0.21 ± 0.05	5.765 ± 0.018	5.6
c-C_3H_2O	$6_{1,6} \rightarrow 5_{1,5}$	79483.519	< 3	...	...	...	< 3	...	...	...
HCCCHO	$9_{0,9} \rightarrow 8_{0,8}$	83775.816	< 2.0	...	...	...	< 2.5	...	...	...
CH_3CN	$6_0 \rightarrow 5_0$	110383.500	6.4 ± 1.9	0.13 ± 0.13	5.902 ± 0.023	3.4	< 6	...	...	...
	$6_1 \rightarrow 5_1$	110381.372	11.8 ± 1.9	0.24 ± 0.08	5.81 ± 0.03	6.2	< 6	...	...	...
	$6_2 \rightarrow 5_2$	110374.989	< 6	...	...	...	< 6	...	...	...
CH_3NC	$5_0 \rightarrow 4_0$	100526.541	3.6 ± 0.7	0.4 ± 0.3	5.87 ± 0.13	5.1	< 2.2	...	...	...
CH_2CHCN	$9_{0,9} \rightarrow 8_{0,8}$	84945.988	< 2.0	...	...	...	< 2.5	...	...	...
	$9_{1,9} \rightarrow 8_{1,8}$	83207.496	< 3	...	...	...	< 2.4	...	...	...
	$10_{0,10} \rightarrow 9_{0,9}$	94276.625	< 3	...	...	...	< 2.4	...	...	...
HCCCN	$9 \rightarrow 8$	81881.461	693.6 ± 1.0	0.395 ± 0.001	5.824 ± 0.001	694	346.7 ± 0.9	0.383 ± 0.002	5.816 ± 0.001	385
	$11 \rightarrow 10$	100076.392	286.5 ± 0.7	0.318 ± 0.001	5.842 ± 0.001	409	96.3 ± 0.8	0.290 ± 0.003	5.837 ± 0.001	120
HCCNC	$8_8 \rightarrow 7_8$	79484.131	16.1 ± 1.1	0.31 ± 0.15	5.83 ± 0.07	15	6.9 ± 0.9	0.29 ± 0.16	5.86 ± 0.06	7.7
	$10_{10} \rightarrow 9_{10}$	99354.250	6.8 ± 0.7	0.31 ± 0.15	5.83 ± 0.07	9.7	< 2.3	...	...	...

Line profiles were fitted using MADCUBA, except for methanol (CH_3OH), cyanoacetilene (HCCCN) and acetonitrile (CH_3CN), where we used CLASS (see Section 3.3.2 for details). ([a]) Uncertainties in the line area are calculated as $\Delta T\sqrt{\Delta v\, \delta v}$, with ΔT the RMS noise nevel, Δv the line width, and δv the velocity resolution of the spectrum. Similarly, upper limits are calculated as $3\ \Delta T\sqrt{\Delta v\, \delta v}$. ([b]) LSR stands for *local standard of rest*. ([c]) This refers to the signal to noise ratio in integrated intensity area. If a certain value has no uncertainty, this means that it had to be fixed so that MADCUBA could fit the LTE model. ([d]) Hyperfine components of the $N = 1 \rightarrow 0$, $K = 0$, $J = 3/2 \rightarrow 1/2$ transition of CH_3O.

From Figs. 3.1.2 and 3.2.1, we find that the line intensities vary for each position: for the N-bearing species (CH_3CN, HCCNC, and HCCCN), the emission is brighter toward the core's centre with respect to the methanol peak. On the contrary, for O-bearing species (CH_3OH and H_2CCO) we find that the emission level is similar for both positions. Note that in the cases of CH_3OH, CH_3CN and HCCCN the fits were obtained with RADEX instead of MADCUBA. In these cases the lines were also fitted with independent gaussians using CLASS, but they are not shown in Figs. 3.1.2 and 3.2.1, as they are independent fits for each transition and there are only made to obtain the parameters of the lines shown in Table 2.

3.3.2 Molecular column densities and excitation temperatures

3.3.2.1 LTE analysis

Using the tool SLIM from MADCUBA, we performed the LTE fitting to the observed line profiles for each species, obtaining the parameters shown in Table 2. As in most cases we detected several lines of the same molecular species, this fitting allows us to obtain its column density (N_{obs}) and excitation temperature (T_{ex}). These parameters are also shown in Table 3.

However, there were cases in which the MADCUBA fitting algorithm did not converge. This can be due because the signal-to-noise ratio is not high enough, or also because we only have one transition detected. In these cases, we had to fix one or more parameters, so that the algorithm could converge. Similarly, for non-detections we had to fix the excitation temperature so that MADCUBA could fit an upper limit for the column density.

In general, when we needed to fix the temperature, we used the fitted value for H_2CCO for oxygen-bearing species, and the fitted one for HCCCN for nitrogen-bearing species, as these are molecules with several transitions which also have high signal-to-noise ratio. These values are, for H_2CCO, 7.7 K and 9.7 K for the dust and methanol peaks, respectively, and, for HCCCN, 6.7 K and 5.3 K. For cases where we had to fix the central velocity of the line, we used a value of v_{LSR} = 5.8 km s^{-1}.

For methyl isocyanide (CH_3NC) toward the core centre and isocyanoacetylene (HCCNC) toward the methanol peak, we only detected 1 transition, so we had to fix the excitation temperature to T_{ex} = 6.7 K for CH_3NC and to T_{ex} = 5.3 K for HCCNC (which are the corresponding values for HCCCN at the dust and methanol peaks, respectively).

For ketene (H_2CCO), we calculated the ortho-to-para ratio (we have detected two ortho transitions and one para transition), by carrying out a separate fit of the transitions and by dividing the resulting column densities of the ortho and para species. We obtain an ortho-to-para ratio of $1.0^{+0.4}_{-0.3}$ for the position of the dust peak and $0.9^{+0.5}_{-0.4}$ for the methanol peak. This ratio is different to what has been previously observed towards the molecular cloud TMC-1 and the pre-stellar core L1689B, where the ratio was $\sim$3 (Ohishi et al., 1991; Bacmann et al.,

Table 3.3.2. Excitation/kinetic temperatures (T), column densities (N_{obs}), and abundances (χ_{obs}) of COMs and COM precursors toward the dust and methanol peaks in L1517B.

	Dust peak			Methanol peak		
Molecule	T (K)	N_{obs} (cm^{-2})	χ_{obs}	T (K)	N_{obs} (cm^{-2})	χ_{obs}
CH_3OH A	10.0	$(4.29 \pm 0.11) \cdot 10^{12}$	$1.23^{+0.21}_{-0.16} \cdot 10^{-10}$	10.0	$(4.96 \pm 0.12) \cdot 10^{12}$	$5.2^{+0.6}_{-0.5} \cdot 10^{-10}$
CH_3OH E	10.0	$4.59^{+2.63}_{-0.25} \cdot 10^{12}$	$1.5^{+0.7}_{-0.3} \cdot 10^{-10}$	10.0	$(5.62 \pm 0.20) \cdot 10^{12}$	$5.9^{+0.7}_{-0.6} \cdot 10^{-10}$
CH_3OH	10.0	$8.9^{+2.6}_{-0.3} \cdot 10^{12}$	$2.7^{+0.7}_{-0.5} \cdot 10^{-10}$	10.0	$(1.058 \pm 0.023) \cdot 10^{13}$	$1.10^{+0.12}_{-0.10} \cdot 10^{-9}$
CH_3O	10 ± 4	$(2.8 \pm 0.9) \cdot 10^{11}$	$8.0^{+2.8}_{-2.5} \cdot 10^{-12}$	7 ± 4	$(1.8 \pm 0.6) \cdot 10^{11}$	$1.9^{+0.7}_{-0.6} \cdot 10^{-11}$
CH_3OCHO	7.7	$< 9 \cdot 10^{11}$	$< 4 \cdot 10^{-11}$	9.7	$< 1.0 \cdot 10^{12}$	$< 1.5 \cdot 10^{-10}$
CH_3OCH_3	7.7	$< 1.0 \cdot 10^{12}$	$< 5 \cdot 10^{-11}$	9.7	$< 5 \cdot 10^{11}$	$< 7 \cdot 10^{-11}$
CH_3CHO	7.7	$< 2.5 \cdot 10^{11}$	$< 1.2 \cdot 10^{-11}$	9.7	$(2.1 \pm 0.4) \cdot 10^{11}$	$2.2^{+0.5}_{-0.4} \cdot 10^{-11}$
t-HCOOH	7.7	$< 3 \cdot 10^{12}$	$< 1.5 \cdot 10^{-10}$	9.7	$< 1.9 \cdot 10^{12}$	$< 3 \cdot 10^{-10}$
c-C_3H_2O	7.7	$< 1.5 \cdot 10^{10}$	$< 7 \cdot 10^{-13}$	9.7	$< 5 \cdot 10^{10}$	$< 7 \cdot 10^{-12}$
H_2CCO	7.7 ± 1.0	$(8.4 \pm 1.3) \cdot 10^{11}$	$2.4^{+0.6}_{-0.4} \cdot 10^{-11}$	10 ± 3	$(7.2 \pm 1.9) \cdot 10^{11}$	$7.6^{+2.2}_{-2.0} \cdot 10^{-11}$
CCCO	7.7	$< 2.4 \cdot 10^{11}$	$< 1.1 \cdot 10^{-11}$	9.7	$< 1.2 \cdot 10^{12}$	$< 1.8 \cdot 10^{-10}$
HCCCHO	7.7	$< 1.6 \cdot 10^{11}$	$< 8 \cdot 10^{-12}$	9.7	$< 1.9 \cdot 10^{11}$	$< 3 \cdot 10^{-11}$
HCCNC	6.7	$(2.4 \pm 1.0) \cdot 10^{11}$	$(7 \pm 3) \cdot 10^{-12}$	5.3	$(1.9 \pm 0.8) \cdot 10^{11}$	$1.9^{+0.9}_{-0.8} \cdot 10^{-11}$
CH_2CHCN	6.7	$< 4 \cdot 10^{10}$	$< 1.8 \cdot 10^{-12}$	5.3	$< 3 \cdot 10^{10}$	$< 5 \cdot 10^{-12}$
CH_3NC	6.7	$(1.6 \pm 0.9) \cdot 10^{10}$	$(5 \pm 3) \cdot 10^{-13}$	5.3	$< 2.3 \cdot 10^{10}$	$< 3 \cdot 10^{-12}$
CH_3CN	10.0	$(2.1 \pm 0.6) \cdot 10^{11}$	$6.0^{+2.1}_{-1.9} \cdot 10^{-12}$	10.0	$< 2.0 \cdot 10^{11}$	$< 3 \cdot 10^{-11}$
HCCCN	10.0	$(5.16 \pm 0.05) \cdot 10^{12}$	$1.48^{+0.25}_{-0.18} \cdot 10^{-10}$	10.0	$3.72^{+0.14}_{-0.13} \cdot 10^{12}$	$(3.9 \pm 0.4) \cdot 10^{-10}$

Temperatures (T) refer to excitation temperatures (T_{ex}) for all the species except for methanol (CH_3OH), cyanoacetilene (HCCCN) and acetonitrile (CH_3CN), where they refer to kinetic temperatures (T_{kin}). We used MADCUBA to derive the molecular parameters from the observations except for methanol, cyanoacetilene and acetonitrile, where we used RADEX. Molecular abundances were calculated using an H_2 column density of $(3.5 \pm 0.5) \times 10^{22}$ cm^{-2} for the dust continuum peak and of $(9.6 \pm 1.0) \times 10^{21}$ cm^{-2} for the position of the methanol peak. For the non-detections and also for some detections, we had to fix the excitation temperature (T_{ex}) so that MADCUBA could fit the column density (N_{obs}).

2012). An exact value of 3 would correspond to the statistical ratio due to the nuclear spin degeneracy. However, lower values are also possible for the lowest temperatures, like in the case of formaldehyde (H_2CO), where an ortho-to-para ratio of 1 would indicate that this molecule is formed under thermal equilibrium at a temperature of ∼10 K (Kahane et al., 1984). We note, however, that we have only detected two ortho lines and one para line of ketene, and therefore the derived ortho-to-para ratios can be subject to significant uncertainties.

3.3.2.2 Non-LTE analysis

For some cases, in particular for methanol (CH_3OH), cyanoacetilene (HCCCN) and acetonitrile (CH_3CN), we have decided to use the non-LTE code RADEX (van der Tak et al., 2007) to infer the column density N_{obs} of the species instead of MADCUBA. In the case of methanol and acetonitrile, we did it because MADCUBA could not fit properly all the transitions, and the errors were greater than 3 σ. For the case of cyanoacetilene (HCCCN), we opted for

RADEX because it allows us to estimate the H_2 number density, as we have two transitions. As input parameters for these calculations, we assumed the linewidth obtained with the MADCUBA fit ($\sim$0.28 km s^{-1}) and a kinetic temperature of 10 K (as inferred for all radii across the L1517B starless core; see Tafalla et al. (2004) and Section 3.5 below). Apendix B.1 describes the procedure used to derive the phyisical properties and their uncertainties from the modeling.

In the case of CH_3OH, we have to distinguish between the A and E species, making a separate fit for each molecule. We used the collisional coefficients with H_2 given by Rabli and Flower (2010). In addition to deriving the column densities reported in Table 3, we also derived the H_2 number density from the observations of methanol E, as this parameter can be determined by the intensity ratio of the detected lines. Unfortunately, for the case of methanol A we only detected one transition so the column density was estimated by assuming the H_2 number density obtained by Tafalla et al. (2004): 2.20×10^5 cm^{-3} for L1517B core centre and 1.23×10^5 cm^{-3} for the location of the methanol peak. For methanol E we could derive the H_2 number density, although the uncertainties are high, obtaining values of $5^{+8}_{-4} \times 10^4$ cm^{-3} for the core's centre and $2.0^{+0.9}_{-0.6} \times 10^5$ cm^{-3} for the methanol peak. These values are consistent, within the uncertainties, with those derived by Tafalla et al. (2004), as we have to take into account the possible uncertainties in the determination of the density profiles from the absorption coefficient of the dust. As for the excitation temperatures, they are close to the assumed kinetic temperature of 10 K, consistent with those expected for the derived H_2 densities, although this is biased by the fact that the lowest temperature for which we have collisional rates is 10 K.
For CH_3CN, we fixed the H_2 number density to the values from Tafalla et al. (2004) (as the signal-to-noise ratio was not high enough to get a good fit). For the core's centre, one of the lines could be fitted within 2 σ, providing a better fit than the one by MADCUBA. For the methanol peak, we did not detect any transition, so we obtained an upper limit to its column density from RADEX.

For HCCCN, we have used the two transitions to derive the column density and the H_2 number density towards both positions. We used the collisonal coefficients with H_2 given by Faure et al. (2016). Fig. B.1.3 shows the results obtained from the non-LTE analysis for the core's centre. As expected the derived H_2 densities and column densities are related. We obtained the HCCCN column densities shown in Table 3 and H_2 densities of $1.80^{+0.04}_{-0.04} \times 10^5$ cm^{-3} and $5.56^{+0.19}_{-0.18} \times 10^4$ cm^{-3} for the dust and methanol peaks, respectively. Although these results present lower uncertainties than those of methanol, our H_2 densities are still consistent with the prediction by Tafalla et al. (2004) taking into account the beam size of our observations and the uncertainties in the determination of the density from dust emission, as we previously discussed. In the following sections, we use the values from Tafalla et al. (2004) because they are more direct than the derivation through RADEX, and because the authors assume the same dust emissivity ($\kappa = 0.005$ cm^2 g^{-1} for 1.2 mm) as that considered by Crapsi et al. (2005) for inferring the CO depletion factor of L1517B, which has been employed in our astrochemical simulations (Section 3.5).

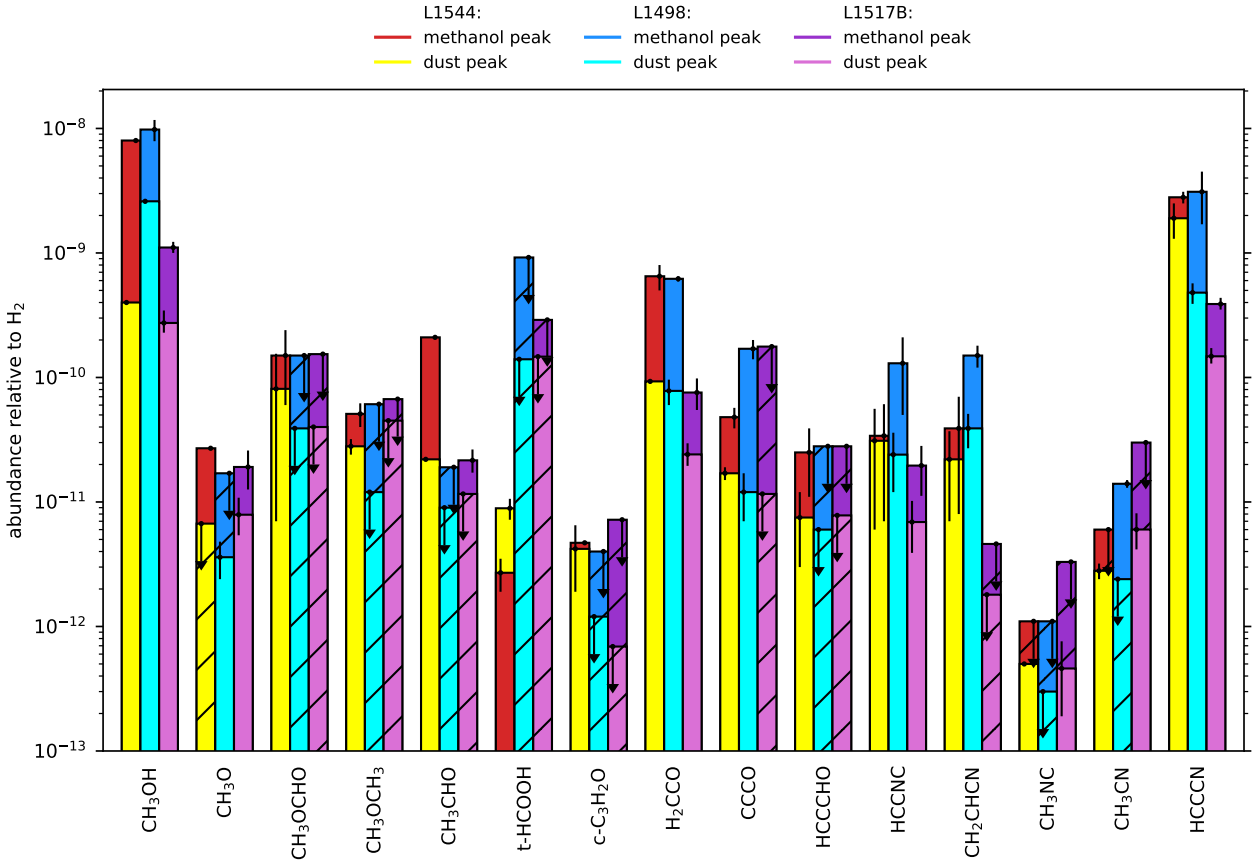


Figure 3.3.1. Bar plot of the abundances of different COMs and COM precursors measured for the cores L1517B, L1498 and L1544, toward the dust peak (lighter colors) and the methanol peak (darker colors) for each of them. Upper limits are indicated both with arrows at the top of the corresponding bar and with stripes. Abundances for L1544 are taken from Jiménez-Serra et al. (2016, 2021), while for L1498 they are taken from Jiménez-Serra et al. (2021). Note that this is not a cumulative bar plot, so the upper edge of each bar, for both colors (that is, for both the centre and the methanol peak, for each core), indicates the abundance of the corresponding species.

3.3.3 Molecular abundances

Once we know the column densities of COMs and COM precursors detected toward L1517B, we can compute the molecular abundances of these species by using the H_2 column densities measured toward the positions of the dust and methanol peaks (Table 3). For the position of the dust peak, the H_2 column density of $(3.5 \pm 0.5) \times 10^{22}$ cm^{-2} was obtained using the 1.2 mm continuum emission observed by Tafalla et al. (2004) with the MAMBO 1 mm bolometer array at IRAM 30 m telescope. For the position of the methanol peak, however, we employed the data from the *Herschel* space telescope at 0.25, 0.35 and 0.5 mm assuming a dust optical depth index of $\beta = 1.5$ (Spezzano et al., 2016); the derived H_2 column density is $(9.6 \pm 1.0) \times 10^{21}$ cm^{-2}. As discussed in Jiménez-Serra et al. (2021), this is the best procedure to probe the total H_2 column density respectively toward the innermost regions (with bolometers at 1.2 mm) and outer shells (with *Herschel*) in starless cores.

From Table 3, we can see that the abundance of the detected species is enhanced toward the position of the methanol peak with respect to the core's centre. For methanol (CH_3OH),

its abundance is enhanced by a factor of ∼4, while for the rest of detected species in both positions (CH_3O, H_2CCO, HCCNC, and HCCCN) the factor of enhancement is ∼3. Acetaldehyde (CH_3CHO) is only detected in the methanol peak, the ratio would be $\gtrsim$2. As for acetonitrile (CH_3CN) and methyl isocianide (CH_3NC) we only find lines toward the dust peak, and the ratio would be $\lesssim$6. Therefore, for these three molecules, the lower/upper limits for the factor of enhancement toward the methanol peak are consistent with those of the molecules detected in both positions. Additionally, the ratio of the abundances of methanol A and E species toward the dust and methanol peak is, respectively, $0.91^{+0.08}_{-0.32}$ and $0.88^{+0.04}_{-0.04}$, which is similar to what has been observed in the L1544 and L1498 cores (Jiménez-Serra et al., 2016, 2021) and to the expected values of ∼ 0.7–1.0 (Wirström et al., 2011).

3.4 Comparison with other cores

In this Section, we compare our results obtained for the starless core L1517B with those of the L1544 and L1498 cores (see Jiménez-Serra et al., 2016, 2021) and also with those of the L1521E core (see Nagy et al., 2019; Scibelli et al., 2021).

3.4.1 Comparison with L1498 and L1544

For L1517B, L1498 and L1544, the same molecular species have been observed toward two positions (i.e., the dust peak and the methanol peak for each of them), which allows us to obtain information about the radial distribution of the abundances of COM and COM precursors within the cores.

Fig. 3.3.1 presents the abundances of the COM and COM precursor molecules observed toward these three cores and for the two measured positions. The abundances measured toward the dust and methanol peaks are plotted respectively in yellow and red for the L1544 pre-stellar core (Jiménez-Serra et al., 2016), in light and dark blue for the L1498 starless core (Jiménez-Serra et al., 2021), and in light and dark purple for the L1517B starless core (this work).

Consistently with L1498 and L1544, the abundances of the species detected toward L1517B tend to be higher toward the position of the methanol peak than toward the dust peak (t-HCOOH toward L1544 is an exception). Fig. 3.3.1 also shows that the L1517B core presents a lower number of detections than L1544 and a similar number to L1498. For the N-bearing species, we have detected only HCCCN and HCCNC toward both positions of L1517B from a total of 5 targeted species, although we have found CH_3CN and CH_3NC toward the dust peak. A significant difference with respect to L1498 and L1544 is that vinyl cyanide (CH_2CHCN) is not detected toward L1517B, showing upper limits to its abundance that are factors $\gtrsim$10 lower than the abundances measured toward L1498 and L1544. Following the same trend, for O-bearing species only four COMs and COM precursor species (CH_3OH,

CH_3O, CH_3CHO, and H_2CCO) out of 10 have been detected toward L1517B, and in the case of CH_3CHO it is only detected toward the methanol peak.

If we compare in detail L1517B and L1498, we can see that the first one has a level of chemical complexity somewhat lower than L1498. Firstly, the abundances of the COMs and COM precursors detected toward L1517B tend to show lower abundances than (or are comparable to) those measured toward L1498. Secondly, if we count the total number of detections for any position, we obtain 13 detections for L1517B and 14 for L1498, which may seem a small difference. However, if we consider the detections and non-detections for each position, the lower level of chemical complexity for the L1517B starless core becomes more apparent.

To quantitatively measure this we have calculated, for each pair of cores, the fraction of molecules with the same state of detection (detected or not detected). That is, for each source, we build a vector with one entry for each molecule and position (15 molecules $\times$ 2 positions, 30 entries), whose value is 1 if there is a detection and 0 if not. Then, we build another vector with the same number of entries, that are 1 if the two input entries (of the two cores) have the same value and 0 otherwise. Finally, we sum up all the elements of this vector and divide it by the number of entries, in order to normalize it. We call the resulting number the *fractional similarity*, whose values lie between 0 and 1. In this way, if for each position and molecule both entries have the same value toward both sources (either a detection or a non-detection), the fractional similarity is 1. On the contrary, if for each position and molecule the entries are different, the resulting similarity is 0. If we compute the fractional similarity between L1517B and L1498 we obtain 0.70 ± 0.03.[5] This indicates that, although the two cores are similar in terms of number of detections, they are truly different regarding their level of chemical complexity since the similarity, clearly less than 1, is produced by 9 different entries in the vectors (that is, in the detections for each molecule and position).

For the sake of completeness, we have also computed the fractional similarity between L1544 and L1517B, which is 0.57 ± 0.03, and between L1544 and L1498, which is 0.43 ± 0.03. This indicates that the three cores show a significantly different level of chemical complexity in terms of detections and non detections, and the two more similar cores would be L1498 and L1517B.

We have used two other methods to evaluate the level of chemical complexity toward the three cores. First, we have computed for each source and position the mean molecular mass of the species weighted by the abundances in a logarithmic scale.[6] This can be viewed as an estimate of the level of complexity of each source and position, as COMs tend to have more atoms with increasing complexity, and thus a greater molecular mass. From Table 3.4.1, we

[5] The value used for the uncertainty is $1/30 \simeq 0.03$, which is the difference in the fractional similarity produced by one entry that is different in the two vectors.

[6] We used logarithmic weights due to the range of the values of our abundances, which encloses several orders of magnitude.

Table 3.4.1. Weighted geometric mean of the abundance with respect to H_2, $\langle \chi_{obs} \rangle$, and weighted mean molecular mass, $\langle m_{molec} \rangle$ for the targeted species in L1517B, L1498, and L1544.

Source		$\langle \chi_{\mathbf{obs}} \rangle$ $(\times 10^{-12})$		$\langle m_{\mathbf{molec}} \rangle$ (g/mol)	
L1517B	dust peak	$7.1^{+1.5}_{-1.5}$	15^{+3}_{-3}	$45.5^{+0.3}_{-0.3}$	$45.61^{+0.20}_{-0.27}$
	meth. peak	23^{+5}_{-5}		$45.7^{+0.3}_{-0.4}$	
L1498	dust peak	$10.5^{+2.1}_{-2.1}$	32^{+5}_{-5}	$46.18^{+0.18}_{-0.24}$	$46.25^{+0.15}_{-0.17}$
	meth. peak	53^{+10}_{-10}		$46.32^{+0.23}_{-0.24}$	
L1544	dust peak	19^{+4}_{-4}	30^{+4}_{-4}	$46.8^{+0.3}_{-0.4}$	$46.46^{+0.15}_{-0.21}$
	meth. peak	42^{+7}_{-7}		$46.13^{+0.16}_{-0.19}$	

The weights for the mean abundance are the molecular masses for each molecule. The weights for the mean molecular mass are the abundance for each molecule in logarithmic scale. The values on the second columns for each magnitude are the arithmetic means of the magnitudes for both positions. To deal with uncertainty propagation and upper limits we made simulations with statistical distributions (see Appendix B.2).

find that the lowest value of the mean molecular mass is obtained for L1517B, while the greatest one is for L1544, with L1498 being close to it. This would indicate that L1517B presents the lowest level of chemical complexity of the three cores, followed by L1498 and L1544. Additionally, we also computed for each source and position the geometric mean of the abundance of the species weighted by the molecular mass.[7] The resulting value represents a measure of the amount of the more complex species in the core, as we are giving larger weights in the mean to more complex species. As shown by Table 3.4.1, L1517B shows the lowest mean abundance, followed by L1498 and L1544, which present barely the same quantity. All of this indicates that the level of chemical complexity in the L1517B starless core seems to be lower than that of L1498, and of L1544.

From all these results, we suggest that L1517B is in a less chemically evolved stage than L1498 and L1544. This is supported by the deuterium fractionation of L1517B, similar to that of L1498, but L1517B shows no evidence of infall motions (Crapsi et al., 2005; Tafalla et al., 2004). Additionally, L1517B shows the lowest level of chemical complexity within the three cores, presenting only the simplest COMs and COM precursors and with lower abundances in most of the cases.

If we take into account the possible time evolution between the three cores (with L1517B being the youngest and L1544 the oldest), we propose a scenario in which the N-bearing molecules would form first in starless cores. Then, as the cores evolve, they would accrete gas from the surrounding molecular cloud, becoming denser and yielding a strong depletion of carbon monoxide (CO). At the pre-stellar core stage, the catastrophic depletion of CO takes place, triggering the formation of O-bearing COMs and COM precursors as observed in L1544 (Vasyunin et al., 2017). Therefore, chemical complexity would increase over time. In the next section, we explore this scenario by modelling the chemistry of N-bearing and

[7] We used the geometric mean due to the range of the values of our abundances.

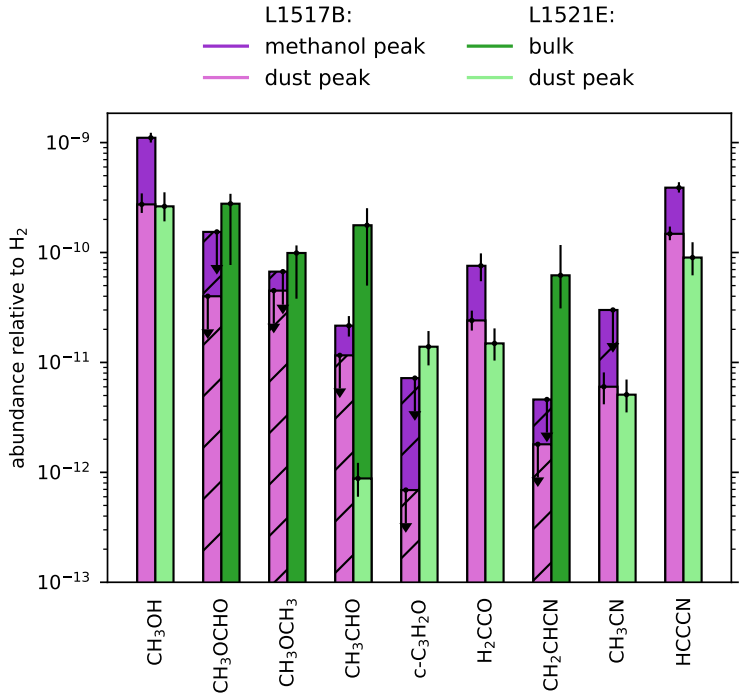


Figure 3.4.1. Bar plot of the abundances of different COMs and COM precursors measured for the young starless cores L1517B (in light and dark purple) and L1521E (in light and dark green). For L1517B, the values are as in Figure 3.3.1. For L1521E, the abundances shown in dark green refer to the *bulk COM emission* observed by Scibelli et al. (2021) toward L1521E with a beam size of ∼70″. The large beam includes not only the dust peak position but also the methanol ring (peak located at 22–29″ from L1521E's dust peak). Abundances from L1521's dust peak are shown in light green and are taken from Nagy et al. (2019). Upper limits are indicated with both arrows and stripes.

O-bearing COMs and COM precursors toward L1517B.

3.4.2 Comparison with L1521E

L1521E is another well-studied starless core located in the Taurus molecular cloud. Like L1517B, L1521E is also considered to be young. Indeed, L1521E shows no evidence of infall motions, and has a central density of 2.7×10^5 cm^{-3} (Tafalla and Santiago, 2004). Its CO depletion factor is $f_{\rm CO} = 4.3 \pm 1.6$ (Nagy et al., 2019), lower than those of L1517B and L1498. Although its deuterium fractionation $N_{\rm N_2D^+}/N_{\rm N_2H^+}$ has not been measured yet, several authors agree in the fact that this core is young (e.g., Tafalla and Santiago, 2004; Kong et al., 2015; Scibelli et al., 2021). Nagy et al. (2019) observed the molecular line emission toward the dust peak of this core (including some COMs and COM precursors) using the IRAM 30 m telescope. More recently, Scibelli et al. (2021) detected four COMs toward L1521E (CH_3OCHO, CH_3OCH_3, CH_3CHO, and CH_2CHCN) using the 12 m ARO (Arizona Radio Observatory) telescope, with a much larger beam size (∼70″ versus 22–31″ for the IRAM 30 m telescope). This implies that while the observations of Nagy et al. (2019) would cover the location of just the dust peak, the observations of Scibelli et al. (2021) would cover both the dust peak and the methanol ring, as its methanol peak is found 22–29″ away from

the core center (Nagy et al., 2019). We thus refer to the observations of Scibelli et al. (2021) as those performed toward the *bulk* of the core.

Among the 15 targeted species measured toward L1517B, there are 9 of them that have also been targeted toward L1521E: six toward the dust peak and four for the bulk of the core (with just one molecule in common for both positions). Fig. 3.4.1 reports the abundances of the common species observed toward L1517B and L1521E. For the data under the *bulk* label, we use the values derived by Scibelli et al. (2021) assuming two source size cases: the lower limits of the error bars refer to the values obtained assuming a beam filling factor of 1, while the error bars upper limits refer to the abundances obtained with the best-fit source size of 35″. Note, however, that vinyl cyanide (CH_2CHCN) is an exception since its abundance was only derived assuming a beam filling factor of 1 (the source size for this molecule could not be estimated; Scibelli et al., 2021).

In general, there is a good agreement between the abundances measured toward the dust peak position of the two cores (see light purple and light green bars in Fig. 3.4.1), with the sole exception of cyclopropenone (c-C_3H_2O). In contrast, the differences are bigger between the bulk abundances of L1521E and the abundances measured toward the dust and methanol peaks of L1517B. However, if we take into account the uncertainties with the error bars upper/lower limits, the abundances of L1521E's bulk are largely consistent with those measured toward both positions of L1517B, with the exception of vinyl cyanide (CH_2CHCN). A possible explanation for this

We finally note that calculations of the fractional similarity and of the geometric means of the abundance and of the molecular mass for L1521E, cannot be performed since the number of data available for this core is rather small.

3.5 Modelling the formation of COMs and COM precursors in L1517B

3.5.1 The model

We have modelled the chemistry of COMs and COM precursors toward the starless core L1517B by using the 0D gas-grain chemical code MONACO (Vasyunin and Herbst, 2013; Vasyunin et al., 2017), which has been successfully applied previously to the starless cores L1544 and L1498 (Jiménez-Serra et al., 2016, 2021). Our goal is to compare the observed molecular abundances with those predicted by the model, and to infer the radial distribution of COMs and COM precursors toward L1517B and its age.

The MONACO chemical code is a rate equations-based, three-phase (gas, ice surface, ice bulk) numerical model that provides the evolution of the fractional abundances of atomic

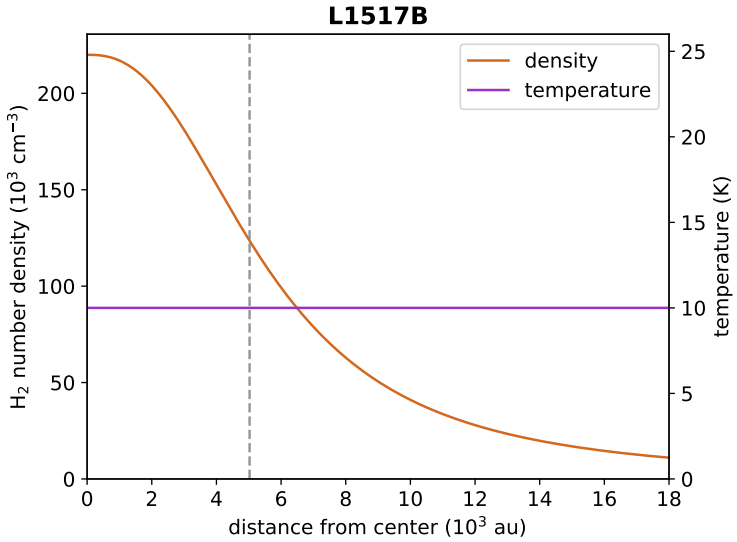


Figure 3.5.1. H_2 gas density and temperature profiles used for modelling the chemistry of COMs and COM precursors toward L1517B. These radial profiles are taken from Tafalla et al. (2004). The vertical gray dashed line indicates the distance of the observed methanol peak with respect to the centre of the core.

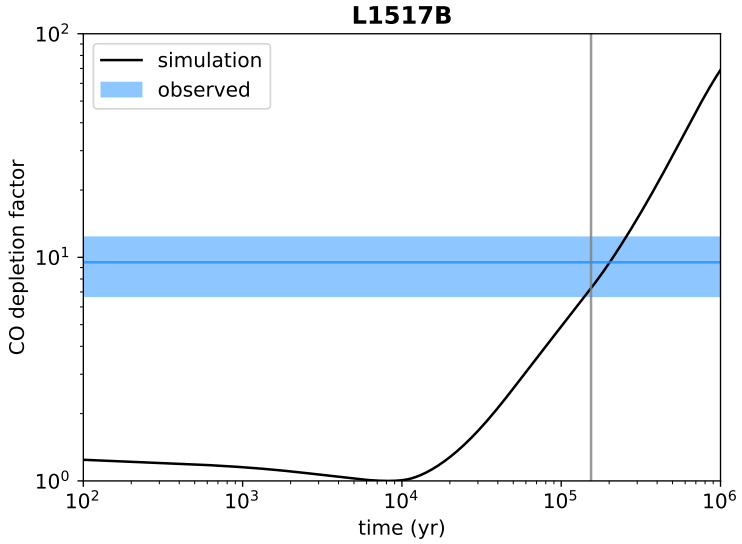


Figure 3.5.2. Evolution of the simulated CO depletion factor over time. The blue line and shaded region show the value $f_{CO} = 9.8 \pm 2.8$, derived by Crapsi et al. (2005), while the gray line indicates the best-fit age for L1517B.

and molecular species with time. To obtain the radial distribution of the abundances of COMs and COM precursors in L1517B, the code is run for a grid of distances on the density and temperature profile of L1571B, assuming physical parameters to be constant in time. The physical structure of the L1517B starless core is obtained from the H_2 gas density profile by Tafalla et al. (2004), and the gas kinetic temperature distribution inferred by the same authors using NH_3 observations. The assumed H_2 gas density profile for L1517B is:

$$n(r) = \frac{2.2 \times 10^5 \text{ cm}^{-3}}{1 + \left(\frac{r}{35''}\right)^{2.5}} , \tag{3.1}$$

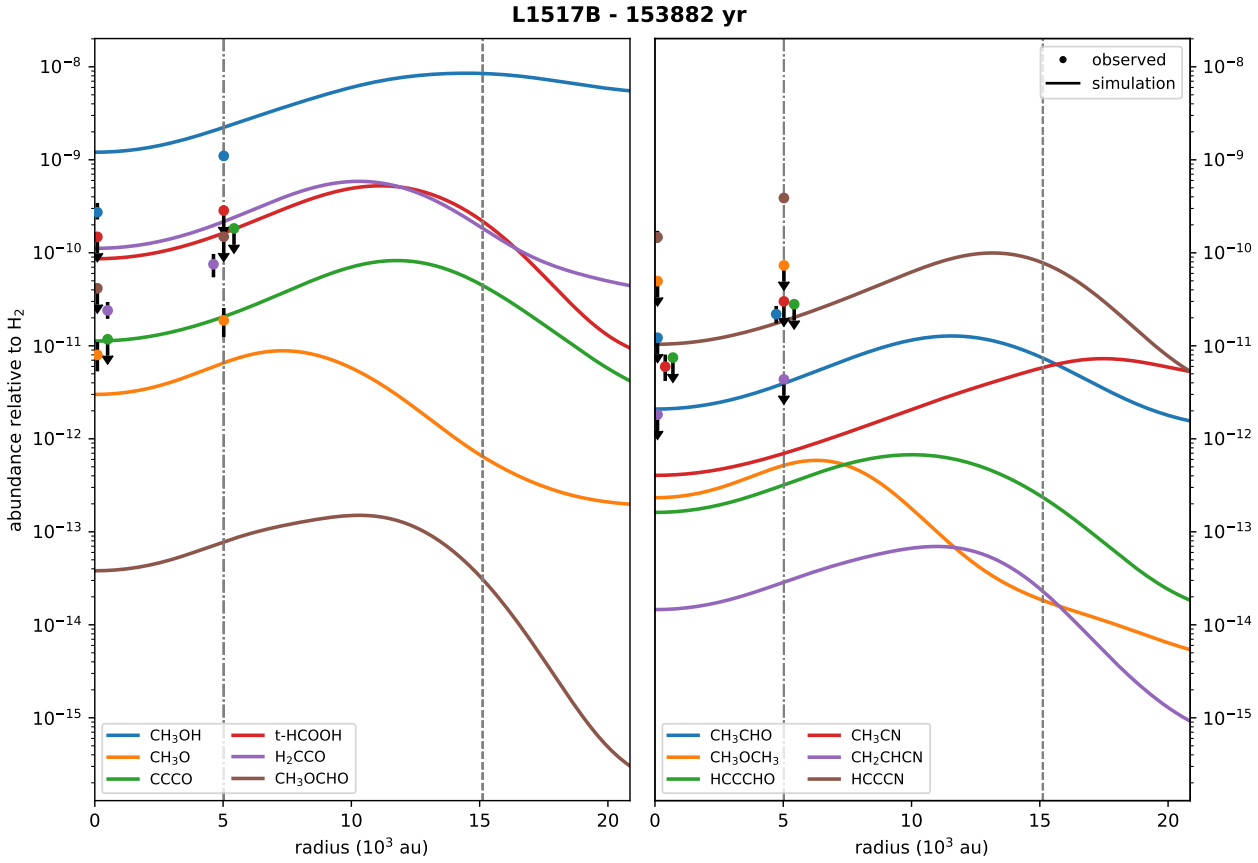


Figure 3.5.3. Radial distribution of the modelled abundances of the COMs and COM precursors observed for L1517B. Colour curves represent the modelled abundances, while dots and vertical arrows represent the observed abundances, and their upper limits, toward the two observed positions in L1517B, the dust and methanol peaks. The position of some of the dots (observed COMs and COM precursor abundances) has been slightly shifted to enhance visibility. The dotted-dashed gray line indicates the position of the observed methanol peak, while the gray dashed line indicates the position of the methanol peak according to the chemical simulations.

where r is the angular distance (in arcseconds). As for the temperature radial profile, we have used a constant temperature of T = 10 K (see Tafalla et al., 2004). Fig. 3.5.1 shows both profiles as a function of distance to the centre of the core. In our model, we assume that the dust and gas temperatures are equal.

Given the low temperatures in starless cores ($T \simeq 10$ K), the precursors of COMs are formed on the surface of dust grains via hydrogenation processes. Once formed, a small fraction of these compounds are non-thermally desorbed from dust grains into the gas phase, where they undergo gas-phase chemical reactions that yield the observed COMs (Vasyunin and Herbst, 2013; Vasyunin et al., 2017). As non-thermal desorption processes, the code includes UV photo-desorption, cosmic-ray-induced desorption (Hasegawa and Herbst, 1993) and reactive (chemical) desorption (Minissale et al., 2016). The chemical network for the O-bearing and N-bearing COMs and precursors is the same as the one used for the L1498 starless core (see the details in Jiménez-Serra et al., 2021). As initial abundances, the model employs the results of a simulation of a diffuse cloud model with constant gas density of 100 cm^{-3} and gas temperature of 20 K for 10^7 years, using the *low metals* initial abundances

EA1 from Wakelam and Herbst (2008).

3.5.2 Comparing simulations and observations

Using the initial conditions explained in Section 3.5.1, we have simulated the chemical evolution of L1517B by running the MONACO code for 10^6 years. To constrain the age of the core, we have used the CO depletion factor of f_{CO} = 9.5 $\pm$ 2.8 measured by Crapsi et al. (2005) toward L1517B, and the location of the methanol peak at a distance of 5000 au. Interestingly, our model cannot reproduce simultaneously the location of the methanol peak at 5000 au and the CO depletion factor of $\sim$7–12. In Fig. 3.5.2, we show the evolution of the predicted CO depletion factor over time and its observed value. Note that the MONACO code considers the telescope beam size when calculating the CO depletion factor. By comparing the model results with the CO and COMs and COM precursor observations, the best agreement is reached after 1.5 $\times$ 10^5 years. The CO depletion factor at this time is $\sim$7.3, which is on the lower edge of the estimated interval for the CO depletion factor (see Fig. 3.5.2).

At this time of chemical evolution, however, the location of the methanol peak in the model does not match the observed one toward L1517B. This is clearly shown in Fig. 3.5.3, where we present the radial distribution of the abundance of COMs and COM precursors modelled for L1517B, together with the measured values and upper limits (see dots and vertical arrows). As shown in Fig. 3.5.3, the location of the methanol peak in the model is at $\sim$15000 au away from the dust peak (vertical dashed line), which is a factor of 3 further away that observed (at $\sim$5000 au; see dashed-dotted vertical line).

The origin of this discrepancy likely arises from the fact that the central H_2 gas density toward L1517B is 2.2 $\times$ 10^5 cm^{-3}, i.e. significantly lower than measured toward the L1544 pre-stellar core. If this density were higher, as in L1544 ($\sim$$10^7$ cm^{-3}; Caselli et al., 2022), the CO depletion factor (9.5 $\pm$ 2.8) would be reached earlier in the simulations, enabling a better match between the modelled and the observed location of the methanol peak at 5000 au. This would also help reconciling the age of the L1517B starless core predicted by the model (of $\sim$1.5$\times$$10^5$ years) with the one expected from observations. As already mentioned, L1517B does not show evidence for gas accretion either toward the innermost regions or from the outer envelope, as observed toward L1498 or L1544, which suggests that L1517B is at an earlier stage of evolution, or at least, at a similar evolutionary stage as L1498 (note that they both have similar central H_2 gas densities and deuterium fractionation values; Tafalla et al., 2004; Crapsi et al., 2005). Another possible explanation of this discrepancy would be a prominent asphericity of the core, although from Fig. 3.1.1 it seems that L1517B is rather spherical.

For the abundances of COMs and COM precursors, Table 5 compares the values between the modelled and observed abundances toward the positions of the dust and methanol peaks in L1517B. All the modelled and observed abundances agree within a factor of 10

Table 3.6.1. Observed and modelled abundances (χ_{obs}, χ_{mod}) for several COMs and COM precursors in L1517B.

	Dust peak			Methanol peak		
Species	χ_{obs}	χ_{mod}	**agreement**	χ_{obs}	χ_{mod}	**agreement**
CH_3OH	$2.7^{+0.7}_{-0.5} \times 10^{-10}$	1.21×10^{-9}	+	$1.11^{+0.12}_{-0.10} \times 10^{-9}$	2.23×10^{-9}	+
CH_3O	$8.0^{+2.9}_{-2.5} \times 10^{-12}$	3.0×10^{-12}	+	$1.9^{+0.7}_{-0.6} \times 10^{-11}$	6.6×10^{-12}	+
CCCO	$< 8 \times 10^{-12}$	1.12×10^{-11}	+	$< 1.3 \times 10^{-10}$	2.05×10^{-11}	+
t-HCOOH	$< 1.0 \times 10^{-10}$	8.6×10^{-11}	+	$< 2.0 \times 10^{-10}$	1.63×10^{-10}	+
H_2CCO	$2.4^{+0.6}_{-0.5} \times 10^{-11}$	1.12×10^{-10}	+	$7.6^{+2.3}_{-2.0} \times 10^{-11}$	2.17×10^{-10}	+
CH_3OCH_3	$< 3 \times 10^{-11}$	2.32×10^{-13}	+	$< 5 \times 10^{-11}$	5.2×10^{-13}	+
CH_3CHO	$< 8 \times 10^{-12}$	2.09×10^{-12}	+	$2.2^{+0.5}_{-0.4} \times 10^{-11}$	4.0×10^{-12}	+
HCCCHO	$< 5 \times 10^{-12}$	1.62×10^{-13}	+	$< 2.1 \times 10^{-11}$	3.2×10^{-13}	+
CH_3CN	$6.0^{+2.1}_{-1.9} \times 10^{-12}$	4.0×10^{-13}	-	$< 3 \times 10^{-11}$	7.0×10^{-13}	+
CH_2CHCN	$< 1.2 \times 10^{-12}$	1.46×10^{-14}	+	$< 3 \times 10^{-12}$	2.9×10^{-14}	+
HCCCN	$1.48^{+0.24}_{-0.18} \times 10^{-10}$	1.04×10^{-11}	-	$3.9^{+0.4}_{-0.4} \times 10^{-10}$	1.85×10^{-11}	-

The modelled and measured abundances agree (+) or not (-) within a factor of 10.

except for CH_3CN in the core's center and HC_3N in both positions. Actually, at earlier times of the simulations, where the simulated methanol peak is closer to the observed one, abundances of those two species are closer to the observed ones. In any case, we should take into account that astrochemical models entail intrinsic uncertainties derived from uncertainties from some of the constants used in the simulation, such as the chemical reaction rates (Vasyunin et al., 2004; Wakelam et al., 2005, 2010).

Comparing the predicted age for L1517B (1.5×10^5 yr) with those predicted for L1498 and L1544, we find that, contrary to what we thought, L1517B would not be the youngest core, as the predicted ages for L1498 and L1544 are, respectively, 9.8×10^4 yr and 1.6×10^5 yr (Jiménez-Serra et al., 2016, 2021).

For L1521E, Scibelli et al. (2021) also used the MONACO code to estimate the age of the core (of $\sim 6 \times 10^4$ yr). However, they could only reproduce the observed abundances of acetaldehyde (CH_3CHO), out of a total of five modelled COMs (with differences in the predicted abundances by factors of 20-160). Therefore, additional aspects in the chemical modelling of these young starless cores such as different initial conditions and/or dynamical evolution of the cores, may be required.

3.6 Evolution in the HC_3N / CH_3CN abundance ratio across low-mass star formation

Recent observational campaigns have been devoted to investigate the chemical composition in COMs (both O-bearing and N-bearing) towards all stages in the process of star formation. In this Section, we compare the information available for several sources for two of

these COMs and COM precursors, the N-bearing species cyanoacetylene (HC_3N) and acetonitrile (CH_3CN), with the abundances obtained towards our limited sample of starless cores (L1544, L1498, L1517B, and L1521E). We have selected these two molecular species because they have been measured systematically in the past few years across a variety of objects, from Class 0/I protostars (Bergner et al., 2017) to Class II protoplanetary discs (see the MAPS ALMA large program; e.g., Ilee et al., 2021), and comets (Biver et al., 2021).

Fig. 3.6.1 reports the values of the column density ratio of HC_3N / CH_3CN for the starless cores L1544, L1498 and L1517B (for both positions measured in these cores), the starless core L1521E in the dust peak, twelve Class 0/I protostellar systems (studied by Bergner et al., 2017), four Class II protoplanetary discs (studied by Ilee et al., 2021), and two comets (studied by Biver et al., 2021, and references therein). To obtain this ratio, we used the column densities for both molecules, except for the comets, for which we employed the abundances of both molecules with respect to water. From Fig. 3.6.1, it seems that there is a trend for the starless cores to present much higher HC_3N / CH_3CN ratios (of $\sim$200–700) than those measured toward Class 0/I protostars and Class II protoplanetary discs, and toward comets (in this case, their HC_3N / CH_3CN abundance ratios reach values even lower than 10). The core L1521E would be an exception, as its ratio is similar to those of the protostars. In any case, the observed general trend suggests that the observed HC_3N / CH_3CN abundance ratio decreases when starless cores enter into the Class 0 phase, remaining approximately constant until the Class II protoplanetary phase, and droping by two orders of magnitude for comets.

The observed range of values for the HC_3N / CH_3CN ratio could be explained by the fact that it is difficult to form HC_3N at late evolutionary stages in the process of low-mass star formation. Indeed, HC_3N is mainly a gas-phase product whose formation is favoured by the low-densities and cold temperatures found in the outer envelopes of pre-stellar systems and protostars (see Section 5.1 in Bergner et al., 2017, for more details). In the same work, the authors report similar CH_3CN abundances for different kinds of sources (young stellar objects, comets, hot cores, and hot corinos), which would explain the decreasing trend in the HC_3N / CH_3CN abundance ratio with time.

3.7 Conclusions

The high-sensitivity spectra obtained for the starless core L1517B reveal a chemical complexity a bit poorer than that observed in the core L1498, and quite lower than in L1544, as it only presents a few and simple O-bearing and N-bearing species, such as CH_3O, H_2CCO, CH_3CN, and HCCCN, and in general with lower abundances than in L1498 and L1544. Quantitatively, the geometric mean of the abundance of the targeted species in L1517B is about half than in L1498 and L1544. Similarly to these latter cores, the targeted molecules are more abundant by a factor of $\sim$3 towards the methanol peak of L1517B, located at $\sim$5000 au from the core's centre.

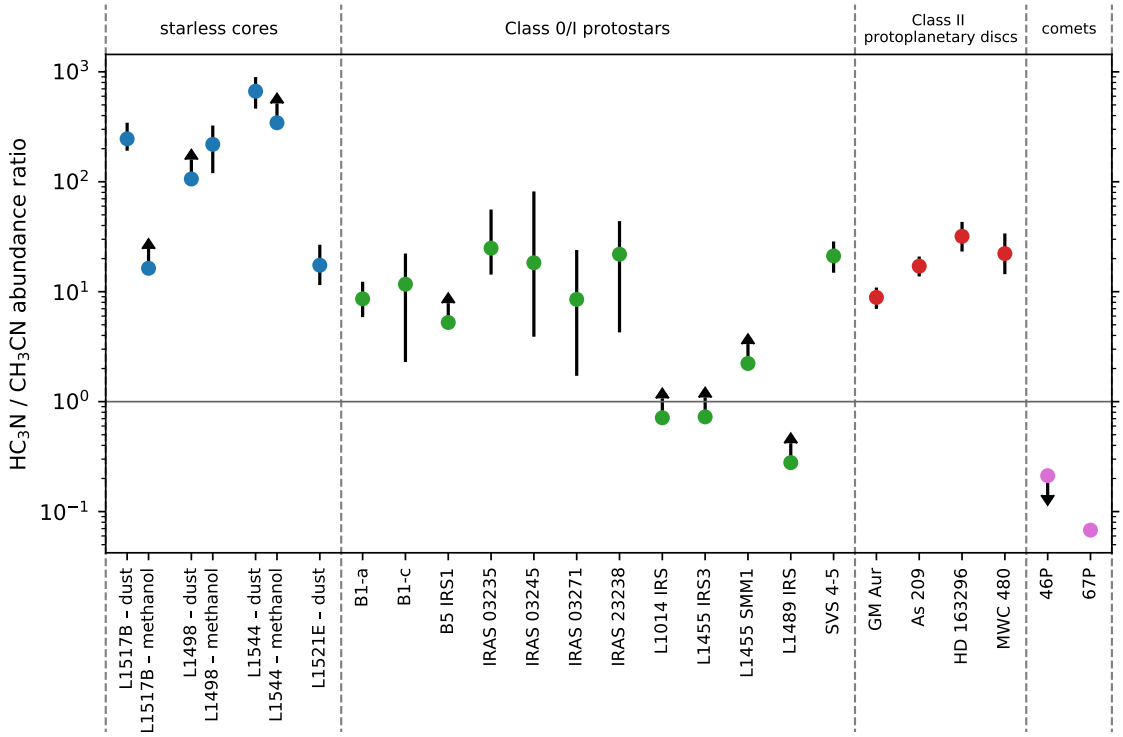


Figure 3.6.1. Abundance ratio between cyanoacetylene (HC_3N) and acetonitrile (CH_3CN) for different sources: the starless cores L1517B, L1498 and L1544 (with both positions, the dust and methanol peaks), the starless core L1521E in the dust peak, 12 Class 0/I protostars, 4 Class II proto-planetary discs, and 2 comets (46P and 67P). The references for each source are: L1517B, this work; L1498, Jiménez-Serra et al. (2021); L1544, Jiménez-Serra et al. (2016); L1521E, Nagy et al. (2019); Class 0/I protostellar systems, Bergner et al. (2017); Class II protoplanetary discs, Ilee et al. (2021); comets, Biver et al. (2021). The points of the plot as well as the original data used to obtain the ratios can be found in Appendix B.3.

We have also modelled the chemical evolution of L1517B and, except for HCCCN and CH_3CN, our simulations agree with the observed abundances and upper limits within a factor of 10. The model also predicts the observed enhancement of the abundances as we move further away from the core's centre, although we overestimate the distance of the methanol peak by a factor of 3. Our model suggests that L1517B is chemically older than L1498, but the absence of infall motions and its relatively poor chemical complexity makes us think that this core is at an earlier evolutionary stage than L1498. Actually, our model would predict a younger age for L1517B than L1498 if the central density of the core were higher.

We propose a scenario in which N-bearing molecules are formed first, followed by O-bearing molecules once the catastrophic depletion of CO takes place. Furthermore, complexity increases with time, with bigger molecules being formed at later stages of the core, as observed in L1544. We note, however, that the starless core L1521E does not seem to follow this trend, although the sample of molecules studied is smaller than for the other three cores, and the beam of the ARO observations is larger. More observational studies of starless and pre-stellar cores are needed in order to clearly determine the influence of the environment, initial conditions and dynamics on the chemical evolution of the cores.

Finally, we have also studied the HC_3N / CH_3CN abundance ratio measured in sources at different stages in the formation of a low-mass stellar system, observing a decreasing trend over time, which could be explained by the adverse conditions of the late stages of the low-mass star formation to form HC_3N.

Chapter 4

Predicting the composition of interstellar ices with machine learning

This section has been adapted from the paper titled *A fast machine learning tool to predict the composition of interstellar ices from infrared absorption spectra*, which will be submitted to the journal Astronomy & Astrophysics (Megías et al., submitted). The complete list of authors is: Andrés Megías Toledano, Izaskun Jiménez Serra, François Dulieu, Julie Vitorino, Belén Maté, David Ciudad, Will R. M. Rocha, Marcos Martínez Jiménez and Jacobo Aguirre.

Abstract

Current observations carried out by the James Webb Space Telescope (JWST) allow to observe the absorption features of icy mantles that cover interstellar dust grains, which are mainly composed of H_2O, CO and CO_2, together with other minor species. Thanks to its sensitivity and spectral resolution, JWST has the potential to observe ice features towards hundreds of sources at different stages along the process of star formation. However, identifying the spectral features of the different species and quantifying the composition of the ice is not trivial, and requires complex spectroscopic analysis. With this work, we aim to reducing the difficulty and the time employed in doing this task. We present a new software analysis tool designed for ice absorption spectra called AICE (*Automatic Ice Composition Estimator*). AICE is based on artificial neural networks, which, given the infrared absorption spectrum of the ice between 2.5 to 10 μm, predicts its fractional composition in terms of H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4. In order to train the model, we have used hundreds of laboratory experiments of ice mixtures from different databases, which have been reprocessed with baseline subtraction and normalization. Once trained, AICE takes less than 1 second to execute and predict the ice composition associated with the observed infrared absorption spectrum, with typical errors of ∼3 % in the derived species fraction. We tested the performance of AICE by using two ice spectra reported towards the highly-extinguished NIR38

and J110621 background stars observed within the JWST Ice Age ERS program, showing a good agreement with previous estimations of the ice composition. The fast and accurate performance of AICE enables the systematic analysis of hundreds of different ice spectra with a modest investment of time, allowing to obtain statistical studies due to its fast execution and simplicity to use. Besides, this model can be enhanced and re-trained with more laboratory data, improving the precision of the predictions and enlarging the list of predicted species.

4.1 Introduction

Interstellar dust particles are known to be constituted by a silicate-rich core, which may also contain carbonaceous material (Draine, 2004; Draine and Li, 2007). With the cold temperatures reached in the interstellar medium (ISM), being as low as $T \simeq 10$ K in molecular clouds and starless cores, volatile compounds present in the gas phase can condensate onto dust grains and react on its surface, forming the so-called icy mantles (Boogert et al., 2015). This process gets accelerated during the prestellar stage, when the core collapses and its central gas densities reach high values $\gtrsim 10^5$–10^6 cm^{-3} (e.g., Caselli, 2002). Apart from the pure depletion itself, which provides carbon monoxide (CO) and atomic species like hydrogen (H), carbon (C), oxygen (O) and nitrogen (N) onto grains, surface chemistry plays an important role, as it drives the formation of other species such as water (H_2O), carbon dioxide (CO_2), methane (CH_4) and methanol (CH_3OH). These species form thanks to the high mobility of H atoms, which hydrogenate simple species like O, C, N or CO yielding H_2O, CH_4, NH_3 or CH_3OH, respectively (Dulieu et al., 2010; Qasim et al., 2020; Redaelli et al., 2023; Watanabe and Kouchi, 2002). Further energetic and thermal processes of the ice may occur along the process of star formation (van Dishoeck et al., 2006; Boogert et al., 2015; Herbst and Garrod, 2022; and references in them), which could induce a complex grain chemical structure containing complex organic molecules (COMs), which can be transferred to planetesimals and cometesimals (Altwegg et al., 2016).

The characterization of the chemical composition of ices is, however, challenging, since it requires observations of ice absorption features in the infrared (IR) domain against bright enough background sources. These signatures are due to the bonds and functional groups present in the molecule (Boogert et al., 2015; Rocha et al., 2022), which cause wide spectral features that are often shared by different species, further complicating the identification of complex solid mixtures due to degenerate solutions.

From ground-based observatories, the main components of the icy mantles have been measured towards different dense molecular clouds and dense cores: water (H_2O), carbon monoxide (CO), carbon dioxide (CO_2), and methanol (CH_3OH) (e.g., Gillett and Forrest, 1973; Whittet et al., 1989; Boogert et al., 2013; Goto et al., 2021). Other minor species with fainter absorption features have also been detected, such as the ^{13}C isotopologue of carbon monoxide (^{13}CO; Boogert et al., 2002), carbonyl sulfide (OCS; Geballe et al., 1985; Palumbo et al., 1997) and the cyanate ion (OCN^-; Lacy et al., 1984; Pontoppidan et al., 2003; van

Broekhuizen et al., 2005). However, achieving successful observations from the ground is not trivial due to the presence of Earth's atmosphere telluric lines and to high noise levels in the infrared.

In this regard, space telescopes such as ISO (Infrared Space Observatory), AKARI and Spitzer enabled higher-quality ice absorption observations (e.g., Schutte et al., 1996; Boogert et al., 2011; Noble et al., 2013), which yielded the detection of solid methane (CH_4), ammonia (NH_3), and the ^{13}C isotopologue of carbon dioxide ($^{13}CO_2$).

Thanks to its great sensitivity and unprecedented spectral resolution, the James Webb Space Telescope (JWST) is providing the most detailed ice absorption spectra ever observed. Yang et al. (2022) provided the first view of the ice composition as seen by the JWST toward a young Class 0 protostar, finding features of H_2O, CO_2, CH_3OH, CH_4 and NH_3 (the CO band at 4.67 μm was not covered). McClure et al. (2023) and Dartois et al. (2024) reported the IR absorption spectra towards the two highly-extinguished background stars NIR38 and J110621 in the Chamaleon I dense molecular cloud, as part of the Early Release Science (ERS) program Ice Age. Besides the major species mentioned above, these authors have also found traces of minor compounds such as OCN^-, OCS and the ^{13}C isotopologes of CO and CO_2 (McClure et al., 2023; Brunken et al., 2024). The dangling bonds of OH have also been identified thanks to JWST IR absorption spectra (Noble et al., 2024), which informs us about the physical structure of the ice.

In all the studies above, the identification of the different species is the result of a detailed and complex spectroscopic analysis, in which, in most cases, the identification has to be carried out band-by-band. In addition, this identification is done by eye and the observed molecular bands are integrated and divided by the experimentally-derived band strengths of pure ices in order to obtain the ice column density (see, e.g., Bouilloud et al., 2015). However, it is well-known that the shape (and even the position and intensity) of the IR molecular bands) are affected by the ice mixture and ice temperature, which adds significant uncertainties to the accurate quantification of the ice chemical composition (Boogert et al., 2015; Rocha et al., 2022).

With the aim of improving and facilitating the determination of the composition of ices, Rocha et al. (2021) presented a machine learning tool called ENIIGMA that uses genetic algorithms to find the combination of experimental ice spectra that best fit the observations. This software has successfully been used for analyzing JWST data in the works of, e.g., McClure et al. (2023) and Rocha et al. (2024), providing a detailed analysis of the possible ice components and the calculation of ice column densities using experimental band strengths. Genetic algorithms, however, are computationally expensive, requiring long computational times to search for all possible ice mixture combinations of laboratory spectra. Given that JWST promises the delivery of ice absorption spectra toward hundreds (if not thousands) background sources (see for instance, the case of the Chamaeleon I cloud in the Ice Age program; McClure et al., 2023), faster and automated alternative approaches need to be developed to efficiently analyse the ice chemical composition of large source samples.

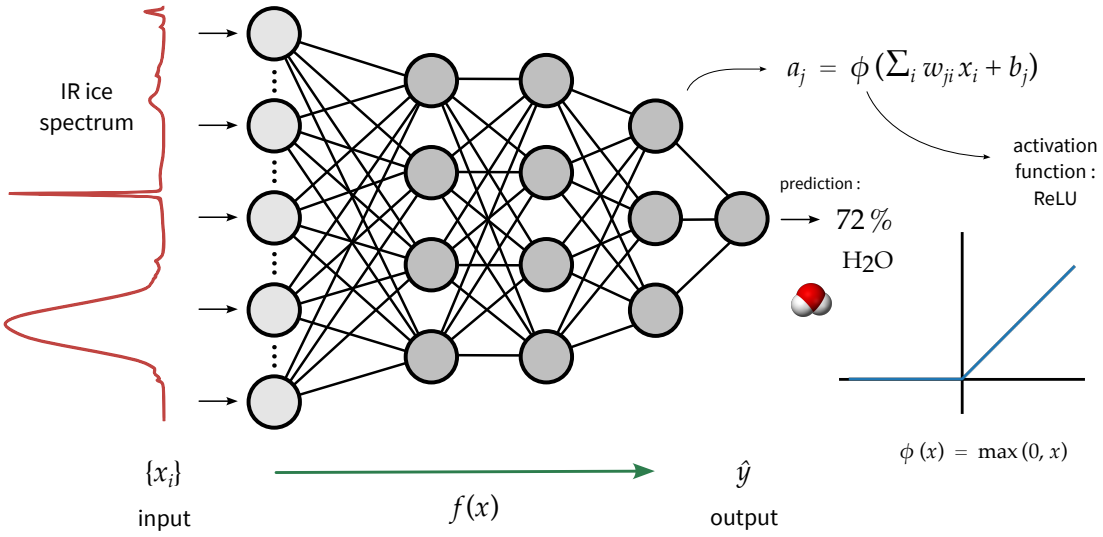


Figure 4.2.1. Scheme of the algorithm behind AICE: artificial neural networks. For each one of the target molecules (H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4), a multilayer perceptron is used to predict the corresponding molecular fraction. The infrared ice spectrum in absorbance is fed to the model, which transforms the spectra points in successive steps or layers; in each one, a linear combination of the previous values (x_i) is computed and then transformed by the activation function (ϕ), obtaining: $a_j = \phi\left(\sum_i w_{ji}\, x_i + b_j\right)$, with w_{ji} and b_j being the weights and biases of the layer j, respectively. Note that the actual neural networks used in AICE are larger than depicted in this scheme (see Section 2.2 and Fig. 4.2.2 for more details).

In this work, we present a new and public machine learning tool, AICE (*Automatic Ice Composition Estimator*), based on artificial neural networks. AICE is a fast and easy-to-use tool programmed in Python language, which has been trained using ice laboratory data. Once trained, AICE allows the analysis of hundreds of ice absorption spectra in matter of minutes (one spectrum is analysed in less than 1 second). Therefore, AICE provides the opportunity to carry out large and complex IR ice absorption spectra in an automatic and systematic way, enabling the study of ices using statistically-significant samples.

The paper is organised as follows. In Section 2 we present in detail our model, describing the architecture and dataset used. In Section 3, we present the results of the training of the model and the results of the test with real JWST data already published. In Section 4, we discuss the different uncertainties that can take place within tool and we compare it with other methods. Finally, in Section 5 we summarize our work and conclude with possible future steps.

4.2 Methodology: the AICE tool

AICE (*Automatic Ice Composition Estimator*) is a Python 3 analysis tool that consists of several artificial neural networks trained on laboratory ice data. AICE is publicly available on GitHub.[1] Given an infrared absorption ice spectra in the range of 2.5–10.2 μm, it predicts the fractional composition of H_2O, CO, CO_2, CH_3OH, NH_3, and CH_4 in the ice. Moreover,

[1] https://github.com/andresmegias/aice/

it also provides information about a representative temperature of the ice although, as we discuss in Section 5, this can be an ill-defined parameter for astrophysical ices.

We note that for the analysis of the IR ice spectrum with AICE, the contribution from silicate absorption should first be removed. To do this, AICE is presented together with a set of several other modules that allow to pre-process the observed astronomical data: (i) merging and rebinning of spectra from different instruments/telescopes, (ii) continuum estimation and conversion to optical depth, and (iii) silicate removal. See Appendix C.1 for more details. In the following sections, we describe how AICE works, providing details about its neural network architecture, its training, and the ice laboratory dataset used for the training.

4.2.1 Artificial neural networks

The mathematical model behind AICE is a set of artificial neural networks of the simplest kind: multi-layer perceptrons. These models are composed of concatenations of linear combinations and non-linear activation functions that transform the input numbers in several steps (or layers), yielding a final output, which can be either a single value or several numbers (Murtagh, 1991; Grosan and Abraham, 2011). Appendix C.2 contains a detailed description of multi-layer perceptrons and how they work.

In a nutshell, an artificial neural network is a mathematical function f with thousands of parameters (commonly known as *weights*) that converts a set of inputs, $\{x_i\}$, into a set of outputs, $\{\hat{y}_i\}$ (LeCun et al., 2015). Thanks to an ingenious optimization method, these parameters can be succesfully tuned to make the neural network mimic a particular operation or task, given enough number of pairs of inputs with their corresponding outputs ($\{x_i\}$, $\{y_i\}$). This process is called *training*, and it is composed of several iterations or *epochs* where we evaluate the difference between the predicted outputs, $\{\hat{y}_i\}$, and the actual outputs, $\{y_i\}$, using a certain error or *loss* function. During the training, thousands of pairs of input/output data are used to make the network learn the desired operation. In this procedure, a small fraction of the training dataset (about 10 %), is typically set aside and used in each epoch to evaluate the performance of the current model on data that has not been 'seen' by the network. After finishing the training, one can apply the trained model to new input data to perform the referred operation or task and obtain a prediction of the output; this process is called *inference*.

4.2.2 Architecture used for AICE

In the case of AICE, the input points, $\{x_i\}$, correspond to the absorbance values of a spectrum given as a function of wavenumber from 4000 to 980 cm^{-1}, in steps of 1 cm^{-1} (corresponding to 3021 points). This input is normalized by its mean value of absorbance, which

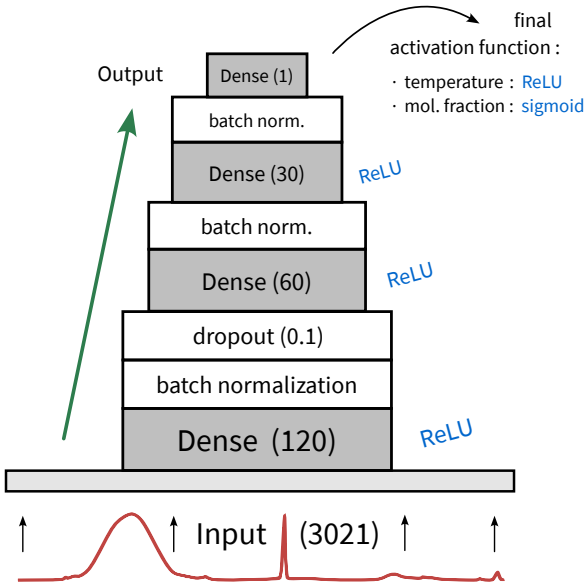


Figure 4.2.2. Diagram of the architecture of the neural networks of AICE: they are multi-layer perceptrons whose layer size decreases from the input spectrum to the output prediction.

allows for a better generalization; this normalization is equivalent to a normalization by area, but with the advantage of being dimensionless.

For the output, we want to obtain predictions for the fractional composition of H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4 in the ice, and of its temperature. Instead of using one single neural network to make these predictions, we opted to use a different neural network for each parameter (i.e., one for each molecular compounds, and an additional one for the ice temperature), since it yielded better results than a single neural network. Therefore, for each neural network we infer a single output value, $\hat{y}$. As the molecular fractional composition has to be a number between 0 and 1, we use a standard logistic sigmoid as the activation function in the output layer, $\phi(x) = 1 \,/\, (1 + e^{-x})$. For the case of the temperature parameter, we just use a ReLU, $\phi(x) = \max(0, x)$, as the temperature (in K) cannot be negative.

Figure 4.2.1 shows a scheme of the general structure of the neural networks used in AICE, while Figure 2 shows the particular architecture employed. We have used 3 hidden layers, which in the case of a multilayer perceptron are called *fully-connected* or *dense* layers. Their size (the number of neurons) starts being 120 for the first hidden layer and decreases progressively to 60 and 30, to end with a final layer that returns a single value after applying the last activation function; in this way, we decrease the number of activations/neurons in each layer from the 3021 points of the input to the single output, which is a common practice when building this kind of models (e.g.: Krizhevsky et al. 2012; Simonyan and Zisserman 2014; He et al. 2016). We have also used batch normalization after each activation function, which in the end yields a total of 372,181 trainable parameters for each neural network (that is, the weights and biases). Dropout was applied only after first hidden layer with a probability of 0.1.

For the loss function, instead of using the mean squared error (MSE), we used the mean

squared logarithmic error (MSLE). If $\hat{y}_k$ are the predictions of the neural network for a certain parameter (fractional molecular composition or temperature) and y_k are the actual corresponding values, through a dataset of m values, the MSLE is defined as:

$$\mathcal{L} = \frac{1}{m}\sum_{i=1}^{m}\left(\ln\left(\hat{y}_k+1\right)-\ln\left(y_k+1\right)\right)^2 \ . \tag{4.1}$$

The reason of using the MSLE instead of the MSE is to improve the accuracy of the predictions when the concentration of a molecule is very low or even zero. Adding +1 in the logarithm avoids having infinite values when either the real or the predicted fraction of a molecule is exactly zero. In the case of temperature, we keep this loss function, as we also want to prioritize the accuracy of the model at low temperatures; e.g., an error of 1 K when the real temperature is 10 K is more important than when the real temperature is 70 K.

We have used the Adam method of gradient descent for training our model (Kingma and Ba, 2014), with an initial learning rate of $\alpha = 0.08$ and exponential decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is multiplied by 2/3 if there is no improvement greater than 10^{-4} in the validation loss after 15 epochs, and the training is stopped either after 160 epochs, or when the learning rate reaches a value of $\alpha = 0.01$ or when there is no improvement in the validation loss after 35 epochs, restoring the weights that yielded the lowest loss. To improve and fasten the training, we randomly split our training subset in batches of 64 elements, and we apply batch normalization after the activation of each hidden layer. See Appendix C.2 to know more details about the mathematics behind our networks. The mentioned models were coded and trained in Python[2] programming language and using the library Keras,[3] which is based on TensorFlow.[4]

4.2.3 Training dataset

The predictions given by our tool AICE rely on the quality of the training dataset used to optimize the models. We have used a total of 564 laboratory infrared spectra with varying chemical composition of the ice and temperature, both of which are known for each spectra. These data are the result of 219 different experiments, from which the majority of them correspond to different ice mixtures. The main ice components include H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4 (our target species). However, note that we have also included spectra from other mixtures with, e.g., formic acid (HCOOH) or acetaldehyde (CH_3CHO), as we want our model to identify those cases in which our main species (H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4) are not present.

[2] https://www.python.org/
[3] https://keras.io/
[4] https://www.tensorflow.org/

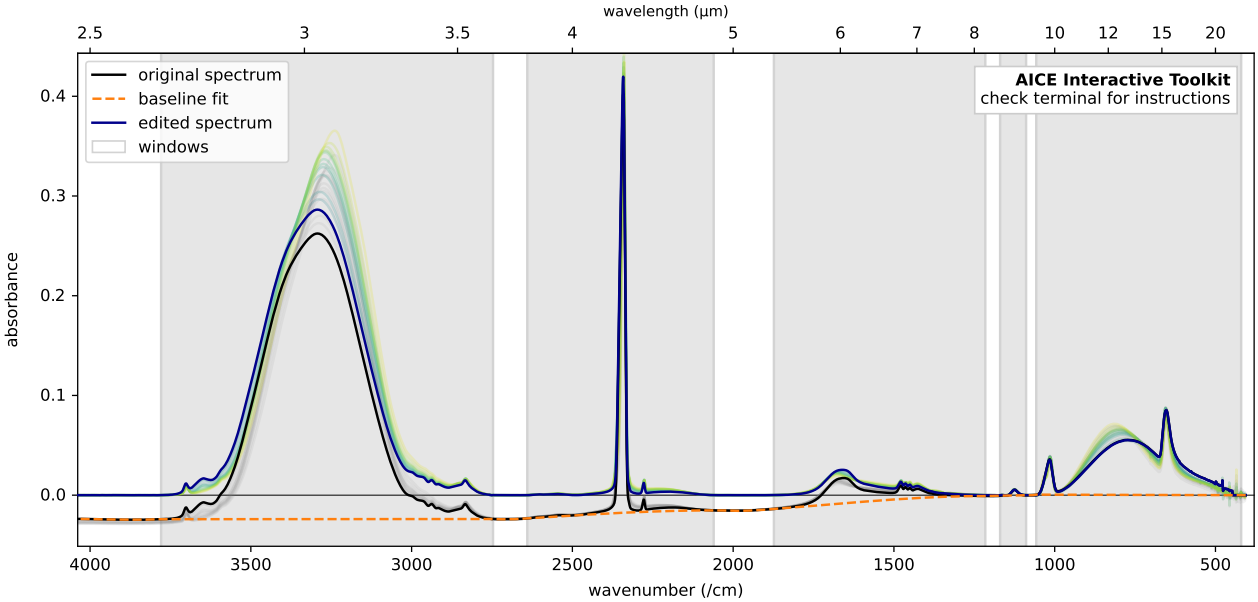


Figure 4.2.3. Example of the baseline reduction for the experiment #35 from LIDA, performed with an interactive Python interface within AICE. The plot shows the spectra corresponding to this experiment ($H_2O:CO_2:CH_3OH$, 9:2:1), for several different temperatures. In black and blue we show the original and reduced spectrum from 10 K, while in light gray and light green/yellow we indicate the spectra for the rest of temperatures measured in the experiment up to 175 K. We selected manually the windows that define the points of the continuum subtraction (white regions) and then the tool performed a baseline fit based on splines.

4.2.3.1 Infrared molecular spectra

From the total of 564 laboratory spectra (from 219 experiments) used for the training, the majority of them (445 spectra, from 132 experiments) come from the Leiden Ice Database for Astrochemistry (LIDA), and are publicly available to visualize and download.[5] This database, presented in detail by Rocha et al. (2022), includes nowadays more than 1100 different spectra of more than 170 different ices, meaning that infrared measurements for multiple temperatures are available. In the majority of the cases, the ice is formed at a low temperature (10–15 K), and the ice is then warmed up until it completely desorbs from the surface used. The majority of the spectra are recorded in transmission mode; only a few are in reflection-absorption (RAIRS), with an angle of reflection of 45°.[6] The infrared range is approximately 4000–400 cm^{-1}, and the resolution varies from 0.25 to 2 cm^{-1}.

Note that we did not use all the spectra available in LIDA, since we carried out a pre-selection and discarded some of them for different reasons:

1. We did not use the data from mixtures containing apolar molecules (such as O_2 or N_2), since they do not present significant infrared features but only produce slight changes in the shape of the absorption features of other species contained in the mixture.

[5] `https://icedb.strw.leidenuniv.nl/`

[6] This could lead to some inconsistencies for the same spectra due to differences in the light polarization of the two methods. Despite this, given that we normalize the spectra, the angle of reflection is 45°(larger differences would appear if light was almost paralell to the surface), and we have a rather small dataset, we decided to still use those experiments.

2. We did not use any processed ice (with UV radiation, for example) or any ice where surface chemistry was involved, since in those cases we do not know the exact final chemical composition.
3. Regarding the temperature, we also discarded any spectra in which any of the chemical compounds was clearly desorbed (i.e., their infrared bands had disappeared with increasing temperature). This was the case of some ternary and quaternary mixtures for which it was evident that the most volatile species starts to desorb at relatively low temperatures and hence, the labeling of the composition is no longer correct.
4. We discarded any spectra with a temperature greater than 100 K. One reason is to reduce the variability in the range of the parameters of the neural network to facilitate the training of the model, as the desorption temperature of the targeted species varies quite a lot. In any case, the only targeted molecules that desorb at higher temperatures than 100 K are water and methanol; for molecular clouds, starless/prestellar cores or young protostellar systems, we expect to have temperatures lower than 100 K.
5. We ruled out spectra from the same experiment with a difference in temperature of less than 8 K, as we want to ensure a significant difference in the shape of the measured spectra. However, we kept the spectrum with the highest recorded temperature for each experiment (not greater than 100 K), to ensure that all extreme cases were covered by the neural networks, within our desired temperature range.

The final number of spectra considered for our training dataset was thus 445, that is, almost the half of the data stored in LIDA.

On top of the LIDA spectra, we have also used a few spectra from other different databases and laboratories to complete our training dataset: 58 spectra from 21 different experiments from NASA's Optical Constants Database (OCdb)[7]; 34 spectra from 32 experiments from NASA's Cosmic Ice Laboratory (CIL)[8]; 30 spectra from 30 experiments from the database of the Universidade do Vale do Paraíba (Univap)[9]; 5 spectra from 5 different experiments from the Experimental Astrophysics Laboratory (Laboratorio di Astrofisica Sperimentale, LASp) in the Astrophysical Observatory of Catania (INAF)[10]; 3 spectra from the same single experiment from the National Synchrotron Radiation Research Center (NSRRC)[11] in Taiwan; and 2 spectra from the same experiment from the Institute of the Structure of Matter (IEM, CSIC) in Spain.[12] All of them have similar experimental properties to the ones from LIDA. Table C.4.1 in Appendix C.4 provides a complete list of all spectra collected for this work.

[7] https://ocdb.smce.nasa.gov/

[8] https://science.gsfc.nasa.gov/691/cosmicice/spectra.html

[9] https://www1.univap.br/gaa/nkabs-database/data.htm. The subsample of spectra that we have used come from two different laboratories: GANIL (Grand Accelerateur National d'Ions Lourds, in France) and LASA (Laboratório de Astroqumica e Astrobiologia da Univap, in Brazil).

[10] https://www.oact.inaf.it/2021/01/24/il-laboratorio-di-astrofisica-sperimentale/.

[11] https://www.nsrrc.org.tw/english/index.aspx

[12] https://www.iem.csic.es/fismol/ices/index_en.shtml

4.2.3.2 Spectra processing and baseline removal

All of these 564 infrared spectra had to be processed and reduced before using them to train our neural networks.

First of all, we checked whether the spectra had been reduced or whether they had baseline issues that could distort the spectrum even in regions with only continuum. In the cases for which baseline issues were evident, we subtracted these baselines from our data by using an interactive Python tool developed by us and included within AICE. This tool allows the user to manually select the continuum regions of the spectra that will be used to fit the baseline, and to subtract it. The baseline fit is done with splines, i.e., a concatenation of polynomials of third order, where the smoothness of the resulting fit can be adjusted. Fig. 4.2.3 shows an example of this procedure. We refer to Appendix C.3 for more details on the method.

Another point that we had to take into account was the presence of absorption bands from residual compounds that were not listed in the composition of the experiment, e.g., solid water present on the IR detector, or gaseous CO_2 or even H_2O present in the chamber of the sample. In addition, curve deformations that are not linked to any absorption signatures can come from experimental artifacts and induce a bias in the recognition of features by AICE. We removed these features by interpolating the spectra with polynomials (splines). Similarly, there can be signals of molecular isotopologes due to the impurity of the experimental samples, like the ^{13}C variants of CO and CO_2; this can indeed be seen in Fig. 3 for $^{13}CO_2$, with the small peak next to the CO_2 absorption line at $\sim$2340 cm^{-1}.

Lastly, there were some regions of some spectra that were too noisy, so we smoothed those parts to obtain spectra with an uniform and low level of noise. In any case, we did not smooth any thin feature or line, as that would deform the shape of the spectrum. An example of the baseline subtraction, removal of artifacts and smoothing is shown in Appendix C.3, along with more detailed explanations.

4.2.3.3 Resampling and normalization

As mentioned before, AICE works with a range of wavenumber values from 4000 to 980 cm^{-1} with a separation of 1 cm^{-1} (3021 points). As not all the spectra have this same spacing, we resampled all spectra to a resolution of 1 cm^{-1}.

After resampling, we normalized the spectra simply dividing by the mean of the absorbance values within the AICE range. This is equivalent to normalizing by the area, as the integrated area of a set of points is proportional to its mean, given a fixed range in wavenumber. In this way, we generalize our data to any combination proportional to the original ice mixture.

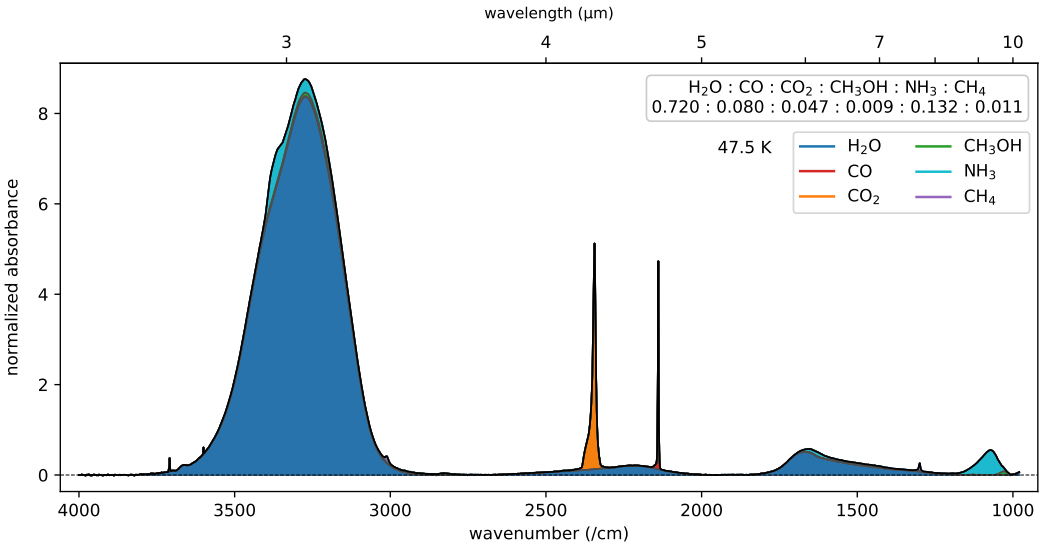


Figure 4.2.4. Example of synthetic generated spectrum (#15), normalized by its mean absorbance value. The colors represent the different molecular contributions. This spectrum represents a layered mixture of H_2O, NH_3, CO, CO_2, CH_4 and CH_3OH, in decreasing order molecular fraction, all of them at a temperature of 17.5 K. The pure ice spectra used to construct this curve come from different experiments from LIDA (see Table C.6.1 in Appendix C.6), at different temperatures close to 17.5 K. The exact curves for each pure ice are the result of interpolating two experimental spectra taken at temperatures lower and higher than the final 17.5 K (e.g., 15 K and 20 K), as explained in Section 4.2.3.4. Note that the main CH_3OH band at 1025 cm^{-1} is overshadowed by the NH_3 band.

4.2.3.4 Data augmentation using pure ice spectra

Out of our 6 targeted species, ammonia and methane present very few experiments in our dataset, which makes the training of our networks a challenging task. In order to tackle this problem, we opted for enlarging our dataset by generating synthetic mixtures using linear combinations of experimental pure ice spectra. To do this, we employed pure ice spectra from LIDA for our 6 target species. For each pure spectrum, we integrated one of the strongest bands, and used the experimental band strengths to derive the column density of the ice. By dividing the absorbance by the column density, we obtain the absorption cross section of each pure spectrum. The cross sections of different species can then be summed up after weighing each contribution by the column density of each molecule. However, as we use a normalization by the mean, we do not need to choose any column density for each species; we just use the relative proportions of the desired ice mixture, and afterwards we divide the final spectrum by its mean (see Fig. 4.2.4 for an example of synthetic spectrum). This process can be repeated for the different temperatures measured for each pure species, obtaining a set of normalized cross sections for each temperature.

In reality, the spectra of ice mixtures are not just a linear combination of the pure ice spectra as there are interactions between molecules that deform the shape, the position and even the intensity of the bands. However, the absorption spectra of a combination of separated layers of different species is actually the linear combination of the pure ice spectra (apart from a tiny contribution of the interfaces, which is insignificant in our dataset, with hundreds of monolayers per ice). Therefore, these synthetic combinations from pure ices can

be seen both as a first approximation to mixed ice spectra and as the spectra corresponding to combinations of ice thick layers with different molecules. Actually, there are some experiments in LIDA that correspond to two superimposed layers of pure ices at the same temperature.

By using the method described above, we generated a total of 282 synthetic spectra containing different ice mixtures of our 6 targeted species. We chose the values of each molecular proportion within the mixture using a random algorithm that ensures that all of the species are uniformly represented (see Appendix C.6 for more details). Regarding the temperatures, care needs to be taken since species such as CO and CH_4 desorb at temperatures $\gtrsim$25 K. In addition, our pure NH_3 data only has measured laboratory spectra up to $\sim$60 K. Therefore, the temperature range for data augmentation needs to reflect these limitations.

We have thus employed the following procedure. For each synthetic mixture, when all species present a concentration higher than 2 %, we restrict the allowed temperature range to the common range of recorded temperatures for the laboratory ice spectra (i.e., between 10 K and 28 K). However, if in the synthetic mixture one of the compounds presents a concentration lower than 2 %, we allow a broader temperature range between 10 K and the minimum value of the highest temperatures recorded for the ice spectra from species with concentrations $\geq$2 %. This helps the neural networks to learn the behavior of high temperature ices for major molecular compounds. In any case, note that species such as CO and CH_4, which desorb in large amounts at temperatures $\gtrsim$30 K, could remain in the ice in small quantities at higher temperatures when mixed (Simon et al., 2019; Kruczkiewicz et al., 2024).

Once the temperature range is defined, we select randomly one temperature within that range. We then perform a linear interpolation of the pure ice spectra of each molecule in the dimension of temperature, to obtain the spectrum of each pure ice at the selected temperature. The final synthetic spectrum is built through the linear combination of the interpolated pure ice spectra using the proportions previously determined for this particular mixture.

Therefore, we have a total dataset of 856 spectra (574+282), which is used to train our neural networks. We stress that a larger number of synthetic spectra in our dataset would spoil the performance of the networks since the synthetic spectra do not account for the molecular interactions in the ice mixture that get reflected in the real molecular band profiles. With our procedure, the number of synthetic spectra represents 1/3 of the total dataset, providing an optimum performance of the networks.

4.2.3.5 Training and validation subsets

In order to train our neural networks, we had to split our final dataset of 574+282 spectra into a training and a validation subset.

- First, we use all of the pure ice spectra for training, as we want to ensure that our models can recognize the simplest case of a pure mixture. This applies to all the pure ices, including the species not targeted by our model.

- Then, we use for training the spectra containing molecules different from the target species, even if the concentration is small. By doing this, we want the networks to learn when the target species are not present in the ice.
- For the rest of the laboratory spectra and the 282 synthetic spectra, we randomly shuffle them and split a portion of them so that the validation dataset constitutes the 20 % of the whole dataset.

Following this procedure, we end up with a total of 685 spectra for training and 171 for validation. However, this number is rather low to properly train an artificial neural network. Furthermore, the data that have been selected to be in the validation subset will not be used to fit the parameters of our models, so we will lose their information for the training of the networks.

To prevent this, we use the following approach regarding the random shuffle of our training dataset. We make 10 different shuffles of our data with fixed random seeds,[13] so that we end up with 10 different pairs of training and validation subsets. Then, we train our neural networks for each of them, obtaining different values for their parameters at the end of the training. Finally, we use all of the 10 variants of our models to obtain a prediction for the input spectra, being the result the mean of the output results. In this way, for each of our target variables we end up with an *ensemble* of neural networks that not only accounts for the different shuffles (or splits) of our training dataset, but also explores the different initializations of the parameters of the neural networks, since they are randomly set at the beginning of the training.

To summarize, we train 7 sets of 10 neural networks, one set for each targeted variable (temperature and molecular fraction for our 6 target species). Each set of models contains 10 variants that are trained with the different shuffles of our whole dataset of 574+282 spectra. The final AICE result is the mean of the output results of the 10 neural networks, for each of the 7 target variables. Additionally, for each prediction, we compute the standard deviation between the 10 model variants, which will be considered as the uncertainty in the AICE prediction.

4.3 Choosing the model hyperparameters

The parameters that define the architecture of the neural network, such as the number of layers and number of neurons per layer, as well as the choice of the loss function (MSE, MSLE...) and training method (the *optimizer*, like Adam) constitute the hyperparameters of the neural network. More examples of them are the batch size and the learning rate.

The choice of the exact values of these parameters starts with an educated guess. In our case, we performed several tests varying some of the values and checking the performance

[13] The random seed is an integer number that defines the sequence of random numbers generated by the computer.

of the models after the training procedure, using the validation dataset for it. This has to be done with several random seeds (thus, different splits of training/validation), to reduce the possibility of biasing our hyperparameters to a specific seed.

For example, we first tried with an additional hidden layer at the beginning with 480 neurons, but after some tests we realized that we could dispense with it, obtaining virtually the same performance with the benefit of reducing the model's size by a factor of 4. We also tried using 1-dimensional convolutional neural networks, a type of models which is commonly used for pattern recognition (LeCun et al., 2015), but we could not reach the performance obtained with our multi-later perceptron, hence we prefered using this simple yet powerful kind of model.

In a similar way, we also tried using a unique neural network with several final outputs, one for each of the predicted variables (6 molecular fractions + ice temperature). However, we could not reach the same performance as the one achieved using individual neural networks, probably due to the inhomogeneity of the training dataset with respect to our parameter space (temperature and molecular fractions). Focusing only on predicting one variable in each neural network reduces the simplicity of the optimization task, yielding a better individual accuracy for each model. Besides, this modular approach should facilitate the extension of our model with additional targeted species in the future, since we would only need to train those additional networks, without retraining the models presented in this work.

Lastly, similar tests were performed when choosing the number of synthetic spectra to be included in the augmented dataset and the minimum separation in temperature in the original laboratory dataset.

4.4 Results

In this Section, we first present the results of the training our model and report the performance of the validation of the networks. We then test AICE on two ice absorption spectra recently obtained with the JWST towards the background stars NIR38 and J110621 in the Chamaeleon I molecular cloud (reported by McClure et al., 2023).

4.4.1 Training and validation

Training of each neural networks takes approximately 17 s, so training each set of 7 neural networks takes about 120 s.[14] Repeating this process 10 times, for 10 different randoms seeds (from 1 to 10), we complete the training of our model AICE in around 20 min.

[14] Processor: Intel i5, 4 cores, 2 GHz; RAM: 16 GB, DDR4, 3.7 GHz.

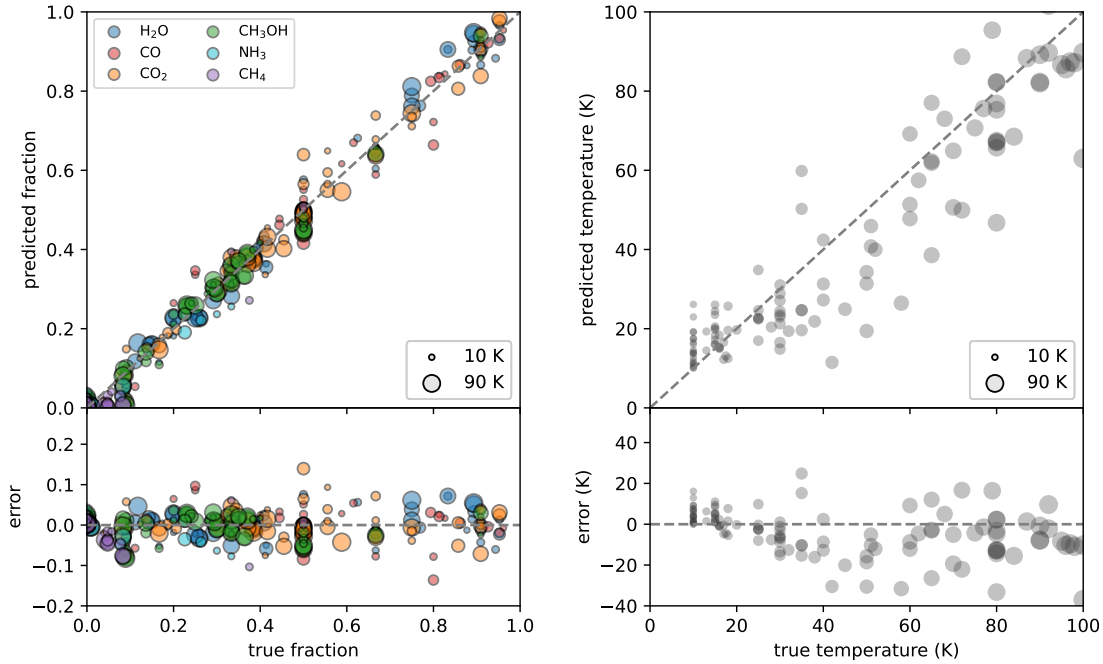


Figure 4.4.1. Performance of our neural networks after the training with a random seed equal to 1, measured on the real laboratory data included in the validation subset (that is, excluding the synthetic spectra). The plots show the predicted versus the true values for the molecular fractions (*left*) and temperature (*right*) for each experiment. Each experiment is represented by 6 points in the left plot (one for each targeted molecule, color coded), and by 1 point in the right plot. In the left plot, temperature is represented by the size of the points.

Figure C.5.1 in Appendix C.5 shows the evolution of the training and validation losses during the training for a certain seed, while Fig. 4.4.1 shows the results of the validation for the same seed. In this plot we excluded the synthetic spectra in order to show the performance obtained for the laboratory data. As it can be seen from Figure 4.4.1, the resulting error for the molecular fraction for the different targeted species typically lies below 10 %, while it is usually below 20 K for temperatures. To be more precise, the root mean squared error (RMSE) can be used to estimate the accuracy of the neural networks in their predictions. The list of root mean squared errors in the validation subset for the different individual neural networks and random seeds (and therefore splits of the dataset in training/validation) is shown in Table 4.4.1. The RMSE in the fractional composition, after averaging the 6 species and 10 seeds/splits, is 2.2 %, while for the temperature we have an average RMSE of 11.5 K.

We also need to bear in mind that one potential source of error comes from uncertainties in the labeling of the laboratory spectra. While the temperature is a very well determined parameter, the determination of the exact composition is not trivial. Sometimes, pre-made gaseous mixtures can be used in the experimental chamber, although the different molecular velocities can lead to a slightly different ice composition. Other times, the IR molecular bands can be integrated and scaled by band strengths from pure ices, but this can led to errors up to 20 %. A more precise method involves knowing the input pressures of each gas in the chamber and carefully calibrating the fluxes and pressures of each species, which is not trivial. Therefore, this uncertainty can get transferred to the training of our models,

Table 4.4.1. Root mean squared errors (RMSE) obtained in the validation of the neural networks after training, for each targeted variable and each random split of the dataset.

	root mean squared error (RMSE) in validation						
split	**fractional composition** (%)						**temp.**
#	H_2O	CO	CO_2	CH_3OH	NH_3	CH_4	(K)
1	2.5	2.7	3.0	2.4	1.2	1.8	11.8
2	2.5	2.6	2.2	2.0	1.5	2.0	10.4
3	2.0	2.8	2.2	2.1	1.6	1.6	11.4
4	2.2	2.8	3.4	1.8	1.8	3.6	9.7
5	2.2	2.7	2.0	2.3	1.4	1.3	12.9
6	2.0	2.6	2.3	2.4	1.5	1.7	10.7
7	2.4	2.4	2.0	2.4	1.6	1.6	11.7
8	2.5	2.2	2.4	2.1	1.5	2.5	11.4
9	2.2	3.1	2.3	2.0	1.6	3.7	11.7
10	2.4	2.7	3.0	2.1	1.3	2.4	13.0

which translates into a dispersion in the predictions. As a clear example, we highlight the cluster of circles at a molecular fraction of 0.5 (50 %) in the left panel of Figure 4.4.1, since many spectra are binary mixtures with ratios 1:1. To explain the large dispersion for some of the molecular fractions, we suspect that either their labeled composition is uncertain, or the neural networks are just making a bigger mistake than usual.[15] Regarding the temperature, it is indeed a delicate variable, since in the experiments used to train AICE, the lower temperature corresponds to the temperature at which the ice was initially deposited, while the following temperatures in increasing order correspond to the annealing of the cold deposited ice. However, the effects of this annealing in the shape, position and intensity of the molecular bands, can be also obtained experimentally by other means, such as by the type of deposition, as discussed in Section 4.5.

As mentioned in the previous section, the final result of AICE will be, for each targeted variable (molecular fractions and temperature), the mean value of the predictions of the 10 variants corresponding to each random seed. Similarly, the uncertainty in the prediction will be the standard deviation of the predictions of those 10 variants, for each targeted variable.

4.4.2 Testing with astronomical ice

Once we have trained and validated our model, we can test how it performs with astronomical data. To do this, we use the JWST spectra reported by McClure et al. (2023) obtained within the IceAge ERS project. It consists of infrared ice absorption spectra measured towardss two background stars in the Chamaeleon I molecular cloud: NIR38 and J110621.[16]

Figure 4.4.2 shows the resulting spectrum for NIR38 in the left panel (black curve) and

[15] During some of the initial trainings of our models, we identified prominent outliers, which turned out to be mislabeled spectra; after correcting the labeling, those points were no longer outliers, showing typical errors $\lesssim$3%. This demonstrates the utility of these models to properly identify the ice composition.

[16] The spectra are available in Zenodo: `https://doi.org/10.5281/zenodo.7501239`

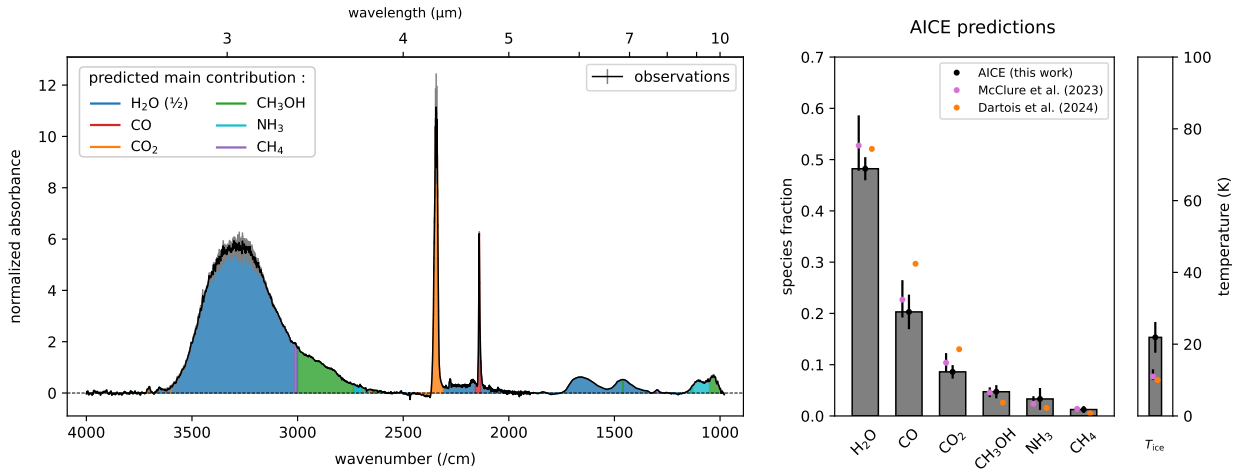


Figure 4.4.2. *Left:* Absorbance spectrum derived by JWST measurements towards the background star NIR38 in the Cha I molecular cloud (in black), and predictions of the molecular contributions in different spectral regions performed by AICE (in colors). *Right:* Direct numerical predictions given by AICE of the fractional molecular composition and temperature of the ices towards NIR38 (grey bars and black points), compared to estimations by other authors (in pink, McClure et al., 2023; in orange, Dartois et al., 2024).

Table 4.4.2. Comparison between the AICE predictions with the results of McClure et al. (2023) and Dartois et al. (2024) for NIR38 and J110621.

prediction	**NIR38**			**J110621**	
	AICE (this work)	**McClure et al. (2023)**	**Dartois et al. (2024)**	**McClure et al. (2023)**	**AICE (this work)**
H_2O (%)	48.2 ± 2.3	53^{+6}_{-5}	52	63^{+3}_{-4}	51 ± 3
CO (%)	20 ± 3	23^{+4}_{-3}	30	$17.5^{+2.7}_{-2.0}$	23 ± 3
CO_2 (%)	8.6 ± 1.3	$10.4^{+1.8}_{-1.8}$	13	$8.3^{+1.2}_{-1.1}$	6.3 ± 1.3
CH_3OH (%)	4.7 ± 1.3	$4.6^{+1.0}_{-0.9}$	2.6	$2.4^{+0.9}_{-0.5}$	3.3 ± 1.4
NH_3 (%)	3.3 ± 2.1	$2.4^{+1.5}_{-0.4}$	1.6	$3.2^{+0.7}_{-0.4}$	2.9 ± 2.0
CH_4 (%)	1.3 ± 0.7	$1.38^{+0.20}_{-0.18}$	0.52	$1.17^{+0.13}_{-0.13}$	1.5 ± 0.7
temperature (K)	22 ± 4	11^{+2}_{-1}	10	11^{+2}_{-1}	20 ± 5
CO / H_2O (%)	42 ± 7	43^{+12}_{-10}	57	28^{+6}_{-4}	45 ± 6
CO_2 / H_2O (%)	18 ± 3	20^{+5}_{-5}	25	$13.2^{+2.6}_{-2.1}$	12 ± 3
CH_3OH / H_2O (%)	10 ± 3	$8.7^{+2.5}_{-2.2}$	5	$3.9^{+1.5}_{-0.8}$	7 ± 3
NH_3 / H_2O (%)	7 ± 4	$4.6^{+2.8}_{-1.1}$	3	$5.1^{+1.3}_{-0.7}$	6 ± 4
CH_4 / H_2O (%)	2.6 ± 1.4	$2.6^{+0.6}_{-0.5}$	1	$1.84^{+0.30}_{-0.24}$	2.8 ± 0.14

Note. Individual molecular fractions are given with respect to the total quantity of all the predicted species; in the case of the two other works, we calculated these fractions with the data available in the corresponding papers. For the temperature, McClure et al. (2023) do not present a unique temperature, but several possible ones ranging from 10 to 16 K; consequently, we adopted the value of 11^{+2}_{-1} K as an estimate of a representative temperature. The estimates by Dartois et al. (2024) are the same for both background stars (i.e., the one of the laboratory spectrum used in their radiative transfer calculations), and thus these values are placed at the center of the table. The fractions with respect to water are derived from the direct results of AICE and those reported by the cited authors, using the Python library RichValues (see Appendix C.7 for more details).

the predictions of AICE for the ice composition in the right panel (grey bars). Using the predicted composition in terms of molecular fractions, AICE is capable of estimating the major contribution for each part of the spectrum. To do so, it uses the molecular pure ice spectra reported in Table C.6.1, from LIDA, to construct a linear combination using the molecular fractions. Then, for each point of the spectrum, it computes the major molecular contribution and assigns a color; an exception is H_2O, which was scaled by a factor of $\frac{1}{2}$ just for visualization purposes, to allow to distinguish weaker features such as the CH_4 molecular band at $\sim$3010 cm^{-1}.

In Figure 4.4.2 (right panel) we also compare the AICE results with the ice fractional composition derived by McClure et al. (2023) using ENIIGMA (Rocha et al., 2021), and Dartois et al. (2024) using radiative transfer calculations.

As shown in Figs. 4.4.2 and 4.4.3, AICE tends to overestimate the temperature of the ice. However, the ice temperature is a delicate variable. Our model was trained with ices at a single temperature, while in molecular clouds we can have a distribution of dust grain temperatures along the line of sight. Furthermore, changes produced by temperature in the shape of the spectra are generally subtle, or even insignificant in some cases such as CO.

One last test consists of adding the AICE fractional compositions to check whether the total contribution lies close to 100 %. The result is 87 ± 5 % for NIR38 and 88 ± 8 % for J110621. The total composition is fairly low, which could be due to an underestimation of CO ice by AICE. Alternatively, the low total composition could also indicate that either the observed spectrum includes additional species different from the 6 molecules for which AICE has been trained, or some of the main targeted species (e.g., H_2O) are slightly underestimated.

As mentioned in Section 4.2.3.5, the uncertainties in the AICE results are estimated as the standard deviation of the predictions of 10 different sets of neural networks (corresponding to different random seeds). We also need to take into account the uncertainties associated with the observed JWST spectra and hence, we propagate the observational uncertainties of the spectrum into the predictions of our model using the Python library RichValues (see Appendix C.7 for more details). Then, we add this contribution quadratically to the standard deviation in the predictions for each variable. However, the observational contribution is negligible for NIR38 and very low for J110621, representing only $\lesssim$10 % of the standard deviations of the neural network models.

4.4.3 Comparison of the AICE predicted column densities

To obtain the predicted AICE column densities, we derive the column density of one of our targeted molecules and scale this value by using the fractional compositions given by AICE. The column density of one molecular species is calculated by integrating one of the molecule's infrared bands in optical depth (or absorbance, in which case we introduce a factor of ln 10) and dividing this value by the corresponding band strength. In our case, we choose the band strength of the main band of CO at 4.67 μm of $1.1 \cdot 10^{-16}$ cm at 14 K

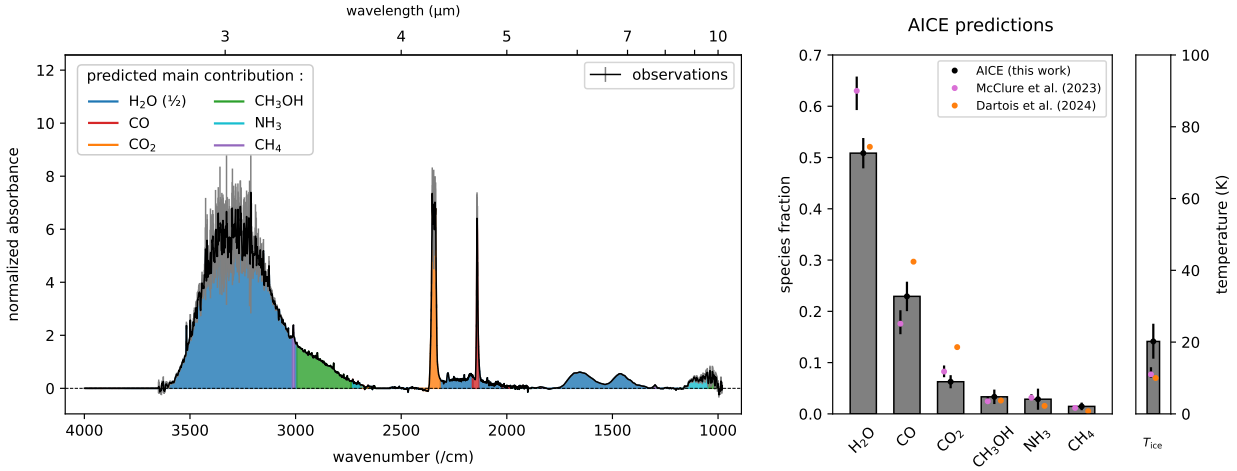


Figure 4.4.3. Same as Fig. 4.4.2 but for the background star J110621: infrared spectra obtained after the reduction and predictions given by AICE.

Table 4.4.3. Comparison of the ice column densities predicted by AICE and ENIIGMA (McClure et al., 2023) for the background stars NIR38 and J110621.

molecule	column density (10^{18} cm^{-2})			
	NIR38		J110621	
	AICE	ENIIGMA	AICE	ENIIGMA
H_2O	$5.5^{+1.0}_{-0.8}$	$6.9^{+1.9}_{-1.1}$	$8.1^{+1.3}_{-1.0}$	$13.4^{+1.3}_{-1.9}$
CO	$2.3^{+0.3}_{-0.3}$	$3.0^{+0.6}_{-0.4}$	$3.7^{+0.5}_{-0.5}$	$3.7^{+0.6}_{-0.4}$
CO_2	$0.98^{+0.25}_{-0.19}$	$1.38^{+0.20}_{-0.20}$	$1.0^{+0.3}_{-0.3}$	$1.74^{+0.21}_{-0.22}$
CH_3OH	$0.54^{+0.18}_{-0.15}$	$0.61^{+0.11}_{-0.11}$	$0.52^{+0.25}_{-0.23}$	$0.51^{+0.19}_{-0.09}$
NH_3	$0.38^{+0.26}_{-0.24}$	$0.30^{+0.22}_{-0.03}$	$0.5^{+0.3}_{-0.3}$	$0.66^{+0.15}_{-0.06}$
CH_4	$0.15^{+0.08}_{-0.08}$	$0.18^{+0.02}_{-0.01}$	$0.24^{+0.12}_{-0.11}$	$0.25^{+0.01}_{-0.03}$

Note. We used the column density of solid CO to scale the fractional compositions predicted by AICE to derive the rest of molecular ice column densities.

We finally stress that, as explained by Rocha et al. (2022), this method can involve uncertainties **up to** ∼20 %. This is due to the fact that molecular band strengths of ice mixtures can be different from the corresponding ones of pure ices, and because they can also change with temperature. Indeed, differences of this order can be seen in the different band strengths reported by Bouilloud et al. (2015) for the same molecular species.

4.4.4 Effects of band saturation

As mentioned in Section 4.4.3, some of the molecular bands may suffer from saturation effects, yielding a flat shape at the bottom of the infrared bands in the observed spectrum. This is particularly critical for J110621 because it presents the highest extinction value (A_V = 95 mag) of the two stars studied by McClure et al. (2023). Therefore, in this section we

Table 4.4.4. Comparison between the predictions of the standard AICE model and the version trained with saturation effects (AICE-s) towardss NIR38 and J110621.

	NIR38		**J110621**	
prediction	**AICE**	**AICE-s**	**AICE**	**AICE-s**
H_2O (%)	48.2 ± 2.3	47 ± 3	51 ± 3	48 ± 3
CO (%)	20 ± 3	20.4 ± 2.4	23 ± 3	24 ± 3
CO_2 (%)	8.6 ± 1.3	9.0 ± 0.9	6.3 ± 1.3	6.6 ± 1.0
CH_3OH (%)	4.7 ± 1.3	3.6 ± 2.1	3.3 ± 1.4	2.6 ± 1.8
NH_3 (%)	3.3 ± 2.1	2.6 ± 0.7	2.9 ± 2.0	2.0 ± 0.7
CH_4 (%)	1.3 ± 0.7	2.0 ± 1.4	1.5 ± 0.7	2.1 ± 1.5
temp. (K)	22 ± 4	26 ± 3	20 ± 5	27 ± 6

evaluate whether our AICE results are affected by band saturation.

A first look at the fractional ice compositions and ice column densities predicted by AICE (see Tables 4.4.2 and 4.4.3), reveals that our predictions for H_2O, CO and CO_2 are consistent with previous calculations (McClure et al., 2023; Dartois et al., 2024). The ENIIGMA results indeed take into account the possible saturation of the bands.

To better assess the sensitivity of our tool to band saturation, we retrained our neural networks with an augmented dataset in which we simulated the effect of band saturation. To do this, we used 7 copies of each spectrum in our training subset, applying a percentage of saturation (from 10 % to 70 %) by cropping the absorbance values to those fractions of the maximum value of absorbance. We then retrained our models for our 7 targeted variables and our 10 random seeds, obtaining a new set of model parameters (see Appendix C.8). The results of the validation are similar to the standard version of AICE, and we do not appreciate any significant increase in the validation error.

The comparison between the AICE standard model and the AICE version with saturation effects is displayed in Table 4.4.4. From this table, we find that both versions (AICE and AICE-s) present consistent results within the uncertainties. However, the predicted fractions of CO and CO_2 by AICE-s are slightly larger, so they agree better with the results by Dartois et al. (2024). The predicted temperature is also a little higher and more uncertain, which might indicate that the inclusion of saturation in the training subset reduces the precision of our neural networks for predicting the temperature. As for the predicted total sum, AICE-s predicts 84 ± 5 % for NIR38 and 85 ± 5 % for J110621, which is similar to the standard AICE model.

We note that this trend is observed for both NIR38 and J110621, even though the latter is expected to be more affected by saturation than NIR38, especially for the CO_2 band at 4.27 μm. In particular, the CO_2 fraction predicted by AICE-s remains very similar, which suggests that the original version of AICE already learns how to identify saturation (at least at these levels of uncertainty).

Table 4.4.5. Comparison between the predictions of the standard AICE and the version trained in a smaller range of 4000-2000 cm^{-1} (AICE-lite) towardss NIR38 and J110621.

	NIR38		**J110621**	
prediction	**AICE**	**AICE-lite**	**AICE**	**AICE-lite**
H_2O (%)	48.2 ± 2.3	54 ± 3	51 ± 3	53 ± 4
CO (%)	20 ± 3	18 ± 3	23 ± 3	20 ± 3
CO_2 (%)	8.6 ± 1.3	8.7 ± 1.3	6.3 ± 1.3	6.1 ± 1.2
CH_3OH (%)	4.7 ± 1.3	6 ± 3	3.3 ± 1.4	5.7 ± 2.5
NH_3 (%)	3.3 ± 2.1	6 ± 3	2.9 ± 2.0	8 ± 5
CH_4 (%)	1.3 ± 0.7	1.5 ± 0.8	1.5 ± 0.7	1.8 ± 1.0
temp. (K)	22 ± 4	15.5 ± 1.5	20 ± 5	16.7 ± 2.0

4.4.5 Effects of changes in the spectral range

Throughout this work we have always used the same spectral range, from 4000 cm^{-1} to 980 cm^{-1} (2.5-10.2 μm). In this section, we evaluate the performance of AICE when the observed spectrum presents a different wavelength coverage, especially if it is smaller.

The comparison between the standard version of AICE and this new version (AICE-lite), shown in Table 4.4.5, reveal that both sets of predictions are consistent within the uncertainties, with the slight exception of the H_2O prediction for NIR38, which is now a little higher. However, we find an uncertainty increase for CH_3OH and NH_3, which can be due to the fact that they do not present significant features in the range of 4000–2000 cm^{-1}. Therefore, although the networks provide a good performance with the AICE-lite version for abundant species such as H_2O, CO and CO_2 (and even for CH_3OH, although the prediction uncertainty is larger), the training of AICE with a broader wavelength coverage provides more accurate results for CH_3OH and NH_3. However, the predictions of the temperature given by AICE-lite are lower and less uncertain, closer to the previous estimates by McClure et al. (2023) and Dartois et al. (2024). This suggests that the shape of the observed spectra for both background stars in the range of 4000–2000 cm^{-1} is compatible with lower temperatures, while the part of 2000–980 cm^{-1} is more compatible with higher temperatures. Lastly, the predicted sums of the six targeted species are 94 ± 6 % for NIR38 and 94 ± 8 % for J110621, which is closer to 100 % than the AICE and AICE-s models.

This shows that AICE can be succesfully retrained to work with different spectral ranges. In fact, we could choose any range, providing that enough experimental data are available. In the case that the spectra to be analysed has a slightly different spectral range than that of the training dataset, or even a different resolution, AICE can still work and predict the ice composition and temperature, but will present larger errors the more the range and resolution of the input spectra differ from those of the training dataset.

4.5 Discussion

We have shown that artificial neural networks can be trained to identify and quantify the molecular composition of ices from infrared absorption spectra. Our tool AICE gives accurate and fast predictions of the ice fractional composition when used to analyse laboratory spectra and also astronomical observations. The obtained results of the ice fractional composition towards the background stars NIR38 and J110621 in Chamaeleon I, observed with the JWST, are in agreement with previous estimates obtained using other methods (see Sections 4.4.2 and 4.4.3). However, previous estimates with ENIIGMA use the band strenghs of pure species to derive ice column densities in the mixtures. As discussed in Section 3.3, in the mixtures the band strenghts could change up to 20 %. AICE has the advantage of implicitly considering band strength variations because the neural networks are trained with experimental ice mixtures, in whose bands (shapes, positions and and intensities) such information is contained. Therefore, the accuracy of the ice fractional composition predicted by the neural networks of AICE depends only on the accuracy of the composition of the experimental ice mixtures used for training.

Starting from the architecture of our model of artificial neural networks, future improvements are foreseen. The most important of them is the augmentation of the ice laboratory dataset, which is essential for an accurate performance of AICE. The list of targeted species could be enlarged to include other complex organic molecules and molecular ions/anions, always providing that enough experimental data exist. Current efforts focus on understanding whether RAIRS spectra (i.e., ice spectra obtained via reflection-absorption, and not transmission) could be used in large amounts to train future versions of AICE (e.g., spectra taken by Kruczkiewicz et al., 2021, or Vitorino et al., 2024). This would allow us to include the information from other databases such as the LERMA database,[17] significantly enlarging current's AICE database. The enlarging of the training dataset could enable the use of more sophisticated architectures for the neural networks that would improve the accuracy of the predictions and reduce the size of the model, while keeping inference times to less than one second.

The comparison with other analysis methods (McClure et al., 2023; Dartois et al., 2024) shows that AICE is considerably faster than ENIIGMA and the radiative-transfer model of Dartois et al. (2022). However, our model does not include any grain size parametrization, as Dartois et al. (2022)'s model does, and it does not predict the ice composition for more complex molecules than CH_3OH, as ENIIGMA does. AICE could be trained with a larger dataset including a wider and more complete range of ice mixtures, and/or it could also include the effects produced by grain growth. Therefore, in principle, AICE could be expanded to included these features. Alternatively, our tool AICE could be implemented within other tools like the ones mentioned above, to provide a first guess of the composition of the ice, and then perform a more detailed calculation with ENIIGMA and Dartois et al.

[17] https://lerma.labo.cyu.fr/DR/index.php

(2022)'s model, to make a precise fit of the observed astronomical spectrum. With respect to the spectra reduction, we note that the continuum determination is performed concatenating polynomials, so it might be prone to errors when obtaining the spectra in optical depth or absorbance. In addition, the removal of the silicate contribution is performed using a synthetic model, which might carry additional errors in the $\sim$ 8–10 μm range, which is crucial to identify CH_3OH, NH_3, and more complex organics. For the comparison of the AICE results presented in Sections 4.4.2 and 4.4.3, we have however followed similar reduction steps to those performed by McClure et al. (2023). In the case of Dartois et al. (2024), these authors follow an inverse procedure, adding the effect of silicates and of the stellar photosphere, which is also subject to some uncertainties.

On the other hand, it is also important to mention that astronomical ices observed with the JWST present contributions from different dust grain populations along the line of sight, which would sum up and result in a combined signal. This combined spectra would differ from the type of spectra used to train AICE, since ices in laboratory are grown in a single sample with uniform temperature. We believe this is one of the reasons behind the overestimate of the ice temperature by AICE for the ices measured towards NIR38 and J110621, since our model was trained with spectra that have a uniform temperature. In fact, there are more considerations that should be taken into account about the temperature of ices.

The spectral signature of an ice (i.e., the shape and position of the bands) is modified by the environment of the vibrators (the atoms), which is itself affected by the structure of the molecular film, i.e., the way in which the molecules organise themselves in relation to each other. For example, for water deposited on a cold surface from the gaseous phase, the structure is very disordered, with a high surface-to-volume ratio, creating a complex topology with closed pores when the layer is large enough. As the temperature rises, annealing occurs, increasing the average number of coordination [18] of each molecule and reducing the water-vacuum interfaces, which compacts the ice. The spectral signature also changes, as the mean number of coordination for each molecules increases (up to almost 4). However, this temperature-induced effect can be reproduced by other mechanisms. All that is needed to induce a change in the structure of the ice is to add energy, or to release energy. This energy can come from UV or even IR radiation (Noble et al., 2020), it can be deposited by energetic particles (electrons or ions) or simply released during a reaction (for example by H+H; Accolla et al., 2011).

Therefore, the temperature variable that is predicted by our tool AICE does not univocally represent the real temperature of the ice, but the minimum temperature to which it has been brought/annealed. In the laboratory, we could obtain the same spectrum by going back down in temperature, and also without increasing the temperature but by depositing it in a different way (by jet, or by forming it; e.g., Noble et al., 2024), or by modifying it by any energetic processing, even if slight. As our temperature variable does not really represent

[18] The number of coordination in a molecule is the number of bounds to its neighbours, not only in a crystalline solid but in a general solid or cluster structure.

the temperature of the ice but rather the way in which it was annealed during the experiments, it is not surprising that our model suggests that the best description is an ice with the same signature as a lightly annealed ice, which is simply due to the formation history of interstellar films, which is not that of a simple deposition at 10 K, but has been partly processed (upon its formation, chemical activity, radiative field and energetic processing).

4.6 Conclusions

We have developed a machine learning model based on artificial neural networks that predicts the ice composition corresponding to an input infrared absorption spectra. The model, which is packed into a user-friendly Python software called AICE, was trained to identify the temperature of the ice and its composition for the most commonly detected molecules, i.e., H_2O, CO, CO_2, CH_3OH, NH_3, and CH_4, using hundreds of experimental ice absorption spectra. The model shows a validation error of $\sim$3 % in for the fractional composition and $\sim$11 K for the temperature. When tested with astronomical observations taken by JWST, our model's predictions are in agreement with previous estimates obtained by other authors (McClure et al., 2023; Dartois et al., 2024) within the uncertainties, with the exception of the temperature, which is slightly overpredicted. The computation time is very low, $\sim$0.1 s per input spectrum, which would allow to automatically analyze hundreds of astronomical spectra measured with JWST and perform a statistically-robust analysis. In addition, AICE is a versatile tool that could also be implemented or used in combination with other ice analysis tools such as ENIIGMA.

Our tests using a dataset that reflects possible saturation effects in the bands of H_2O, CO and CO_2 show that AICE is little sensitive to these effects since it learns to recognize the actual ice composition by focusing on the whole shape of the molecular band (including the band wings) and not on the relative peak intensity only. Additional tests reveal that AICE performance is optimal for the determination of the ice composition for species such as H_2O, CO, CO_2 and CH_3OH, when using a reduced spectral range similar to the one provided by the JWST instruments NIRCam and NIRSpec.

Lastly, we also include an interactive toolkit within AICE, which can be used by laboratory astrochemists to remove the experimental baselines and possible contaminations present in laboratory ice absorbance spectra, in an easy and semi-automated way. This could help enlarging the list of clean and reduced infrared spectra available in public databases.

Chapter 5

Astrochemical simulations of molecular clouds for the Ice Age project

This section has been adapted from the paper *Modelling CH_3OH and hydride formation in the JWST Ice Age era*, submitted the journal Astronomy & Astrophysics (Jiménez-Serra et al., submitted). In particular, this chapter focuses on the work that has been carried out as a part of this thesis: the UCLCHEM simulations and the best model determination using the mean absolute error (MAE). The complete list of authors of the paper is: Izaskun Jiménez Serra, Andrés Megías Toledano, Joseph Salaris, Herma Cuppen, Angèle Taillard, Miwha Jin, Valentine Wakelam, Anton I. Vasyunin, Paola Caselli, Yvonne J. Pendleton, Emmanuel Dartois, Jennifer A. Noble, Serena Viti, Katerina Borscheva, Robin T. Garrod, Thanja Lamberts, Helen Fraser, Gary Melnick, Melissa McClure, Will R. M. Rocha, Maria N. Drozdovskaya, and Dariusz C. Lis.

Abstract

Recent JWST observations have measured the ice chemical composition toward two highly-extinguished background stars, NIR38 and J110621, in the Chamaeleon I molecular cloud. They suggest that CH_3OH ice in this cloud could have formed in a water-rich environment where the catastrophic depletion of CO has not yet occurred. This contrasts with the idea of CH_3OH ice being predominantly formed via CO hydrogenation. Laboratory experiments and quantum chemical calculations have proposed alternative grain-surface reactions for the formation of CH_3OH ice in water-rich environments. Unfortunately, no chemical modelling has been carried out thus far to test their efficiency. In addition, it remains unexplored how the efficiencies of the proposed mechanisms depend on the astrochemical code employed. We model the ice chemistry in the Chamaeleon I cloud to establish the dominant formation processes of CH_3OH, CO, CO_2, and of the hydrides CH_4 and NH_3 (besides H_2O). By using a set of state-of-the-art astrochemical codes (MAGICKAL, MONACO, Nau-

tilus, UCLCHEM, and KMC simulations), we can test the effects of the different code architectures (rate equation vs. stochastic codes) and of the assumed ice chemistry (diffusive vs. non-diffusive). We consider a grid of models with different gas densities, dust temperatures, visual extinctions, and cloud-collapse lengths. Besides the successive hydrogenation of CO, the codes chemical networks have been augmented to include the alternative processes for CH_3OH ice formation in water-rich environments (i.e., the chain of reactions starting from CH_4 + OH, and C + H_2O). Our models show that the JWST ice observations are better reproduced for gas densities $\geq 10^5$ cm^{-3} and collapse time-scales $\geq 10^5$ yr. CH_3OH ice formation occurs predominantly (>99 %) via CO hydrogenation. The contribution of reactions CH_4 + OH and C + H_2O, is negligible. The CO_2 ice may form either via CO + OH or CO + O depending on the code. However, KMC simulations reveal that both mechanisms are efficient despite the low rate constant of the CO + O surface reaction. CH_4 is largely underproduced for all codes except for UCLCHEM, for which a higher amount of atomic C is available during the translucent cloud phase of the models. Large differences in the predicted abundances are found at very low dust temperatures ($T_{\rm dust}$ < 12 K) between diffusive and non-diffusive chemistry codes. This is due to the fact that non-diffusive chemistry takes over diffusive chemistry at such low $T_{\rm dust}$. This could explain the rather constant ice chemical composition found in Chamaeleon I and other dense cores despite the different visual extinctions probed.

5.1 Introduction

Interstellar ices represent the main reservoir of organic and volatile components in molecular clouds and star-forming regions (see, e.g., the review by Boogert et al., 2015, and references therein). Ices grow during the collapse of molecular clouds, largely influencing the chemical composition of protoplanetary disks and, ultimately, of planetesimals and cometesimals (see, e.g., works by Quénard et al., 2018; Coutens et al., 2020). Indeed, a fraction of the chemical compounds measured in comet 67P/Churyumov-Gerasimenko by the ESA *Rosetta* mission could be of prestellar origin, which suggests that they could have formed during the dense molecular cloud phase (Altwegg et al., 2016; Drozdovskaya et al., 2019, 2021).

The composition of ices in molecular clouds and star-forming regions is studied through absorption spectroscopy of interstellar ices in the near and mid infrared (IR) against bright background sources (see, e.g., works by Whittet et al., 1983; Pendleton et al., 1999; Boogert et al., 2013; Goto et al., 2018). However, these observations have been scarce because they are challenging from the ground due to telluric features and high sky-background noise in the IR. In addition, dense molecular clouds and dense cloud cores can have very high levels of visual extinction ($A_{\rm V} \geq 20$ mag), for which detailed ice absorption spectroscopic observations in the near- and mid-IR require IR-bright background field stars and very high sensitivity.

The James Webb Space Telescope (JWST), with its unprecedented sensitivity and high

spectral resolving power (up to ∼3000), is providing the most comprehensive view of the chemical composition of interstellar ices during the initial conditions of star and planet formation. The Ice Age Early Release Science (ERS) program has obtained ice absorption spectra toward hundreds of field stars located behind the Chameleon I dense molecular cloud (McClure et al., 2023). The goal of the project is to establish the chemical composition of interstellar ices concurrent to the process of Solar-type system formation from the molecular cloud stage, to prestellar cores, young stellar objects (Class 0 and I sources), and, ultimately, proto-planetary disks (Class II-type sources; McClure et al., 2023; Sturm et al., 2023, 2024; Dartois et al., 2024; Noble et al., 2024). The first results have reported the ice chemical composition toward the most highly-extinguished background stars ever observed to date in molecular clouds, NIR38 and J110621, with pre-flight estimated visual extinctions of A_V = 60 mag and A_V = 95 mag, respectively (McClure et al., 2023), although these values are subject to large uncertainties (see Dartois et al., 2024; Noble et al., 2024:and Section 5.4.4). These observations reveal that the ice composition is very similar toward the two lines-of-sight: carbon monoxide (CO) ice is ∼30% and ∼45% by number, respectively, relative to water (H_2O); and CO is more abundant than carbon dioxide (CO_2; between 13% and 20% with respect to water) and methanol (CH_3OH; between 4 and 9%). Hydrides such as methane (CH_4) and ammonia (NH_3) comprise only up to 5% of ice with respect to water.

In this work, we model the chemical composition of the ices observed with the JWST toward the background stars NIR38 and J110621 within the Ice Age ERS program. For this modelling effort, we use five different astrochemical codes to identify the features (or chemical processes) within these models that best reproduce the observed ice chemical compositions.

5.2 The models

Our aim is to model the ice chemical formation and evolution in the Chameleon I dense molecular cloud using different astrochemical codes. These codes are built following different philosophies: we have both rate equation codes such as MAGICKAL (Garrod, 2019), MONACO (Vasyunin and Herbst, 2013; Vasyunin et al., 2017), Nautilus (Ruaud et al., 2016), and UCLCHEM (Holdship et al., 2017); as well as the stochastic Kinetic Monte Carlo (KMC) model of Cuppen and Herbst (2007). These codes also include different levels of complexity for the formation and chemistry of the ices and, thus, the idea of this work is to test the features from these astrochemical codes that best reproduce the JWST observations obtained toward NIR38 and J110621, which are field stars with the highest extinction observed thus far (McClure et al., 2023). For instance, while MAGICKAL, MONACO and KMC include diffusive and non-diffusive chemistry, Nautilus and UCLCHEM only include diffusive chemistry. On the other hand, Nautilus and UCLCHEM include some non-thermal desorption mechanisms than are not considered in the other codes. As part of this thesis, I have carried out the chemical modelling using the UCLCHEM code. In the following I describe this code, report

the grid of physical parameters explored, and present the results obtained with it.

5.2.1 UCLCHEM

UCLCHEM (Holdship et al., 2017) is a time-dependent gas-grain chemical model based on the rate equations approach, and it is a public open source code.[1] Gas phase reactions are included as two-body reactions, including interactions with cosmic rays and UV photons. Gas species can accrete onto grain surfaces, which are modelled with two phases: the surface and the bulk. Diffusive chemistry is considered as described in Quénard et al. (2018), using a constant ratio between diffusion and binding energies of 0.5. However, UCLCHEM does not include non-diffusive processes in the ices. Both thermal and non-thermal desorption mechanisms are considered in this code. For thermal desorption, UCLCHEM follows the formalism of Viti et al. (2004), while for the non-thermal desorption processes UCLCHEM includes desorption produced by the energy released in H_2 formation, incident cosmic rays, UV photo-desorption, cosmic-ray-induced UV photo-desorption (as described in the work by Holdship et al., 2017) and chemical reactive desorption following the formalism of Minissale et al. (2016).

The default grain chemical network is based on the one developed by Quénard et al. (2018), and the gas-phase chemical network is taken from the UMIST 2013 database (McElroy et al., 2013). This grain network was updated by adding some new reactions involving CH_3OH, CO_2 and HOCO (hydrocarboxil radical), and by changing the activation energy of reaction $H_2CO + O \rightarrow CO_2 + H_2$ from 0 to 335 K (Minissale et al., 2015). The reactions involving HOCO are listed in Table 5.2.1, and they have been extracted from the works by Goumans et al. (2008), Goumans and Andersson (2010), Arasa et al. (2013), Ruaud et al. (2015), Qasim et al. (2019), and Molpeceres et al. (2023). For the formation of methanol, besides the classical route of multiple CO hydrogenation reactions in the ice, we also include the following chemical processes: 1) the chain of neutral-neutral grain surface reactions $CH_4 + OH \rightarrow CH_3 + H_2O$ and $CH_3 + OH \rightarrow CH_3OH$ proposed by Qasim et al. (2018); and 2) the reaction $C + H_2O \rightarrow H_2CO$ and subsequent hydrogenation reactions proposed by Molpeceres et al. (2021) and Potapov et al. (2021).

5.2.2 Physical parameters

The chemistry of the Chamaeleon I molecular cloud is modelled considering its evolution from its translucent phase to the dense cloud stage. To do this, we first consider a *phase 0* in which the chemistry of a translucent cloud is modelled for a duration of 10^6 yr assuming that the cloud is static and adopting a number density of total H nuclei of $n_H = 2\times10^3$ cm^{-3}, a dust and gas temperature $T_{dust} = T_{gas} = 15$ K, visual extinction $A_V = 2$ mag, UV radiation field $G_0 = 1$ Hab, and a cosmic-ray ionization rate $\zeta = 1.3 \times 10^{-17}$ s^{-1}.

[1] https://uclchem.github.io/

Table 5.2.1. Grain-surface reactions added in UCLCHEM for HOCO and CO_2.

Reaction			Energy barrier (K)
$CO + O \rightarrow$	CO_2		2500
$CO + OH \rightarrow$	HOCO	(50%)	150
	$H + CO_2$	(50%)	150
$N + HOCO \rightarrow$	$NH + CO_2$	(50%)	0
	$OH + OCN$	(50%)	0
$O + HOCO \rightarrow$	$OH + CO_2$		0
$H + HOCO \rightarrow$	$H_2 + CO_2$	(70%)	0
	$CO + H_2O$	(20%)	0
	HCOOH	(10%)	0

The third reaction, forming H and CO_2, was already present in the default chemical network of UCLCHEM, but we updated the energy barrier from 1000 K to 150 K (Ruaud et al., 2015).

The chemical evolution of the cloud collapse is modeled in a subsequent *phase 1*, in which the output gas and ice (both surface and bulk) abundances from phase 0 are used as input abundances for the following simulations. For the cloud collapse, we assume a free-fall collapse treatment with the gas density ($n_{\rm H}$) evolving with time (t) as (Brown et al., 1988; Rawlings et al., 1992):

$$\frac{\mathrm{d}\, n_{\rm H}}{\mathrm{d}\, t} = b\ n_{\rm H} \left(\frac{n_{\rm H}}{n_{\rm H,0}}\right)^{\frac{1}{3}} \left[6 \cdot 4\pi G\, m_{\rm H}\, n_{\rm H,0} \left(\left(\frac{n_{\rm H}}{n_{\rm H,0}}\right)^{\frac{1}{3}} - 1\right)\right]^{\frac{1}{2}}, \tag{5.1}$$

where $n_{\rm H,0}$ is the initial hydrogen nuclei central density, $m_{\rm H}$ is the mass of a hydrogen nucleus, and G is the gravitational constant. We also include a factor b in the equation so that we can test the effects of an accelerated collapse in the chemistry of the ices (Holdship et al., 2017). The collapse finishes when $n_{\rm H}$ reaches the final density values at the final collapse times shown in Table 5.2.2. As the density $n_{\rm H}$ evolves during the collapse, the visual extinction $A_{\rm V}$ increases as a function of density (and, therefore, time), following equation 1 by Jiménez-Serra et al. (2018).

Table 5.2.2 reports the grid of physical parameters considered for the models of the two field stars NIR38 and J110621. The assumed visual extinctions are $A_{\rm V}$ = 60 mag and $A_{\rm V}$ = 95 mag, respectively; however, these values may be lower (see Dartois et al., 2024: and Section 5.4.4). The considered gas/dust temperatures are 14 K for NIR38 and 12 K for J110621, as inferred from the Herschel-based dust temperatures maps of Väisälä et al. (2014).[2] For UCLCHEM, $T_{\rm dust}$ is kept constant at either 12 K or 14 K (see Table 5.2.2). The gas temperature is set to be the same as the dust temperature in all runs and for all codes.

[2] The uncertainties associated with these $T_{\rm dust}$ values are as high as $\pm$0.6 K, after considering a β range between 1.8 and 2.2 and the uncertainties due to photometric noise (Väisälä et al., 2014).

Table 5.2.2. Physical parameters assumed for the grid of astrochemical models.

	NIR38				J110621			
model #	$n_{\rm H}$ ($\rm cm^3$)	$T_{\rm dust}$ (K)	$A_{\rm V}$ (mag)	$t_{\rm col}$ (yr)	$n_{\rm H}$ ($\rm cm^3$)	$T_{\rm dust}$ (K)	$A_{\rm V}$ (mag)	$t_{\rm col}$ (yr)
1	2×10^4	14	60	10^4	2×10^4	12	95	10^4
2	2×10^4	14	60	10^5	2×10^4	12	95	10^5
3	2×10^4	14	60	10^6	2×10^4	12	95	10^6
4	2×10^5	14	60	10^4	2×10^5	12	95	10^4
5	2×10^5	14	60	10^5	2×10^5	12	95	10^5
6	2×10^5	14	60	10^6	2×10^5	12	95	10^6
7	2×10^6	14	60	10^4	2×10^6	12	95	10^4
8	2×10^6	14	60	10^5	2×10^6	12	95	10^5
9	2×10^6	14	60	10^6	2×10^6	12	95	10^6

Note. $T_{\rm gas}$ is assumed to be the same as $T_{\rm dust}$ in all models.

The final $n_{\rm H}$ densities at the end of the phase 1 collapse stage are 2×10^4, 2×10^5 and 2×10^6 $\rm cm^{-3}$. These densities have been selected from the H_2 volume gas densities estimated by Belloche et al. (2011) toward the Cha1-C1 and Cha1-C2 cores (see their Table 6), which are consistent with the results of Väisälä et al. (2014). These densities lie between some 10^5 and 10^6 $\rm cm^{-3}$. For completeness, we also explore the lower-density case with $n_{\rm H} = 2\times10^4$ $\rm cm^{-3}$.

For the final collapse times, we assume time-scales of 10^4, 10^5 and 10^6 yr. Free-fall collapse times [3] correspond to $\sim10^6$ yr. The collapse times of 10^4 and 10^5 yr are tested by modifying the b parameter in equation 5.1, which adopts values from 1 (free-fall, $\sim$ 10^6 yr) to 120 (accelerated collapse, $\sim$ 10^4 yr). The three different collapse times assumed in our grid of models are included to test the effects of a fast/slow collapse on the chemical composition of the ices. Note that short time-scales of the order of 10^4 yr are typically not observed for prestellar cores (Keto and Caselli, 2010). However, accelerated collapse is considered in this study to investigate the possible effects of gravitational collapse driven by stellar feedback from nearby star formation. The cosmic-ray ionization rate and UV radiation field assumed for the collapse stage in all models are, respectively, $\zeta = 1.3 \times 10^{-17}\,\rm s^{-1}$ and $G_0 = 1$ Hab.

As initial elemental abundances, we consider the low-metal abundance case (the EA1 case) from Wakelam and Herbst (2008). All models consider a constant grain size of 0.1 μm. The output fractional abundances are given with respect to the number density of total H nuclei. The conversion from ice molecular abundances to ice molecular column densities is performed using the column densities of total H nuclei toward NIR38 and J110621 calculated as $N_{\rm H} = 1.6 \times 10^{21} A_{\rm V}$ $\rm cm^{-2}$.[4]

[3] Free-fall collapse times are calculated as $t_{\rm col} \simeq 4 \times 10^7 / (n_{\rm H,0}/\rm cm^{-3})^{1/2}$ yr, with $n_{\rm H,0}$ the initial gas cloud density.

[4] This expression differs only by a factor of $\sim$1.1 with respect to the one reported by Lacy et al. (2017) of $N_{\rm H} = 2 \times 0.9 \times 10^{21} A_{\rm V}$ $\rm cm^{-2}$.

Table 5.3.1. Observed column densities and column density ratios with respect to H_2O toward NIR38 and J110621 (McClure et al., 2023)

	NIR38 (A_V = 60 mag)		J110621 (A_V = 95 mag)	
Molecule	N (10^{18} cm^{-2})	$N/N(H_2O)$ (in %)	N (10^{18} cm^{-2})	$N/N(H_2O)$ (in %)
H_2O	$6.9^{+1.9}_{-1.1}$	100	$13.4^{+1.3}_{-1.8}$	100
CO	$3.0^{+0.6}_{-0.4}$	43^{+12}_{-10}	$3.7^{+0.6}_{-0.4}$	28^{+6}_{-4}
CO_2	$1.38^{+0.20}_{-0.20}$	20^{+5}_{-5}	$1.74^{+0.21}_{-0.22}$	$13.1^{+2.5}_{-2.0}$
CH_3OH	$0.61^{+0.11}_{-0.11}$	$8.7^{+2.5}_{-2.1}$	$0.51^{+0.19}_{-0.09}$	$3.9^{+1.5}_{-0.8}$
NH_3	$0.30^{+0.22}_{-0.03}$	$4.6^{+3.0}_{-1.1}$	$0.66^{+0.15}_{-0.06}$	$5.0^{+1.3}_{-0.7}$
CH_4	$0.180^{+0.017}_{-0.013}$	$2.6^{+0.5}_{-0.6}$	$0.250^{+0.010}_{-0.030}$	$1.84^{+0.30}_{-0.24}$

5.3 Results

In this Section, we analyze the results obtained with the astrochemical code UCLCHEM. We present the evolution of the chemical composition of the ices and the main processes that lead to their formation for the grid of models described in the previous section. We focus on some of the molecular species observed with JWST toward NIR38 and J110621: H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4. The model results are thus compared to the observational data reported in McClure et al. (2023) and presented in Table 5.3.1.

5.3.1 Static phase

Figure 5.3.1 shows the time evolution of the abundances with respect to H_2 for different species resulting from the static phase stage of the UCLCHEM simulation. For the gas-phase we only plot the relative abundances of atomic C and CO, since these two species are essential to understanding the amount of C that will be locked into CO, CO_2, and CH_4 ices during the phase 1 of the simulations.

It is clear that atomic C has been efficiently converted into CO by the end of the phase 0 stage with a CO gas-phase abundance that goes up to $\sim 7 \times 10^{-5}$ with respect to the total H nuclei at the end of the simulations. This will largely affect the subsequent production of CO_2 and CH_4 during the phase 1 stage of the models. The most abundant ice species is H_2O, followed by CO and CH_4. However, the abundance of CO_2 is very low through all the simulations.

This relatively low rate of ice formation is explained by the high efficiency for non-thermal desorption in UCLCHEM, due mainly to the impact of UV photons at these low visual extinctions (A_V = 2 mag). In particular, this value is 1 desorbed molecules per 10 UV photons. In the other astrochemical codes, non-thermal desorption is not that efficient, so there will be differences in the ice growth rate.

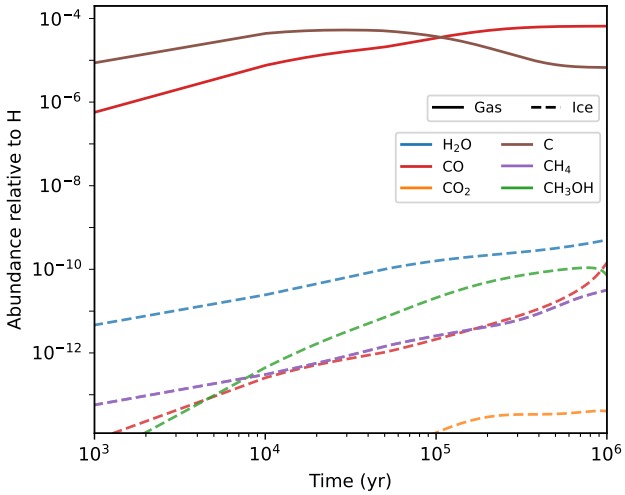


Figure 5.3.1. Time evolution of the gas and ice molecular abundances for most relevant species predicted by UCLCHEM for the phase 0 runs. Solid lines indicate gas-phase abundances, while dashed lines show ice abundances. Colors indicate the different molecular species as shown in the legend.

5.3.2 Collapse phase

In Figure 5.3.2, we present the evolution of the relevant species both in the ice and in the gas phase for some example cases of the phase 1 runs (collapse stage). We have selected two extreme cases for the two modelled stars (NIR38 in the top and J110621). On the left, we show the case of the lowest final density and the fastest collapse ($n_{\rm H} = 10^4\,{\rm cm}^{-3}$, $t_{\rm col} = 10^4\,{\rm yr}$), while on the right we have the highest final density and the slowest collapse ($n_{\rm H} = 10^6\,{\rm cm}^{-3}$, $t_{\rm col} = 10^6\,{\rm yr}$).

For the case of a fast collapse, we observe just a constant ice growth (in logarithmic scale) with a slight gas depletion over time. This suggests that the ice grow is proportional (in logarithmic scale) to the increase in column density. In any case, this physical scenario of fast collapse and low final density does not seem very realistic. For the case of a slow collapse with a higher final density, the growth of the ices mainly occurs at the end of the collapse phase with the rapid increase of the density; again, it seems that the increase of the gas density is the main driver of ice growth.

In Figures 5.3.3 and 5.3.4 we summarize the obtained ice column densities at the end of the collapse phase for the different models run with UCLCHEM. We see an underproduction of H_2O, which is probably due to the relatively low ice abundances at the end of the static phase. For CO, CO_2 and CH_3OH, the predicted ice abundances are either underestimated or overestimated depending on the model. The overproduction of CO_2 is detrimental for CH_3OH formation as the majority of atomic C is locked into the inert CO_2 species.

According to our models, the main formation route of CH_3OH on grains at the high visual extinctions of the field stars NIR38 and J110621 is the hydrogenation of CO on dust grain surfaces. Regarding CO_2, its formation mainly occurs on the surface of dust grains via the reaction CO + OH $\rightarrow$ H + CO_2. We find no difference in the final ice abundance of CO, CO_2 and CH_3OH between the case of adding the reactions C + H_2O $\rightarrow$ H_2CO and CH_4 +

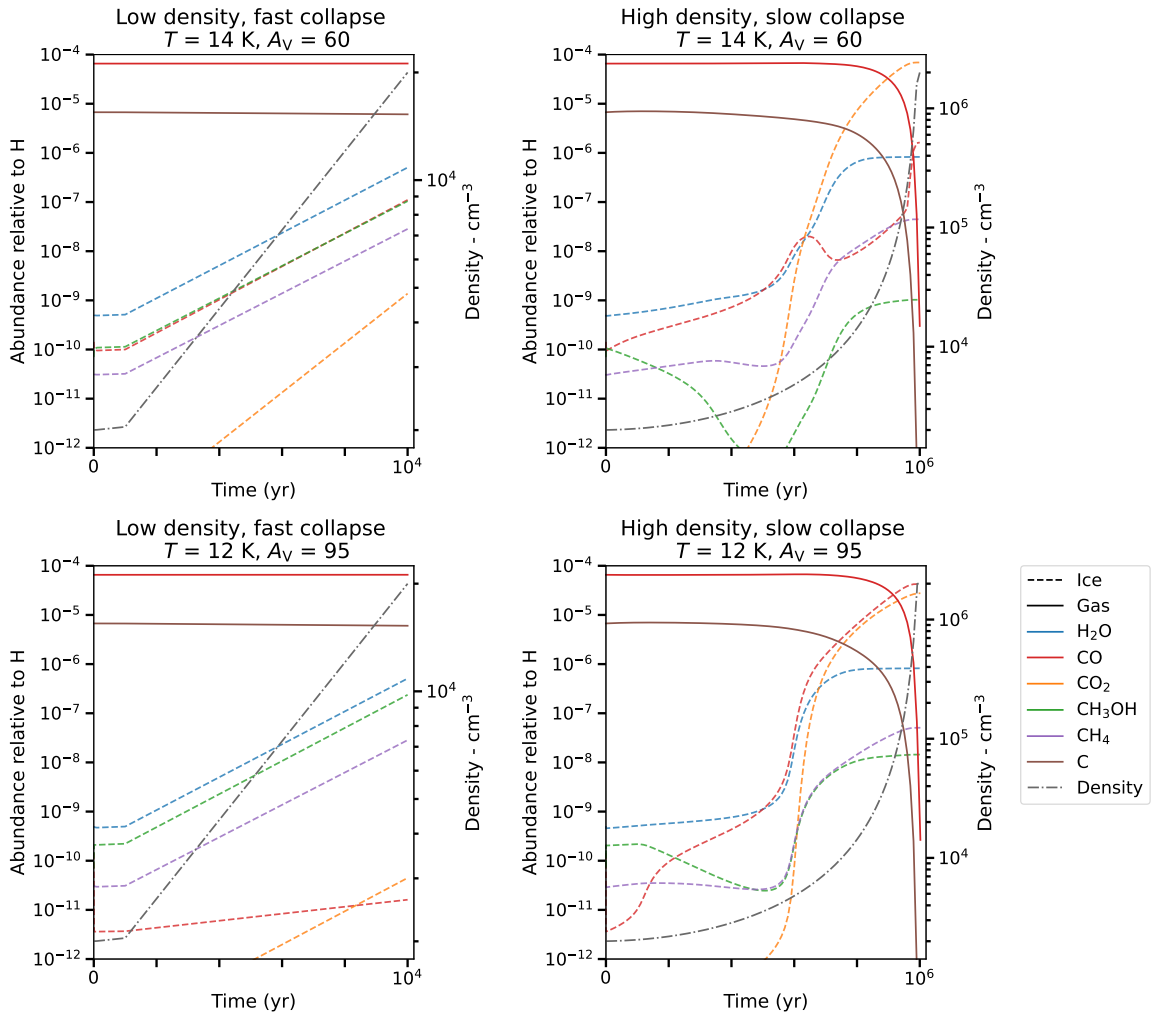


Figure 5.3.2. The time evolution of gas (solid) and ice (dashed) abundances in UCLCHEM during the collapse 'phase 1' stage. Models with T = 14 K and A_v = 60 mag correspond to NIR38, while models with T = 12 K and A_v = 95 mag correspond to J110621.

OH $\rightarrow$ CH_3 + H_2O to the chemical network and the case of not adding them. For NH_3 and CH_4, UCLCHEM tends to overpredict the ice abundance of these molecules by factors $\sim$2–8, obtaining better results for CH_4.

Overall, the models that best reproduce the observed ice compositions are those with intermediate collapse times ($t_{col} = 10^5$ yr) and intermediate or high final densities ($n_H = 10^5, 10^6$ cm^{-3}), although there are larger mismatches for the field star J110621.

5.3.3 Best fit models

To evaluate the performance of the models in a systematic way, we used the same algorithm for all models. First, we select the models that predict H_2O column densities closer to the Ice Age observations, that is, the models with the lowest difference or error. Then, we calculate the mean absolute error (MAE) between the predicted and observed abundances with respect to H_2O on a logarithmic scale.

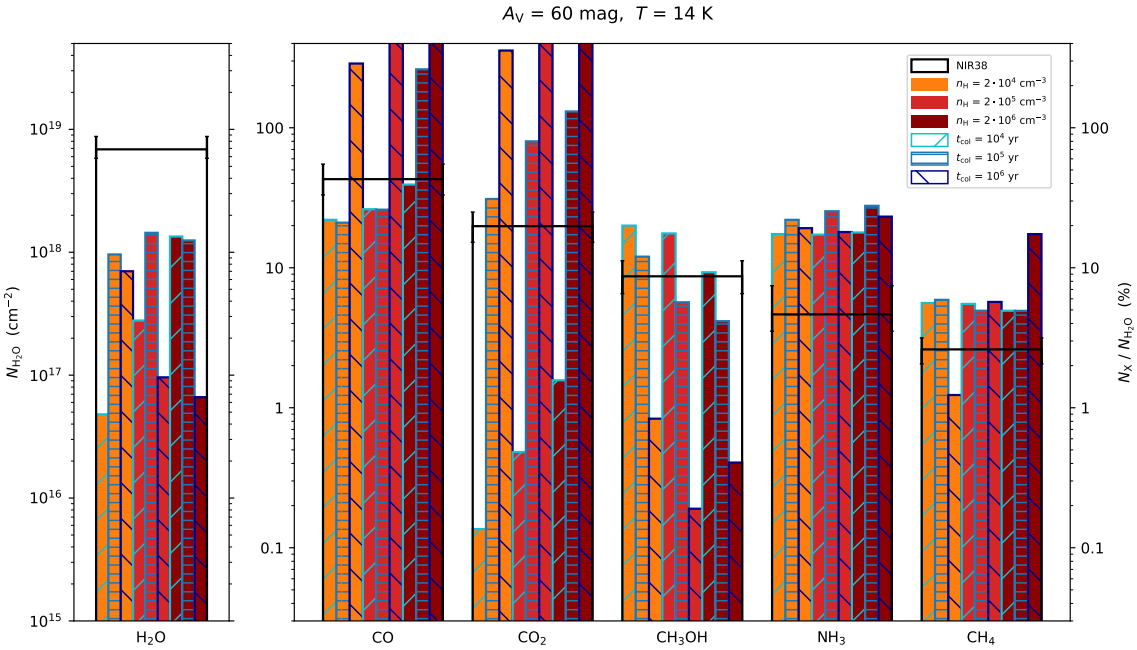


Figure 5.3.3. Ice compositions predicted by UCLCHEM for NIR38 for the models listed in Table 5.2.2. *Left panel:* Predicted water ice column densities. *Wide right panel:* Ice column density ratios with respect to water for the remaining species. Histograms are colored and drawn as indicated in the upper right part of the panel. The visual extinction and dust temperatures assumed in the models are reported in the upper part of the panel. Black wide histograms report the observed Ice Age column density of water and ice column density ratios obtained by McClure et al. (2023).

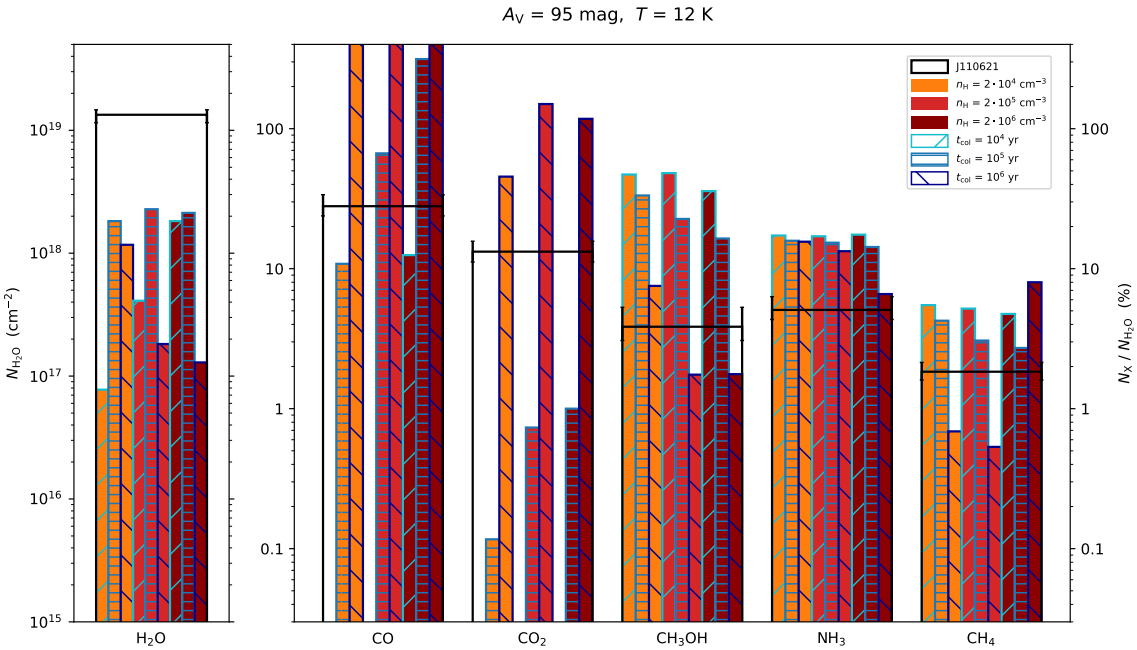


Figure 5.3.4. Ice compositions predicted by UCLCHEM for J110621. Details are the same as in Figure 5.3.3. Missing bars indicate ice abundance ratios ≤ 0.03 relative to H_2O.

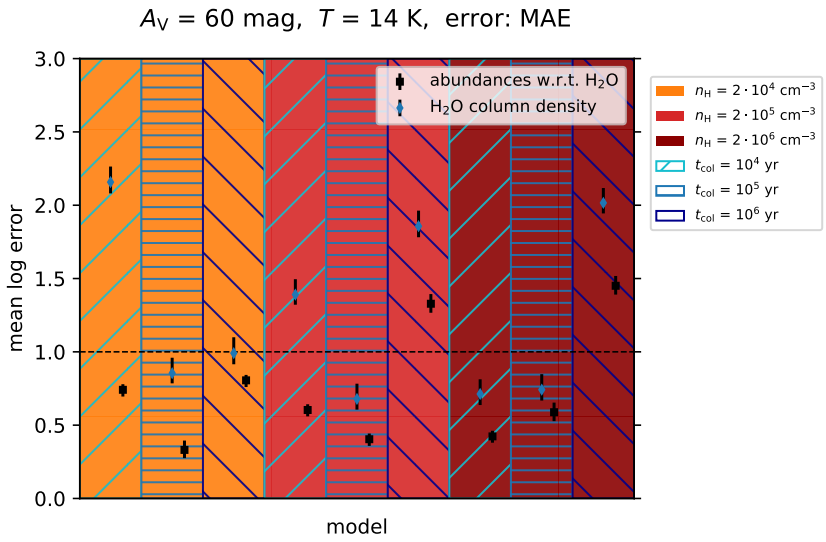


Figure 5.3.5. Logarithmic errors between the predicted ice compositions by the different models of UCLCHEM for $A_V = 60$ with $T = 14$ K. Errors for only the H_2O column densities are indicated with blue rhombus, while black squares mark the errors calculated with the abundances with respect to water, both superimposed to the bars that represent the models in previous figures.

The equation used for the MAE is:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| \log_{10}(y_i) - \log_{10}(\hat{y}_i) \right| , \tag{5.2}$$

where y_i represents the observed value of species i (in abundance with respect to H_2O), $\hat{y}_i$ is the value predicted by the astrochemical model, and N is the number of species observed with the JWST. We propagate the uncertainty in the values reported by McClure et al. (2023) to derive the uncertainty in the MAE.[5] The best model corresponds to the one that presents the lowest MAE value within the estimated uncertainty coming from the observations.

Figure 5.3.5 shows the H_2O error and the MAE for all the NIR38 models using UCLCHEM, while Figure 5.3.6 presents the same but for J110621. For NIR38, we can see that there are four models with the lowest value of MAE (within the uncertainties), but the one with the lowest final density (2×10^4 cm^{-3}) presents an error for H_2O column density which is too high and clearly larger than the rest, so we discard it. Then, the best models are the ones with intermediate collapse times ($t_{col} = 10^5$ yr) and intermediate/high final densities ($n_H = 2 \times 10^5$, 2×10^6 cm^{-3}), and also the one with high final density and fast collapse. However, if we now repeat the same procedure but with J110621 (Fig. 5.3.6), we obtain the same models except for the last one with fast collapse. Therefore, the two best models for UCLCHEM are, for both sources, the ones with intermediate collapse times and intermediate/high final densities (second and fifth models from the right). From these two models, the ones with high intermediate final density (for both NIR38 and J110621) are slightly better (with lower MAE), so we choose those ones as the best model for UCLCHEM. Therefore, for UCLCHEM

[5] This propagation is done automatically using the Python library RichValues: `https://pypi.org/project/richvalues`. For more details, see Appendix C.7

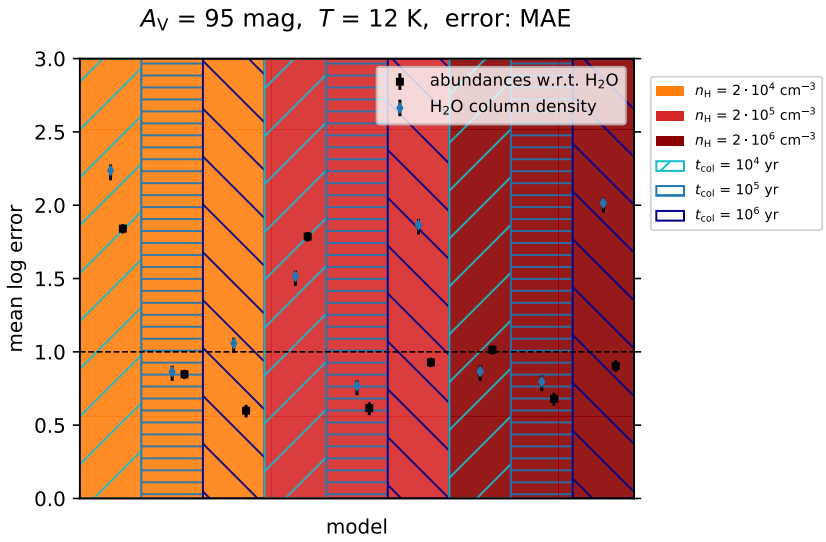


Figure 5.3.6. Same as Fig. 5.3.5, but for J110621 ($A_V = 95$). Here we see that the third model from the right shows a MAE value too high, which makes us think that the two best models for the two sources are the ones with intermediate collapse times and intermediate/high final densities (second and figfth from the right).

the best model is the one with intermediate final density (2×10^5 cm^{-3}) and intermediate collapse time (t_{col} = 10^5 yr). For NIR38 this model presents a MAE of $0.40^{+0.04}_{-0.05}$, while for J110621 the MAE is $0.62^{+0.04}_{-0.05}$.

We also note that we performed some tests using the root mean squared error (RMSE) and found similar results. The main difference is that the RMSE penalizes the models that have bigger discrepancies between observations and predictions for some molecules. Therefore it would yield a larger error if one species is very wrongly predicted, while favouring models with a more uniform error between the different species predictions. For the RMSE, the equation used would be:

$$\text{RMSE} = \left(\frac{1}{N} \sum_{i=1}^{N} \left(\log_{10}(y_i) - \log_{10}(\hat{y}_i) \right)^2 \right)^{\frac{1}{2}} . \tag{5.3}$$

5.3.4 Models with a lower temperature

In the model runs explored above, we have considered T_{dust} = 12 K and 14 K as estimated from *Herschel*-based dust temperature maps. However, it is well known that these maps are representative of the outer and slightly warmer layers of molecular clouds (see, e.g., works by Hocuk et al., 2017; Clément et al., 2023). In order to test the influence of the dust temperature on the ice chemical composition toward the field stars NIR38 and J110621, we have repeated the simulations for the best-fit models, but assuming a dust temperature of T_{dust} = 10 K. The results can be seen in Figures 5.3.7 and 5.3.8.

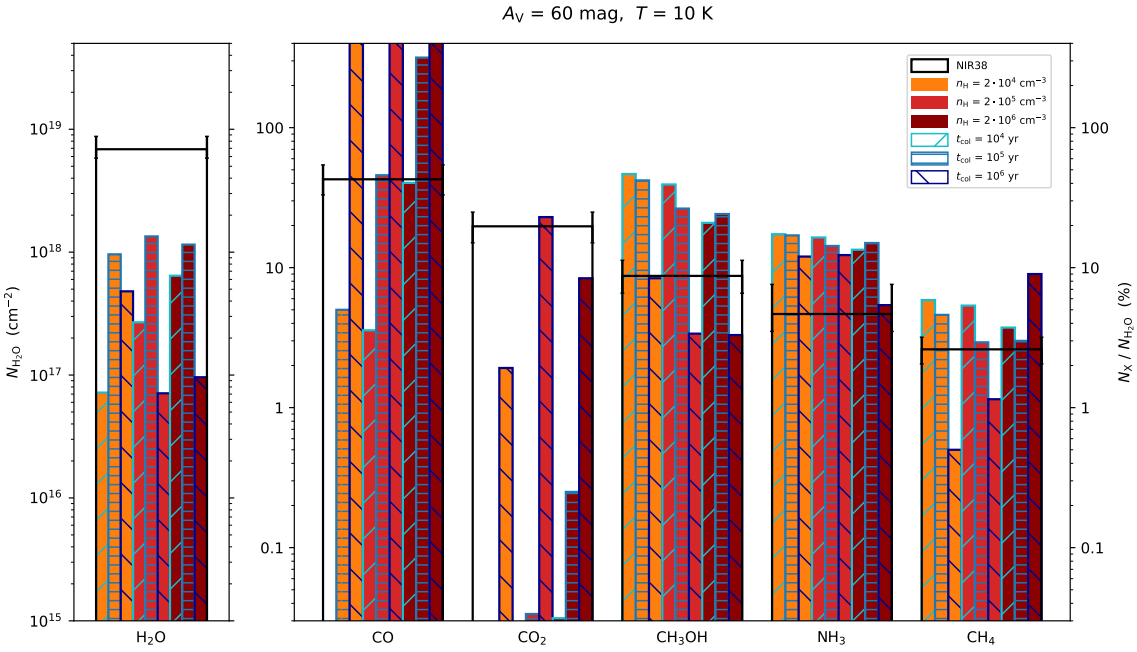


Figure 5.3.7. Ice compositions predicted by UCLCHEM for J110621 with T_{dust} = 10 K. Details are the same as in Figure 5.3.3. Missing bars indicate ice abundance ratios ≤ 0.03 relative to H_2O.

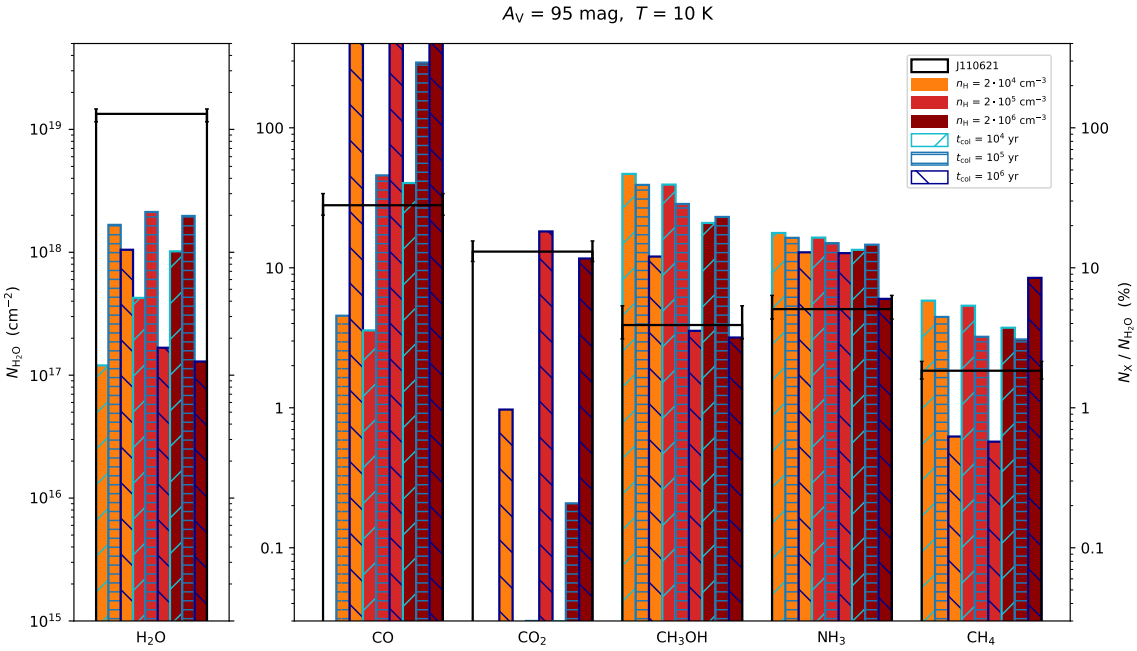


Figure 5.3.8. Ice compositions predicted by UCLCHEM for J110621 with T_{dust} = 10 K. Details are the same as in Figure 5.3.3. Missing bars indicate ice abundance ratios ≤ 0.03 relative to H_2O.

The simulations for both sources are quite similar, since their visual extinctions are both quite high, which produces a supression of the chemistry related to UV photons. For both sources, the decrease of $T_{\rm dust}$ in the UCLCHEM simulations induces a drastic reduction in the formation of CO_2 as well as an increase in the formation of CH_3OH. This behaviour was hinted from the results for J110621 (A_V = 95, T = 12 K), for which we already observed an underproduction of CO_2 due to the decrease in the efficiency of the reaction CO + OH $\rightarrow$ CO_2 + H. This decrease seems to be produced just because of the lower temperature and hence lower mobility of the OH radicals in the dust surface. Additionally, we see an increase of H atoms at the end of the collapse for the model with $T_{\rm dust}$ = 10 K with respect to the other models with higher temperatures, which could reduce the space for OH radicals to move and react with CO.

5.4 Discussion

5.4.1 Formation of CH_3OH ice

The formation routes of methanol considered in our simulations are: 1) successive CO hydrogenation in the reaction chain CO $\rightarrow$ HCO $\rightarrow$ H_2CO $\rightarrow$ CH_3O $\rightarrow$ CH_3OH (Watanabe and Kouchi, 2002); 2) the reaction chain CH_4 + OH $\rightarrow$ CH_3 + H_2O and CH_3 + OH $\rightarrow$ CH_3OH (Hiraoka et al., 1995; Qasim et al., 2018; Lamberts et al., 2022); and 3) the formation of formaldehyde through C + H_2O and subsequent hydrogenation processes (Potapov et al., 2021; Molpeceres et al., 2021).

The most efficient formation mechanism of methanol for the high extinction conditions probed toward the lines of sight of NIR38 and J110621, is the hydrogenation of CO on grain surfaces. We note, however, that we do not exclude that the other two mechanisms of methanol formation (namely, the chain of reactions starting from CH_4 + OH and C + H_2O) are efficient at lower extinctions and densities, but that the dominant mechanism toward the densest regions in the Chamaeleon I cloud is CO hydrogenation.

One may think that the low efficiency of the reaction chain CH_4 + OH in the formation of CH_3OH ice is due to the underproduction of CH_4 in the models. However, note that, even when CH_4 is efficiently formed, as in the case of the UCLCHEM code, the CH_3OH ice formation reaction CH_3 + OH $\rightarrow$ CH_3OH contributes $\leq 1.4\,\%$ to the total CH_3OH ice produced for the best models of NIR38 and J110621.

5.4.2 Conversion of CO into CO_2

In our best models, CO_2 is overproduced as compared to the CO_2 ice abundances measured with the JWST. Note that the abundances in McClure et al. (2023) could be a factor of $\sim$1.45

higher, due to uncertainties in the determination of the CO_2 4.7 μm band strength.[6] The overproduction of CO_2 is a well-known effect that was pointed out by, e.g., Garrod and Pauly (2011) and, more recently, by Clément et al. (2023). Indeed, the formation of CO_2 ice in the ices is strongly dependent on T_{dust} and it is extremely efficient once $T_{dust} \geq 12$ K. However, note that the reaction mechanisms responsible for this CO_2 ice overproduction is different depending on the astrochemical code used.

The main possible formation routes of CO_2 in the ice are: 1) the reaction CO + OH → CO_2 + H (Ruaud et al., 2015); 2) the reaction of CO + O → CO_2 (Minissale et al., 2016); 3) the reaction H_2CO + O → CO_2 + H_2 (Minissale et al., 2015; here the H_2 is released to the gas phase); and 4) the set of reactions involving HOCO listed in Table 5.2.1. We see that for the UCLCHEM code the dominant mechanism for CO_2 ice formation is the reaction CO + OH → CO_2 + H. Additionally, when T_{dust} = 14 K we also see a contribution of H_2CO + O → CO_2 + H_2. The importance of this reaction increases with the collapse, creating approximately one third of the total CO_2 ice, with the two other thirds being produced by the reaction of CO with OH.

However, as shown in Jimenez-Serra et al. (submitted), new KMC calculations indicate that the set of reactions involving HOCO are also important, concluding that all of the mentioned formation routes should be taken into accout in future astrochemical modelling of interstellar ices. Indeed, the experimental work by Ishibashi et al. (2024) shows that the reaction CO + OH → CO_2 + H is quite improbable, since it involves an intermediate product which is indeed an excited state of the hydrocarboxil radical (HOCO*), and which is more probable to get stabilized into HOCO through energy dissipation of HOCO* to the water ice in the dust grain. This could explain why we get an overproduction of CO_2 in our UCLCHEM simulations.

5.4.3 Formation of hydrides in the ice: H_2O, NH_3 and CH_4

As shown in Section 5.3, the models tend to underproduce H_2O. The models that generate the largest amounts of water are the ones with a slow-collapse phase and with high-densities, reproducing the observed water ice abundances within a factor of 10.

Similarly, CH_4 also tends to present higher abundances in all model runs as compared to the ones measured by the Ice Age observations. As seen in Figures 5.3.3 and 5.3.4, the results from UCLCHEM reproduce within factors of 2–4 the CH_4 ice abundances observed toward NIR38 and J110621. The formation of CH_4 ice is intimately linked to the formation of CO_2 ice since both molecular species compete for the conversion of atomic C depleted onto dust grains. This conversion, however, is expected to take place at lower extinctions during the translucent cloud (phase 0) stage (Garrod and Pauly, 2011). As explained in Section 5.3.1, the amount of gas-phase atomic C available for depletion in the translucent cloud (phase

[6] The band strength used in McClure et al. (2023) was 1.1×10^{-16} cm as revisited by Bouilloud et al. (2015), but a value of 7.6×10^{-17} cm can be found in earlier works (Gerakines et al., 1995).

0) stage is roughly one order of magnitude higher in the UCLCHEM models than for the other codes. For NH_3, its formation in the ice occurs predominantly via the hydrogenation of atomic N. The models reproduce fairly well the ice abundances observed with the JWST, which differ by typically factors of a few.

We finally note that lowering the temperature from 12 K / 14 K to 10 K in our simulations has, in general, little to no effect on the resulting ice abundances of H_2O, CH_4 and NH_3.

5.4.4 Uncertainties in the determination of the extinction toward NIR38 and J110621

The visual extinction values first reported by McClure et al. (2023) toward the lines-of-sight of NIR38 and J110621 (A_V = 60 mag and A_V = 95 mag, respectively), were pre-flight estimates. A more recent analysis of the NIRSpec spectra obtained with the JWST toward these two field stars, has however yielded A_K values of 4.1 for NIR38 and 5.6 mag for J110621 (Dartois et al., 2024). If we assume A_V/A_K = 8.4, these values correspond to visual extinctions of A_V = 34 mag and A_V = 47 mag, respectively. McClure et al. (2023) used the pre-launch extinctions determined photometrically under the assumption that both background stars were both G or K giants. Instead, Dartois et al. (2024) fitted the photospheric absorption features to determine that both stars were K7 dwarfs.

The new A_V values are therefore factors of $\sim$2 lower than those reported in McClure et al. (2023) and than those used in our chemical modelling simulations. However, we note that we do no expect significant differences for these lower visual extinctions because UV photochemistry (the only one that could be affected by the lower extinction values) is inefficient at $A_V > 10$ mag. The only impact is expected for the predicted column densities of water ice reported in Figures 5.3.3 and 5.3.4, whose values would be lower by roughly a factor of $\sim$2. The ice column density ratios would not suffer any change because they are calculated directly from the ice abundances predicted by the models.

5.5 Conclusions

In this work, we have used the state-of-the-art astrochemical code UCLCHEM to model the ice abundances measured with the JWST within the Ice Age ERS program toward the highly-extinguished background stars NIR38 and J110621 in the line of sight of the Chamaeleon I dense molecular cloud (McClure et al., 2023). To do this, we have considered a small grid of physical parameters covering gas densities $n_H = 2\times10^4$, 2×10^5 and 2×10^6 cm^{-3}, dust temperatures of T_{dust} = 10, 12 and 14 K, and cloud collapse time-scales of $t_{col} = 10^4$, 10^5 and 10^6 yr. The comparison of the model predictions with the observed JWST ice abundances toward NIR38 and J110621 allow us to understand the chemical processes behind the ice formation of CH_3OH (and hence, of CO and CO_2, which are tightly related to the ice chemistry

of CH_3OH) and of hydride species such as H_2O, CH_4 and NH_3. The main conclusions of our work are listed below:

- The best match between the predicted and the observed ice abundances is found for the models with intermediate final density ($n_H = 2 \times 10^5$ cm^{-3}) and intermediate collapse models with ($t_{col} = 10^5$ yr). Low-density ($n_H = 2 \times 10^4$ cm^{-3}), fast models ($t_{col} = 10^4$ yr) are ruled out because their collapse time-scales are shorter that the depletion time-scales preventing the efficient growth of the ices. These results nicely match the densities and time-scales required by grain growth models to grow molecular cloud dust grains to sizes of 0.9 μm as measured toward the Chamaeleon I cloud.
- Our models find that CH_3OH ice formation toward the highly-extinguished lines-of-sight of NIR38 and J110621 in the Chamaeleon I cloud, predominantly occurs via the successive hydrogenation of CO. The contribution of other proposed CH_3OH formation mechanisms, such as the reaction chains starting with CH_4 + OH or with C + H_2CO, is minor for the high-extinction conditions probed toward the NIR38 and J110621 field stars.
- The CO_2 ice abundance with respect to H_2O is underproduced for J110621. For UCLCHEM, the main formation route of CO_2 is the reaction between CO and OH in dust grain surfaces. However, recent KMC simulations (Jiménez-Serra et al., submitted) reveal that other reactions involving HOCO should be considered given that they contribute equally to the formation of CO_2 ice as a result of the higher mobility of atomic O on the grain as compared to OH.
- The formation of H_2O ice tends to be underpredicted in all models, although the predicted values lie within a factor of 10 with respect to the JWST observations. Both CH_4 and NH_3 abundances with respect to H_2O tend to be overestimated although they are fairly well reproduced for the best-fit models.

Our chemical modelling does not consider the effects of grain growth and/or ice morphology (Dartois et al., 2024; Noble et al., 2024), cloud dynamics (Clément et al., 2023) or the presence of multiple ice components along the line-of-sight. Future chemical modelling will have to be carried out to assess the importance of all these effects in detail. Furthermore, spatial information on the gas-phase abundances of these major ice compounds will be obtained in the near future to complement the JWST Ice Age observations, which will help us to constrain key parameters in the modelling of the chemistry of ices toward the Chamaeleon I molecular cloud.

Chapter 6

Discussion

Through this tesis we have explored the chemistry of molecular clouds and starless/prestellar cores both in gas and solid phase, combining astronomical observations with models and tools that allow to study the level of chemical complexity in those environments.

In Chapter 3 we have proposed a hypothetical scenario of chemical evolution from static starless cores to collapsing and more chemically rich prestellar cores, as based on our observations of the cores L1517B, L1498 and L1544, with L1517B being the youngest core and L1544, the most evolved one of the three. The first one, whose recent observations are presented in this thesis (Megías et al., 2023), shows a low level of chemical complexity, with only relatively small organic molecules containing oxygen and specially nitrogen atoms, like the methoxy radical (CH_3O), ketene (H_2CCO), acetonitrile (CH_3CN) and cyanoacetilene (HC_3N). These kind of molecules would be the first ones formed in starless cores, due mainly to the depletion of atomic nitrogen, oxygen and little CO onto dust grains. COM precursors would then form on the surface of dust grains to be subsequently released into the gas phase via chemical reactive desorption. COMs would finally form later on through gas phase reactions, according to our modelling (Megías et al., 2023). In a next stage where depletion of CO is more prominent, more complex organic molecules –specially containing oxygen– would start to be formed, which is observed in the more evolved starless core L1498 (Jiménez-Serra et al., 2021), with molecules such as the tricarbon monoxide radical (C_3O) and vinyl cyanide (CH_2CHCN). When entering to the collapsing phase of a prestellar core, more complex organic species could be finally formed, such as dimethylether (CH_3OCH_3) or methyl formate (CH_3OCHO), as observed in the prestellar core L1544, as a consequence of the catastrophycal depletion of CO onto dust grains. (Jiménez-Serra et al., 2016). In contrast, this hypothetical evolution does not seem to be followed in the case of the young starless core L1521E, which also presents complex organic species such as the two last mentioned (Nagy et al., 2019; Scibelli et al., 2021). However, observations of deuterated species as still lacking, which may reveal the actual evolutionary stage of this core; i.e., if some deuteration were detected, this would imply a later evolutionary stage than originally thought, since a higher level of deuteration is associated with more evolved cores, as explained in Section

1.4 (gaseous CO destroys deuterium, so the CO depletion into dust grains enhances the chemistry of deuterated species in the gas phase). Additionally, and regarding deuteration as well, the observation of deuterated COMs would be very interesting because this would allow us to compare the level of deuteration in COMs with those measured in protoplanetary discs and in asteroids and in comet 67P/Churimov-Gerasimenko. Similar values would suggest a chemical inheritance across all these stages in the formation a Solar-type system.

The astrochemical simulations of L1517B carried out with the software MONACO show that its predictions are overall consistent with the observed abundances in gas-phase. In these models, a critical parameter is the efficiency of the process of chemical reactive desorption. Although a theoretical expression has been proposed for this parameter based on a limited number of experiments (Minissale et al., 2016), it has been argued that this process may not be as efficient as initially thought (Wakelam et al., 2017). In our models, we assume that chemical reactive desorption ejects 1% of the COMs present in the ice. However, this value may be different depending on the species and on the type of surface (this efficiency is believed to be lower in H_2O surfaces as compared to CO-rich ice surfaces; Vasyunin et al., 2017). Therefore it remains unclear how efficient this mechanism is in reality. More laboratory experiments are needed to clearly establish this matter, although the experimental determination of this parameter is very challenging (Minissale et al., 2016; Cazaux et al., 2016).

In recent years, a new trend has been observed in astrochemistry for which the formation of COMs is efficient on dust grain surfaces even at the cold temperatures of starless/prestellar cores via a new proposed mechanism: non-diffusive radical-radical grain surface reactions (Jin and Garrod, 2020; Garrod et al., 2022). This process involves that radicals are formed in-situ in the same pore of the dust grains and therefore, radical diffusion is not required for a COM to be formed. This process is much more efficient than diffusive chemistry, especially at very low temperatures, and it dominates grain surface chemistry when included in astrochemical models (Jiménez-Serra et al., submitted). This is so that large abundances of COMs are predicted to be formed on grain surfaces via this mechanism. Therefore, in order to have the full picture of the chemistry of starless and prestellar cores, and to compare it with astrochemical models, we would need to expand the astronomical observations with ice absorption spectroscopy in the infrared, as it has been done with James Webb space telescope (JWST; McClure et al., 2023). In fact, the observed large COMs in bodies of the Solar system like meteorites (Cronin et al., 1979; Cronin and Pizzarello, 1983, 1999; Glavin et al., 1999; Koga et al., 2021, 2024), the asteroid Ryugu (Oba et al., 2023) and the comet 67P/Churyumov-Gerasimenko (Altwegg et al., 2016) correspond to the solid-phase reservoir, which is inherited from the icy layers of dust grains that get processed and survive all the star-formation process, which starts in the starless and prestellar core phase. Therefore, in order to constrain and better understand this chemical inheritance between the different stages of the low-mass star formation process, we would need more observations of ice absorption spectroscopy. Even though JWST observations for the Ice Age project have targeted the molecular cloud Chamaleon I, finding CH_3OH –the simplest COM and a precursor of

larger COMs–, there were no evidence of infall motions in the targeted positions, so there should not be prestellar cores in such positions. Therefore, future JWST observations toward prototypical prestellar cores such as L1544 are needed to understand how the ice chemical complexity builds up during these initial stages in the process of a Solar-type system, and to characterize the chemical composition that will be inherited by the protostellar stages.

In order to improve our understanding of the ice chemical composition as observed with JWST, the tool AICE (*Automatic Ice Composition Estimator*) has been developed as a part of this thesis, which is explained in Chapter 4. This underlying machine-learning algorithm –based on artificial neural networks– has been trained on hundreds of laboratory experiments with different ice composition and temperature. This novel tool allows to quickly predict the ice composition and temperature corresponding to the input infrared absorption spectrum, in terms of the main ice components (H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4). However, note that the neural network model could be trained for many other species as long as infrared (IR) laboratory spectra become available. As part of this thesis work, I have carried out a short stay at the LERMA laboratory (now called LIRA) in the University of Cergy-Paris, within the group of Fancois Dulieu, who is an expert in laboratory experiments in interstellar ice analogues. One of the goals of this stay was to establish whether the IR ice aborption spectra obtained with the RAIRS technique (ice absorption measurements carried out in reflexion-absorption) are comparable to those obtained in transmission, which is what JWST observes. The results are still preliminary but show that RAIRS spectra do not show large differences with respect to transmission spectra, which would enlarge significantly the amount of available IR laboratory spectra of interstellar ices that could be used for AICE. In addition, we note that hundreds of interstellar ice absorption spectra are expected to be published in the following months and years thanks to JWST observations (like those of the Ice Age project) and therefore AICE will enable statistically relevant studies of the ice chemical composition in molecular clouds and dark cloud starless/prestellar cores.

Lastly, in Chapter 5 we present astrochemical simulations performed with the software UCLCHEM as a part of this thesis, within the context of the international Ice Age collaboration. These simulations have been performed in order to understand the formation process of methanol (CH_3OH) ice in molecular dark clouds. McClure et al. (2023) proposed that CH_3OH could be formed in a water (H_2O) -rich environment. However, the modelling presented in this thesis clearly demonstrates that CH_3OH forms through the successive hydrogenation of CO present in the icy layers of dust grains; it is worth mentioning that this scenario is compatible with either an ice that contains a pure CO layer or an ice that contains a mixture including CO where the newly adsorbed molecules of CO would get hydrogenated. The reason for this is that in the translucent cloud phase (the phase 0 in our models) atomic C rapidly gets converted into CO in the gas phase, and hence, during the collapse phase of the cloud (phase 1 in our models), there is very little atomic C left in the gas phase that can adsorb onto the surface and react with H_2O or that can produce CH_4. Because of this, the formation of CH_3OH starting from reactants such as methane (CH_4) or atomic carbon (C) are negligible in our models. This is also consistent with the analysis of the Ice Age data by

Dartois et al. (2024), which reveal that the H_2O broad wing observed in the ice absorption spectra toward longer wavelengths is not due to water-ammonia (H_2O-NH_3) mixtures but to grain growth. In our modelling, a grid of chemical models with different physical initial conditions are compared with ice observations of the Chamaleon I molecular cloud, to understand and assess the capacity of astrochemical models to explain the astronomical observations. Additionally, in the paper corresponding to these collaboration (Jiménez-Serra et al., submitted) the authors show the results of different astrochemical codes (UCLCHEM, MONACO, MAGICKAL, Nautilus, and Kinetic Monte-Carlo), where the results from the different models are compared between them and with the observations. Since each astrochemical code treats the different chemical processes slightly different, we can extract common general trends provided by all the codes and also infer the chemical processes that might have a special importance by identifying the models that best reproduce the observations. In this regard, Jiménez-Serra et al. (submitted) conclude that all codes predict that the main formation path of CH_3OH is the hydrogenation of carbon monoxide (CO) in the surfaces of dust grains. Another interesting result is the importance of non-diffusive chemistry, which should be better studied and treated in astrochemical codes. In any case, we should remind that astrochemical codes based in rate equations depend on the correct determination of many rate equations, either experimentally or computationally, but there are many of them that are subject to significant uncertainties, which can affect the results of the simulations. The Kinetic Monte-Carlo method also depends on reaction rates, although in a more indirect way. Additionally, all those methods depend on other physical-chemical constants like the binding energies, that are also susceptible of significant uncertainties. Therefore, we should bear in mind that we are trying to model complex systems, which are subject to our level of precision in our prior knowledge of the physical-chemical parameters that impact our system. In order to improve the results of this kind of simulations it is therefore essential to keep characterising those parameters, either experimentally or theoretically.

To finish this chapter, we would like to comment a few points about the possible future steps in the study of the chemistry of molecular clouds and starless/prestellar cores. As already mentioned, JWST is very well suited for observing ices in interstellar dust grains thanks to its great sensitivity and spectral resolution. Current and future observations of infrared absorption spectroscopy will surely enlarge our knowledge about the ice chemistry in different stages of the formation of a stellar system. However, to identify larger molecules such as COMs unambiguously will probably need a huge ammount of integration time, since those molecules are predicted to have quite low abundances (by theoretical considerations and by astrochemical models). With respect to other future instruments, on the one hand the ELT (Extremy Large Telescope, in Chile) could provide a way of to deepen in the interstellar ice composition through infrared absorption spectroscopy, thanks to its instrument METIS (*Mid-infrared Extremely Large Telescope Imager and Spectrograph*).[1] This instrument will provide very high resolving power ($\mathcal{R} \sim 100\,000$) spectroscopy in the range of 3–5 μm. Even

[1] https://elt.eso.org/instrument/METIS/

if it is not a very wide range, and despite the fact that telluric lines obstruct the view of actual astronomical features, its great sensitivity and resolution could help to identify subtle features of very low abundant species as COMs (e.g., the C-H stretching mode at 3.37 μm of CH_3CH_2OH). On the other hand, instruments like the ngVLA[2] (Next Generation Very Large Array, in the United States) and the SKA[3] (Square Kilometre Array, in South Africa and Australia) could produce very detailed maps of gas-phase emission of COMs towards molecular clouds and starless/prestellar cores, thanks to its excellent spatial resolution (as low as 0.08 mas (milliarcseconds) at 93 GHz for ngVLA and 1 mas at 20 GHz for SKA). High-angular resolution observations of the molecular line emission in prestellar cores will allow us to establish the small-scale structure of the gas, identifying inhomogeneities and hence the locations and chemistry of the 'seeds' of star formation (Caselli et al., 2019, 2022).

Lastly, regarding the future of machine learning applied to astrochemistry, there are a few lines of work which could be enhanced. Similarly as we have done with AICE, artificial neural networks can be used to analyse radio astronomical spectra and identify the present molecules in gas phase. This is indeed a line of work that is being explored by Kessler et al. (in prep.), from the Laboratoire d'Astrophysique de Bordeaux (LAB). Barrientos et al. (2021) also used neural networks in conjunction with the MADCUBA software to constrain the physical parameters of the molecular emission of abundant molecules such as CO, HCO^+, SiO or CH_3CN. Another use of machine learning in this field is the classification of molecular emission maps of molecular clouds or starless/prestellar cores, as it has being done by Giers et al. (in prep.), from the Max Planck Institute for Extraterrestrial Physics (MPE), with unsupervised clustering methods. Another interesting use of these algorithms is the emulation of astrochemical codes, as it is the case of Emulchem (de Mijolla et al., 2019) and Chemulator (Holdship et al., 2021). Both of them consist of neural networks trained to replicate (emulate) the behaviour and results of the astrochemical code UCLCHEM. In this way, instead of solving the set of differential equations corresponding to the chemical network of our system along time, the neural networks directly output the evolution of the chemical abundances with time, producing similar results in timescales of orders of magnitudes faster. Similarly, Molpeceres et al. (2023) used artificial neural networks to solve model interatomic potentials and perform relatively fast molecular dynamic simulations of ices in dust grains. This approach of emulation could also be used, for example, for the radiative transfer model of Dartois et al. (2024), since it is computationally quite expensive, although the generation of the training dataset would require large amounts of time. Appart from the mentioned approaches, machine learning could surely be used in several more ways, given the versatility of artificial neural networks –including deep learning– and the variety of other machine learning methods.

[2] https://ngvla.nrao.edu/
[3] https://www.skao.int/

Chapter 7

Conclusions and perspectives

This last chapter of the thesis lists in more detail the results and conclusions obtained within Chapters 3, 4 and 5, and presents several possible research lines to be studied in the future as a continuation of the work presented in this thesis.

7.1 Conclusions

We list here the main concusions of this thesis, as extracted from Chapters 3, 4 and 5, as well as some additional conclusions.

7.1.1 Chapter 3: Chemical complexity in starless and prestellar cores

This chapter presented new radio astronomical observations of complex organic molecules (COMs) measured toward the starless core L1517B. These observations were compared to those previously obtained toward the starless core L1498 and the prestellar core L1544, as well as with other observations of the starless core L1521E . Astrochemical simulations were also performed to model the COM chemistry of core L1517B using the software MONACO. The main conclusions of this work are:

- The core L1517B shows a lower level of chemical complexity than L1498 and L1544.
- The abundances of L1517B and L1521E are largely consistent, although L1521E shows a wider range of COMs.
- The chemical age inferred by the astrochemical model MONACO for L1517B is larger than for L1498 due to its high carbon monoxide (CO) depletion. However, the low level of chemical complexity found in L1517B, added to the absence of infall motions and the low level of deuteration, suggest that it is in an early evolutionary stage.
- We propose an hypothetical scenario of chemical evolution in which complexity increases with time along the process of collapse of a starless/prestellar core. In this sce-

nario, nitrogen-bearing molecules would be formed first, followed by oxigen-bearing molecules once the catastrophic depletion of CO takes place.

7.1.2 Chapter 4: Predicting the composition of astronomical ices with machine learning

This chapter presents the tool AICE (*Automatic Ice Composition Estimator*), a tool coded in Python and based in artificial neural networks that predicts the chemical composition of astronomical ices from infrared ice absorption spectra. The main conclusions are:

- The results of AICE show a relatively low error in the validation step (~3 % in fractional composition and ~11 K in temperature).
- When applied to astronomical observations, the AICE predictions are consistent with previous estimates done with other analysis tools such as ENIIGMA, which are computationally expensive.
- The tool is very fast making its predictions (~0.1 s in a typical laptop), which will allow to analyze hundreds of astronomical observations taken by James Webb Space Telescope (JWST).
- The tool can be improved to include larger species such as complex organic molecules (COMs), by enlarging the dataset of laboratory experiments.

7.1.3 Chapter 5: Astrochemical simulations of molecular clouds for the Ice Age project

This chapter presents astrochemical simulations performed using the software UCLCHEM in the context of the Ice Age collaboration, for the Chamaleon I molecular cloud. In these simulations, we compare the predicted ice abundances with the observations obtained with the James Webb space telescope (JWST) toward two highly-extinguished background stars toward this cloud. The main conclusions are:

- The best-fit models are the ones with intermediate collapse times (10^5 yr) and intermediate final densities ($2 \cdot 10^5$ cm^{-3}).
- Our models show that the main formation route of solid methanol (CH_3OH) is the hydrogenation of carbon monoxide (CO) in the surface of dust grains. Alternative routes involved methane (CH_4) and atomic carbon (C) but are not efficient at the physical conditions of the molecular cloud found toward the two background stars.
- The abundances of solid water ice (H_2O) are always underestimated, those of carbon monoxide (CO_2) are underestimated for the background star J110621, and those of CH_4 and NH_3 are relatively well reproduced.

7.1.4 Additional conclusions

Through this thesis we have seen the utility and importance of having software tools that help to process analyze astronomical data in an easy and automatic way. Indeed, several tools have been developed as a part of this thesis: the GILDAS-CLASS Automated Pipeline, that allows to use the tool CLASS –within the software GILDAS– in an automated way and process and reduce single-pointing radio astronomical spectra – note that this tool has been used also by other group in the data reduction of Yebes 40 m data (Scibelli et al., 2024); the script RADEX Line Fitter; that uses the non-LTE software RADEX to derive the physical parameters that best fit the observed gas-phase line intensities; the machine-learning tool AICE (*Automatic Ice Composition Estimator*), which offers tools to quicly estimate the main ice composition and temperature of hundreds of spectra; and the Python library RichVales, which offers a user-friendly and automated tool to work with values with associated undertainties and upper/lower limits, and to properly propagate the uncertainties even for cases of highly uncertain quantities. All these codes are free software and have been published in the Astrophysics Source Code Library (ASCL) and on GitHub.

7.2 Wrap-up

This section presents a summary (or *wrap-up*) of the main results and conclusions obtained in this thesis, highlighting the new contributions made to the field.

First of all, we have presented new astrochemical observations of a starless core (L1517B) and made a comparison with other starless and prestellar cores (L1498, L1544, L1521E). As a result, we propose an hypothetical scenario of chemical evolution between starless and prestellar cores: simple nitrogen-bearing molecules would form first, followed by oxygen-bearing molecules once the catastrophic depletion of carbon monoxide (CO) takes place, and with complexity increasing with time. However, deep observations of the chemistry of other starless cores should be performed in order to better clarify the possible evolutionary path followed by these objects. We also observe that the ratio between cyanoacetilene (HC_3N) and acetonitrile (CH_3CN) could be used as a proxy of the evolutionary stage for starless/prestellar cores, protostellar objects and comets.

Secondly, we have presented a machine learning tool (AICE) to analyse observations of interstellar ices of molecular clouds or starless/prestellar cores, which predicts the ice composition in terms of its six main components (H_2O, CO, CO_2, CH_3OH, NH_3 and CH_4). This tool is faster than other available methods, and in principle more precise in predicting the exact fraction of ice mixtures, although it does not take into account the grain size distribution and it cannot predict larger molecules such as COMs yet. Fortunately, it could be expanded and enhanced in the future to tackle these limitations.

Then, we have performed simulations with the astrochemical software UCLCHEM within the Ice Age collaboration, comparing them with astronomical observations of ices towards

a molecular cloud (Chamaleon I). As a result, we find that the models with collapse times of 0.1 Myr and final densities of $2 \cdot 10^5$ cm^{-3} are the ones that best reproduce the observations. The models also show that methanol (CH_3OH) is mainly formed in grains through the hydrogenation of CO.

Lastly, in the process of this doctoral project, there have been additional public software tools that have been developed to analyse and process scientific data and that can be useful for other researchers: the GILDAS-CLASS Automated Pipeline, RADEX Line Fitter and the RichValues Python package.

7.3 Perspectives

In this thesis, we have presented some works to try to understand the building up of the chemical complexity in starless and prestellar cores. Although we have advanced in our understanding of the formation of COMs in starless/prestellar cores, a comprehensive observational effort is still needed to understand how common chemical complexity is across multiple molecular clouds in the Galaxy – note that the cores studied in this thesis belong to the Taurus Molecular Cloud – and to obtain a more detailed view on the spatial distribution of COMs in starless/prestellar cores – both in the gas phase and in ices – on the level of chemical complexity that can be achieved across cores and molecular cloud environments. In the following, we present several lines of study that could be investigated as a future work.

7.3.1 Maps of COM emission in the prestellar core L1544

To study the spatial distribution of COMs (both O-bearing and N-bearing) in starless/prestellar cores, in this thesis we used two single-pointing spectra obtained toward two positions in the observed starless/prestellar cores. However, these observations provide limited information about asymmetries and the actual distribution of the emission of COMs belonging to different molecular families. We have current ongoing observations of L1544 with the IRAM 30 m radio telescope (P.I.: Izaskun Jiménez-Serra) to map the emission of COMs across this prestellar core. This dataset will provide the first high-sensitivity maps of the weak emission of O-bearing COMs such as acetaldehyde (CH_3CHO) and of N-bearing COMs such as vinyl cyanide (CH_2CHCN), providing essential information to constrain chemical modelling already carried out with the MONACO code. These maps will be complemented with single-pointing integrations carried out with JWST toward three background sources already observed from the ground by Goto et al. (2021), for which a detection of methanol (CH_3OH) was reported, and for which a second tentative detection of CH_3OH was also hinted. In this way, we will have a complete view of the interplay between the gas-phase and solid-phase chemistry occurring at such low temperatures in the interstellar medium.

7.3.2 Larger sample of starless cores

In this thesis we presented studies and observations of 4 starless/prestellar cores, which were located in the same molecular cloud complex in Taurus. Therefore, the chemistry may be biased by the initial conditions of the initial molecular cloud and differences may arise from different environments for other clouds. A larger sample of cores studied in detail will be carried out in the future, targeting as many COMs as possible to enhance our picture of how chemical complexity evolves from the initial conditions in molecular clouds to the whole collapse process from starless cores to prestellar cores. Indeed, current collaborations with researcher Samantha Scibelli (National Radio Astronomy Observatory, Virginia) are increasing the sample of studied cores increasing the statistics about the detection of COMs in these cold objects (Scibelli et al., 2024), although they are not yet studied with as much detail (especially chemically speaking) as L1517B, L1498 and L1544. The cores studied by Scibelli et al., 2024 are not located in the Taurus molecular cloud but in the Perseus molecular cloud, although in the work of Scibelli and Shirley, 2020 the authors did study cores in Taurus, and a subset of them (14) were studied in more detail and we know more about their physical structure and evolutionary stages (Scibelli et al., 2023).

7.3.3 Statistical sample of the ice chemical composition as observed with JWST

As already mentioned, hundreds of ice absorption spectra taken by James Webb Space Telescope (JWST) are expected to become available in the following months, as part of the Ice Age collaboration. Indeed, currently there is a submitted paper by Smith et al. which presents new data for over a hundred new positions on the Chamaleon I molecular cloud. These will also be analyzed with our tool AICE and the results will be compared with the previous published analysis to understand the accuracy of AICE when used with statistically relevant samples. Additionally, data from JWST stored in their public archives will be downloaded and explored in search for more ice absorption spectra, using the tool AICE to predict the ice composition. Lastly, AICE could be expanded to consider larger species by using more experimental spectra obtained in laboratory.

Bibliography

M. Accolla, E. Congiu, F. Dulieu, G. Manicò, H. Chaabouni, E. Matar, H. Mokrane, J. L. Lemaire, and V. Pirronello. Changes in the morphology of interstellar ice analogues after hydrogen atom exposure. *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)*, 13(17):8037, 2011.

M. Agúndez, N. Marcelino, B. Tercero, C. Cabezas, P. de Vicente, and J. Cernicharo. O-bearing complex organic molecules at the cyanopolyyne peak of TMC-1: Detection of C_2H_3CHO, C_2H_3OH, $HCOOCH_3$, and CH_3OCH_3. *Astronomy & Astrophysics*, 649:L4, 2021.

M. Agúndez, J. C. Loison, K. M. Hickson, V. Wakelam, R. Fuentetaja, C. Cabezas, N. Marcelino, B. Tercero, P. de Vicente, and J. Cernicharo. Detection of ethanol, acetone, and propanal in TMC-1 new O-bearing complex organics in cold sources. *Astronomy & Astrophysics*, 673:A34, 2023a.

M. Agúndez, N. Marcelino, B. Tercero, and J. Cernicharo. Aromatic cycles are widespread in cold clouds. *Astronomy & Astrophysics*, 677:L13, 2023b.

M. Agúndez, N. Marcelino, B. Tercero, I. Jiménez-Serra, and J. Cernicharo. Abundance and excitation of molecular anions in interstellar clouds. *Astronomy & Astrophysics*, 677:A106, 2023c.

M. Agúndez, C. Bermúdez, C. Cabezas, G. Molpeceres, Y. Endo, N. Marcelino, B. Tercero, J. C. Guillemin, P. de Vicente, and J. Cernicharo. The rich interstellar reservoir of dinitriles: Detection of malononitrile and maleonitrile in TMC-1. *Astronomy & Astrophysics*, 688:L31, 2024.

K. Altwegg, H. Balsiger, A. Bar-Nun, J. J. Berthelier, A. Bieler, P. Bochsler, C. Briois, U. Calmonte, M. R. Combi, H. Cottin, J. De Keyser, F. Dhooghe, B. Fiethe, S. A. Fuselier, S. Gasc, T. I. Gombosi, K. C. Hansen, M. Haessig, A. Jackel, E. Kopp, A. Korth, L. Le Roy, U. Mall, B. Marty, O. Mousis, T. Owen, H. Reme, M. Rubin, T. Semon, C. Y. Tzou, J. H. Waite, and P. Wurz. Prebiotic chemicals–amino acid and phosphorus–in the coma of comet 67P/Churyumov-Gerasimenko. *Science Advances*, 2(5):e1600285, 2016.

P. André. Nearby protoclusters as laboratories for understanding star formation on galactic scales. *Astrophysics and Space Science*, 281(1):51, 2002.

P. André. Interstellar filaments and star formation. *Comptes Rendus Geoscience*, 349(5):187, 2017.

P. André, J. Di Francesco, D. Ward-Thompson, S. I. Inutsuka, R. E. Pudritz, and J. E. Pineda. From filamentary networks to dense cores in molecular clouds: Toward a new paradigm

for star formation. In H. Beuther, R. S. Klessen, C. P. Dullemond, and T. Henning, editors, *Protostars and Planets VI*, page 27, 2014.

C. Arasa, M. C. van Hemert, E. F. van Dishoeck, and G. J. Kroes. Molecular Dynamics Simulations of CO_2 Formation in Interstellar Ices. *Journal of Physical Chemistry A*, 117: 7064, 2013.

A. Bacmann, V. Taquet, A. Faure, C. Kahane, and C. Ceccarelli. Detection of complex organic molecules in a prestellar core: A new challenge for astrochemical models. *Astronomy & Astrophysics*, 541:L12, 2012.

P. Baldi. Gradient descent learning algorithm overview: a general dynamical systems perspective. *IEEE Transactions on Neural Networks*, 6(1):182, 1995.

N. Balucani, C. Ceccarelli, and V. Taquet. Formation of complex organic molecules in cold objects: The role of gas-phase reactions. *Monthly Notices of the Royal Astronomical Society: Letters*, 449(1):L16, 2015.

A. Barrientos, J. Holdship, M. Solar, S. Martín, V. M. Rivilla, S. Viti, J. Mangum, N. Harada, K. Sakamoto, S. Muller, K. Tanaka, Y. Yoshimura, K. Nakanishi, R. Herrero-Illana, S. Mühle, R. Aladro, S. Aalto, C. Henkel, and P. Humire. Towards the prediction of molecular parameters from astronomical emission lines using neural networks. *Experimental Astronomy*, 52(1-2):157, 2021.

A. Belloche, K. M. Menten, C. Comito, H. S. P. Müller, P. Schilke, J. Ott, S. Thorwirth, and C. Hieret. Detection of amino acetonitrile in Sgr B2(N). *Astronomy & Astrophysics*, 492(3): 769, 2008.

A. Belloche, F. Schuller, B. Parise, P. André, J. Hatchell, J. K. Jørgensen, S. Bontemps, A. Weiß, K. M. Menten, and D. Muders. The end of star formation in Chamaeleon I?. A LABOCA census of starless and protostellar cores. *Astronomy & Astrophysics*, 527:A145, 2011.

A. Belloche, H. S. P. Müller, K. M. Menten, P. Schilke, and C. Comito. Complex organic molecules in the interstellar medium: IRAM 30 m line survey of Sagittarius B2(N) and (M). *Astronomy & Astrophysics*, 559:A47, 2013.

E. A. Bergin and W. D. Langer. Chemical evolution in preprotostellar and protostellar Cores. *The Astrophysical Journal*, 486(1):316, 1997.

E. A. Bergin and M. Tafalla. Cold dark clouds: The initial conditions for star formation. *Annual Review of Astronomy & Astrophysics*, 45(1):339, 2007.

J. B. Bergner, K. I. Öberg, R. T. Garrod, and D. M. Graninger. Complex organic molecules toward embedded low-mass protostars. *The Astrophysical Journal*, 841(2):120, 2017.

S. E. Bisschop, Fuchs, G. W., Boogert, A. C. A., van Dishoeck, E. F., and Linnartz, H. Infrared spectroscopy of hcooh in interstellar ice analogues. *Astronomy & Astrophysics*, 470(2):749, 2007.

N. Biver, D. Bockelée-Morvan, J. Boissier, R. Moreno, J. Crovisier, D. C. Lis, P. Colom, M. A. Cordiner, S. N. Milam, N. X. Roth, B. P. Bonev, N. Dello Russo, R. J. Vervack, and M. A. DiSanti. Molecular composition of comet 46P/Wirtanen from millimetre-wave spectroscopy. *Astronomy & Astrophysics*, 648:A49, 2021.

L. Bizzocchi, P. Caselli, S. Spezzano, and E. Leonardo. Deuterated methanol in the pre-stellar core l1544. *Astronomy & Astrophysics*, 569:A27, 2014.

A. A. Boogert, P. A. Gerakines, and D. C. Whittet. Observations of the icy universe. *Annual Review of Astronomy and Astrophysics*, 53:541, 2015.

A. C. A. Boogert, W. A. Schutte, F. P. Helmich, A. G. G. M. Tielens, and D. H. Wooden. Infrared observations and laboratory simulations of interstellar CH_4 and SO_2. *Astronomy & Astrophysics*, 317:929, 1997.

A. C. A. Boogert, G. A. Blake, and A. G. G. M. Tielens. High-resolution 4.7 micron Keck/NIRSPEC spectra of protostars. II. Detection of the ^{13}CO isotope in icy grain mantles. *The Astrophysical Journal*, 577(1):271, 2002.

A. C. A. Boogert, T. L. Huard, A. M. Cook, J. E. Chiar, C. Knez, L. Decin, G. A. Blake, A. G. G. M. Tielens, and E. F. van Dishoeck. Ice and dust in the quiescent medium of isolated dense cores. *The Astrophysical Journal*, 729(2):92, 2011.

A. C. A. Boogert, J. E. Chiar, C. Knez, K. I. Öberg, L. G. Mundy, Y. J. Pendleton, A. G. G. M. Tielens, and E. F. van Dishoeck. Infrared spectroscopic survey of the quiescent medium of nearby clouds. I. Ice formation and grain growth in Lupus. *The Astrophysical Journal*, 777(1):73, 2013.

S. Bottinelli, C. Ceccarelli, B. Lefloch, J. P. Williams, A. Castets, E. Caux, S. Cazaux, S. Maret, B. Parise, and A. G. G. M. Tielens. Complex molecules in the hot core of the low-mass protostar NGC 1333 IRAS 4A. *The Astrophysical Journal*, 615(1):354, 2004.

M. Bouilloud, N. Fray, Y. Bénilan, H. Cottin, M. C. Gazeau, and A. Jolly. Bibliographic review and new measurements of the infrared band strengths of pure molecules at 25 K: H_2O, CO_2, CO, CH_4, NH_3, CH_3OH, HCOOH and H_2CO. *Monthly Notices of the Royal Astronomical Society*, 451(2):2145, 2015.

P. D. Brown, S. B. Charnley, and T. J. Millar. A model of the chemistry in hot molecular cores. *Monthly Notices of the Royal Astronomical Society*, 231:409, 1988.

N. G. C. Brunken, W. R. M. Rocha, E. F. van Dishoeck, R. Gutermuth, H. Tyagi, K. Slavicinska, P. Nazari, S. T. Megeath, N. J. Evans, II, M. Narang, P. Manoj, A. E. Rubinstein, D. M. Watson, L. W. Looney, H. Linnartz, A. Caratti o Garatti, H. Beuther, H. Linz, P. Klaassen, C. A. Poteet, S. Federman, G. Anglada, P. Atnagulov, T. L. Bourke, W. J. Fischer, E. Furlan, J. Green, N. Habel, L. Hartmann, N. Karnath, M. Osorio, J. Muzerolle Page, R. Pokhrel, R. Rahatgaonkar, P. Sheehan, T. Stanke, A. M. Stutz, J. J. Tobin, L. Tychoniec, S. Wolk, and Y.-L. Yang. JWST observations of $^{13}CO_2$ ice. Tracing the chemical environment and thermal history of ices in protostellar envelopes. *Astronomy & Astrophysics*, 685:A27, 2024.

A. C. Carnall. SpectRes: A fast spectral resampling tool in Python. arXiv:1705.05165, 2017.

P. Caselli. Deuterated molecules as a probe of ionization fraction in dense interstellar clouds. *Planetary and Space Science*, 50(12):1133, 2002.

P. Caselli and C. Ceccarelli. Our astrochemical heritage. *The Astronomy and Astrophysics Review*, 20:56, 2012.

P. Caselli, C. M. Walmsley, A. Zucconi, M. Tafalla, L. Dore, and P. C. Myers. Molecular ions in L1544. I. Kinematics. *The Astrophysical Journal*, 565(1):331, 2002.

P. Caselli, L. Bizzocchi, E. Keto, O. Sipilä, M. Tafalla, L. Pagani, L. E. Kristensen, F. F. S. van der Tak, C. M. Walmsley, C. Codella, B. Nisini, Y. Aikawa, A. Faure, and E. F. van Dishoeck. NH_3 (1_0-0_0) in the pre-stellar core L1544. *The Astrophysical Journal*, 603:L1, 2017.

P. Caselli, J. E. Pineda, B. Zhao, M. C. Walmsley, E. Keto, M. Tafalla, A. Chacón-Tanarro, T. L. Bourke, R. Friesen, D. Galli, and M. Padovani. The Central 1000 au of a pre-stellar core revealed with ALMA. I. 1.3 mm continuum O observations. *The Astrophysical Journal*, 874(1):89, 2019.

P. Caselli, J. E. Pineda, O. Sipilä, B. Zhao, E. Redaelli, S. Spezzano, M. J. Maureira, F. Alves, L. Bizzocchi, T. L. Bourke, A. Chacón-Tanarro, R. Friesen, D. Galli, J. Harju, I. Jiménez-Serra, E. Keto, Z.-Y. Li, M. Padovani, A. Schmiedeke, M. Tafalla, and C. Vastel. The central 1000 au of a prestellar core revealed with ALMA. II. Almost complete freeze-out. *The Astrophysical Journal*, 929(1):13, 2022.

S. Cazaux, S. Morisset, M. Spaans, and A. Allouche. When sticking influences H_2 formation. *Astronomy & Astrophysics*, 535:A27, 2011.

S. Cazaux, M. Minissale, F. Dulieu, and S. Hocuk. Dust as interstellar catalyst. II. How chemical desorption impacts the gas. *Astronomy & Astrophysics*, 585:A55, 2016.

C. Ceccarelli, P. Caselli, E. Herbst, A. G. G. M. Tielens, and E. Caux. Extreme deuteration and hot corinos: The earliest chemical signatures of low-mass star formation. In B. Reipurth, D. Jewitt, and K. Keil, editors, *Protostars and Planets V*, page 47, 2007.

C. Ceccarelli, P. Caselli, F. Fontani, R. Neri, A. López-Sepulcre, C. Codella, S. Feng, I. Jiménez-Serra, B. Lefloch, J. E. Pineda, C. Vastel, F. Alves, R. Bachiller, N. Balucani, E. Bianchi, L. Bizzocchi, S. Bottinelli, E. Caux, A. Chacón-Tanarro, R. Choudhury, A. Coutens, F. Dulieu, C. Favre, P. Hily-Blant, J. Holdship, C. Kahane, A. Jaber Al-Edhari, J. Laas, J. Ospina, Y. Oya, L. Podio, A. Pon, A. Punanova, D. Quenard, A. Rimola, N. Sakai, I. R. Sims, S. Spezzano, V. Taquet, L. Testi, P. Theulé, P. Ugliengo, A. I. Vasyunin, S. Viti, L. Wiesenfeld, and S. Yamamoto. Seeds of life in space (SOLIS): The Organic composition diversity at 300-1000 au scale in Solar-type star-forming regions. *The Astrophysical Journal*, 850(2):176, 2017.

J. Cernicharo, N. Marcelino, E. Roueff, M. Gerin, A. Jiménez-Escobar, and G. M. M. Caro. Discovery of the methoxy radical, CH_3O, toward B1: Dust grain and gas-phase chemistry in cold dark clouds. *The Astrophysical Journal*, 759(2):L43, 2012.

J. Cernicharo, M. Agúndez, C. Cabezas, N. Marcelino, B. Tercero, J. R. Pardo, J. D. Gallego, F. Tercero, J. A. López-Pérez, and P. de Vicente. Discovery of CH_2CHCCH and detection of HCCN, HC_4N, CH_3CH_2CN, and, tentatively, CH_3CH_2CCH in TMC-1. *Astronomy & Astrophysics*, 647:L2, 2021a.

J. Cernicharo, M. Agúndez, C. Cabezas, B. Tercero, N. Marcelino, J. R. Pardo, and P. de Vicente. Pure hydrocarbon cycles in TMC-1: Discovery of ethynyl cyclopropenylidene, cyclopentadiene, and indene. *Astronomy & Astrophysics*, 649:L15, 2021b.

J. Cernicharo, B. Tercero, M. Agúndez, C. Cabezas, R. Fuentetaja, N. Marcelino, and P. de Vicente. CN and CCH derivatives of ethylene and ethane: Confirmation of the detection of CH_3CH_2CCH in TMC-1. *Astronomy & Astrophysics*, 686:A139, 2024.

A. Clément, A. Taillard, V. Wakelam, P. Gratier, J. C. Loison, E. Dartois, F. Dulieu, J. A. Noble, and M. Chabot. Astrochemical models of interstellar ices: History matters. *Astronomy & Astrophysics*, 675:A165, 2023.

J. J. Condon and S. M. Ransom. *Essential Radio Astronomy*. 2016.

E. Congiu, A. Sow, T. Nguyen, S. Baouche, and F. Dulieu. A new multi-beam apparatus for the study of surface chemistry routes to formation of complex organic molecules in space. *Review of Scientific Instruments*, 91(12):124504, 2020.

A. Coutens, B. Commerçon, and V. Wakelam. Chemical evolution during the formation of a protoplanetary disk. *Astronomy & Astrophysics*, 643:A108, 2020.

A. Crapsi, P. Caselli, C. M. Walmsley, P. C. Myers, M. Tafalla, C. W. Lee, and T. L. Bourke. Probing the evolutionary status of starless cores through N_2H^+ and N_2D^+ observations. *The Astrophysical Journal*, 619(1):379, 2005.

J. R. Cronin and S. Pizzarello. Amino acids in meteorites. *Advances in Space Research*, 3(9):5, 1983.

J. R. Cronin and S. Pizzarello. Amino acid enantiomer excesses in meteorites: Origin and significance. *Advances in Space Research*, 23(2):293, 1999.

J. R. Cronin, S. Pizzarello, and C. B. Moore. Amino Acids in an Antarctic Carbonaceous Chondrite. *Science*, 206(4416):335, 1979.

H. Cuppen and E. Herbst. Simulation of the formation and morphology of ice mantles on interstellar grains. *The Astrophysical Journal*, 668(1):294, 2007.

A. Dalgarno and J. H. Black. Molecule formation in the interstellar gas. *Reports on Progress in Physics*, 39(6):573, 1976.

E. Dartois, J. A. Noble, N. Ysard, K. Demyk, and M. Chabot. Influence of grain growth on CO_2 ice spectroscopic profiles. Modelling for dense cores and disks. *Astronomy & Astrophysics*, 666:A153, 2022.

E. Dartois, J. A. Noble, P. Caselli, H. J. Fraser, I. Jiménez-Serra, B. Maté, M. K. McClure, G. J. Melnick, Y. J. Pendleton, T. Shimonishi, Z. L. Smith, J. A. Sturm, A. Taillard, V. Wakelam, A. C. A. Boogert, M. N. Drozdovskaya, J. Erkal, D. Harsono, V. J. Herrero, S. Ioppolo, H. Linnartz, B. A. McGuire, G. Perotti, D. Qasim, and W. R. M. Rocha. Spectroscopic sizing of interstellar icy grains with JWST. *Nature Astronomy*, 8(3):359, Mar 2024.

A. L. F. de Barros, E. F. da Silveira, S. Pilling, A. Domaracka, H. Rothard, and P. Boduch. Processing of low carbon content interstellar ice analogues by cosmic rays: implications for the chemistry around oxygen-rich stars. *Monthly Notices of the Royal Astronomical Society*, 438(3):2026, 2014.

D. de Mijolla, S. Viti, J. Holdship, I. Manolopoulou, and J. Yates. Incorporating astrochemistry into molecular line modelling via emulation. *Astronomy & Astrophysics*, 630:A117, 2019.

M. De Simone, C. Ceccarelli, C. Codella, B. E. Svoboda, C. Chandler, M. Bouvier, S. Yamamoto, N. Sakai, P. Caselli, C. Favre, L. Loinard, B. Lefloch, H. B. Liu, A. López-Sepulcre, J. E. Pineda, V. Taquet, and L. Testi. Hot corinos chemical diversity: Myth or reality? *The Astrophysical Journal Letter*, 896(1):L3, 2020.

C. Dominik, M. Min, and R. Tazaki. OpTool: Command-line driven tool for creating complex dust opacities. Astrophysics Source Code Library, 2021.

B. T. Draine. Interstellar dust grains. *Annual Review of Astronomy & Astrophysics*, 41:241, 2003.

B. T. Draine. Interstellar Dust. In *Origin and Evolution of the Elements*, page 317, 2004.

B. T. Draine. *Physics of the interstellar and intergalactic medium*. 2011.

B. T. Draine and A. Li. Infrared emission from interstellar dust. IV. The silicate-graphite-PAH model in the post-Spitzer era. *The Astrophysical Journal*, 657(2):810, 2007.

M. N. Drozdovskaya, E. F. van Dishoeck, M. Rubin, J. K. Jørgensen, and K. Altwegg. Ingredients for solar-like systems: protostar IRAS 16293-2422 B versus comet 67P/Churyumov-Gerasimenko. *Monthly Notices of the Royal Astronomical Society*, 490(1): 50, 2019.

M. N. Drozdovskaya, I. R. H. G. Schroeder I, M. Rubin, K. Altwegg, E. F. van Dishoeck, B. M. Kulterer, J. De Keyser, S. A. Fuselier, and M. Combi. Prestellar grain-surface origins of deuterated methanol in comet 67P/Churyumov-Gerasimenko. *Monthly Notices of the Royal Astronomical Society*, 500(4):4901, 2021.

F. Dulieu, L. Amiaud, E. Congiu, J. H. Fillion, E. Matar, A. Momeni, V. Pirronello, and J. L. Lemaire. Experimental evidence for water formation on interstellar dust grains by hydrogen and oxygen atoms. *Astronomy & Astrophysics*, 512:A30, 2010.

P. Ehrenfreund, A. C. A. Boogert, P. A. Gerakines, A. G. G. M. Tielens, and E. F. van Dishoeck. Infrared spectroscopy of interstellar apolar ice analogs. *Astronomy & Astrophysics*, 328:649, 1997.

P. Ehrenfreund, O. Kerkhof, W. A. Schutte, A. C. A. Boogert, P. A. Gerakines, E. Dartois, L. D'Hendecourt, A. G. G. M. Tielens, E. F. van Dishoeck, and D. C. B. Whittet. Laboratory studies of thermally processed H_2O-CH_3OH-CO_2 ice mixtures and their astrophysical implications. *Astronomy & Astrophysics*, 350:240, 1999.

C. P. Endres, S. Schlemmer, P. Schilke, J. Stutzki, and H. S. Müller. The Cologne Database for Molecular Spectroscopy, CDMS, in the Virtual Atomic and Molecular Data Centre, VAMDC. *Journal of Molecular Spectroscopy*, 327:95, 2016.

A. Faure, F. Lique, and L. Wiesenfeld. Collisional excitation of HC_3N by para- and ortho-H_2. *Monthly Notices of the Royal Astronomical Society*, 460(2):2103, 2016.

H. J. Fraser and E. F. van Dishoeck. SURFRESIDE: A novel experiment to study surface chemistry under interstellar and protostellar conditions. *Advances in Space Research*, 33(1): 14, 2004.

K. Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20(3):121, 1975.

P. A. B. Galli, L. Loinard, H. Bouy, L. Sarro, G. N. Ortiz-León, S. A. Dzib, J. Olivares, M. Heyer, J. Hernandez, C. R.-Z. niga, M. Kounkel, and K. Covey. Structure and kinematics of the Taurus star-forming region from Gaia-DR2 and VLBI astrometry. *Astronomy & Astrophysics*, 630:A137, 2019.

Ó. Gálvez, B. Maté, V. J. Herrero, and R. Escribano. Spectroscopic Effects in CH_4/H_2O Ices. *The Astrophysical Journal*, 703(2):2101, 2009.

R. T. Garrod. A new modified-rate approach for gas-grain chemical simulations. *The Astrophysical Journal*, 491(1):239, 2008.

R. T. Garrod. Simulations of ice chemistry in cometary nuclei. *The Astrophysical Journal*, 884 (1):69, 2019.

R. T. Garrod and T. Pauly. On the formation of CO_2 and other interstellar ices. *The Astrophysical Journal*, 735(1):15, 2011.

R. T. Garrod, I. H. Park, P. Caselli, and E. Herbst. Are gas-phase models of interstellar chemistry tenable? The case of methanol. *Faraday Discuss.*, 133:51, 2006.

R. T. Garrod, V. Wakelam, and E. Herbst. Non-thermal desorption from interstellar dust grains via exothermic surface reactions. *Astronomy & Astrophysics*, 467(3):1103, 2007.

R. T. Garrod, S. L. W. Weaver, and E. Herbst. Complex chemistry in star-forming regions: An expanded gas-grain warm-up chemical model. *The Astrophysical Journal*, 682(1):283, 2008.

R. T. Garrod, M. Jin, K. A. Matis, D. Jones, E. R. Willis, and E. Herbst. Formation of complex organic molecules in hot molecular cores through non-diffusive grain-surface and ice-mantle chemistry. *The Astrophysical Journal*, 259(1):1, 2022.

T. R. Geballe, F. Baas, J. M. Greenberg, and W. Schutte. New infrared absorption features due to solid phase molecules containing sulfur in W 33A. *Astronomy & Astrophysics*, 146: L6, 1985.

P. A. Gerakines and R. L. Hudson. A modified algorithm and open-source computational package for the determination of infrared optical constants relevant to astrophysics. *The Astrophysical Journal*, 901(1):52, 2020.

P. A. Gerakines, W. A. Schutte, J. M. Greenberg, and E. F. van Dishoeck. The infrared band strengths of H_2O, CO and CO_2 in laboratory simulations of astrophysical ice mixtures. *Astronomy & Astrophysics*, 296:810, 1995.

P. A. Gerakines, W. A. Schutte, and P. Ehrenfreund. Ultraviolet processing of interstellar ice analogs. I. Pure ices. *Astronomy & Astrophysics*, 312:289, 1996.

F. C. Gillett and W. J. Forrest. Spectra of the Becklin-Neugebauer point source and the Kleinmann-Low nebula from 2.8 to 13.5 microns. *The Astrophysical Journal*, 179:483, 1973.

D. P. Glavin, J. L. Bada, K. L. F. Brinton, and G. D. McDonald. Amino acids in the Martian meteorite Nakhla. *Proceedings of the National Academy of Science*, 96(16):8835, 1999.

P. F. Goldsmith and W. D. Langer. Population diagram analysis of molecular line emission. *The Astrophysical Journal*, 517(1):209, 1999.

M. Goto, J. D. Bailey, S. Hocuk, P. Caselli, G. B. Esplugues, S. Cazaux, and M. Spaans. The first frost in the Pipe nebula. *Astronomy & Astrophysics*, 610:A9, 2018.

M. Goto, A. I. Vasyunin, B. M. Giuliano, I. Jiménez-Serra, P. Caselli, C. G. Román-Zúñiga, and J. Alves. Water and methanol ice in L 1544. *The Astrophysical Journal*, 651:A53, 2021.

T. P. M. Goumans and S. Andersson. Tunnelling in the O + CO reaction. *Monthly Notices of the Royal Astronomical Society*, 406(4):2213, 2010.

T. P. M. Goumans, M. A. Uppal, and W. A. Brown. Formation of CO2 on a carbonaceous surface: a quantum chemical study. *Monthly Notices of the Royal Astronomical Society*, 384 (3):1158, 2008.

T. Grassi, S. Bovino, D. R. G. Schleicher, J. Prieto, D. Seifried, E. Simoncini, and F. A. Gianturco. KROME - a package to embed chemistry in astrophysical simulations. *Monthly Notices of the Royal Astronomical Society*, 439(3):2386, 2014.

A. Greve, Kramer, C., and Wild, W. The beam pattern of the iram 30-m telescope - (a reflector with several surface error distributions). *Astronomy & Astrophysics Supplement Series*, 133(2):271, 1998.

C. Grosan and A. Abraham. *Artificial Neural Networks*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

T. Hama and N. Watanabe. Surface processes on interstellar amorphous solid water: Adsorption, diffusion, tunneling reactions, and nuclear-spin conversion. *Chemical Reviews*, 113(12):8783, 2013.

T. I. Hasegawa and E. Herbst. Three-phase chemical models of dense interstellar clouds: Gas, dust particle mantles and dust particle surfaces. *Monthly Notices of the Royal Astronomical Society*, 263(3):589, 1993.

T. I. Hasegawa, E. Herbst, and C. M. Leung. Models of Gas-Grain Chemistry in Dense Interstellar Clouds with Complex Organic Molecules. *The Astrophysical Journal Supplement*, 82:167, 1992.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR*, page 1, 2016.

E. Herbst and R. T. Garrod. Synthetic approaches to complex organic molecules (COMs) in the cold interstellar medium. *Frontiers in Astronomy and Space Sciences*, 8:209, 2022.

E. Herbst and W. Klemperer. The formation and fepletion of molecules in dense interstellar clouds. *The Astrophysical Journal*, 185:505, 1973.

E. Herbst and E. F. van Dishoeck. Complex organic interstellar molecules. *Annual Review of Astronomy and Astrophysics*, 47(1):427, 2009.

K. Hiraoka, A. Yamashita, Y. Yachi, K. Aruga, T. Sato, and H. Muto. Ammonia formation from the reactions of H atoms with N atoms trapped in a solid N_2 matrix at 10–30 K. *The Astrophysical Journal*, 443:363, 1995.

S. Hocuk, L. Szűcs, P. Caselli, S. Cazaux, M. Spaans, and G. B. Esplugues. Parameterizing the interstellar dust temperature. *Astronomy & Astrophysics*, 604:A58, 2017.

J. Holdship, S. Viti, I. Jiménez-Serra, A. Makrymallis, and F. Priestley. UCLCHEM: A gas-grain chemical Code for clouds, cores, and C-shocks. *The Astronomical Journal*, 154(1): 38, 2017.

J. Holdship, S. Viti, C. Codella, J. Rawlings, I. Jiménez-Serra, Y. Ayalew, J. Curtis, A. H., J. Lawrence, S. Warsame, and S. Horn. Observations of CH_3OH and CH_3CHO in a sample of protostellar outflow sources. *The Astrophysical Journal*, 880(2):138, 2019.

J. Holdship, S. Viti, T. J. Haworth, and J. D. Ilee. Chemulator: Fast, accurate thermochemistry for dynamical models through emulation. *Astronomy & Astrophysics*, 653:A76, 2021.

D. J. Hollenbach and A. G. G. M. Tielens. Photodissociation regions in the interstellar medium of galaxies. *Reviews of Modern Physics*, 71(1):173, 1999.

J. M. Hollis, F. J. Lovas, and P. R. Jewell. Interstellar glycolaldehyde: The first sugar. *The Astrophysical Journal*, 540(2):L107, 2000.

J. M. Hollis, F. J. Lovas, A. J. Remijan, P. R. Jewell, V. V. Ilyushin, and I. Kleiner. Detection of acetamide (CH_3CONH_2): The largest interstellar molecule with a peptide bond. *The Astrophysical Journal*, 643(1):L25, 2006.

D. M. Hudgins, S. A. Sandford, L. J. Allamandola, and A. G. G. M. Tielens. Mid- and far-infrared spectroscopy of ices: Optical constants and integrated absorbances. *The Astrophysical Journal*, 86:713, 1993.

R. L. Hudson, P. A. Gerakines, and R. F. Ferrante. IR spectra and properties of solid acetone, an interstellar and cometary molecule. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 193:33, 2018.

J. D. Ilee, C. Walsh, A. S. Booth, Y. Aikawa, S. M. Andrews, J. Bae, E. A. Bergin, J. B. Bergner, A. D. Bosman, G. Cataldi, L. I. Cleeves, I. Czekala, V. V. Guzmán, J. Huang, C. J. Law, R. L. Gal, R. A. Loomis, F. Ménard, H. Nomura, K. I. Öberg, C. Qi, K. R. Schwarz, R. Teague, T. Tsukagoshi, D. J. Wilner, Y. Yamato, and K. Zhang. Molecules with ALMA at Planet-forming Scales (MAPS). IX. Distribution and properties of the large organic molecules HC_3N, CH_3CN, and c-C_3H_2. *The Astrophysical Journal Supplement Series*, 257(1): 9, 2021.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, page 448, Lille, France, 2015. PMLR.

A. Ishibashi, G. Molpeceres, H. Hidaka, Y. Oba, T. Lamberts, and N. Watanabe. Proposed importance of HOCO chemistry: Inefficient formation of CO_2 from CO and OH reactions on ice dust. *The Astrophysical Journal*, 976(2):162, 2024.

A. Jiménez-Escobar, A. Ciaravella, C. Cecchi-Pestellini, C. H. Huang, N. E. Sie, Y. J. Chen, and G. M. Muñoz Caro. X-Ray Photo-desorption of H_2O:CO:NH_3 Circumstellar Ice Analogs: Gas-phase Enrichment. *The Astrophysical Journal*, 868(1):73, 2018.

I. Jiménez-Serra, J. Martín-Pintado, A. Rodríguez-Franco, and S. Martín. Grain evolution across the shocks in the L1448-mm outflow. *The Astrophysical Journal Letters*, 627(2):L121, 2005.

I. Jiménez-Serra, Q. Zhang, S. Viti, J. Martín-Pintado, and W. J. de Wit. Chemical Segregation toward Massive Hot Cores: The AFGL2591 Star-forming Region. *The Astrophysical Journal*, 753(1):34, 2012.

I. Jiménez-Serra, A. I. Vasyunin, P. Caselli, N. Marcelino, N. Billot, S. Viti, L. Testi, C. Vastel, B. Lefloch, and R. Bachiller. The spatial distribution of complex organic molecules in the L1544 pre-stellar core. *The Astrophysical Journal*, 830(1):L6, 2016.

I. Jiménez-Serra, S. Viti, D. Quénard, and J. Holdship. The chemistry of phosphorus-bearing molecules under energetic phenomena. *The Astrophysical Journal*, 862(2):128, 2018.

I. Jiménez-Serra, A. I. Vasyunin, S. Spezzano, P. Caselli, G. Cosentino, and S. Viti. The complex organic molecular content in the L1498 starless core. *The Astrophysical Journal*, 917(1):44, 2021.

M. Jin and R. T. Garrod. Formation of complex organic molecules in cold interstellar environments through nondiffusive grain-surface and ice-mantle chemistry. *The Astrophysical Journal Supplement Series*, 249(2):26, 2020.

J. K. Jørgensen, R. Visser, N. Sakai, E. A. Bergin, C. Brinch, D. Harsono, J. E. Lindberg, E. F. van Dishoeck, S. Yamamoto, S. E. Bisschop, and M. V. Persson. A recent accretion burst in the low-mass protostar IRAS 15398-3359: ALMA imaging of its related chemistry. *The Astrophysical Journal*, 779(2):L22, 2013.

C. Kahane, M. A. Frerking, W. D. Langer, P. Encrenaz, and R. Lucas. Measurement of the formaldehyde ortho to para ratio in three molecuar clouds. *Astronomy & Astrophysics*, 137:211, 1984.

T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7, 1989.

E. Keto and P. Caselli. Dynamics and depletion in thermally supercritical starless cores. *Monthly Notices of the Royal Astronomical Society*, 402(3):1634, 2010.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. art. arXiv:1412.6980, 2014.

T. Koga, E. T. Parker, H. L. MCLain, J. C. Aponte, J. E. Elsila, J. P. Dworkin, D. P. Glavin, H. Naraoka, S. Sandford, and Sandford. Extraterrestrial hydroxy amino acids in CM and CR carbonaceous chondrites. *Meteoritics & Planetary Science*, 56(5):1005, 2021.

T. Koga, Y. Takano, Y. Oba, H. Naraoka, and N. Ohkouchi. Abundant extraterrestrial purine nucleobases in the Murchison meteorite: Implications for a unified mechanism for purine synthesis in carbonaceous chondrite parent bodies. *Geochimica et Cosmochimica Acta*, 365: 253, 2024.

S. Kong, P. Caselli, J. C. Tan, V. Wakelam, and O. Sipilä. The deuterium fractionation timescale in dense cloud cores: A parameter space exploration. *The Astrophysical Journal*, 804(2):98, 2015.

L. E. Kristensen and M. M. Dunham. Protostellar half-life: New methodology and estimates. *Astronomy & Astrophysics*, 618:A158, 2018.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

F. Kruczkiewicz, J. Vitorino, E. Congiu, P. Theulé, and F. Dulieu. Ammonia snow lines and ammonium salts desorption. *Astronomy & Astrophysics*, 652:A29, 2021.

F. Kruczkiewicz, F. Dulieu, A. V. Ivlev, P. Caselli, B. M. Giuliano, C. Ceccarelli, and P. Theulé. Comprehensive laboratory constraints on thermal desorption of interstellar ice analogues. *Astronomy & Astrophysics*, 686:A236, 2024.

J. H. Lacy, F. Baas, L. J. Allamandola, S. E. Persson, P. J. McGregor, C. J. Lonsdale, T. R. Geballe, and C. E. P. van de Bult. 4.6 micron absorption features due to solid phase CO and cyano group molecules toward compact infrared sources. *The Astrophysical Journal*, 276:533, 1984.

J. H. Lacy, C. Sneden, H. Kim, and D. T. Jaffe. H_2, CO, and dust absorption through cold molecular clouds. *The Astrophysical Journal*, 838(1):66, 2017.

T. Lamberts, G. Fedoseev, M. C. van Hemert, D. Qasim, K.-J. Chuang, J. C. Santos, and H. Linnartz. Methane formation in cold regions from carbon atoms and molecular hydrogen. *The Astrophysical Journal*, 928(1):48, 2022.

H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin. Exploring strategies for training deep neural networks. *Journal of Machine Learning Research*, 10(1):1, 2009.

Y. LeCun, L. Bottou, G. B. Orr, and K. R. Müller. Efficient backprop. In *Neural Networks: Tricks of the Trade*, page 9. Springer, Berlin, Heidelberg, 1998.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

X. Liu. *Organic Chemistry I.* Kwantlen Polytechnic University, 2021.

Á. López-Gallifa, V. M. Rivilla, M. T. Beltrán, L. Colzi, C. Mininni, Á. Sánchez-Monge, F. Fontani, S. Viti, I. Jiménez-Serra, L. Testi, R. Cesaroni, and A. Lorenzani. The GUAPOS project - V: The chemical ingredients of a massive stellar protocluster in the making. *Monthly Notices of the Royal Astronomical Society*, 529(4):3244, 2024.

K. L. Luhman. The formation and early evolution of low-mass stars and brown dwarfs. *Annual Review of Astronomy and Astrophysics*, 50:65, 2012.

M.-M. Mac Low and R. S. Klessen. Control of star formation by supersonic turbulence. *Reviews of Modern Physics*, 76(1):125, 2004.

J. G. Mangum and Y. L. Shirley. How to calculate molecular column density. *Publications of the Astronomical Society of the Pacific*, 127(949):266, 2015.

N. Marcelino, J. Cernicharo, M. Agúndez, E. Roueff, M. Gerin, J. Martín-Pintado, R. Mauersberger, and C. Thum. Discovery of interstellar propylene (CH_2CHCH_3): Missing links in interstellar gas-phase chemistry. *The Astrophysical Journal*, 665(2):L127, 2007.

S. Martín, J. Martín-Pintado, C. Blanco-Sánchez, V. M. Rivilla, A. Rodríguez-Franco, and F. Rico-Villas. Spectral Line Identification and Modelling (SLIM) in the MAdrid Data CUBe Analysis (MADCUBA) package - Interactive software for data cube analysis. *Astronomy & Astrophysics*, 631:A159, 2019.

T. G. Mayerhöfer, S. Pahlow, and J. Popp. The Bouguer-Beer-Lambert law: Shining light on the obscure. *Chemphyschem*, 21(18):2029, 2020.

M. K. McClure, W. R. M. Rocha, K. M. Pontoppidan, N. Crouzet, L. E. U. Chu, E. Dartois, T. Lamberts, J. A. Noble, Y. J. Pendleton, G. Perotti, D. Qasim, M. G. Rachid, Z. L. Smith, F. Sun, T. L. Beck, A. C. A. Boogert, W. A. Brown, P. Caselli, S. B. Charnley, H. M. Cuppen, H. Dickinson, M. N. Drozdovskaya, E. Egami, J. Erkal, H. Fraser, R. T. Garrod, D. Harsono, S. Ioppolo, I. Jiménez-Serra, M. Jin, J. K. Jørgensen, L. E. Kristensen, D. C. Lis, M. R. S. McCoustra, B. A. McGuire, G. J. Melnick, K. I. Öberg, M. E. Palumbo, T. Shimonishi, J. A. Sturm, E. F. van Dishoeck, and H. Linnartz. An Ice Age JWST inventory of dense molecular cloud ices. *Nature Astronomy*, 7(4):431, Apr 2023.

D. McElroy, C. Walsh, A. J. Markwick, M. A. Cordiner, K. Smith, and T. J. Millar. The UMIST database for astrochemistry 2012. *Astronomy & Astrophysics*, 550:A36, 2013.

M. W. McElwain, L. D. Feinberg, M. D. Perrin, M. Clampin, C. M. Mountain, M. D. Lallo, C.-P. Lajoie, R. A. Kimble, C. W. Bowers, C. C. Stark, D. S. Acton, C. Atkinson, B. Barinek, A. Barto, S. Basinger, T. Beck, M. D. Bergkoetter, M. Bluth, R. A. Boucarut, G. R. Brady, K. J. Brooks, B. Brown, J. Byard, L. Carey, M. Carrasquilla, D. Chae, D. Chaney, P. Chayer, T. Chonis, L. Cohen, H. J. Cole, T. M. Comeau, M. Coon, E. Coppock, L. Coyle, B. H. Dean, K. J. Dziak, M. Eisenhower, N. Flagey, R. Franck, B. Gallagher, L. Gilman, T. Glassman, J. J. Green, J. Grieco, S. Haase, T. J. Hadjimichael, J. G. Hagopian, W. G. Hahn, G. F. Hartig, K. A. Havey, W. L. Hayden, R. Hellekson, B. Hicks, S. T. Holfeltz, J. M. Howard, J. A. Huguet, B. Jahne, L. A. Johnson, J. D. Johnston, A. S. Jurling, J. R. Kegley, S. Kennard, R. A. Keski-Kuha, J. S. Knight, B. A. Kulp, J. S. Levi, M. B. Levine, P. Lightsey, R. A. Luetgens, J. C. Mather, G. W. Matthews, A. G. McKay, K. I. Mehalick, M. Meléndez, G. E. Mosier, J. Murphy, E. P. Nelan, M. B. Niedner, D. M. Nol, C. M. Ohara, R. G. Ohl, E. Olczak, S. B. Osborne, S. Park, C. Perrygo, L. Pueyo, D. C. Redding, M. W. Regan, P. Reynolds, R. Rifelli, J. R. Rigby, D. Sabatke, B. N. Saif, T. R. Scorse, B.-J. Seo, F. Shi, N. Sigrist, K. Smith, J. S. Smith, E. C. Smith, S. T. Sohn, H. P. Stahl, R. Telfer, T. Terlecki, S. C. Texter, D. Van Buren, J. M. Van Campen, B. Vila, M. F. Voyton, M. Waldman, C. B. Walker, N. Weiser, C. Wells, G. West, T. L. Whitman, E. Wolf, and T. P. Zielinski. The James Webb space telescope mission: Optical telescope element design, eevelopment, and performance. *Publications of the Astronomical Society of the Pacific*, 135 (1047):058001, 2023. ISSN 1538.

B. A. McGuire, A. M. Burkhardt, C. N. Shingledecker, S. V. Kalenskii, E. Herbst, A. J. Remijan, and M. C. McCarthy. Detection of interstellar HC_5O in TMC-1 with the Green Bank Telescope. *The Astrophysical Journal Letters*, 843(2):L28, 2017.

B. A. McGuire, A. M. Burkhardt, S. Kalenskii, C. N. Shingledecker, A. J. Remijan, E. Herbst, and M. C. McCarthy. Detection of the aromatic molecule benzonitrile (c-C_6H_5CN) in the interstellar medium. *Science*, 359(6372):202, 2018.

B. A. McGuire, A. M. Burkhardt, R. A. Loomis, C. N. Shingledecker, K. L. Kelvin Lee, S. B. Charnley, M. A. Cordiner, E. Herbst, S. Kalenskii, E. Momjian, E. R. Willis, C. Xue, A. J. Remijan, and M. C. McCarthy. Early science from GOTHAM: Project overview, methods,

and the detection of interstellar propargyl cyanide ($HCCCH_2CN$) in TMC-1. *The Astrophysical Journal Letters*, 900(1):L10, 2020.

A. Megías, I. Jiménez-Serra, J. Martín-Pintado, A. I. Vasyunin, S. Spezzano, P. Caselli, G. Cosentino, and S. Viti. The complex organic molecular content in the L1517B starless core. *Monthly Notices of the Royal Astronomical Society*, 519(2):1601, 2023.

T. J. Millar, C. Walsh, M. A. Cordiner, R. Ní Chuimín, and E. Herbst. Hydrocarbon anions in interstellar clouds and circumstellar envelopes. *The Astrophysical Journal Letters*, 662(2): L87, 2007.

T. J. Millar, C. Walsh, M. Van de Sande, and A. J. Markwick. The UMIST Database for Astrochemistry 2022. *Astronomy & Astrophysics*, 682:A109, 2024.

M. Minissale, J.-C. Loison, S. Baouche, H. Chaabouni, E. Congiu, and F. Dulieu. Solid-state formation of CO_2 via the H_2CO + O reaction. *Astronomy & Astrophysics*, 577:A2, 2015.

M. Minissale, F. Dulieu, S. Cazaux, and S. Hocuk. Dust as interstellar catalyst. I. Quantifying the chemical desorption process. *Astronomy & Astrophysics*, 585:A24, 2016.

M. Minissale, A. Moudens, S. Baouche, H. Chaabouni, and F. Dulieu. Hydrogenation of CO-bearing species on grains: Unexpected chemical desorption of CO. *Monthly Notices of the Royal Astronomical Society*, 458(3):2953, 2016.

P. Modica and M. E. Palumbo. Formation of methyl formate after cosmic ion irradiation of icy grain mantles. *Astronomy & Astrophysics*, 519:A22, 2010.

G. Molpeceres, J. Kästner, G. Fedoseev, D. Qasim, R. Schömig, H. Linnartz, and T. Lamberts. Carbon atom reactivity with amorphous solid water: H_2o-catalyzed formation of $h_2$co. *The Journal of Physical Chemistry Letters*, 12(44):10854, 2021.

G. Molpeceres, J. Enrique-Romero, and Y. Aikawa. Cracking the puzzle of CO_2 formation on interstellar ices. *Astronomy & Astrophysics*, 677:A39, 2023.

G. Molpeceres, V. Zaverkin, K. Furuya, Y. Aikawa, and J. Kästner. Reaction dynamics on amorphous solid water surfaces using interatomic machine-learned potentials. Microscopic energy partition revealed from the P + H $\rightarrow$ PH reaction. *Astronomy & Astrophysics*, 673:A51, 2023.

M. H. Moore and R. L. Hudson. Infrared spectra of organic molecules in water ice mixtures relevant to comets and the ISM. In *Division for Planetary Sciences Meeting Abstracts*, volume 29 of *Division for Planetary Sciences Meeting Abstracts*, page 29.17, 1997.

M. H. Moore, R. F. Ferrante, R. L. Hudson, and J. N. Stones. Ammonia water ice laboratory studies relevant to outer Solar System surfaces. *Icarus*, 190(1):260, 2007.

G. M. Muñoz Caro and R. Escribano, editors. *Laboratory Astrophysics*, volume 451 of *Astrophysics and Space Science Library*. 2018.

F. Murtagh. Multilayer perceptrons for classification and regression. *Neurocomputing*, 2(5): 183, 1991.

P. C. Myers. Filamentary structure of star-forming complexes. *The Astrophysical Journal*, 700 (2):1609, 2009.

Z. Nagy, S. Spezzano, P. Caselli, A. Vasyunin, M. Tafalla, L. Bizzocchi, D. Prudenzano, and E. Redaelli. The chemical structure of the very young starless core L1521E. *Astronomy & Astrophysics*, 630:A136, 2019.

V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML 2010*, page 807, 2010.

J. A. Noble, H. J. Fraser, Y. Aikawa, K. M. Pontoppidan, and I. Sakon. A Survey of H_2O, CO_2, and CO ice features toward background stars and low-mass young stellar objects using AKARI. *The Astrophysical Journal*, 775(2):85, 2013.

J. A. Noble, E. Michoulier, C. Aupetit, and J. Mascetti. Influence of ice structure on the soft UV photochemistry of PAHs embedded in solid water. *Astronomy & Astrophysics*, 644: A22, 2020.

J. A. Noble, H. J. Fraser, Z. L. Smith, E. Dartois, A. C. A. Boogert, H. M. Cuppen, H. J. Dickinson, F. Dulieu, E. Egami, J. Erkal, B. M. Giuliano, B. Husquinet, T. Lamberts, B. Maté, M. K. McClure, M. E. Palumbo, T. Shimonishi, F. Sun, J. B. Bergner, W. A. Brown, P. Caselli, E. Congiu, M. N. Drozdovskaya, V. J. Herrero, S. Ioppolo, I. Jimenez-Serra, H. Linnartz, G. J. Melnick, B. A. McGuire, K. I. Oberg, G. Perotti, D. Qasim, W. R. M. Rocha, and R. G. Urso. Detection of the elusive dangling OH ice features at 2.7 μm in Chamaeleon I with JWST NIRCam. *Nature Astronomy*, 8:1169, 2024.

Y. Oba, Y. Takano, J. P. Dworkin, and H. Naraoka. Ryugu asteroid sample return provides a natural laboratory for primordial chemical evolution. *Nature Communications*, 14:3107, 2023.

K. I. Öberg and E. A. Bergin. Astrochemistry and compositions of planetary systems. *Physics Reports*, 893:1, 2021.

K. I. Öberg, H. J. Fraser, A. C. A. Boogert, S. E. Bisschop, G. W. Fuchs, E. F. van Dishoeck, and H. Linnartz. Effects of CO_2 on H_2O band profiles and band strengths in mixed H_2O:CO_2 ices. *Astronomy & Astrophysics*, 462(3):1187, 2007.

K. I. Öberg, S. Bottinelli, J. K. Jørgensen, and E. F. van Dishoeck. A cold complex chemistry toward the low-mass protostar B1-b: Evidence for complex molecule production in ices. *The Astrophysical Journal*, 716(1):825, 2010.

K. I. Öberg, V. V. Guzmán, K. Furuya, C. Qi, Y. Aikawa, S. M. Andrews, R. Loomis, and D. J. Wilner. The comet-like composition of a protoplanetary disk as revealed by complex cyanides. *Nature*, 520(7546):198, 2015.

M. Ohishi, K. Kawaguchi, N. Kaifu, W. M. Irvine, Y. C. Minh, S. Yamamoto, and S. Saito. The ortho to para ratio for ketene in TMC-1. volume 16, page 387. Astronomical Society of the Pacific, 1991.

Y. Ozaki, K. B. Beć, Y. Morisawa, S. Yamamoto, I. Tanabe, C. W. Huck, and T. S. Hofer. Advances, challenges and perspectives of quantum chemical approaches in molecular spectroscopy of the condensed phase. *Chemical Society Reviews*, 50:10917–10954, 2021.

P. Palmeirim, P. André, J. Kirk, D. Ward-Thompson, D. Arzoumanian, V. Könyves, P. Didelon, N. Schneider, M. Benedettini, S. Bontemps, J. Di Francesco, D. Elia, M. Griffin, M. Hennemann, T. Hill, P. G. Martin, A. Men'shchikov, S. Molinari, F. Motte, Q. Nguyen

Luong, D. Nutter, N. Peretto, S. Pezzuto, A. Roy, K. L. J. Rygl, L. Spinoglio, and G. L. White. Herschel view of the Taurus B211/3 filament and striations: evidence of filamentary growth? *Astronomy & Astrophysics*, 550:A38, 2013.

M. E. Palumbo, T. R. Geballe, and A. G. G. M. Tielens. Solid carbonyl sulfide (OCS) in dense molecular clouds. *The Astrophysical Journal*, 479(2):839, 1997.

Y. J. Pendleton, A. G. G. M. Tielens, A. T. Tokunaga, and M. P. Bernstein. The interstellar 4.62 micron band. *The Astrophysical Journal*, 513(1):294, 1999.

H. M. Pickett, R. L. Poynter, E. A. Cohen, M. L. Delitsky, J. C. Pearson, and H. S. P. Müller. Submillimiter, millimeter, and microwave spectral line catalog. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 60(5):883, 1998.

S. Pilling, E. Seperuelo Duarte, E. F. da Silveira, E. Balanzat, H. Rothard, A. Domaracka, and P. Boduch. Radiolysis of ammonia-containing ices by energetic, heavy, and highly charged ions inside dense astrophysical environments. *Astronomy & Astrophysics*, 509: A87, 2010.

S. Pilling, E. Seperuelo Duarte, A. Domaracka, H. Rothard, P. Boduch, and E. F. da Silveira. Radiolysis of H_2O:CO_2 ices by heavy energetic cosmic ray analogs. *Astronomy & Astrophysics*, 523:A77, 2010.

S. Pilling, D. P. P. Andrade, E. F. da Silveira, H. Rothard, A. Domaracka, and P. Boduch. Formation of unsaturated hydrocarbons in interstellar ice analogues by cosmic rays. *Monthly Notices of the Royal Astronomical Society*, 423(3):2209, 2012.

S. Pilling, W. R. M. Rocha, F. M. Freitas, and P. A. da Silva. Photochemistry and desorption induced by X-rays in water rich astrophysical ice analogs: implications for the moon Enceladus and other frozen space environments. *RSC Advances*, 9(49):28823, 2019.

K. M. Pontoppidan, H. J. Fraser, E. Dartois, W. F. Thi, E. F. van Dishoeck, A. C. A. Boogert, L. d'Hendecourt, A. G. G. M. Tielens, and S. E. Bisschop. A 3-5 μm VLT spectroscopic survey of embedded young low mass stars I. Structure of the CO ice. *Astronomy & Astrophysics*, 408:981, 2003.

W. Portugal, S. Pilling, P. Boduch, H. Rothard, and D. P. P. Andrade. Radiolysis of amino acids by heavy and energetic cosmic ray analogues in simulated space environments: α-glycine zwitterion form. *Monthly Notices of the Royal Astronomical Society*, 441(4):3209, 2014.

A. Possolo, C. Merkatas, and O. Bodnar. Asymmetrical uncertainties. *Metrologia*, 56(4): 045009, 2019.

A. Potapov, S. A. Krasnokutski, C. Jäger, and T. Henning. A new "non-energetic" route to complex organic molecules in astrophysical environments: The C + $H_2O \rightarrow H_2CO$ solid-state reaction. *The Astrophysical Journal*, 920(2):111, 2021.

A. Punanova, A. Vasyunin, P. Caselli, A. Howard, S. Spezzano, Y. Shirley, S. Scibelli, and J. Harju. Methanol mapping in cold cores: Testing model predictions. *The Astrophysical Journal*, 927(2):213, 2000.

D. Qasim, K. J. Chuang, G. Fedoseev, S. Ioppolo, A. C. A. Boogert, and H. Linnartz. Formation of interstellar methanol ice prior to the heavy CO freeze-out stage. *Astronomy & Astrophysics*, 612:A83, 2018.

D. Qasim, Lamberts, T., He, J., Chuang, K.-J., Fedoseev, G., Ioppolo, S., Boogert, A. C. A., and Linnartz, H. Extension of the HCOOH and CO_2 solid-state reaction network during the CO freeze-out stage: inclusion of H_2CO. *Astronomy & Astrophysics*, 626:A118, 2019.

D. Qasim, G. Fedoseev, K. J. Chuang, J. He, S. Ioppolo, E. F. van Dishoeck, and H. Linnartz. An experimental study of the surface formation of methane in interstellar molecular clouds. *Nature Astronomy*, 4:781, 2020.

D. Quénard, J. D. Ilee, I. Jiménez-Serra, D. H. Forgan, C. Hall, and K. Rice. The fate of formamide in a fragmenting protoplanetary disk. *The Astrophysical Journal*, 868(1):9, 2018.

D. Quénard, I. JimÃ©nez-Serra, S. Viti, J. Holdship, and A. Coutens. Chemical modelling of complex organic molecules with peptide-like bonds in star-forming regions. *Monthly Notices of the Royal Astronomical Society*, 474(2):2796, 2018.

D. Rabli and D. R. Flower. The rotational excitation of methanol by molecular hydrogen. *Monthly Notices of the Royal Astronomical Society*, 406(1):95, 2010.

M. G. Rachid, Terwisscha van Scheltinga, J., Koletzki, D., and Linnartz, H. Infrared spectra of complex organic molecules in astronomically relevant ice mixtures - II. Acetone. *Astronomy & Astrophysics*, 639:A4, 2020.

M. G. Rachid, Brunken, N., de Boe, D., Fedoseev, G., Boogert, A. C. A., and Linnartz, H. Infrared spectra of complex organic molecules in astronomically relevant ice mixtures - IV. Methylamine. *Astronomy & Astrophysics*, 653:A116, 2021.

M. G. Rachid, Rocha, W. R. M., and Linnartz, H. Infrared spectra of complex organic molecules in astronomically relevant ice mixtures - V. Methyl cyanide (acetonitrile). *Astronomy & Astrophysics*, 665:A89, 2022.

J. M. C. Rawlings, T. W. Hartquist, K. M. Menten, and D. A. Williams. Direct diagnosis of infall in collapsing protostars - I. The theoretical identification of molecular species with broad velocity distributions. *Monthly Notices of the Royal Astronomical Society*, 255:471, 1992.

J. M. C. Rawlings, D. A. Williams, S. Viti, and C. Cecchi-Pestellini. A radical route to interstellar propylene formation. *Monthly Notices of the Royal Astronomical Society: Letters*, 436(1):L59, 2013.

E. Redaelli, L. Bizzocchi, P. Caselli, O. Sipilä, V. Lattanzi, B. M. Giuliano, and S. Spezzano. High-sensitivity maps of molecular ions in L1544 - I. Deuteration of N_2H^+ and HCO^+ and primary evidence of N_2D^+ depletion. *Astronomy & Astrophysics*, 629:A15, 2019.

E. Redaelli, O. Sipilä, M. Padovani, P. Caselli, D. Galli, and A. V. Ivlev. The cosmic-ray ionisation rate in the pre-stellar core L1544. *Astronomy & Astrophysics*, 656:A109, 2021.

E. Redaelli, S. Bovino, P. Sanhueza, K. Morii, G. Sabatini, P. Caselli, A. Giannetti, and S. Li. The core population and kinematics of a massive clump at early stages: An Atacama Large Millimeter/submillimeter Array view. *The Astrophysical Journal*, 936(2):169, 2022a.

E. Redaelli, A. Chacón-Tanarro, P. Caselli, M. Tafalla, J. E. Pineda, S. Spezzano, and O. Sipilä. A large ($\approx$ 1 pc) contracting envelope around the prestellar core L1544. *The Astrophysical Journal*, 941(2):168, 2022b.

E. Redaelli, L. Bizzocchi, P. Caselli, and J. E. Pineda. Nitrogen fractionation in ammonia and its insights into nitrogen chemistry. *Astronomy & Astrophysics*, 674:L8, 2023.

V. M. Rivilla, L. Colzi, I. Jiménez-Serra, J. Martín-Pintado, A. Megías, M. Melosso, L. Bizzocchi, Á. López-Gallifa, A. Martínez-Henares, S. Massalkhi, B. Tercero, P. de Vicente, J.-C. Guillemin, J. García de la Concepción, F. Rico-Villas, S. Zeng, S. Martín, M. A. Requena-Torres, F. Tonolo, S. Alessandrini, L. Dore, V. Barone, and C. Puzzarini. Precursors of the RNA World in Space: Detection of (Z)-1,2-ethenediol in the Interstellar Medium, a Key Intermediate in Sugar Formation. *The Astrophysical Journal Letters*, 929(1): L11, 2022a.

V. M. Rivilla, J. García De La Concepción, I. Jiménez-Serra, J. Martín-Pintado, L. Colzi, B. Tercero, A. Megías, Á. López-Gallifa, A. Martínez-Henares, S. Massalkhi, S. Martín, S. Zeng, P. De Vicente, F. Rico-Villas, M. A. Requena-Torres, and G. Cosentino. Ionize Hard: Interstellar PO+ Detection. *Frontiers in Astronomy and Space Sciences*, 9:829288, 2022b.

V. M. Rivilla, M. Sanz-Novo, I. Jiménez-Serra, J. Martín-Pintado, L. Colzi, S. Zeng, A. Megías, Á. López-Gallifa, A. Martínez-Henares, S. Massalkhi, B. Tercero, P. de Vicente, S. Martín, D. San Andrés, M. A. Requena-Torres, and J. L. Alonso. First Glycine Isomer Detected in the Interstellar Medium: Glycolamide ($NH_2C(O)CH_2OH$). *The Astrophysical Journal Letters*, 953(2):L20, 2023.

W. Rocha and S. Pilling. Determination of optical constants *n* and *k* of thin films from absorbance data using Kramers-Kronig relationship. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 123:436, 2014.

W. R. M. Rocha, S. Pilling, A. L. F. de Barros, D. P. P. Andrade, H. Rothard, and P. Boduch. Infrared complex refractive index of astrophysical ices exposed to cosmic rays simulated in the laboratory. art. arXiv:1609.04684, 2016.

W. R. M. Rocha, G. Perotti, L. E. Kristensen, and J. K. Jørgensen. Fitting infrared ice spectra with genetic modelling algorithms - Presenting the ENIIGMA fitting tool. *Astronomy & Astrophysics*, 654:A158, 2021.

W. R. M. Rocha, M. G. Rachid, B. Olsthoorn, E. F. van Dishoeck, M. K. McClure, and H. Linnartz. LIDA: The Leiden Ice Database for Astrochemistry. *Astronomy & Astrophysics*, 668:A63, 2022.

W. R. M. Rocha, van Dishoeck, E. F., Ressler, M. E., van Gelder, M. L., Slavicinska, K., Brunken, N. G. C., Linnartz, H., Ray, T. P., Beuther, H., Caratti o Garatti, A., Geers, V., Kavanagh, P. J., Klaassen, P. D., Justtanont, K., Chen, Y., Francis, L., Gieser, C., Perotti, G., Tychoniec, Ł., Barsony, M., Majumdar, L., le Gouellec, V. J. M., Chu, L. E. U., Lew, B. W. P., Henning, Th., and Wright, G. JWST Observations of Young protoStars (JOYS+): Detecting icy complex organic molecules and ions - I. CH_4, SO_2, $HCOO^-$, OCN^-, H_2CO, HCOOH, CH_3CH_2OH, CH_3CHO, CH_3OCHO, and CH_3COOH. *Astronomy & Astrophysics*, 683:A124, 2024.

M. Ruaud, J. C. Loison, K. M. Hickson, P. Gratier, F. Hersant, and V. Wakelam. Modelling complex organic molecules in dense regions: Eley-Rideal and complex induced reaction. *Monthly Notices of the Royal Astronomical Society*, 447(4):4004, 2015.

M. Ruaud, V. Wakelam, and F. Hersant. Gas and grain chemical composition in cold cores as predicted by the Nautilus three-phase model. *Monthly Notices of the Royal Astronomical Society*, 459(4):3756, 2016.

S. A. Sandford and L. J. Allamandola. Condensation and Vaporization Studies of CH 3OH and NH 3 Ices: Major Implications for Astrochemistry. *The Astrophysical Journal*, 417:815, 1993.

S. A. Sandford, L. J. Allamandola, A. G. G. M. Tielens, and G. J. Valero. Laboratory studies of the infrared spectral properties of CO in astrophysical ices. *The Astrophysical Journal*, 329:498, 1988.

M. Sanz-Novo, V. M. Rivilla, I. Jiménez-Serra, J. Martín-Pintado, L. Colzi, S. Zeng, A. Megías, Á. López-Gallifa, A. Martínez-Henares, S. Massalkhi, B. Tercero, P. de Vicente, S. Martín, D. San Andrés, and M. A. Requena-Torres. Discovery of the Elusive Carbonic Acid (HOCOOH) in Space. *The Astrophysical Journal*, 954(1):3, 2023.

M. Sanz-Novo, V. M. Rivilla, I. Jiménez-Serra, J. Martín-Pintado, L. Colzi, S. Zeng, A. Megías, Á. López-Gallifa, A. Martínez-Henares, S. Massalkhi, B. Tercero, P. de Vicente, D. San Andrés, S. Martín, and M. A. Requena-Torres. Interstellar Detection of O-protonated Carbonyl Sulfide, $HOCS^+$. *The Astrophysical Journal*, 965(2):149, 2024a.

M. Sanz-Novo, V. M. Rivilla, H. S. P. Müller, I. Jiménez-Serra, J. Martín-Pintado, L. Colzi, S. Zeng, A. Megías, Á. López-Gallifa, A. Martínez-Henares, B. Tercero, P. de Vicente, D. San Andrés, S. Martín, and M. A. Requena-Torres. Discovery of thionylimide, HNSO, in space: The first N-, S-, and O-bearing interstellar molecule. *The Astrophysical Journal Letters*, 965(2):L26, 2024b.

W. A. Schutte, A. G. G. M. Tielens, D. C. B. Whittet, A. Boogert, P. Ehrenfreund, T. de Graauw, T. Prusti, E. F. van Dishoeck, and P. Wesselius. The 6.0 and 6.8 μm absorption features in the spectrum of NGC 7538: IRS9. *The Astrophysical Journal*, 315:L333, 1996.

S. Scibelli and Y. Shirley. Prevalence of complex organic molecules in starless and prestellar cores within the Taurus molecular cloud. *The Astrophysical Journal*, 891(1):73, 2020.

S. Scibelli, Y. Shirley, A. Vasyunin, and R. Launhardt. Detection of complex organic molecules in young starless core L1521E. *Monthly Notices of the Royal Astronomical Society*, 504(4):5754, 2021.

S. Scibelli, Y. Shirley, A. Schmiedeke, B. Svoboda, A. Singh, J. Lilly, and P. Caselli. 3D radiative transfer modelling and virial analysis of starless cores in the B10 region of the Taurus molecular cloud. *Monthly Notices of the Royal Astronomical Society*, 521(3):4579, 2023.

S. Scibelli, Y. Shirley, A. Megías, and I. Jiménez-Serra. Survey of complex organic molecules in starless and pre-stellar cores in the Perseus molecular cloud. *Monthly Notices of the Royal Astronomical Society*, 533(4):4104, 2024.

O. M. Shalabiea and J. M. Greenberg. Bistability and dust/gas chemical modelling in dark interstellar clouds. *Astronomy & Astrophysics*, 296:779, 1995.

A. Simon, K. I. Öberg, M. Rajappan, and P. Maksiutenko. Entrapment of CO in CO_2 ice. *The Astrophysical Journal*, 883(1):21, 2019.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv e-prints*, art. arXiv:1409.1556, 2014.

O. Sipilä, K. Silsbee, and P. Caselli. A revised description of the cosmic ray induced desorption of interstellar ices. *The Astrophysical Journal*, 922(2):126, 2021.

K. Slavicinska, Rachid, M. G., Rocha, W. R. M., Chuang, K.-J., van Dishoeck, E. F., and Linnartz, H. The hunt for formamide in interstellar ices - A toolkit of laboratory infrared spectra in astronomically relevant ice mixtures and comparisons to ISO, Spitzer, and JWST observations. *Astronomy & Astrophysics*, 677:A13, 2023.

S. Spezzano, L. Bizzocchi, P. Caselli, J. Harju, and S. S. Brünken. Chemical differentiation in a prestellar core traces non-uniform illumination. *Astronomy & Astrophysics*, 592:L11, 2016.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929, 2014.

J. A. Sturm, M. K. McClure, T. L. Beck, D. Harsono, J. B. Bergner, E. Dartois, A. C. A. Boogert, J. E. Chiar, M. A. Cordiner, M. N. Drozdovskaya, S. Ioppolo, C. J. Law, H. Linnartz, D. C. Lis, G. J. Melnick, B. A. McGuire, J. A. Noble, K. I. Öberg, M. E. Palumbo, Y. J. Pendleton, G. Perotti, K. M. Pontoppidan, D. Qasim, W. R. M. Rocha, H. Terada, R. G. Urso, and E. F. van Dishoeck. A JWST inventory of protoplanetary disk ices. The edge-on protoplanetary disk HH 48 NE, seen with the Ice Age ERS program. *Astronomy & Astrophysics*, 679:A138, 2023.

J. A. Sturm, M. K. McClure, D. Harsono, J. B. Bergner, E. Dartois, A. C. A. Boogert, M. A. Cordiner, M. N. Drozdovskaya, S. Ioppolo, C. J. Law, D. C. Lis, B. A. McGuire, G. J. Melnick, J. A. Noble, K. I. Öberg, M. E. Palumbo, Y. J. Pendleton, G. Perotti, W. R. M. Rocha, R. G. Urso, and E. F. van Dishoeck. A JWST/MIRI analysis of the ice distribution and PAH emission in the protoplanetary disk HH 48 NE. *arXiv e-prints*, art. arXiv:2407.09627, 2024.

T. Suhasaria, G. A. Baratta, S. Ioppolo, H. Zacharias, and M. E. Palumbo. Solid CO_2 in quiescent dense molecular clouds. Comparison between Spitzer and laboratory spectra. *Astronomy & Astrophysics*, 608:A12, 2017.

D. F. Swinehart. The beer-lambert law. *Journal of Chemical Education*, 39(7):333, 1962.

I. M. Taban, Schutte, W. A., Pontoppidan, K. M., and van Dishoeck, E. F. Stringent upper limits to the solid NH_3 abundance towards W 33A from near-IR spectroscopy with the Very Large Telescope. *Astronomy & Astrophysics*, 399(1):169, 2003.

M. Tafalla and J. Santiago. L1521E: The first starless core with no molecular depletion. *Astronomy & Astrophysics*, 414(3):L53, 2004.

M. Tafalla, P. C. Myers, P. Caselli, and C. M. Walmsley. On the internal structure of starless cores - I. Physical conditions and the distribution of CO, CS, N_2H^+, and NH_3 in L1498 and L1517B. *Astronomy & Astrophysics*, 416(1):191, 2004.

M. Tafalla, J. Santiago-García, P. C. Myers, P. Caselli, C. M. Walmsley, and A. Crapsi. On the internal structure of starless cores - II. A molecular survey of L1498 and L1517B. *Astronomy & Astrophysics*, 455(2):577, 2006.

J. Terwisscha van Scheltinga, Ligterink, N. F. W., Boogert, A. C. A., van Dishoeck, E. F., and Linnartz, H. Infrared spectra of complex organic molecules in astronomically relevant ice matrices - I. Acetaldehyde, ethanol, and dimethyl ether. *Astronomy & Astrophysics*, 611: A35, 2018.

A. G. G. M. Tielens. *The physics and chemistry of the interstellar medium*. 2005.

M. S. Väisälä, J. Harju, M. J. Mantere, O. Miettinen, R. S. Sault, C. M. Walmsley, and J. B. Whiteoak. High-resolution ammonia mapping of the very young protostellar core Chamaeleon-MMS1. *Astronomy & Astrophysics*, 564:A99, 2014.

F. A. van Broekhuizen, K. M. Pontoppidan, H. J. Fraser, and E. F. van Dishoeck. A 3-5 μm VLT spectroscopic survey of embedded young low mass stars II. Solid OCN^-. *Astronomy & Astrophysics*, 441(1):249, 2005.

F. A. van Broekhuizen, Groot, I. M. N., Fraser, H. J., van Dishoeck, E. F., and Schlemmer, S. Infrared spectroscopy of solid CO-CO_2 mixtures and layers. *Astronomy & Astrophysics*, 451(2):723, 2006.

F. F. S. van der Tak, J. H. Black, F. L. Schöier, D. J. Jansen, and E. F. van Dishoeck. A computer program for fast non-LTE analysis of interstellar line spectra - With diagnostic plots to interpret observed line intensity ratios. *Astronomy & Astrophysics*, 468(2):627, 2007.

E. F. van Dishoeck, B. Jonkheid, and M. C. van Hemert. Photoprocesses in protoplanetary disks. *Faraday Discussions*, 133:231, 2006.

C. Vastel, C. Ceccarelli, B. Lefloch, and R. Bachiller. The origin of complex organic molecules in prestellar cores. *The Astrophysical Journal*, 795(1):L2, 2014.

A. I. Vasyunin and E. Herbst. Reactive desorption and radiative association as possible drivers of complex molecule formation in the cold interstellar medium. *The Astrophysical Journal*, 769(1):34, 2013.

A. I. Vasyunin, A. M. Sobolev, D. S. Wiebe, and D. A. Semenov. Influence of uncertainties in the rate constants of chemical reactions on astrochemical modeling results. *Astronomy Letters*, 30(8):566, 2004.

A. I. Vasyunin, P. Caselli, F. Dulieu, and I. Jiménez-Serra. Formation of complex molecules in prestellar cores: A multilayer approach. *The Astrophysical Journal*, 842(1):33, 2017.

S. Viti, M. P. Collings, J. W. Dever, M. R. S. McCoustra, and D. A. Williams. Evaporation of ices near massive stars: models based on laboratory temperature programmed desorption data. *Monthly Notices of the Royal Astronomical Society*, 354(4):1141, 2004.

S. Viti, I. Jiménez-Serra, J. A. Yates, C. Codella, M. Vasta, P. Caselli, B. Lefloch, and C. Ceccarelli. L1157-B1: Water and Ammonia as Diagnostics of Shock Temperature. *The Astrophysical Journal Letters*, 740(1):L3, 2011.

J. Vitorino, J. C. Loison, V. Wakelam, E. Congiu, and F. Dulieu. Sulphur storage in cold molecular clouds: the case of the $NH_4{}^+SH^-$ salt on interstellar dust grains. *Monthly Notices of the Royal Astronomical Society*, 533(1):52, 2024.

V. Wakelam and E. Herbst. Polycyclic aromatic hydrocarbons in dense cloud chemistry. *The Astrophysical Journal*, 680(1):371, 2008.

V. Wakelam, F. Selsis, E. Herbst, and P. Caselli. Estimation and reduction of the uncertainties in chemical models: Application to hot core chemistry. *Astronomy & Astrophysics*, 444(3):883, 2005.

V. Wakelam, I. W. M. Smith, E. Herbst, J. Troe, W. Geppert, H. Linnartz, K. Öberg, E. Roueff, M. Agúndez, P. Pernot, H. M. Cuppen, J. C. Loison, and D. Talbi. Reaction networks for interstellar chemical modelling: Improvements and challenges. *Space Science Reviews*, 156 (1):13, 2010.

V. Wakelam, E. Bron, S. Cazaux, F. Dulieu, C. Gry, P. Guillard, E. Habart, L. Hornekær, S. Morisset, G. Nyman, V. Pirronello, S. D. Price, V. Valdivia, G. Vidali, and N. Watanabe. H_2 formation on interstellar dust grains: The viewpoints of theory, experiments, models and observations. *Molecular Astrophysics*, 9:1, 2017.

C. Walsh, R. A. Loomis, K. I. Öberg, M. Kama, M. L. R. van 't Hoff, T. J. Millar, Y. Aikawa, E. Herbst, S. L. Widicus Weaver, and H. Nomura. First detection of gas-phase methanol in a protoplanetary disk. *The Astrophysical Journal Letters*, 823(1):L10, 2016.

N. Watanabe and A. Kouchi. Efficient formation of formaldehyde and methanol by the addition of hydrogen atoms to CO in H_2O-CO ice at 10 K. *The Astrophysical Journal*, 571 (2):L173, 2002.

D. C. B. Whittet, M. F. Bode, A. J. Longmore, D. W. T. Baines, and A. Evans. Interstellar ice grains in the Taurus molecular clouds. *Nature*, 303(5914):218, 1983.

D. C. B. Whittet, A. J. Adamson, W. W. Duley, T. R. Geballe, and A. D. McFadzean. Infrared spectroscopy of dust in the Taurus dark clouds: solid carbon monoxide. *Monthly Notices of the Royal Astronomical Society*, 241:707, 1989.

J. P. Williams and L. A. Cieza. Protoplanetary disks and their evolution. *Annual Review of Astronomy and Astrophysics*, 49(1):67, 2011.

T. L. Wilson, K. Rohlfs, and S. Hüttemeister. *Tools of Radio Astronomy*. 2013.

E. S. Wirström, W. D. Geppert, Å. Hjalmarson, C. M. Persson, J. H. Black, P. Bergman, T. J. Millar, M. Hamberg, and E. Vigren. Observational tests of interstellar methanol formation. *Astronomy & Astrophysics*, 533:A24, 2011.

Y.-L. Yang, J. D. Green, K. M. Pontoppidan, J. B. Bergner, L. I. Cleeves, N. J. E. II, R. T. Garrod, M. Jin, C. H. Kim, J. Kim, J.-E. Lee, N. Sakai, C. N. Shingledecker, B. Shope, J. J. Tobin, and E. F. van Dishoeck. CORINOS. I. JWST/MIRI Spectroscopy and Imaging of a Class 0 Protostar IRAS 15398-3359. *The Astrophysical Journal Letters*, 941(1):L13, 2022.

Appendix A

Chapter 2

This appendix corresponds to Chapter 2, that is, to the methodology of the thesis.

A.1 Automated GILDAS/CLASS Pipeline

Here we explain briefly how to use the Automated GILDAS/CLASS Pipeline, a bundle of Python and Julia scripts that allows to use the tool CLASS from the software GILDAS in an automated way, in order to reduce single-pointing radio spectra.

It requires a configuration file written in plain text, with the YAML formatting style, containing the information about the data files to be reduced and the parameters of the reduction. As mentioned in the main text, this pipeline has several working modes, being the most important ones the selection, line search, reduction, averaging and merging. The first three modes will be explained in the following, as well as more details about the configuration tile.

A.1.1 Configuration file

First of all, one should write the configuration file, a plain text file written in the YAML formatting style. It must contain the name of the data files that contain the spectra to be processed, including some metadata stored in the CLASS files: the source name, the *line* name (refering to the central frequency of the setup), and the *telescope* name (refering to the exact receiver from the instrument used). [1]Additionally, it can contain several options related to the different working modes of the pipeline.

Figure A.1.1 shows an example configuration file. In this file, we have indivated two input files, each one corresponding to one day of observations at the IRAM 30 m telescope.

[1] Usually, the same instrument can observe simulatenously in different spectral regions close to the central frequency, using different receivers.

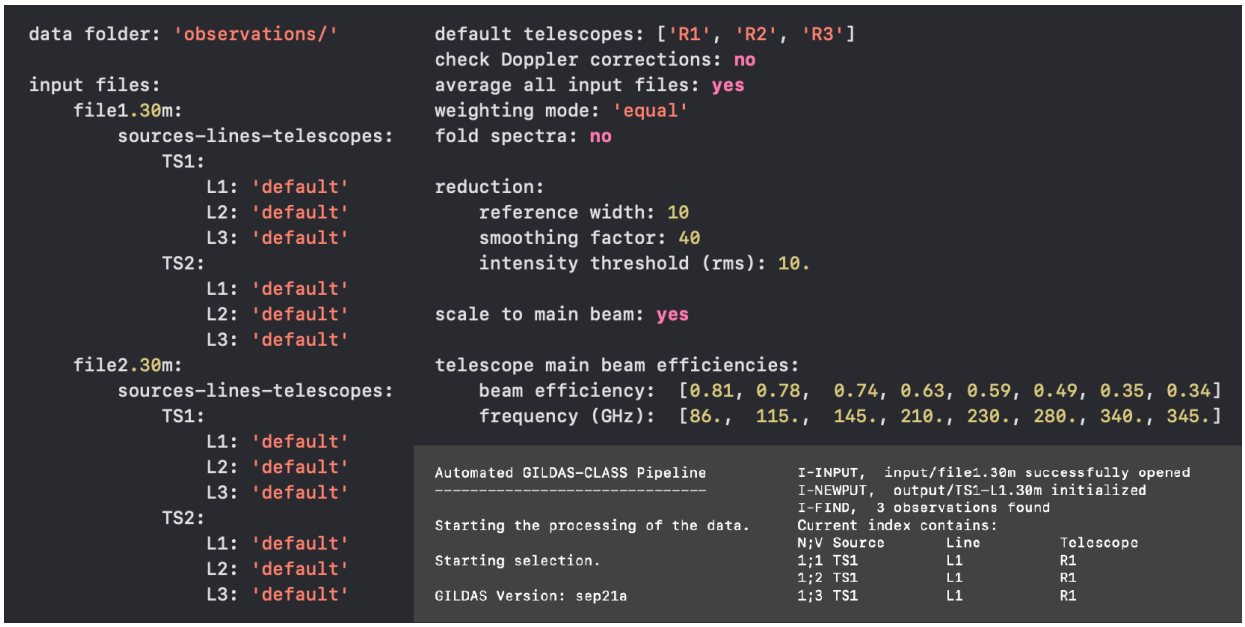


Figure A.1.1. Screenshots of an example configuration file for the Automated GILDAS/CLASS Pipeline (*left*, *upper right*), and of first messages shown in the terminal after the running of the selection mode (*lower right*).

There are two observed sources (TS1, TS2), with three frequency setups (L1, L2, L3) and three receivers (R1, R2, R3). Then, there are some options regarding the selection, line search and reduction modes.

A.1.2 Selection mode

After running the selection mode with the configuration file shown in Figure A.1.1, several CLASS files are created, one for each source and for each frequency setup: TS1-L1.30m, TS1-L2.30m, TS1-L3.30m, TS2-L1.30m, TS2-L2.30m, TS2-L3.30m. In each file, all the corresponding observations of the two observing files have been combined and averaged. For each frequency setup, there are still observations corresponding to the different receivers (R1, R2, R3), hence having a total of $2 \times 3 \times 3 = 18$ spectra, 3 for each of the 6 output files. This process is done is just a few seconds. In this time, CLASS is being called automatically, showing the output of each command in the terminal; the very first messages can be seen in Figure A.1.1 (*lower right*), divided in two columns. This constitutes a decrease in the time needed to perform these operations, in comparison with using just CLASS.

A.1.3 Line search and reduction

These modes allow for the baseline fit and subtraction for each spectrum contained in the files produced after the selection mode. Some parameters regarding this mode can be changed in the configuration file, under the clasification of *reduction* (Fig. A.1.1):

- *reference width*: estimate of the typical width of the lines present in the spectrum;

- *smoothing factor*: size of the smoothing filters performed to fit the baseline, which should be at least of the order of channels corresponding to the typical line width;
- *intensity threshold*: reference value with respect to the RMS noise of the continuum that will be used as a threshold to determine the presence of a line.

These parameters are used both for the line search and for the reduction.

The line search mode allows to automatically identify the most significant lines present in the spectra, so that they can be masked and the baseline can be properly fitted and substracted in the reduction mode. The algorithm is explained below:

1. As a first approximation of the baseline, a median filter is applied with a size equal to the input smoothing factor. A median filter is equivalent to a rolling mean (or rolling average) but applying the median instead of the mean: for each intensity point in the spectrum, we take the intensity values before and after it until we get a number of values equal to the filter size (including the central point); then, we compute the median and substitute the original intensity value by this median.
2. A sigma clip is applied to the difference between the original spectrum and the estimated baseline, to identify the lines. In a sigma clip, we perform several iterations (2 or 3) of the following steps: we compute the standard deviation (*sigma*) of the input spectrum; then, we filter out the values that are greater than a number of times the standard deviation (in our case, the *intensity threshold*). Hence, the values that are too distant from the estimated baseline are identified as lines. From these points, a set of frequency windows (start and end of the line) are saved in a text file.
3. We perform a new estimation of the baseline using polynomials, fitting the original spectrum ignoring the frequency windows with lines.
4. We repeat steps 2 and 3 a certain number of times –by default, just 2 iterations are performed in total.

The baseline estimation using polynomials is performed using *splines*, that is, a concatenation of polynomials. For that, the Python library SciPy is used,[2] in particular its function `UnivariateSpline` from the module `interpolate`. This function requires a smooth parameter (s) to define the number of polynomials to be concatenated together (or the number of union points, the *knots*). The condition to determine the number of knots is that the sum of the squared errors between the spline and the original curve must be less or equal to the smooth parameter. Hence, we use the following procedure to compute s: we perform a rolling median with a size equal to the input smoothing factor, and then we compute the sum of the squared errors between the original curve and the smoothed one. In this way, the spline will have a similar shape than the smoothed spectrum using the median filter.

This baseline estimation is used both for the automatic line search and for the final baseline fit during the reduction mode. In the first case, it helps to estimate the significant lines and derive the corresponding frequency windows, which will be written in a text file. In

[2] `https://scipy.org/`

the reduction mode these frequency windows are used to estimate the baseline and show preliminary results, but the windows can be manually modified thanks to an interactive mode that allows zooming and exploring the spectrum and clicking with the mouse to select a new window or delete an existing one.

After executing the line search mode with the example configuration file (Fig. A.1.1), a total of 129 lines for all the 18 spectra are detected, in just a few seconds. To inspect them, we can directly run the reduction mode, where the interactive mode will allow the user to modify the automatically selected frequency windows. Figure 2.3.4 shows the plot from the interactive mode for the very first synthetic spectrum of the example. As you can see, several regions have been identified as lines. However, the regions at the start and the end of the spectrum have been erroneously identified as containing lines. They can be manually removed with the cursor and then the reduction can be applied again. Similarly, the group of three lines between 64 and 66 GHz can be splitted in three frequency windows, one for each line, for a better baseline fitting. After reviewing the 18 spectra, a total of 6 files will be created: TS1-L1-r.30m, TS1-L2-r.30m, TS1-L3-r.30m, TS2-L1-r.30m, TS2-L2-r.30m, TS2-L3-r.30m (*r* for *reduced*).

A.2 Example chemical network

A simple chemical network involved in the formation of water (H_2O) is presented in this section, as an example for the reaction rate chemical models. As seen in Figure A.2.1, there are many interactions between the elements involved in the chemical network, which are not only atoms and molecules, but also electrons (e^-), cosmic rays (cr) and gamma rays (γ). As a result of these interactions, an uncertainty in the determination of one reaction rate could have significant effects on the abundance determination of some species. That is why it is usual in astrochemical model to compare observations with simulations in a logarithmic scale, focusing in the order of magnitude rather in exact numerical coincidences. The corresponding single chemical reactions involved in this simple exacmple are listed below:

$$O^+ + H_2 \longrightarrow OH^+ + H \ , \tag{A.1}$$
$$H_2O^+ + H_2 \longrightarrow H_3O^+ + H \ , \tag{A.2}$$
$$O^+ + H_2 \longrightarrow OH^+ + H \ , \tag{A.3}$$
$$OH^+ + H_2 \longrightarrow H_2O^+ + H \ , \tag{A.4}$$
$$H_2O^+ + H_2 \longrightarrow H_3O^+ + H \ , \tag{A.5}$$
$$H_3O^+ + e^- \longrightarrow H_2O + H \ , \tag{A.6}$$
$$H_2O^+ + e^- \longrightarrow OH + H \ , \tag{A.7}$$
$$H_2O^+ + e^- \longrightarrow O + H_2 \ , \tag{A.8}$$

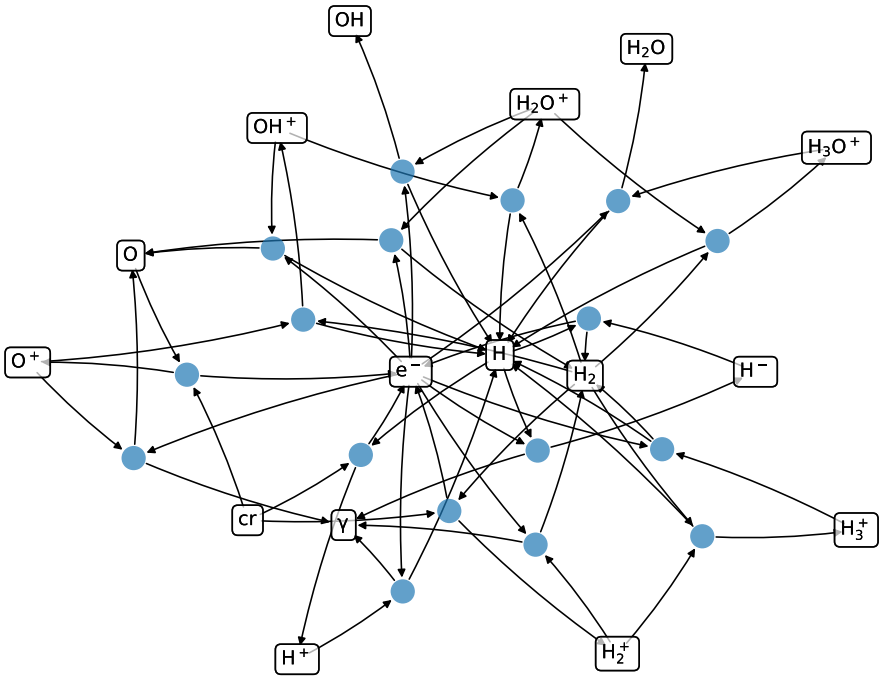


Figure A.2.1. Graph representing a set of chemical reactions involving the formation of water (H_2O) in gas phase. Each blue circle represents a single reaction: for each of them, the reactants have arrows that end in the blue circle, and other arrows stem from the blue circles to the corresponding products. The graph was created with Python's library NetworkX (`https://networkx.org/`) using the Kamada-Kawai layout (Kamada and Kawai, 1989).

$$OH^+ + e^- \longrightarrow O + H \ , \tag{A.9}$$
$$O^+ + e^- \longrightarrow O + \gamma \ , \tag{A.10}$$
$$O + cr \longrightarrow O^+ + e^- \ , \tag{A.11}$$
$$H_2 + cr \longrightarrow H_2^+ + e^- \ , \tag{A.12}$$
$$H + cr \longrightarrow H^+ + e^- \ , \tag{A.13}$$
$$H_2 + H_2^+ \longrightarrow H_3^+ + H \ , \tag{A.14}$$
$$H_3^+ + e^- \longrightarrow H_2 + H \ , \tag{A.15}$$
$$H^- + H \longrightarrow H_2 + e^- \ , \tag{A.16}$$
$$H_2^+ + e^- \longrightarrow H_2 + \gamma \ , \tag{A.17}$$
$$H + e^- \longrightarrow H^- + \gamma \ , \tag{A.18}$$
$$H^+ + e^- \longrightarrow H + \gamma \ . \tag{A.19}$$

This example chemical network was extracted from the 2021 KROME School (KROME is an open-source rate-equation astrochemical software (Grassi et al., 2014):`https://www.kromepackage.org/`.); the mentioned reactions were explained by Herbst and Klemperer (1973) and Dalgarno and Black (1976).

Appendix B

Chapter 3

This appendix corresponds to Chapter 3, that is, for the paper titled *The complex organic molecular content in the L1517B starless core*, published in the journal Monthly Notices of the Royal Astronomical Society (Megías et al., 2023).[1]

B.1 Derivation of the methanol column density and its uncertainty with RADEX

In this Section, we explain the procedure we have followed to determine the column densities of methanol, acetonitrile and cyanoacetilene, and their uncertainties, using the non-LTE molecular excitation code RADEX (van der Tak et al., 2007). It is a radiative transfer model that allows us to predict the line intensity of the selected molecule from the following parameters:

- **Molecule**. It has to be included in the RADEX molecule list, therefore having collisional cross-section derived, experimentally or theoretically.
- **Spectral range**. Frequency range for the transitions to be predicted for the selected molecule.
- **Excitation conditions**.
  - Molecular hydrogen number density, $n_{\mathrm{H_2}}$.
  - Kinetic temperature of the molecule, T_{kin}.
  - Background temperature, T_{bg}.
- **Radiative transfer parameters**.
  - Column density, N.
  - Line width, Δv.

For each species, the results of the calculation would be a set of intensity lines, $\{\tilde{I}_i\}$, where i is the index of the line. In our case, the molecules are methanol (CH_3OH) A and E,

[1] https://doi.org/10.1093/mnras/stac3449

acetonitrile (CH_3CN), and cyanoacetilene (HC_3N), and we have observations for their lines in a specific spectral range (see Table 2), so we have a set of observed intensities, $\{I_i\}$. We have an uncertainty for each line (the RMS noise), so we will also have a set of uncertainties $\{\Delta I_i\}$. We use the frequency range of the observed lines, plus a little margin, for the RADEX calculations. The rest of the parameters except the column density (and optionally the H_2 number density) are known (see Section 3.3.2 for more details):

- Molecular hydrogen number density (not fixed for the case of methanol E and cyanoacetilene):
  - Core's centre: $n_{H_2} = 2.20 \times 10^5$ cm^{-3} (Tafalla et al., 2004).
  - Methanol peak: $n_{H_2} = 1.24 \times 10^5$ cm^{-3} (Tafalla et al., 2004).
- Background temperature: $T_{bg} = 2.73$ K (temperature of the cosmic microwave background).
- Kinetic temperature: $T_{kin} = 10$ K (Tafalla et al., 2004).
- Line width: $\Delta v \sim 0.3$ km s^{-1} (we choose for each case the result of the MADCUBA fit).

Our goal is to optimize the column density and optionally also the H_2 number density (when more than one transition is available) to minimize the difference between the observed lines, $\{I_i\}$, and the predicted ones, $\{\hat{I}_i\}$, taking into account the uncertainties, $\{\Delta I_i\}$. To measure that difference, we define a loss function, $\mathcal{L}$, which we choose to be the chi-square (χ^2) error:

$$\mathcal{L} = \sum_i \left(\frac{\hat{I}_i - I_i}{\Delta I_i} \right)^2 . \tag{B.1}$$

In this way we are measuring the quadratic error weighted with the squared inverse of the uncertainty, so that a difference between the observations and the model contributes less to the loss if the uncertainty on the line intensity is greater.

Once we have properly defined our loss function, we can minimize it with respect to the column density for each of our 8 cases (four species, methanol A and E, cyanoacetilene, and acetonitrile; and two positions, dust and methanol peaks) and also with respect to the H_2 number density for methanol E and cyanoacetilene. We did this with Python,[2] using the methods COBYLA and Nelder-Mead within the function minimize from the library SciPy.[3]

Finally, we have to estimate the uncertainty for the optimized value/s of our parameter/s. To do so, we take a threshold for the loss, greater than the minimized value, which defines a lower and an upper uncertainty. In order to find a proper threshold, we create 10^5 sets of variations of the observed intensities, $\{\{I_i\}_j\}$, so that the value of each intensity, $I_{i,j}$, comes from a normal distribution with mean I_i and standard deviation ΔI_i. Then, we calculate the loss for the optimized model with equation B.1 for each set of intensities $\{I_i\}_j$, obtaining a distribution of losses. We choose the threshold so that the loss values lower than such a threshold are the 68.27 percent of the total values (similarly to the definition of 1 σ confidence interval, but for an asymmetric distribution with a minimum value). The threshold thus

[2] https://www.python.org/
[3] https://scipy.org/

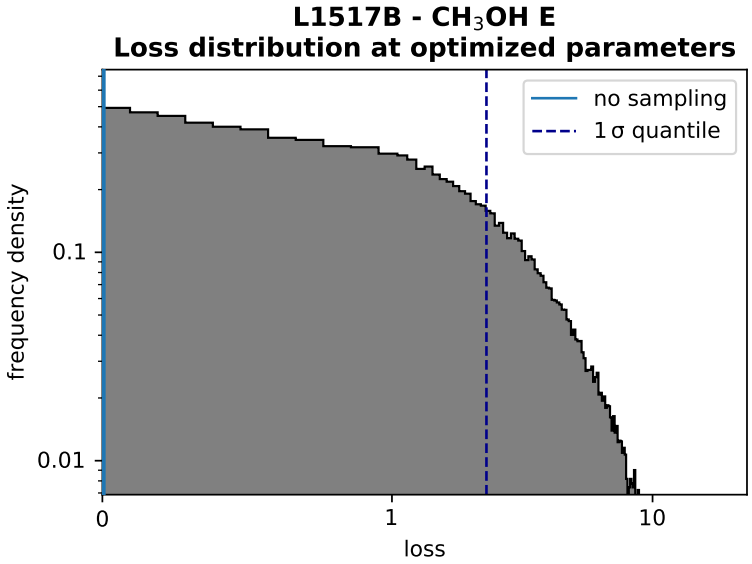


Figure B.1.1. Histogram of the distribution of losses obtained from the intensities of the observed lines and their uncertainties, for the case of methanol E in the dust peak of L1517B. The scale in the horizontal axis is symmetrical-logarithmic with threshold 1, that is, it is linear from 0 to 1 and logarithmic beyond 1. The blue line on the left shows the value of the loss without taking into account the uncertainties of the observed lines (Equation B.1), which is almost 0; and the blue dashed line indicates the 68.27 percentile.

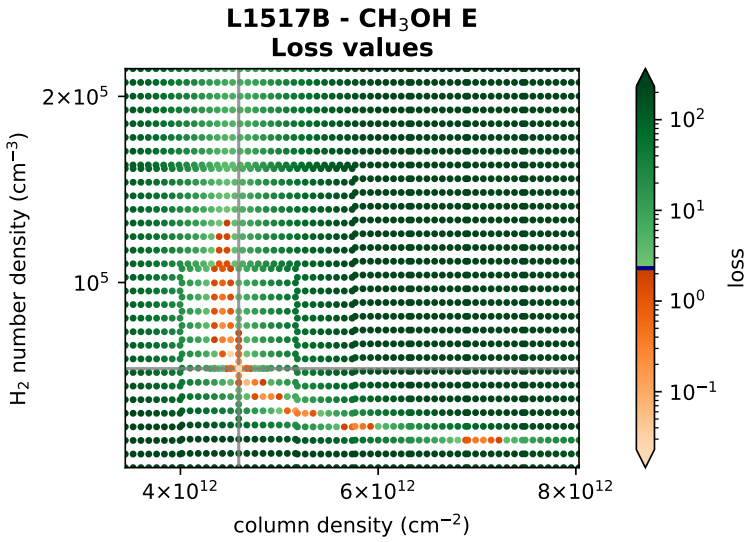


Figure B.1.2. Plot of the loss ($\mathcal{L}$) versus the column density (N) and the H_2 number density (n_{H_2}) for a range of values around the optimized values that minimize the loss function, for the case of methanol E in the dust peak of L1517B. The gray lines indicate the optimized values for N and n_{H_2}. The color map is composed by two ranges of colors in order to highlight the points in which the loss ($\mathcal{L}$) is lower than the threshold value (marked in blue in the color bar), which will define the uncertainties for both optimized parameters.

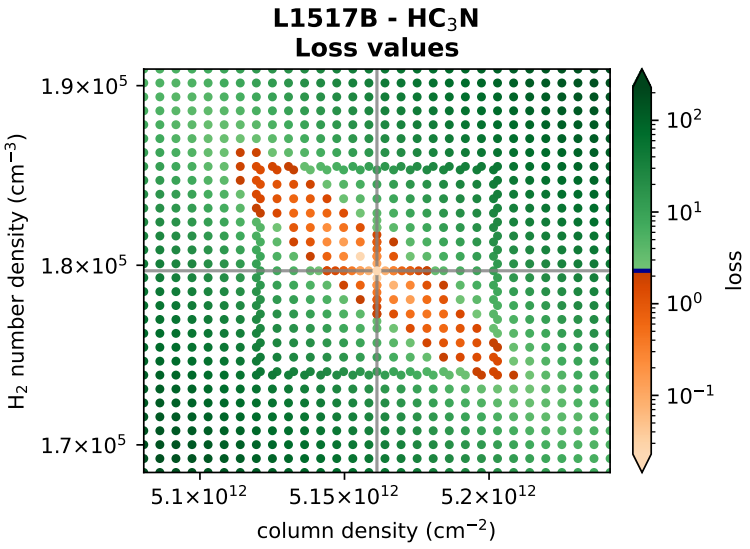


Figure B.1.3. Same as Fig. B.1.2, but for cyanoacetilene (HC_3N) in the dust peak of L1517B.

corresponds to the 68.27 percentile (see Fig. B.1.1).

Then, we have to explore our parameter space around the minimized values in order to find when the loss function reaches our threshold value. If we only fit the column density, we only have to calculate the loss function in the surroundings of the minimized value until we reach the threshold value. However, if we also want to minimise the H_2 number density, we have to explore a bidimensional parameter space. We opted to make an adaptive grid that starts calculating the loss values at the surroundings of the obtained minimum and continues enlarging its size until it encloses all the loss values minor to the threshold with a reasonable margin. This way, we obtained Figs. B.1.2 and B.1.3.

Following this procedure, we have written a Python script that uses RADEX and analyses our data in an automated way.[4] In this way, we obtained the column density values for methanol, acetonitrile and cyanoacetilene in L1517B shown in Table 3.3.2.

B.2 Propagation of uncertainties and upper limits

Let's consider a variable x with a central value μ and an uncertainty σ. There can be different interpretations or definitions of the central value and the uncertainty with respect to the probability density function (PDF) associated with the variable. The most typical one is to consider that μ is the mean of the PDF of x and σ is its standard deviation. Another possible interpretation is that μ is the median of the PDF of x and $(\mu - \sigma, \mu + \sigma)$ is the 1 σ confidence interval (68.27 %) around μ. Another valid interpretation would be that μ is the mode of the PDF of x and the interval $(\mu - \sigma, \mu + \sigma)$ is the credibility interval around μ. In RichValues, we prefer to base the PDFs in quantiles/percentiles, therefore interpreting μ as the median

[4] `https://github.com/andresmegias/radex-python/`

of the PDF (the 50 % percentile) and $(\mu - \sigma, \mu + \sigma)$ as the 1 σ confidence interval (68.27 %). One of the reasons for this choice is that the quantiles (including the median) are robust to outliers. Another reason is that the median and quantiles transform directly through strictly increasing or decreasing functions; that is, the median of the distribution resulting of applying a function f to an input distribution of values $\{x_i\}$ is just the function applied to the median of the original distribution, $\text{median}\, f(\{x_i\}) = f(\text{median}\, \{x_i\}) = f(\mu_i)$.[5] Lastly, this interpretation allows for asymmetric uncertainties.

Then, let's consider a set of n variables $\{x_j\}$. To propagate the uncertainties through a function f applied to the input variables, we can draw a sample of a large number of values (e.g., $\gtrsim 10^4\sqrt{n}$) of each variable x_j, apply the function to each of the elements of the samples, and then obtain a central value and an uncertainty for the resulting distribution. To do so, we need two things: an appropiate PDF for converting each variable x_i to a distribution of values, and a proper method to obtain a central value and an uncertainty from the resulting distribution. We will use median as the central value and the 1 σ confidence interval with respect to it (that includes 68.27 % of the distribution) to obtain the lower and upper uncertainties. As for the PDF, if the domain of the variable x is $(-\infty, \infty)$, a proper function is the normal distribution:

$$\text{PDF}(x) \;=\; \frac{1}{\sqrt{\tau}\,\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad , \tag{B.2}$$

with $\tau \equiv 2\pi$.[6] However, if the domain of the variable is not $(-\infty, \infty)$, this function would be incorrect. Therefore, we have to use another function as the PDF.

Let's suppose a domain (b_1, b_2). We define the left and right amplitudes, a_1 and a_2, as the distances between the limits of the domain and the median, that is, $a_j = |b_j - \mu|$ for $j = 1, 2$. Now, as these amplitudes can be different, we will split our desired PDF in two halfs, one for $x \leq \mu$ and other for $x > \mu$. Then, we will use an amplitude a as a reference, which must be greater than the uncertainty, $a > \sigma$. If the amplitude is quite greater than the uncertainty, $a \gg \sigma$, a good PDF would be just the normal distribution truncated to the domain (b_1, b_2). However, for amplitudes closer to the uncertainty, it would be considerably incorrect, as the truncation shifts the median of the distribution and modifies the confidence intervals, and thus the uncertainties. To fix this, we make the following variable change:[7]

$$x - \mu \;\;\rightarrow\;\; \tilde{x} - \mu \;\equiv\; \frac{4}{\tau}\, a\, \tan\left(\frac{\tau}{4}\,\frac{x-\mu}{a}\right) \quad . \tag{B.3}$$

Using this new variable $\tilde{x}$ with a normal distribution, we are able to compress the original domain of $(-\infty, \infty)$ to $(-a, a)$. However, we get two disadvantages: firstly, the normalization constant is now different, and secondly, the relation between the parameter of the standard

[5] The same would be valid for any quantile, since strictly increasing or decreasing functions preserve order.

[6] See: `https://tauday.com/tau-manifesto`.

[7] A better variable change was eventually found, as explained in Appendix C.7. With the updated variable change, the formulas and the interpretation of the PDFs are simplified. However, the original variable change is explained here for the sake of consistency with the original paper.

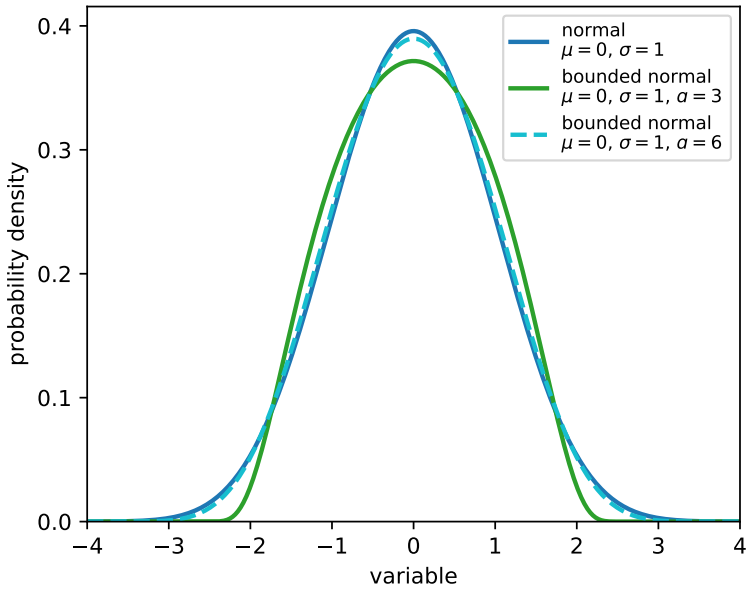


Figure B.2.1. Probability density functions for a normal distribution and two bounded normal distributions with diferent domains.

deviation of the original normal distribution and the 1 σ confidence interval (from which we define the uncertainty, σ) is now different; therefore, we should rename the standard deviation of the original normal distribution to s, which we will call *width*. The first change is not a problem, as the PDF will be normalized computationally for each case. As for the second one, we have characterized computationally the relation between σ and s. It happens that s/σ decreases with a/σ, having $s/\sigma \simeq 2.65$ when $a/\sigma = 2$ and $s/\sigma \to 1$ when $a/\sigma \to \infty$. We have then a relation between the width, the uncertainty and the amplitude, that is, $s = s(\sigma, a)$.

Therefore, the resulting PDF would be the following:

$$\mathrm{PDF}(x) \propto \exp\left(-\frac{1}{2}\left(\frac{\frac{4}{\tau}a\tan\left(\frac{\tau}{4}\frac{x-\mu}{a}\right)}{s(\sigma,a)}\right)^2\right) . \tag{B.4}$$

We observed that s/σ increases quasi-exponentially as a/σ approaches a minimum value of ~ 1.47. Moreover, we discovered that if $a/\sigma \lesssim 1.7$, the estimated relation $s(\sigma, a)$ starts to be incorrect because of the increasing dispersion in s. Therefore, we must use another PDF for the cases in which $a/\sigma \lesssim 1.7$. Actually, for that limit case the shape of the PDF is almost a uniform distribution between $-a$ and a. We can model this PDF as a trapezoidal function with a rectangular core and triangles in the edges. By doing so, one can easily demonstrate that this kind of PDF would only work if a/σ is greater than the inverse of the integrated area corresponding to the 1 σ confidence interval (~ 0.683); that is, $a/\sigma \gtrsim 1.46$, which is consistent with the found asymptote of the calculated relationship $s(\sigma, a)$.

We then choose to only use this distribution when $a \geq 2\sigma$, for the sake of simplicity and following a conservative approach. For $a \gg \sigma$, this PDF tends to a normal PDF, so we use this general PDF instead of a truncated gaussian for this regime (see Figure B.2.1). We call

this distribution a *bounded normal distribution*.

Finally, for $\sigma < a < 2\sigma$, we use a shifted and mirrored lognormal distribution whose PDF would be:

$$f(x) = \frac{1}{\sqrt{\tau}\, a \left|\ln\left(1-\frac{\sigma}{a}\right)\right|} \frac{1}{1-\frac{|x-\mu|}{a}} \exp\left(-\frac{1}{2}\left(\frac{\ln\left(1-\frac{|x-\mu|}{a}\right)}{\ln\left(1-\frac{\sigma}{a}\right)}\right)^2\right) \quad . \tag{B.5}$$

This PDF meets our requirements: the median is equal to μ and the uncertainty associated with the 1 σ confidence interval is σ. We call this distribution a *mirrored lognormal distribution*.

We have defined our PDFs for the case of a variable x with a central value μ and an uncertainty σ, building the final PDF with two halves with amplitude $a = a_j$ for $j = 1, 2$. In case we had lower and upper uncertainties, σ_1 and σ_2, we should just replace σ by σ_1 for the left half of the PDF and by σ_2 for the right half.

Lastly, we should also address the case of a variable with an upper/lower limit or even a finite interval. Let's consider an interval (x_1, x_2), which may be finite or infinite. If it is finite, we choose a uniform distribution between x_1 and x_2 as the corresponding PDF. But if it is infinite, we choose a log-uniform distribution with finite thresholds for 0 and $\pm\infty$, which we set to $\pm 10^{-90}$ and $\pm 10^{90}$. For example, for an interval of $(-100, \infty)$, we would build a sample $\{x_-\}$ from a uniform distribution between -90 and 2 and a sample $\{x_+\}$ from a uniform distribution between -90 and 90. Our final distribution would be the joining of the samples of $\{-10^{\{x_-\}}\}$ and $\{10^{\{x_+\}}\}$.

To sum up, if we have a set of variables with central values μ_i and uncertainties σ_1, σ_2, we first build distributions $\{x_i\}$ using the mentioned PDFs. Then, we apply the function to the distributions, $f(\{x_i\})$, obtaining a new distribution. Finally, we use an algorithm to detect if the distribution corresponds to an interval (that can be an upper/lower limit) or a defined value with uncertainties, and derive the corresponding parameters.

Using this approach, we have written a Python library that allows to deal with values with uncertainties and upper/lower limits, performing the uncertainty propagation automatically.[8] We have used it throughout the calculations of this paper, e.g., for obtaining the values shown in Fig. 3.6.1.

B.3 Abundance ratio of cianoacetylene and acetonitrile

Table C1 shows the abundance ratio of cyanoacetylene (HC_3N) and acetonitrile (CH_3CN) through different sources, from starless cores to comets. The propagation of the uncertainties and the upper limits in the column density is done through simulations using statistical distributions (see Appendix B.2).

[8] `https://pypi.org/project/richvalues/`

Table C1. Column densities of HC_3N and CH_3CN (N) and abundance ratio of this molecules (χ_{HC_3N} / χ_{CH_3CN}) for the starless cores L1517B, L1498, L1544, and L1521E, 12 protostellar systems, 4 protoplanetary discs, and 2 comets.

Source			N_{HC_3N} (cm^{-2})	N_{CH_3CN} (cm^{-2})	χ_{HC_3N} / χ_{CH_3CN}
starless cores	L1517B	dust peak	$(5.16 \pm 0.05) \times 10^{13}$	$(2.1 \pm 0.6) \times 10^{11}$	250^{+100}_{-50}
		methanol peak	$3.72^{+0.14}_{-0.13} \times 10^{12}$	$< 2.0 \times 10^{11}$	> 17
	L1498	dust peak	$(1.6 \pm 0.3) \times 10^{13}$	$< 8 \times 10^{10}$	> 100
		methanol peak	$(2.2 \pm 1.0) \times 10^{13}$	$(1.0 \pm 0.1) \times 10^{11}$	220^{+100}_{-100}
	L1544	dust peak	$(1.0 \pm 0.3) \times 10^{14}$	$(1.5 \pm 0.2) \times 10^{11}$	670^{+230}_{-200}
		methanol peak	$(4.2 \pm 0.4) \times 10^{13}$	$< 9.1 \times 10^{10}$	> 310
	L1521E	dust peak	$(8.4 \pm 2.5) \times 10^{12}$	$(4.8 \pm 1.4) \times 10^{11}$	18^{+9}_{-6}
protostellar systems	B1-a		$(4.2 \pm 1.2) \times 10^{12}$	$(4.9 \pm 1.1) \times 10^{11}$	9^{+3}_{-3}
	B1-c		$(4 \pm 3) \times 10^{12}$	$(3.5 \pm 0.6) \times 10^{11}$	12^{+11}_{-9}
	B5 IRS1		$(3.2 \pm 1.2) \times 10^{12}$	$< 1.7 \times 10^{11}$	> 5
	IRAS 03235		$(3.9 \pm 1.0) \times 10^{12}$	$(1.6 \pm 0.9) \times 10^{11}$	25^{+31}_{-11}
	IRAS 03245		$(4 \pm 3) \times 10^{12}$	$(1.9 \pm 1.5) \times 10^{11}$	19^{+69}_{-15}
	IRAS 03271		$(1.7 \pm 1.4) \times 10^{12}$	$(1.9 \pm 1.1) \times 10^{11}$	9^{+14}_{-7}
	IRAS 23238		$(3.2 \pm 2.6) \times 10^{12}$	$(1.4 \pm 0.3) \times 10^{11}$	22^{+21}_{-18}
	L1014 IRS		$(4.4 \pm 3.6) \times 10^{11}$	$< 6 \times 10^{10}$	> 0.7
	L1455 IRS3		$(6 \pm 5) \times 10^{11}$	$< 9 \times 10^{10}$	> 0.7
	L1455 SMM1		$(2.4 \pm 1.9) \times 10^{12}$	$< 1.1 \times 10^{11}$	> 2.2
	L1489 IRS		$(3.5 \pm 2.4) \times 10^{12}$	$< 1.4 \times 10^{11}$	> 0.3
	SVS 4-5		$(1.1 \pm 0.3) \times 10^{13}$	$(5.2 \pm 0.9) \times 10^{11}$	21^{+8}_{-6}
protoplanetary discs	GM Aur		$1.9^{+0.4}_{-0.4} \times 10^{13}$	$2.1^{+0.2}_{-0.1} \times 10^{12}$	$8.8^{+2.0}_{-1.9}$
	As 209		$2.9^{+0.5}_{-0.5} \times 10^{13}$	$1.7^{+0.2}_{-0.2} \times 10^{12}$	17^{+4}_{-3}
	HD 163296		$7^{+3}_{-2} \times 10^{13}$	$2.3^{+0.2}_{-0.2} \times 10^{12}$	32^{+11}_{-9}
	MWC 480		$8^{+4}_{-3} \times 10^{13}$	$3.5^{+0.2}_{-0.2} \times 10^{12}$	22^{+11}_{-8}
Source			χ_{HC_3N} / χ_{H_2O}	χ_{CH_3CN} / χ_{H_2O}	χ_{HC_3N} / χ_{CH_3CN}
comets	46P		$< 3 \times 10^{-5}$	$(1.7 \pm 0.01) \times 10^{-4}$	< 0.21
	67P		4×10^{-6}	5.9×10^{-5}	0.068

References for each source: L1517B: this work; L1498: Jiménez-Serra et al. (2021); L1544: Jiménez-Serra et al. (2016); L1521E: Nagy et al. (2019); protostellar systems: Bergner et al. (2017); protoplanetary discs: Ilee et al. (2021); comets: Biver et al. (2021).

Appendix C

Chapter 4

This appendix corresponds to Chapter 4, that is, for the paper titled *A fast machine learning tool to predict the composition of astronomical ices from infrared absorption spectra*, published in the journal Astronomy & Astrophysics (Megías et al., 2024).

C.1 Astronomical ice spectra

In astronomical observations of ices, we normally measure the flux of light received towards a background source, $\mathcal{F}$, or alternatively the spectral flux. To convert it to optical depth (τ) or absorbance (A), we need to know or estimate the non-absorbed continuum, $\mathcal{F}_0$, to then apply the logarithm of the ratio of fluxes:

$$\tau \;=\; \ln\left(\mathcal{F}_0/\mathcal{F}\right) \;=\; A\ln(10) \;\;. \tag{C.1}$$

The continuum can be estimated by different means, although a common practice is to fit a black-body curve or a polynomial curve (e.g.: Rocha et al., 2021; McClure et al., 2023). In this scale, the contribution of several macroscopic layers of ice combines additively, which allows to derive the quantity of ice (column density, N) integrating the area of certain bands and scaling by the experimentally derived areas (band strengths, S):

$$N \;=\; \frac{\int \tau(\tilde{\nu})\,\mathrm{d}\tilde{\nu}}{S} \;=\; \frac{\ln(10)}{S}\int A(\tilde{\nu})\,\mathrm{d}\tilde{\nu} \quad , \tag{C.2}$$

where $\tilde{\nu}$ is the wavenumber, that is, the inverse of the wavelength (λ), $\tilde{\nu} = 1/\lambda$. Before searching for molecular features, one more step needs to be done: subtraction of the absorption due to the silicate-rich core of the dust grains, which has a noticeable effect in the $\sim$8–12 μm range; this can be done with polynomials or with silicate models, like the ones provided in the software OpTool (Dominik et al., 2021).

These steps can be performed within AICE using its three pre-processing modules, that we used to reduce the spectra of NIR38 and J110621 taken by JWST by McClure et al. (2023).

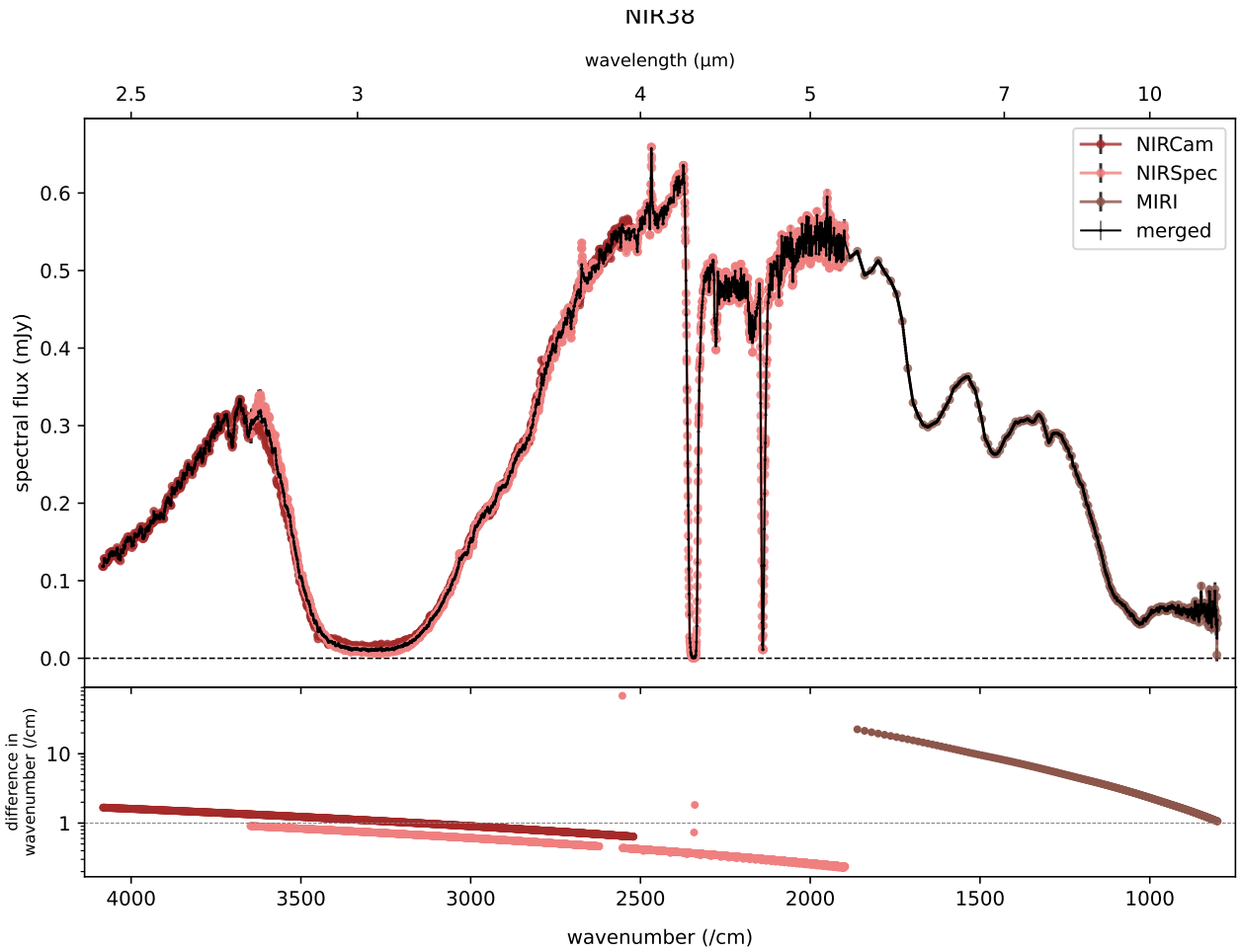


Figure C.1.1. Spectra taken by JWST towards the background star NIR38 in the molecular cloud Cha 1, with its three instruments, and resulting merged spectrum with the desired resolution of 1 cm^{-1}. Lower pannel shows the spacing in wavenumber between each point of the spectra, to visualize the resolution of the data.

The first module allows to merge spectra coming from different instruments to a single spectrum expressed in wavenumber with a certain resolution (Fig. C.1.1). The rebinning of the spectra is done with the help of the SpecRes Python tool by Carnall (2017). For the case of the measurements obtained with the MIRI instrument, which have a poorer resolution (LSR mode), we supersampled the spectra to get a resolution of 1 cm^{-1}, increasing artificially the uncertainties to make them consistent with the original data.

The second one allows to fit several polynomials to the observed spectrum, in logarithmic scale and in different wavelength regions, to construct an estimation of the continuum, in order to convert the spectra from flux to absorbance (Fig. C.1.2). This is the same method as used by McClure et al. (2023), although we do not obtain exactly the same continuum due to our different wavenumber spacing and a slightly different selection of fitting regions. The propagation of the uncertainties from flux to absorbance is done with the Python library RichValues (see Appendix C.7)

Lastly, the third module of pre-processing allows to fit a silicate contribution to the absorbance spectrum in certain regions with no expected ice bands (Fig. C.1.3), using the

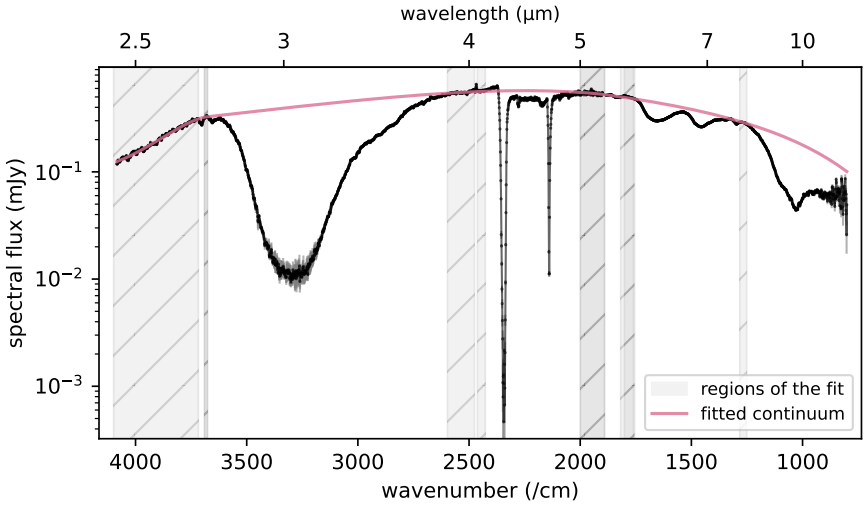


Figure C.1.2. Spectrum taken by JWST towards the background star NIR38 in the molecular cloud Cha 1, and fitted continuum, used to compute the absorbance spectrum. The fit was performed joining several polinomials in the dashed regions and extrapolating to the rest.

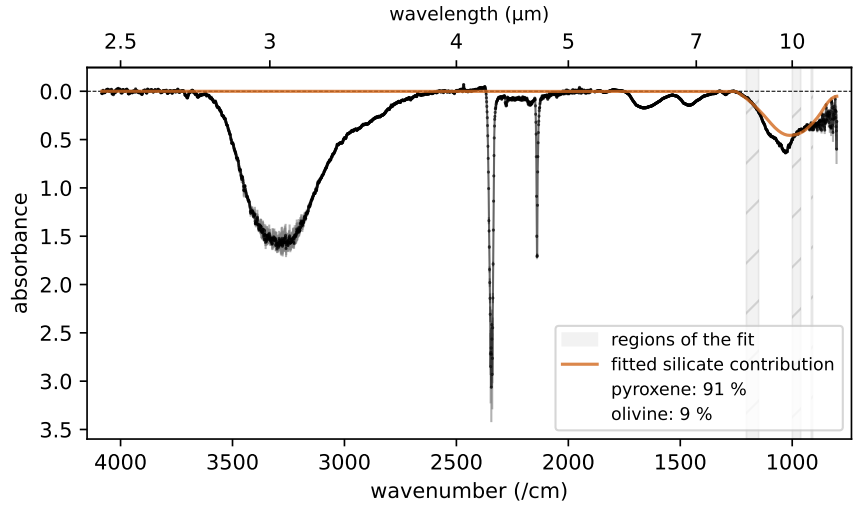


Figure C.1.3. Spectrum in absorbance derived from data by JWST towards the background star NIR38 in the molecular cloud Cha 1, and fitted silicate contribution, which has to be subtracted to obtain the ice spectrum. The fit was performed using the software OpTool, in the dashed regions.

software OpTool (Dominik et al., 2021). We produce a synthetic spectrum and scale and shift it to fit the observed spectrum in the fitted regions, using a grain size of 1 μm and a composition that include a combination of pyroxene and olivine, as done by McClure et al. (2023); however, we fit the different fractions of both types of compositions instead of using fixed compositions (the results are similar).

C.2 Artificial neural networks

This section explains in more detail the definition of artificial neural networks, focusing on multi-layer perceptrons, and how the training procedure works.

Mathematical definition

A multi-layer perceptron is a mathematical function based on the concatenation of linear combinations and non-linear activation functions (Murtagh, 1991). Starting from a set of n numeric inputs, $\{x_i\}$, this function f applies a set of operations to them in several steps or layers, ending with a set of output values, $\{\hat{y}_j\} = f(\{x_i\})$. In the first layer, the inputs are converted to a set of intermediate numbers, $\{a_j\}$; each one of these values is the result of a different linear combination of the input parameters, plus a non-linear function ϕ applied to it (Grosan and Abraham, 2011):

$$a_j = \phi\left(\sum_{i=1}^{n} w_{ji}\, x_i + b_j\right) , \tag{C.3}$$

where w_{ji} are the weights of the linear combination j, and b_j are just different offsets for each of them, called *biases*. The activation function ϕ can adopt several forms, but a common use is the rectified linear unit (ReLU), defined as $\phi(x) = \max(0, x)$ (Fukushima, 1975; Nair and Hinton, 2010).

Each one of the mentioned operations constitute the *neurons*, and the $\{a_j\}$ are the activations of each neuron; all of them constitute the first *hidden layer* of the neural network. Then, the same operation (equation C.3) is applied to the first layer as input values, but with different values of w_{ji} and b_j, obtaining as a result a second hidden layer. Therefore, the activations of the layer l can be generalized as:

$$a_{lj} = \phi\left(\sum_{i=1}^{n_{l-1}} w_{lji}\, a_{l-1,i} + b_{lj}\right) , \tag{C.4}$$

where n_{l-1} is the number of activations of the previous layer, and having that $a_{0,i} = x_i$. After a certain number L of steps or layers, we end up with a final set of numbers, $\{\hat{y}_i\}$, which constitute the final output of the function f, the output layer. In this last set of operations, where $a_{L,j} = \hat{y}_j$, the activation function can adopt different shapes depending on the desired format for the output. For example, if the outputs have to be between 0 and 1, a sigmoid function can be used: $\phi(x) = 1 \,/\, (1 + \mathrm{e}^{-x})$; in another case, if the outputs have to be positive, we can still use a ReLU as activation function.

Training of the neural network

The artificial neural networks is a function f with several parameters: the weights w_{lji} and the biases b_{lj}, where the indices refer to the neuron j of the layer l and the previous activation i. We can see that even a small model with a few layers and neurons would have plenty of free parameters. Thanks to its multiple parameters and to the non-linear activation functions, artificial neural networks have a huge versatility to realize several different operations and transformations to the input data. In particular, they can be 'trained' to imitate and learn a particular transformation or operation as long as we have several examples

of it (Larochelle et al., 2009).

For the sake of simplicity, let us assume that the output of the neural network is a single value, $\hat{y}$, and that the input is a set of values that we group into a vector $\boldsymbol{x}$, so that $\hat{y} = f(\boldsymbol{x})$. We start from a group of m inputs, $\{\boldsymbol{x}\}$, and we also know the corresponding m values after the transformation, $\{y\}$. The set of pairs $(\boldsymbol{x}, y)$ is called *training dataset*. We want the neural network f to return each value y_k when applied to the corresponding vector $\boldsymbol{x}_k$. This translates into an optimization problem, where we want to minimize the difference between the predictions of the model, $\{\hat{y}_k\} = f(\{\boldsymbol{x}_k\})$, and the reference values, $\{y_k\}$. To quantify this difference, we define an error function, or loss function ($\mathcal{L}$), which can be the known mean squared error (MSE):

$$\mathcal{L} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_k - y_k)^2 \ , \tag{C.5}$$

where the sum ranges over the examples in our training dataset. The process of optimizing the parameters of f (the weights and biases) to minimize the loss $\mathcal{L}$ is called *training*. This is analogous to a linear fit to a set of m points (x, y), in which case the size of both the input and the output is 1. In case that both the input and the output are vectors, $(\boldsymbol{x}, \boldsymbol{y})$, we would use equation C.5 in a vectorial form, calculating the squared norm of the vectorial difference between $\hat{\boldsymbol{y}}_k$ and $\boldsymbol{y}_k$, $||\hat{\boldsymbol{y}}_k - \boldsymbol{y}_k||^2$.

In the training of a neural network, its parameters are initialized randomly and then modified in consecutive steps or *epochs*, to explore the parameter space and find values that minimize the loss function when applying f to the different input values, $\{\boldsymbol{x}_i\}$. This process of finding the minimum of the loss function is known as *gradient descent*, and in neural networks this is possible thanks to the *backpropagation* algorithm (Baldi, 1995; LeCun et al., 1998). Gradient descent comprises several variations of the same method with the same objective. In all of them, there is a parameter called 'learning rate' that controls how large are the changes in the parameters of the network after each training epoch.

Lastly, there is an additional step that must be done before starting the training. In order to avoid overfitting, the neural network should not be trained using the whole training dataset. Instead, we randomly select a small fraction of it (e.g., $\sim$10 %) and separate it from the rest; this subset is called *validation dataset*, and will be used to quantify the progress in the training of our model. While the fitting will be performed on the training subset, the loss function will be also evaluated on the validation subset, informing about the performance of the model f in new or unseen data. The training steps will be repeated in several epochs that will reduce the loss function in both subsets, on average, until the validation loss stops improving significantly.

Additional improvements

We list here some additional procedures that can improve the training of neural networks, and that we used to train AICE:

- *Add random noise to the input.* Adding a bit of Gaussian noise to the input in every epoch of the training, and renewing it in each iteration, usually helps the neural network to achieve a better generalization, preventing it from memorizing specific patterns on the data that are not really representative of the dataset.
- *Split the training subset in mini-batches.* We evaluate the loss function and apply the gradient descent to update the network parameters for each mini-batch (a group of data elements), covering all the dataset in several steps, which still define the epochs. In this way, we accelerate the training, since we increase number of times the optimization is performed.
- *Use batch normalization.* This means that, for each mini-batch, we standarize the output of a layer (activations, a), by subtracting its mean along the batch (μ_a) and dividing by its standard deviation (σ_a), and then we apply a linear transformation with two learnable parameters (Ioffe and Szegedy, 2015). The transformed activation would be:

$$a' = \gamma_a \frac{a - \mu_a}{\sqrt{\sigma_a^2 + \epsilon}} + \beta_a \ , \tag{C.6}$$

  where γ_a and β_a are the new trainable parameters (a weight and a bias) and ϵ is a small term introduced to avoid division by 0 (in our case, $\epsilon = 10^{-3}$). For inference (after the training), the mean and variance in the activations for the whole training dataset are used as μ_a and σ_a^2, since the split in batches is only done during training. This is applied for all the activations of the layer, therefore having a different set of (μ_a, σ_a, γ_a, β_a) parameters for each neuron in the layer (for each activation a).
- *Apply dropout to some activations.* This consists on setting a random fraction of the activations of a certain layer to 0, with a certain probability, which can help to reduce overfitting (Srivastava et al., 2014). In our case, we only use dropout after the first hidden layer, with a low probability of 0.1.
- *Reduce the learning rate along the training.* The learning rate defines the size of the changes in the parameters along the training of the networks. Reducing this value after a certain number of epochs with no significant reduction of the loss can help to improve the optimization of the parameters.

C.3 Baseline removal

In order to remove the experimental baselines from our laboratory training dataset, and for some additional processing, we have used a tool included within AICE, that we call AICE Interactive Toolkit.

This software is a Python 3 script that allows to read absorbance spectra in plain text files, plotting the input spectrum (or spectra) and allowing the user to interact via the mouse and keyboard to perform different actions. In the following, we present an example of the use of this tool to process the experiment #12 from LIDA, corresponding to a mixture of H_2O : HCOOH (91:9), published originally by Bisschop et al. (2007).

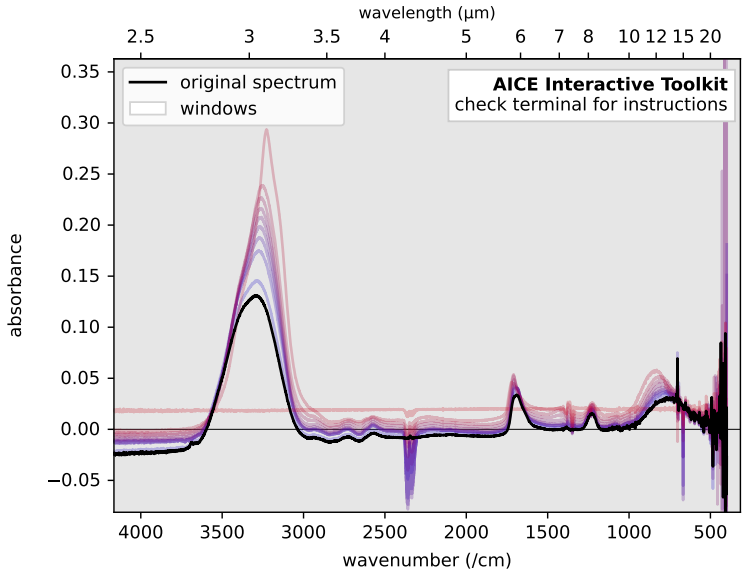


Figure C.3.1. Spectra obtained in the experiment #12 of LIDA, from a mixture of H_2O : HCOOH (91:9), as seen with our AICE Interactive Toolkit. The curve in black is the spectrum at 15 K, and the rest of coloured curves correspond to spectra with increasing temperature, with a step of 15 K until 165 K (which is almost flat, as both species have desorbed). The spectra were originally published by Bisschop et al. 2007.

Figure C.3.1 shows the spectra as seen after launching our interactive tool from the terminal and loading the input data. The black curve is the spectrum at 15 K, while the coloured curves correspond to the spectra recorded at different temperatures, from 30 K to 165 K. We can see that each spectrum has a different offset or baseline, that should be removed. Before that, we have to exclude the data corresponding to more than 100 K since we focus on spectra with temperatures of $\leq$100 K. Additionally, we can spot some contamination and noise between 2500 and 2000 cm^{-1} and at the very right of the spectra, which will be corrected after the baseline subtraction.

In order to fit the baseline for all the spectra in this experiment, we first select several spectral windows where the spectra should be flat, i.e., with no absorption. After manually selecting the windows with the cursor, we can now press R to fit a baseline in those windows (for each recorded temperature) and reduce the spectra removing the fitted curves. The mathematical form of the fitted Gaussian is a spline, that is, a concatenation of polynomials, of third order. We use the function/class UnivariateSpline from Python's package SciPy, which requires as an input a smooth factor. In order to compute it, we require a different smooth parameter, s, that can be input by the user (by default, this parameter is $s = 25$). We first smooth our spectra in the selected windows applying a rolling median with a size of s. Then, we compute the sum of the squared errors between the smoothed curve and the original spectrum, in the selected windows; this is the smooth factor that we input to UnivariateSpline. In this way, the condition to be satisfied to select the number of polynomials for the spline (the number of *knots*) is that the sum of the squared errors between the spline and the original spectrum (in the selected windows) is less or equal to the sum of squared errors previously calculated (between the spectrum and its smoothed version). Equivalently, we could use the mean squared error (MSE) instead of the sum of the squared errors.

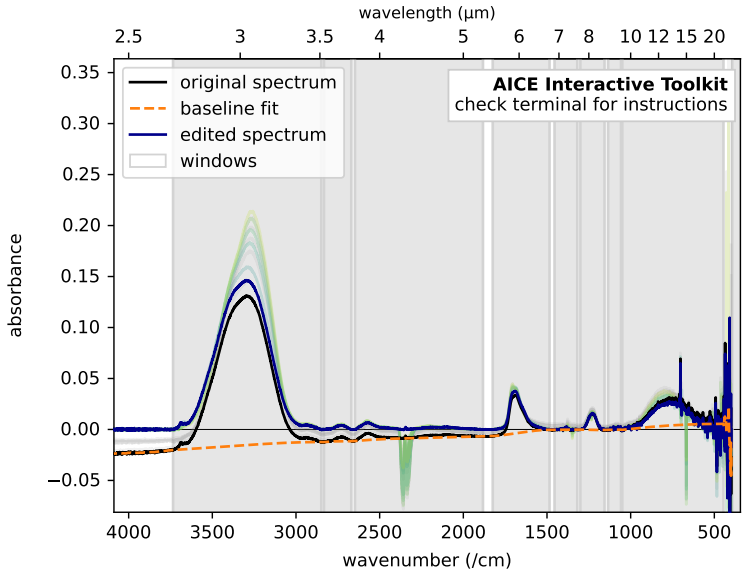


Figure C.3.2. Same as Fig. C.3.1, after removal of the spectra with temperatures greater than 100 K and after baseline subtraction. Now the curve in orange shows the fitted baseline (in the white spectral windows) for the spectrum at 15 K (black curve), and the blue curve shows the corresponding reduced spectrum. The rest of coloured curves are the reduced spectra for the rest of temperatures.

Figure C.3.2 shows the result obtained after performing these steps, with the default value of $s = 25$. We can now see that the reduced spectra have an absorption of zero in the selected regions. At the rightmost part of the spectrum, the baseline presents too much variation, but we do not try to fix this since we are interested in the 4000 to 980 cm^{-1} range.

We have performed the most important steps to reduce the spectra from this experiment. However, additional corrections should be applied to the data. Again in the rightmost part of the spectra, too much noise can be seen. Since we do not have very thin absorption features in the spectra, we can apply a rolling mean to smooth the data in all the spectral range; otherwise, we should apply this excluding the regions with thin features. Therefore, we applied a smoothing with a size of 7 channels, pressing Tab to remove the windows selected for the baseline fit and then Shift+S, introducing 7 as the size of the smooth parameter.

After that, we focus on two regions where signs of contamination can be seen. The most noticeable one appears at around 2400–2300 cm^{-1}, centered in around 2340–2330 cm^{-1}. This is certainly very similar to the main absorption band of CO_2, so it seems that it is due to some gaseous CO_2 present in the experimental chamber. Similarly, we see a pattern around 1400–1300 cm^{-1}. The left half of it (1400–1350 cm^{-1}) corresponds to an absorption band of HCOOH, but the second half (1350–1300 cm^{-1}) seems to be an artifact or a contamination whose origin we cannot identify. Additionally, there is an artifact at $\sim$1230 cm^{-1}. In order to remove all of these features, we select with the cursor the three spectral windows where these contaminations/artifacts are present. Then, by pressing X, our tool removes the data in the selected regions and interpolates the spectra in the window edges using polynomials. Fig. C.3.3 shows the resulting spectra, zoomed into the region of interest.

Once this is done, the spectra processing is virtually done. Optionally, we can artificially

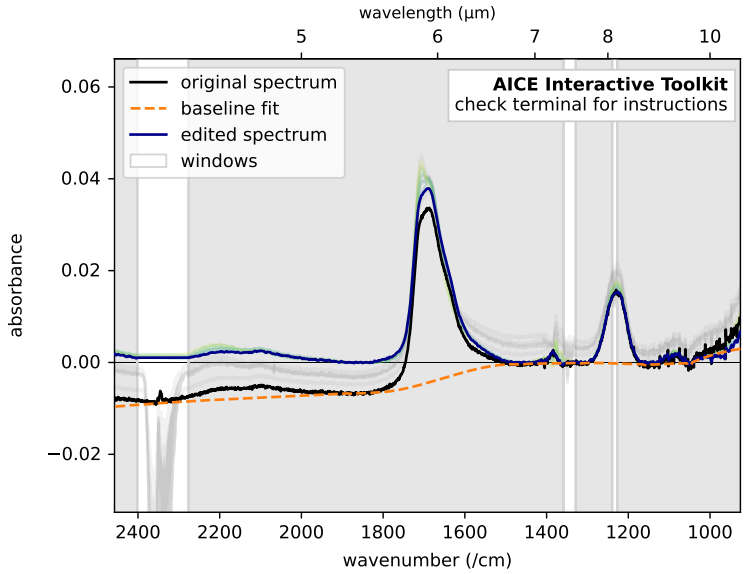


Figure C.3.3. Same as Fig. C.3.2 in a zoomed region, after a smoothing the spectra with a size of 7 and interpolating two contaminated regions. The curve in black shows the original spectrum at 15 K, while the curves in gray represent the original spectra at different temperatures.

add random noise in the last spectral windows to avoid having a fully smooth curve in the interpolated regions. That can be done by pressing N, and also S if too much noise have been added. We could also remove a negative feature at ∼1380 cm^{-1} for the curve at 90 K, pressing X again or even 0 to set the values in the selected regions to zero. In principle, we could also resample the data to have our desired wavenumber range of 4000 to 980 cm^{-1} with a spacing of 1 cm^{-1} within our software, but we performed this step independently with a different Python script, in order to automatize it and make it faster. Therefore, to exit AICE Interactive Toolkit and save the reduced spectra, we press Ctrl+Enter. A log file will also be created containing the list of actions performed to the spectra and some additional information about the input and output files.

C.4 List of training dataset

Table C.4.1 shows the list of all the experiments that constitute our whole training dataset, stating the composition of the ice and the original resolution, as well as the reference. Note that for the majority of the experiments we have several spectra available corresponding at different temperatures.

Table C.4.1. List of experiments contained in our training dataset, containing information such as the original database, the composition and the scientific reference.

Database	#	Composition	Proportion	Resol. (cm^{-1})	Reference
LIDA	1	HCOOH	1	0.96	Bisschop et al. (2007)
LIDA	3	HCOOH + CH_3OH	10 + 90	0.96	Bisschop et al. (2007)
LIDA	4	HCOOH + CO	11 + 89	0.96	Bisschop et al. (2007)
LIDA	5	HCOOH + H_2O	20 + 80	0.96	Bisschop et al. (2007)
LIDA	6	HCOOH + H_2O	34 + 66	0.96	Bisschop et al. (2007)

Continued...

... Continued.

Database	#	Composition	Proportion	Resol. (cm^{-1})	Reference
LIDA	7	$HCOOH + H_2O$	50 + 50	0.96	Bisschop et al. (2007)
LIDA	8	$HCOOH + H_2O + C^{18}O_2$	6 + 67 + 27	0.96	Bisschop et al. (2007)
LIDA	9	$HCOOH + H_2O + CH_3OH$	6 + 68 + 26	0.96	Bisschop et al. (2007)
LIDA	10	$HCOOH + H_2O + CO$	8 + 62 + 30	0.96	Bisschop et al. (2007)
LIDA	11	$HCOOH + C^{18}O_2$	9 + 91	0.96	Bisschop et al. (2007)
LIDA	12	$HCOOH + H_2O$	9 + 91	0.96	Bisschop et al. (2007)
LIDA	13	H_2O	1	1.9	Öberg et al. (2007)
LIDA	14	H_2O	1	1.9	Öberg et al. (2007)
LIDA	15	$C^{18}O_2$	1	1.9	Öberg et al. (2007)
LIDA	16	$C^{18}O_2$	1	1.9	Öberg et al. (2007)
LIDA	17	$H_2O + C^{18}O_2$	1 + 1	1.9	Öberg et al. (2007)
LIDA	18	$H_2O + C^{18}O_2$	1 + 2	1.9	Öberg et al. (2007)
LIDA	19	$H_2O + C^{18}O_2$	2 + 1	1.9	Öberg et al. (2007)
LIDA	20	$H_2O + C^{18}O_2$	2 + 1	1.9	Öberg et al. (2007)
LIDA	21	$H_2O + C^{18}O_2$	1 + 1	1.9	Öberg et al. (2007)
LIDA	22	$H_2O + C^{18}O_2$	1 + 2	1.9	Öberg et al. (2007)
LIDA	23	$H_2O + C^{18}O_2$	4 + 1	1.9	Öberg et al. (2007)
LIDA	25	$H_2O + C^{18}O_2$	1 + 1	1.9	Öberg et al. (2007)
LIDA	26	$H_2O + CO_2$	10 + 1	0.96	Ehrenfreund et al. (1999)
LIDA	27	$CO_2 + H_2O$	1 + 1	1.8	Ehrenfreund et al. (1999)
LIDA	28	$CO_2 + CH_3OH$	10 + 1	0.96	Ehrenfreund et al. (1999)
LIDA	29	$CO_2 + CH_3OH$	3 + 1	0.96	Ehrenfreund et al. (1999)
LIDA	30	$CO_2 + CH_3OH$	2 + 1	0.96	Ehrenfreund et al. (1999)
LIDA	31	$CO_2 + CH_3OH$	1 + 1	1.4	Ehrenfreund et al. (1999)
LIDA	32	$CO_2 + CH_3OH$	1 + 2	0.96	Ehrenfreund et al. (1999)
LIDA	33	$CO_2 + CH_3OH$	1 + 3	0.96	Ehrenfreund et al. (1999)
LIDA	34	$CO_2 + CH_3OH$	1 + 10	0.96	Ehrenfreund et al. (1999)
LIDA	35	$H_2O + CH_3OH + CO_2$	9 + 1 + 2	1.9	Ehrenfreund et al. (1999)
LIDA	36	$H_2O + CH_3OH + CO_2$	0.2 + 0.6 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	37	$H_2O + CH_3OH + CO_2$	0.4 + 0.6 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	38	$H_2O + CH_3OH + CO_2$	1 + 0.6 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	39	$H_2O + CH_3OH + CO_2$	0.7 + 0.7 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	40	$H_2O + CH_3OH + CO_2$	0.8 + 0.9 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	41	$H_2O + CH_3OH + CO_2$	1 + 1 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	42	$H_2O + CH_3OH + CO_2$	0.7 + 1 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	43	$H_2O + CH_3OH + CO_2$	0.6 + 1 + 0.8	1.9	Ehrenfreund et al. (1999)
LIDA	44	$H_2O + CH_3OH + CO_2$	1.2 + 0.7 + 1	0.96	Ehrenfreund et al. (1999)
LIDA	45	$H_2O + CH_3OH + CO_2$	0.5 + 1 + 1	0.96	Ehrenfreund et al. (1999)
LIDA	47	$H_2O + CH_3OH + CO_2$	0.2 + 0.5 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	48	$H_2O + CH_3OH + CO_2$	0.3 + 0.5 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	49	$H_2O + CH_3OH + CO_2$	0.3 + 0.7 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	50	$H_2O + CH_3OH + CO_2$	1.1 + 1.2 + 1	0.96	Ehrenfreund et al. (1999)
LIDA	51	$H_2O + CH_3OH + CO_2$	0.9 + 0.3 + 1	1.9	Ehrenfreund et al. (1999)
LIDA	52	$H_2O + CH_3OH + CO_2 + NH_3$	0.7 + 0.7 + 1 + 0.7	0.96	Ehrenfreund et al. (1999)
LIDA	53	$H_2O + CH_3OH + CO_2 + CH_4$	0.6 + 0.7 + 1 + 0.1	0.96	Ehrenfreund et al. (1999)
LIDA	55	CH_4	1	0.24	Fraser and van Dishoeck (2004)
LIDA	57	$CO + HCOOH$	1 + 10	0.48	Fraser and van Dishoeck (2004)
LIDA	58	CH_3OH	1	0.24	Fraser and van Dishoeck (2004)
LIDA	59	$CO + CO_2$	2 + 1	0.24	Fraser and van Dishoeck (2004)
LIDA	60	$CO + CO_2$	1 + 1	0.24	Fraser and van Dishoeck (2004)
LIDA	61	CO	1	0.24	Fraser and van Dishoeck (2004)
LIDA	62	$CO + CO_2$	1 + 1	0.24	Fraser and van Dishoeck (2004)
LIDA	63	CO_2	1	0.24	Fraser and van Dishoeck (2004)
LIDA	64	$CO + CH_3OH$	1 + 1	0.24	Fraser and van Dishoeck (2004)
LIDA	65	$CO + CH_3OH$	4 + 1	0.48	Fraser and van Dishoeck (2004)
LIDA	68	$CO + CH_4$	1 + 1	0.48	Fraser and van Dishoeck (2004)

Continued...

... Continued.

Database	#	Composition	Proportion	Resol. (cm^{-1})	Reference
LIDA	69	CO + HCOOH	1 + 10	0.48	Fraser and van Dishoeck (2004)
LIDA	72	CO + CO_2	1 + 1	0.24	van Broekhuizen et al. (2006)
LIDA	73	CO_2 + CO	2 + 1	0.24	van Broekhuizen et al. (2006)
LIDA	74	CO + CO_2	4 + 5	0.24	van Broekhuizen et al. (2006)
LIDA	75	CO	1	0.24	van Broekhuizen et al. (2006)
LIDA	76	CO_2	1	0.24	van Broekhuizen et al. (2006)
LIDA	77	CO_2 + CO	10 + 1	0.48	van Broekhuizen et al. (2006)
LIDA	78	CO + CO_2	1 + 10	0.48	van Broekhuizen et al. (2006)
LIDA	79	CO + CO_2	2 + 1	0.24	van Broekhuizen et al. (2006)
LIDA	80	CO_2 + CO	1 + 1	0.24	van Broekhuizen et al. (2006)
LIDA	81	CO + CO_2	1 + 1	0.24	van Broekhuizen et al. (2006)
LIDA	82	CH_3CHO	1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	83	H_2O + CH_3CHO	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	84	CO + CH_3CHO	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	86	CO + CH_3OH + CH_3CHO	20 + 20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	87	CH_3CH_2OH	1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	88	H_2O + CH_3CH_2OH	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	89	CO + CH_3CH_2OH	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	90	CH_3OH + CH_3CH_2OH	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	91	CO + CH_3OH + CH_3CH_2OH	20 + 20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	92	CH_3OCH_3	1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	93	H_2O + CH_3OCH_3	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	94	CO + CH_3OCH_3	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	95	CH_3OH + CH_3OCH_3	20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	96	CO + CH_3OH + CH_3OCH_3	20 + 20 + 1	0.96	Terwisscha van Scheltinga et al. (2018)
LIDA	97	C_3H_6O	1	0.48	Rachid et al. (2020)
LIDA	98	C_3H_6O + H_2O	1 + 5	0.48	Rachid et al. (2020)
LIDA	99	C_3H_6O + H_2O	1 + 20	0.48	Rachid et al. (2020)
LIDA	100	C_3H_6O + CO	1 + 5	0.48	Rachid et al. (2020)
LIDA	101	C_3H_6O + CO	1 + 20	0.48	Rachid et al. (2020)
LIDA	102	C_3H_6O + CO_2	1 + 5	0.48	Rachid et al. (2020)
LIDA	103	C_3H_6O + CO_2	1 + 20	0.48	Rachid et al. (2020)
LIDA	104	C_3H_6O + CH_3OH	1 + 5	0.48	Rachid et al. (2020)
LIDA	105	C_3H_6O + CH_3OH	1 + 20	0.48	Rachid et al. (2020)
LIDA	106	C_3H_6O + H_2O + CO_2	1 + 2.5 + 2.5	0.48	Rachid et al. (2020)
LIDA	107	C_3H_6O + H_2O + CO_2	1 + 10 + 10	0.48	Rachid et al. (2020)
LIDA	108	C_3H_6O + CO + CH_3OH	1 + 2.5 + 2.5	0.48	Rachid et al. (2020)
LIDA	109	CH_3NH_2	1	0.48	Rachid et al. (2021)
LIDA	111	C_3H_6O + CO + CH_3OH	1 + 10 + 10	0.48	Rachid et al. (2020)
LIDA	114	SO_2	1	0.96	Boogert et al. (1997)
LIDA	115	CH_3OH + SO_2	1 + 1	0.96	Boogert et al. (1997)
LIDA	116	NH_3	1	0.96	Taban et al. (2003)
LIDA	118	CO + CO_2	100 + 4	0.96	Ehrenfreund et al. (1997)
LIDA	119	CO + CO_2	100 + 8	0.96	Ehrenfreund et al. (1997)
LIDA	120	CO + CO_2	100 + 16	0.96	Ehrenfreund et al. (1997)
LIDA	121	CO + CO_2	100 + 21	0.96	Ehrenfreund et al. (1997)
LIDA	122	CO + CO_2	100 + 23	0.96	Ehrenfreund et al. (1997)
LIDA	123	CO + CO_2	100 + 26	0.96	Ehrenfreund et al. (1997)
LIDA	124	CO + CO_2	100 + 70	2.0	Ehrenfreund et al. (1997)
LIDA	140	CO_2	1	0.48	Ehrenfreund et al. (1997)
LIDA	141	H_2O + CO	1 + 100	0.96	Ehrenfreund et al. (1997)
LIDA	142	H_2O + CO_2	100 + 14	0.96	Ehrenfreund et al. (1997)
LIDA	143	H_2O + CO + CO_2	1 + 50 + 56	0.96	Ehrenfreund et al. (1997)
LIDA	151	H_2O + CO_2	1 + 6	0.96	Ehrenfreund et al. (1997)
LIDA	152	H_2O + CO_2	1 + 10	0.96	Ehrenfreund et al. (1997)
LIDA	153	H_2O + CO_2	1 + 100	0.96	Ehrenfreund et al. (1997)

Continued...

... Continued.

Database	#	Composition	Proportion	Resol. (cm^{-1})	Reference
LIDA	159	H_2CO	1	0.96	Gerakines et al. (1996)
LIDA	160	OCS	1	2.0	Heuritsch et al. (non-published)
LIDA	161	CH_3OH	1	2.0	Terwisscha van Scheltinga et al. (2018)
LIDA	162	$HCOOCH_3$	1	2.0	Terwisscha van Scheltinga et al. (2018)
LIDA	165	CH_3CN	1	1.0	Rachid et al. (2022)
LIDA	166	CH_4	1	2.0	Rachid et al. (2022)
LIDA	167	$CH_3NH_2 + H_2O$	1 + 5	0.5	Rachid et al. (2021)
LIDA	168	$CH_3NH_2 + H_2O$	1 + 20	0.5	Rachid et al. (2021)
LIDA	169	$NaCOOH + H_2O$	8 + 100	2.0	Gálvez et al. (2009)
LIDA	170	$NH_4COOH + H_2O$	7 + 100	2.0	Gálvez et al. (2009)
LIDA	171	$NH_4Cl + H_2O$	9 + 100	2.0	Gálvez et al. (2009)
LIDA	175	NH_2CHO	1	1.0	Slavicinska et al. (2023)
CIL	12	$H_2O + CO$	20 + 1	2.0	Moore and Hudson (1997)
CIL	11	$H_2O + CO_2$	10 + 1	2.0	Moore and Hudson (1997)
CIL	10	$H_2O + CH_4$	20 + 1	2.0	Moore and Hudson (1997)
CIL	5	$H_2O + CH_3CH_3$	20 + 1	2.0	Moore and Hudson (1997)
CIL	8	$H_2O + CH_3CH_2CH_3$	22 + 1	2.0	Moore and Hudson (1997)
CIL	2	$H_2O + CH_3CH_2CH_2CH_3$	20 + 1	2.0	Moore and Hudson (1997)
CIL	4	$H_2O + CH_2CH_2$	20 + 1	2.0	Moore and Hudson (1997)
CIL	7	$H_2O + CH_2CHCH_3$	17 + 1	2.0	Moore and Hudson (1997)
CIL	3	$H_2O + CHCH$	15 + 1	2.0	Moore and Hudson (1997)
CIL	6	$H_2O + CH_3CCH$	20 + 1	2.0	Moore and Hudson (1997)
CIL	1	$H_2O + CH_2CCH_2$	13 + 1	2.0	Moore and Hudson (1997)
CIL	9	$H_2O + CH_2CHCHCH_2$	20 + 1	2.0	Moore and Hudson (1997)
CIL	21	$H_2O + CH_3OH$	5 + 1	0.5	Moore and Hudson (1997)
CIL	14	$H_2O + CH_3CH_2OH$	20 + 1	2.0	Moore and Hudson (1997)
CIL	x	$H_2O + OHCH_2CH_2OH$	20 + 1	2.0	Moore and Hudson (1997)
CIL	15	$H_2O + H_2CO$	8 + 1	2.0	Moore and Hudson (1997)
CIL	1	$H_2O + CH_3COH$	20 + 1	2.0	Moore and Hudson (1997)
CIL	1	$H_2O + CH_3(CO)CH_3$	20 + 1	2.0	Moore and Hudson (1997)
CIL	x	$H_2O + HCOOH$	9 + 1	2.0	Moore and Hudson (1997)
CIL	1	$H_2O + CH_3COOH$	10 + 1	2.0	Moore and Hudson (1997)
CIL	19	$H_2O + CH_3OCHO$	20 + 1	2.0	Moore and Hudson (1997)
CIL	1	$H_2O + CH_3OCH_3$	20 + 1	2.0	Moore and Hudson (1997)
CIL	13	$H_2O + c\text{-}C_2H_4O$	20 + 1	2.0	Moore and Hudson (1997)
CIL	22	$H_2O + N_2O$	10 + 1	2.0	Moore and Hudson (1997)
CIL	17	$H_2O + HCN$	3 + 1	2.0	Moore and Hudson (1997)
CIL	16	$H_2O + H_2O_2$	97 + 3	2.0	Moore and Hudson (1997)
CIL	24	$H_2O + OCS$	20 + 1	2.0	Moore and Hudson (1997)
CIL	N1(a,b,c)	NH_3	1	4.0	Moore et al. (2007)
CIL	N1(d,e,f)	$H_2O + NH_3$	5 + 1	4.0	Moore et al. (2007)
CIL	N9a	$NH_3 + H_2O$	2 + 1	4.0	Moore et al. (2007)
CIL	N10(a,c,e)	$NH_3 + H_2O$	2 + 1	4.0	Moore et al. (2007)
CIL	N13b	$H_2O + NH_3$	7 + 1	4.0	Moore et al. (2007)
OCdb	1–4	$CO + CH_4$	20 + 1	1.0	Hudgins et al. (1993)
OCdb	5–8	CO + OCS	20 + 1	1.0	Hudgins et al. (1993)
OCdb	9–11	$CO_2 + CH_4$	20 + 1	1.0	Hudgins et al. (1993)
OCdb	23–28	$H_2O + CH_3OH + CO + NH_3$	100 + 10 + 1 + 1	2.0	Hudgins et al. (1993)
OCdb	29–34	$H_2O + CH_3OH + CO + NH_3$	100 + 50 + 1 + 1	2.0	Hudgins et al. (1993)
OCdb	37–41	CH_3OH	1	2.0	Hudgins et al. (1993)
OCdb	42–44	CH_4	1	1.0	Hudgins et al. (1993)
OCdb	45–48	CO_2	1	2.0	Hudgins et al. (1993)
OCdb	49–51	OCS	1	2.0	Hudgins et al. (1993)
OCdb	52–53	$CO_2 + CH_4$	20 + 1	1.0	Hudgins et al. (1993)
OCdb	54–56	$CO_2 + OCS$	20 + 1	1.0	Hudgins et al. (1993)
OCdb	57–62	$H_2O + CH_4$	20 + 1	3.0	Hudgins et al. (1993)

Continued...

... Continued.

Database	#	Composition	Proportion	Resol. (cm^{-1})	Reference
OCdb	63–67	H_2O + OCS	20 + 1	3.0	Hudgins et al. (1993)
OCdb	68–73	H_2O + OCS	2 + 1	2.0	Hudgins et al. (1993)
OCdb	74–79	H_2O	1	8.0	Hudgins et al. (1993)
OCdb	89	CO	1	1.0	Sandford et al. (1988)
OCdb	90–91	CH_3COCH_3	1	1.0	Hudson et al. (2018)
OCdb	92–93	CH_4	1	0.5	Gerakines and Hudson (2020)
OCdb	94–95	CH_3OH	1	1.0	Gerakines and Hudson (2020)
OCdb	96–97	CO_2	1	0.5	Gerakines and Hudson (2020)
OCdb	98–99	N_2O	1	0.5	Gerakines and Hudson (2020)
Univap	G1	CO	1	...	Rocha and Pilling (2014); Ehrenfreund et al. (1999)
Univap	G2	CO_2	1	...	Rocha and Pilling (2014); Pilling et al. (2010)
Univap	G3	NH_3	1	...	Rocha and Pilling (2014); Pilling et al. (2012)
Univap	G4	SO_2	1	...	Rocha and Pilling (2014)
Univap	S2	α-Gly	1	...	Rocha and Pilling (2014); Portugal et al. (2014)
Univap	L1	H_2O	1	...	Rocha and Pilling (2014)
Univap	L3	$CH_3(CO)CH_3$	1	...	Rocha and Pilling (2014)
Univap	L4	CH_3CN	1	...	Rocha and Pilling (2014)
Univap	L5	CH_3COOH	1	...	Rocha and Pilling (2014)
Univap	L6	HCOOH	1	...	Rocha and Pilling (2014)
Univap	L7	CH_3CH_2OH	1	...	Rocha and Pilling (2014)
Univap	L8	CH_3OH	1	...	Rocha and Pilling (2014)
Univap	L9	c-C_6H_{12}	1	...	Rocha and Pilling (2014); Pilling et al. (2012)
Univap	M2	H_2O + CO_2	9 + 1	...	Rocha and Pilling (2014); Pilling et al. (2010)
Univap	M3	H_2O + CO_2	1 + 1	...	Rocha and Pilling (2014); Pilling et al. (2010)
Univap	M5	H_2O + NH_3 + CO	1 + 0.5 + 0.4	...	Rocha and Pilling (2014); Pilling et al. (2010)
Univap	M6	H_2O + NH_3 + c-C_6H_{12}	1 + 0.3 + 0.7	...	Rocha and Pilling (2014); Pilling et al. (2012)
Univap	M7	H_2O + CO_2 + NH_3 + CH_4	9 + 1 + 1 + 1	...	Rocha and Pilling (2014)
Univap	M8	H_2O + CO_2 + NH_3 + SO_2	10 + 1 + 1 + 1	...	Rocha and Pilling (2014)
Univap	D1a	H_2O + CO_2	1 + 1	...	Rocha et al. (2016); Pilling et al. (2010)
Univap	D2a	H_2O + CO_2	10 + 1	...	Rocha et al. (2016); Pilling et al. (2010)
Univap	D3a	H_2O + CH_4	1 + 0.6	...	Rocha et al. (2016)
Univap	D4a	H_2O + CH_4	10 + 0.6	...	Rocha et al. (2016)
Univap	D5a	H_2O + NH_3	1 + 0.5	...	Rocha et al. (2016); Pilling et al. (2010)
Univap	D8a	H_2O + NH_3 + CO	1 + 0.6 + 0.4	...	Rocha et al. (2016); Pilling et al. (2010)
Univap	D9a	H_2O + NH_3 + cC_6H_6	1 + 0.3 + 0.7	...	Rocha et al. (2016); Pilling et al. (2012)
Univap	D10a	H_2O + CO_2 + CH_4	10 + 1 + 1	...	Rocha et al. (2016); Pilling et al. (2019)
Univap	D11a	H_2O + H_2CO + CH_3OH	100 + 0.2 + 0.8	...	Rocha et al. (2016); de Barros et al. (2014)
Univap	D12a	H_2O + NH_3 + CO_2 + CH_4	10 + 1 + 1 + 1	...	Rocha et al. (2016); Pilling et al. (2019)
Univap	D13a	H_2O + NH_4 + CO_2 + CH_4	10 + 1 + 1 + 1	...	Rocha et al. (2016); Pilling et al. (2019)
LASp	1	H_2O + CO	8 + 1	1.0	Suhasaria et al. (2017)
LASp	2	H_2O + CO	25 + 1	1.0	non-published
LASp	3	H_2O + CO	2.5 + 1	1.0	non-published
LASp	4	CO + CH_3OH	1.6 + 1	1.0	Modica and Palumbo (2010)
LASp	5	H_2O + CH_3OH	1 + 1	1.0	non-published
IEM	1	H_2O + CH_4	10 + 3	2.0	Gálvez et al. (2009)
NSRRC	1	H_2O + CO + NH_3	1.5 + 1.3 + 1	1.0	Jiménez-Escobar et al. (2018)

Note. When the spectral resolution is missing, we assumed it to be the difference in wavenumber present in the data files.

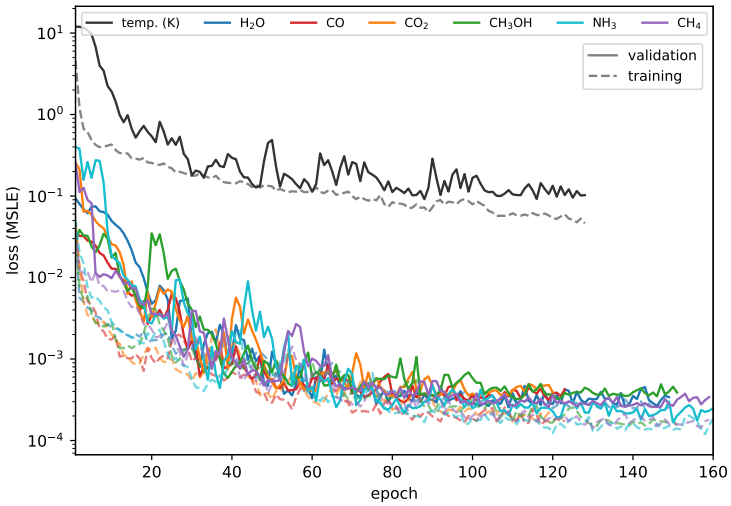


Figure C.5.1. Evolution of the training of the set of neural networks with a random seed equal to 1.

C.5 Evolution of the training

Figure C.5.1 shows the evolution of the training for the set of neural networks with seed equal to 1. We can see how the losses decrease rapidly at the beginning of the training and slowly stabilize after a maximum amount of 160 epochs. In all cases, the loss in the validation subset is larger than the loss in the training subset, since the optimization of the model's parameters is done with respect to the training subset. The noisy patterns of the curves are due to the use of mini-batches in the gradient descent. The final weights saved after the training process, for each model, are the ones of the epoch that yielded the minimum loss in the validation subset.

C.6 Our augmented dataset

Our training dataset consists of 574 experimental spectra plus 282 synthetic spectra. This last part was constructed using pure ice spectra from LIDA, that were normalized by its integrated column density and then combined linearly to form the 282 spectra.

Table C.6.1 shows information about the experiments corresponding these pure ice spectra and the integrated band used to derive the corresponding column density. All of these experiments provided spectra of pure ices for several temperatures, ranging from 15 K (10 K for NH_3) to 100 K (or the desorption temperature of the molecule). Once the ice column density is derived for the spectra with lowest temperature (10 K for NH_3 and 15 K for the rest), we divide each spectra in absorbance by the column density, obtaining the absorption cross-section (cm^2).

Now, we can create a linear combination of our six pure ice cross-sections, $\sigma_{m,T}(\tilde{\nu})$,[1] to construct a synthetic spectrum, in terms of the wavenumber, $\tilde{\nu}$. Since we will further normal-

[1] $m = H_2O, CO, CO_2, CH_3OH, CH_4, NH_3$.

Table C.6.1. Spectra and band strengths used to construct our augmented dataset of linear combinations of pure ice spectra.

Molecule	Experiment	Reference	Integrated band (cm^{-1})	Band strength (cm)	References for the band strength
H_2O	LIDA #13	Öberg et al. (2007)	3330 ± 400	$2.2 \cdot 10^{-16}$	Gerakines et al. (1995), Bouilloud et al. (2015)
CO	LIDA #61	Fraser and van Dishoeck (2004)	2139 ± 30	$1.4 \cdot 10^{-17}$	Gerakines et al. (1995), Bouilloud et al. (2015)
CO_2	LIDA #76	van Broekhuizen et al. (2006)	2343 ± 26	$1.1 \cdot 10^{-16}$	Gerakines et al. (1995), Bouilloud et al. (2015)
CH_3OH	LIDA #58	Fraser and van Dishoeck (2004)	1027 ± 50	$1.8 \cdot 10^{-17}$	Hudgins et al. (1993), Bouilloud et al. (2015)
NH_3	LIDA #116	Taban et al. (2003)	1070 ± 80	$2.1 \cdot 10^{-17}$	Sandford and Allamandola (1993), Bouilloud et al. (2015)
CH_4	LIDA #55	Fraser and van Dishoeck (2004)	1303 ± 30	$8.4 \cdot 10^{-18}$	Boogert et al. (1997), Bouilloud et al. (2015)

The position of the integrated peaks is given as the center plus/minus an approximate width. Band strengths were derived from the bibliographic work by Bouilloud et al. 2015, taking the ones corresponding to the closest temperature to the experimental spectra from LIDA. In this way, we used the band strengths reported by Rocha et al. 2022 for the LIDA release paper.

ize the resulting spectra by their mean absorbance value, we will not care about the absolute values of the coefficients of the linear combinations, $\{c_m\}$, so we will define them between 0 and 1 and normalized so that their sum is equal to 1, for the sake of simplicity. We want to obtain a curve $y_{\boldsymbol{c},T}(\tilde{\nu})$ corresponding to the spectrum with temperature T and coefficients $\boldsymbol{c}$ = (c_{H_2O}, c_{CO}, c_{CO_2}, c_{CH_3OH}, c_{NH_3}, c_{CH_4}). Therefore, our resulting spectrum will be:

$$y_{\boldsymbol{c},T}(\tilde{\nu}) \;=\; \sum_m c_m \, \sigma_{m,T}(\tilde{\nu}) \;\; . \tag{C.7}$$

Note, however, that the temperatures T at which the experimental spectra of the pure ices were recorded are not always the same. Therefore, we should find a way to make an interpolation to obtain the pure ice spectra at the desired temperature.

For this, we create a linear interpolation along the variable T of the available experimental cross-sections, for each targeted molecule, and inside the temperature range for each pure ice. For example, for H_2O we have curves for 15 K, 45 K, 75 K, and 105 K. If T = 20 K, we will interpolate the curves at 15 K and 45 K, obtaining a result close to the cross-section at 15 K; if T = 10 K, we will just use the value of 15 K.

Once we have all our interpolated cross-sections in the dimension of temperature for all our six molecules, we can do a linear combination of them to obtain a synthetic mixture. However, choosing the possible values of the coefficients is not trivial.[2] First of all, the sum of the six coefficients must be always 1, $\sum_m c_m = 1$. Then, we are interested in representing all the possible combinations of our six molecules in an homogeneous way, and exploring all the parameter space, i.e., having a significant number of values of the coefficients between the whole range from 0 to 1. To do so, we will use the following algorithm:

- We start with the molecule m_1. We will set its coefficient c_1 from a uniform distribution between 0 and 1.
- Then, we go for the molecule m_2. We will draw its coefficient c_2 from a uniform distribution between 0 and $1 - c_1$.

[2] A simple approach would be to generate 6 random numbers from a uniform distribution between 0 and 1, and divide them by the sum of all of them. However, in this way the probability of obtaining a normalized coefficient greater than ∼0.5 is very low.

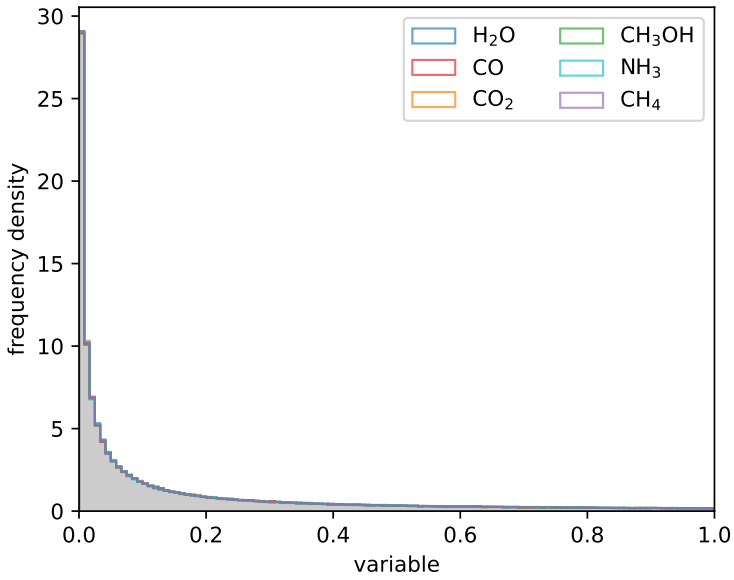


Figure C.6.1. Histogram of the generated coefficients using the algorithm mentioned in the text, with a number of combinations of 240 000. In gray we can see the histogram of the values of all the coefficients, while in colours there are plotted for each molecule separately (they are very similar, hence barely distinguishable).

- We repeat the previous step until the second to last molecule, m_5. The coefficient will be drawn from a uniform distribution between 0 and $1 - c_1 - c_2 - c_3 - c_4$.
- The coefficient for the last molecule m_6 will be directly $c_6 = 1 - c_1 - c_2 - c_3 - c_4 - c_5$.

In this way, we have obtained a combination of coefficients (c_1, c_2, c_3, c_4, c_5, c_6). To obtain a different combination of parameters, we just start again the algorithm, obtaining a new initial coefficient c_1 and, in each step, the rest of the coefficients.

With the previous algorithm, we will obtain a set of coefficients, but they will not constitute an homogeneous mixture of the six molecules. In order to achieve this, we can just repeat the algorithm six times shifting the starting molecule, and using always the same number of obtained coefficients and the same order in the succession of species (in our case: H_2O, CO, CO_2, CH_3OH, NH_3, CH_4).

This is how we created our synthetic dataset of 282 spectra (creating 47 spectra for each of the 6 iterations of the algorithm), starting with a random seed equal to 1. If we create a larger number of spectra, we can assess that the obtained sets of coefficients $\{c\}$ is homogeneous, by plotting an histogram of the populations of each coefficient c_m. Fig. C.6.1 shows the obtained plot for 240 000 synthetic coefficients. Indeed, the differences between molecules are negligible, and they arise just from the fact that we use a limited sample of coefficients.

C.7 Propagation of uncertainties

Throughout the calculations of the paper, there has been several times in which we had to calculate the propagation of uncertainties: for converting the JWST spectra from spectral flux

to absorbance, for propagating this last uncertainty in absorbance through the predictions of the neural networks, and also to perform the ratio of the abundances (predicted by AICE or by other works) with respect to H_2O.

In all these cases, we have used the approach of the Python library RichValues,[3] similarly as used by Megías et al. (2023). It consists of using probability density functions (PDFs) to represent the quantities with uncertainties, and then using a bootstrap iteration to propagate numerically the uncertainties through any desired function. In the following, we explain in detail the mathematical basis of this approach.

Using PDFs to represent numbers with uncertainties

Let's consider a variable x with a central value m and an uncertainty s, $m \pm s$. In RichValues, we prefer to base the PDFs in quantiles/percentiles, therefore interpreting m as the median of the PDF (the 50 % percentile) and $(m - s,\ m + s)$ as the 1 σ confidence interval (68.27 %) around m. One of the reasons for this choice is that the percentiles or quantiles (including the median) are robust to outliers. Another reason is that the median and quantiles transform directly through strictly increasing or decreasing functions; that is, the median of the distribution resulting of applying a function f to an input distribution of values $\{x_i\}$ is just the function applied to the median of the original distribution: median $f(\{x_i\})$ = $f(\text{median}\,\{x_i\}) = f(m_i)$.[4] Lastly, this interpretation allows for asymmetric uncertainties. This approach is indeed what is used by Possolo et al. (2019) in their work about asymmetric uncertainties.

Then, let's consider a set of n variables $\{x_j\}$. To propagate the uncertainties through a function f applied to the input variables, we can draw a sample of a large number of values (e.g., $\gtrsim 10^4\sqrt{n}$) of each variable x_j, apply the function to each of the elements of the samples, and then obtain a central value and an uncertainty for the resulting distribution. This constitutes a type of Monte Carlo simulation, as explained by Possolo et al. (2019), that allows to account for asymmetric uncertainties in the input values. To do so, we need two things: an appropriate PDF for converting each variable x_j to a distribution of values, and a proper method to obtain a central value and an uncertainty from the resulting distribution. We will use the median as the central value and the 1 σ confidence interval with respect to it (that includes 68.27 % of the distribution) to obtain the lower and upper uncertainties. As for the PDF, if the domain of the variable x is $(-\infty,\ \infty)$, a proper function is the well-known Gaussian:

$$\text{PDF}(x) \;=\; \frac{1}{\sqrt{\tau}\,s}\exp\left(-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right) \quad , \tag{C.8}$$

with $\tau \equiv 2\pi$. In this case, m is at the same time the mean, the median and the mode of

[3] `https://pypi.org/project/richvalues/`

[4] The same would be valid for any quantile, since strictly increasing or decreasing functions preserve order.

the distribution; as for s, it is the standard deviation while it also defines the 1 σ (68.27 %) confidence and credibility intervals. However, if the domain of the variable is not $(-\infty, \infty)$, this function would be incorrect; for example, if the domain is $[0, \infty)$, like the case of molecular abundances, a normal distribution would yield a non-zero probability of negative values, with greater effect the larger the relative uncertainty is.[5] Therefore, we should use another function as the PDF.

Let's suppose a domain (b_1, b_2), which defines the possible range of values of our variable. We define the left and right amplitudes, a_1 and a_2, as the distances between the limits of the domain and the median, that is, $a_1 = |m - b_1|$, $a_2 = |b_2 - m|$. Now, as these amplitudes can be different, we will split our desired PDF in two halves, one for $x \leq m$ and other for $x > m$. Then, we will use an amplitude a as a reference, which must be greater than the uncertainty, $a > s$. If the amplitude is quite greater than the uncertainty, $a \gg s$, a good PDF would be just the normal distribution truncated to the domain (b_1, b_2). However, for amplitudes closer to the uncertainty, it would be considerably incorrect, as the truncation shifts the median of the distribution and modifies the confidence intervals, and thus the uncertainties.[6] To fix this, we make the following variable change:

$$\frac{x-\mu}{a} \;\rightarrow\; \frac{\tilde{x}-\mu}{a} \;\equiv\; \operatorname{arctanh}\left(\frac{x-\mu}{a}\right) . \tag{C.9}$$

Using this new variable $\tilde{x}$ with a normal distribution, we are able to compress the original domain of $(-\infty, \infty)$ to $(-a, a)$. Making the corresponding variable changes (see the user's guide,[7] section 8.1.2, for more details), we obtain the following formula for the PDF:

$$\mathrm{PDF}(x) \;=\; \frac{1}{\sqrt{\tau}\, a \operatorname{arctanh}(\sigma/a)} \frac{\exp\left(-\frac{1}{2}\left(\frac{\operatorname{arctanh}\left(\frac{x-\mu}{a}\right)}{\operatorname{arctanh}(\sigma/a)}\right)^2\right)}{1-\left(\frac{x-\mu}{a}\right)^2} , \tag{C.10}$$

We call the corresponding distribution a *bounded normal distribution*, and it meets our requirements: the median is equal to μ, the uncertainty associated with the 1 σ confidence interval (68.27 %) is σ, and its range of values is $(\mu - a, \mu + a)$. Since this distribution is symmetric, μ is also the mean, although σ is no longer the standard deviation.

As we can see in Fig. C.7.1, it resembles a normal distribution but with a restricted domain. In fact, for small uncertainties ($s \ll a$) this PDF quickly tends to a Gaussian. On the other hand, for large uncertainties ($s \sim a/2$), the shape of the PDF starts to resemble that of a uniform distribution. Larger normalized uncertainties should not be common, since it would mean a very uninformative value, although the formula could still be used (see the user's guide [29] for more details).

[5] Actually, the usual analytic formulas for uncertainty propagation are no longer correct in this case, although they are a good approximation if the relative uncertainty is low; this is what it is done in the library RichValues.

[6] It also shifts the mean and modifies the standard deviation.

[7] `https://github.com/andresmegias/richvalues/blob/main/userguide.pdf`

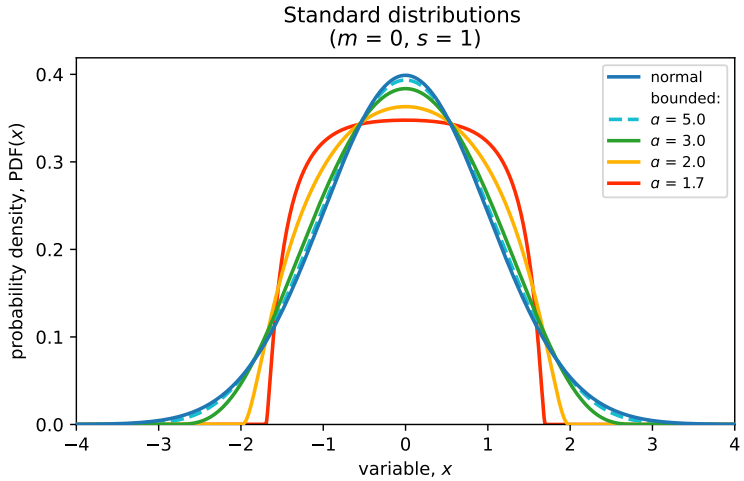


Figure C.7.1. Probability density functions for a standard normal distribution and several standard bounded normal distributions with different domains.

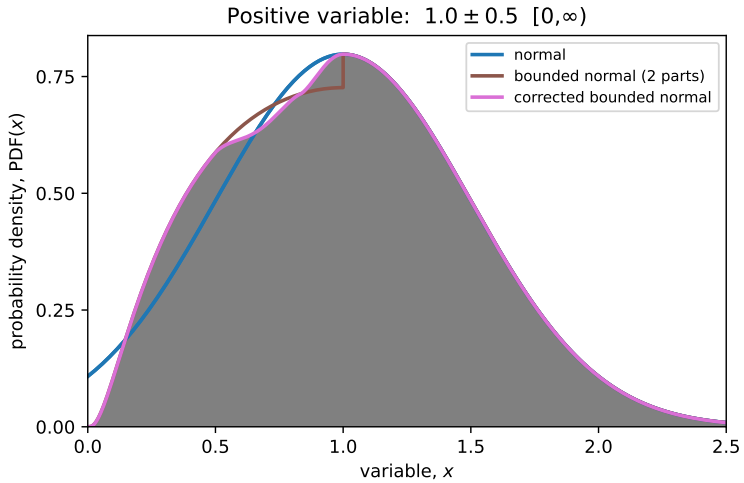


Figure C.7.2. Different probability density functions representing a positive variable with a value of 5 ± 3.

We have defined our PDF for the case of a variable x with a central value m, an uncertainty s, and a domain of $(m - a_1, m + a_2)$, building the final PDF with two halves with amplitudes a_1 and a_2, respectively. In case we had lower and upper uncertainties, s_1 and s_2, we should just replace s by s_1 for the left half of the PDF and by s_2 for the right half.

Interpolation for asymmetric distributions

In general, we can have one half of the PDF for $x < \mu$ with amplitude a_1 and uncertainty σ_1, and another half of the PDF for $x \geq \mu$ with amplitude a_2 and uncertainty σ_2. In most cases, there will be a gap in the union of the two functions, as the value at $x = \mu$ will be different for each half. To avoid this discontinuity, which seems unnatural, we can add a correction to the PDF next to μ based on the cosine function. We will not explain it here (for more details, check the user's guide [6]), but we can see an example of it in Fig. C.7.2. We call this versatile

PDF an *asymmetric bounded Gaussian*. Note that in this asymmetric case, μ is no longer the mean, and neither σ_1 nor σ_2 are equal to the standard deviation.

Alternative PDFs for asymmetric uncertainties and others

In some cases when the amplitudes (a_1, a_2) are large enough and there is a low or medium asymmetry in the uncertainties (s_1, s_2), we can use alternative functions to model the PDF of the rich value: the PDFs of a split normal distribution, a log-normal distribution and a generative extreme value (GEV) distribution. We use these three PDFs to represent values with asymmetric uncertainties using the same reasoning as in the work by Possolo et al. (2019).

In these cases with asymmetries in the uncertainty, the asymmetric bounded Gaussian can produce a bimodal curve, with a second peak due to the interpolation near the median to join the two halves of the bounded Gaussians. If this is the case, one can use these alternative PDFs to avoid that slight bimodality (although it is a PDF that perfectly meets our requirements with respect to to μ, σ_1, σ_2, a_1, a_2). However, the use of these alternative PDFs comes with an increase of computational cost, since the parameters of these PDFs are not directly (μ, σ_1, σ_2, a_1, a_2), therefore one has to make an optimization to find the proper PDF parameters for each case. This is why this option is disabled by default in RichValues. In any case, you can check the user's guide [6] to learn more about these alternative functions and how to use them.

Lastly, our approach with PDFs can also be used to address the case of a variable with an upper/lower limit or even a finite interval, although for the calculations of this paper it is not needed. If it were the case, uniform distributions would be used (either in linear or logarithmic scale), with finite boundaries for infinity ($\pm 10^{90}$) and for zero ($\pm 10^{-90}$)

C.8 Effect of band saturation

Band saturation can occur whether there is a very high column density of material or whether the sensitivity of the instrument does not allow to differentiate the non-zero flux in an absorption feature. In order to explicitly take into account this effect in the training of our neural networks, we have made a test applying data augmentation simulating the effect of saturation.

For every spectrum in our training subset, we have made 7 copies of it, with different levels of saturation, ranging from 10 % to 70 %. To simulate saturation, we crop the curve at values greater than that fraction of the maximum absorbance value, and then we normalize with the new mean absorbance value. An example can be see in Fig. C.8.1. Note that after the new normalization after performing the cropping, the whole spectrum gets higher values of absorbance. This could explain why the standard version of AICE already provides

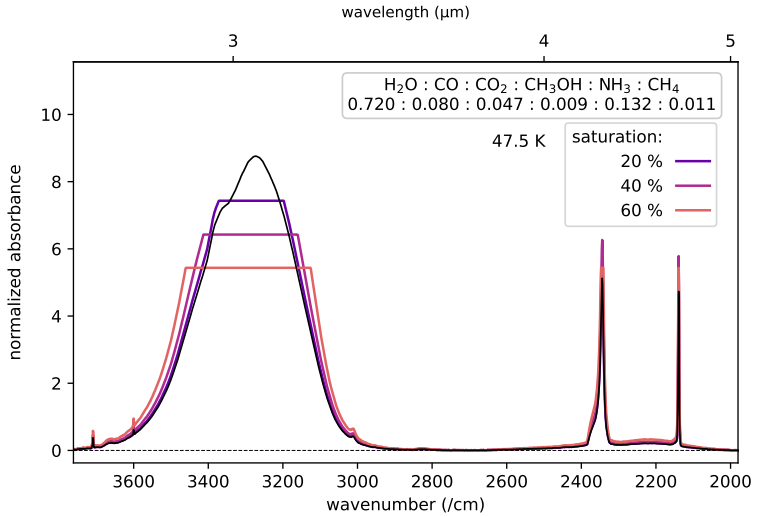


Figure C.8.1. Synthetic spectrum #15 with different levels of simulated saturation: zero (in black) and 20–60 % (in colours).

good estimations of the H_2O, CO and CO_2 fractions –and specially CO_2– for J110621 (see Section 4.4.4).

The explained process is applied only on the training subset, not in the validation subset, for each one of the different splittings for the different random seeds. We reduce the batch size to 32, but then we enlarge it within a factor of 8, so that in every optimization step, the same group of original 32 spectra are taken into account in this new training. This makes the training process only around a 50 % larger in time, instead of a higher increase. After it, we check the performance of our models in the validation (Fig. C.8.2), obtaining a good correlation between the predicted and actual values, for the molecular fraction and the temperature.

After the training with our 10 variants with different random seems, the model parameters (model weights) are stored in a file, in a way that allows to the user to select either the standard version of AICE of this alternative version. In Section 4.4.4 we test this new version with astronomical observations, comparing it with the normal version of our tool. In short, the results by both models are compatible within the uncertainties, although we see some systematic differences. In any case, it does not seem that the standard AICE models are underpredicting any molecular fraction due to band saturation.

C.9 Retraining with a smaller spectral range

We retrained our neural networks with a smaller spectral range of 4000-2000 cm^{-1} (2.5-5.0 μm). Despite of the reduction of the number of points, we obtained a good correlation between the predicted and actual values in the validation, as shown in Fig. C.9.1, although with a higher error. The model parameters were stored in a file, that allowed us to simply compare the predictions of this new version of AICE with the standard one (see Table 4.4.5).

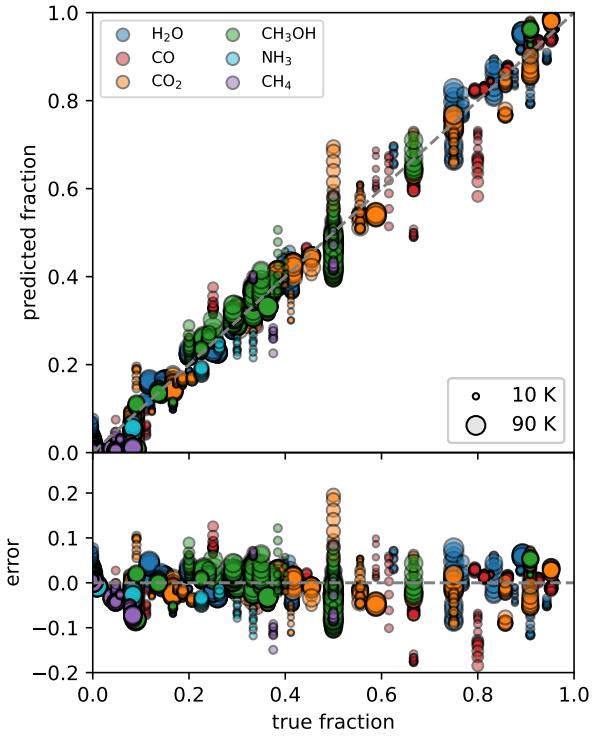


Figure C.8.2. Result of the validation for our neural networks after the training with simulated saturation, for the random seed equal to 1. The synthetic spectra from linear combinations of pure ices are not included in this plot, but we do include the copies of the spectra with simulated saturation; this is why there are many clusters of points with the same colour and value of true molecular fraction. This is equivalent to Fig. 4.4.1.

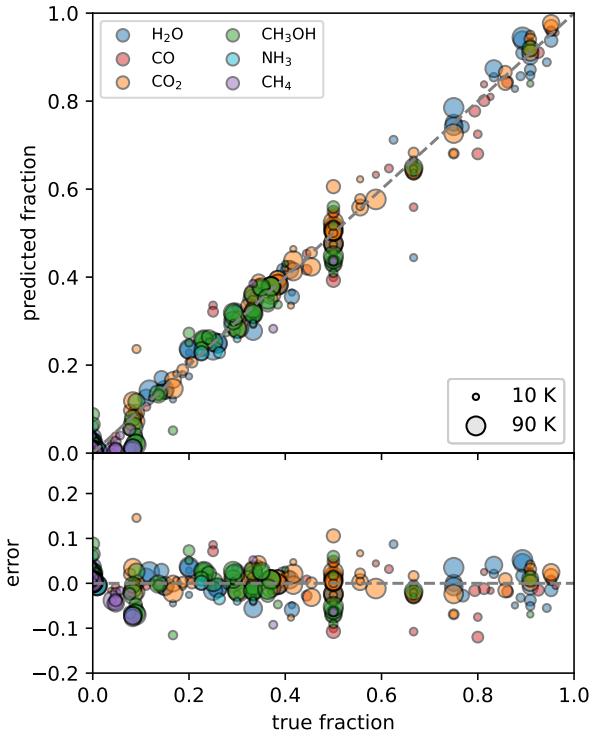


Figure C.9.1. Result of the validation for our neural networks after the training in a smaller spectral range (4000-2000 cm^{-1}) for the random seed equal to 1, without including the synthetic spectra from linear combinations of pure ices. This is equivalent to Fig. 4.4.1.

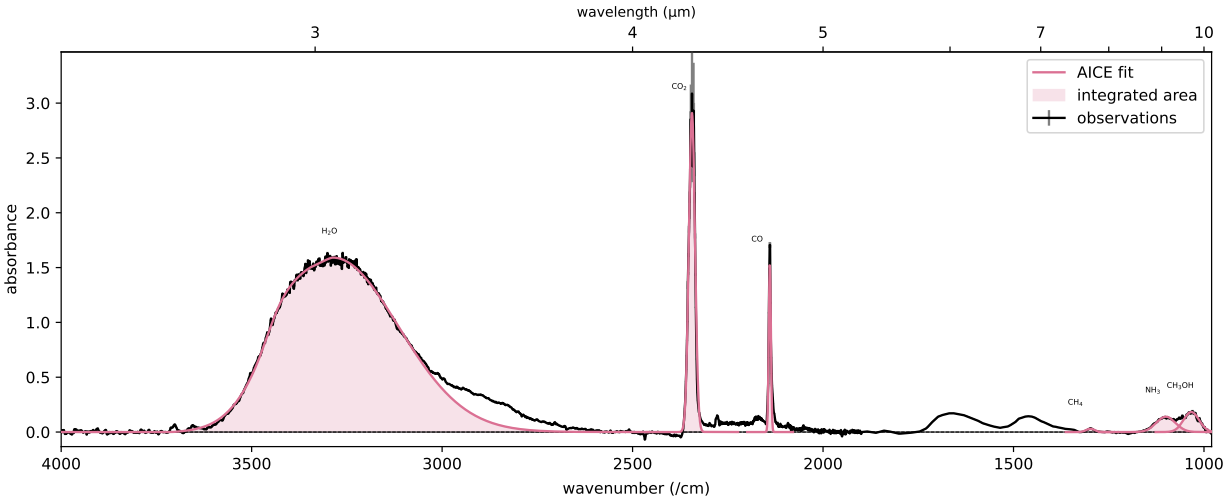


Figure C.10.1. Spectra taken by JWST towards the background star NIR38 in the molecular cloud Cha 1, converted to absorbance scale after fitting the stellar continuum and the silicate contribution. In pale red are the fitted Gaussians by the band integrator included in AICE.

C.10 Automatic band integrator

The tool AICE includes a module that allows to make a simple fit based on Gaussians to individual molecular bands, writing an initial estimation of the fit parameters in a configuration file. It also allows to exclude the central part of the curve in order to take into account possible saturation effects on the spectrum to be integrated. Figure C.10.1 shows the result of this fit for the spectrum observed towards NIR38 by McClure et al. (2023).

List of Tables

3.1.1 Spectral information of the observations of L1517B. 60
3.3.1 Transitions targeted in our L1517B observations. 64
3.3.2 Derived parameters for the observed lines. 66
3.4.1 Derived weighted averages for L1517B, L1498 and L1544. 71
3.6.1 Observed and modelled abundances of COMs in L1517B. 77

4.4.1 Validation errors of AICE. 96
4.4.2 Results of AICE and other codes for NIR38 and J110621. 97
4.4.3 Ice column densities predicted by AICE and ENIIGMA for NIR38 and J110621. 99
4.4.4 Predictions of AICE and AICE-s for NIR38 and J110621. 100
4.4.5 Predictions of AICE and AICE-lite for NIR38 and J110621. 101

5.2.1 Grain-surface reactions added in UCLCHEM for HOCO and CO_2. 109
5.2.2 Physical parameters of the astrochemical models. 110
5.3.1 Observed column densities toward NIR38 and J110621. 111

C.4.1 List of experiments in the training dataset of AICE. 179
C.6.1 Laboratory spectra used to construct our synthetic dataset. 185

List of Figures

1.1.1 View of the Milky Way taken by the spacecraft Gaia. 2
1.1.2 Image of the molecular cloud Cha I taken by JWST. 4
1.1.3 Image of the B211/B213 filament in TMC obtained by Herschel. 5
1.2.1 Scheme of the process of low-mass star formation. 6
1.2.2 Molecular abundances w.r.t. methanol towards a protostar and a comet. . . . 8
1.3.1 Most common type of gas-phase chemical reactions in the ISM. 9
1.3.2 Most common type of solid-phase chemical reactions in the ISM. 10
1.4.1 Molecular transitions found towards the methanol peak of L1544. 12

2.1.1 Scheme of the range of the electromagnetic spectrum. 16
2.1.2 Scheme of the different types of molecular transitions. 16
2.1.3 Plot of the atmospheric opacity with respect to the wavelength of radiation. . 17
2.1.4 Rotational transitions of thionylimide observed in G0693. 18
2.1.5 Schematic view of examples of molecular vibrations. 19
2.1.6 Spectra from JWST towards two positions in the Cha I molecular cloud. . . . 20
2.1.7 Infrared absorption spectra of water, methanol and ethanol at 15 K. 21
2.2.1 Scheme of the propagation of radiation through a material. 22
2.2.2 Specific intensity emitted by a black body at several temperatures. 24
2.2.3 Absorptions and emissions in a two-level system dominated by radiation. . . 26
2.2.4 Screenshots and logo of the software MADCUBA. 31
2.2.5 Scheme of the absorption and emission processes in a two-level system. . . . 33
2.2.6 Some screenshots of Radex Line Fitter, as a simple use example. 35
2.3.1 Scheme of a prototypical antenna pattern. 40
2.3.2 Scheme of the frequency switching method. 43

2.3.3 Screenshots of the software GILDAS and CLASS. 45
2.3.4 Result of the automatic line search and the reduction for an example spectra. 47
2.4.1 Photograph of the IRAM 30m telescope in Sierra Nevada. 49
2.4.2 Photography of James Webb Space Telescope. 50
2.5.1 Screenshots from UCLCHEM's website, including its logo. 53

3.1.1 H_2 column density map of L1517B from *Herschel*. 59
3.1.2 Plots of targeted lines towards L1517B's dust peak. 61
3.2.1 Plots of targeted lines towards L1517B's methanol peak. 62
3.3.1 Bar plot of the observed abundances for L1517B, L1498 and L1544. 68
3.4.1 Bar plot of the observed abundances in L1517B and L1521E. 72
3.5.1 Density and temperature profiles used for L1517B. 74
3.5.2 Evolution of the simulated CO depletion factor over time. 74
3.5.3 Modelled distribution of COMs in L1517B. 75
3.6.1 HC_3N / CH_3CN abundance ratio for different sources. 79

4.2.1 Scheme of an artificial neural network. 84
4.2.2 Architecture of the neural networks of AICE. 86
4.2.3 Example of baseline reduction for laboratory spectra. 88
4.2.4 Example of a synthetic infrared spectrum. 91
4.4.1 Performance of AICE in the validation dataset. 95
4.4.2 Predictions of AICE for NIR38. 97
4.4.3 Predictions of AICE for J110621. 99

5.3.1 Results of UCLCHEM for the static phase. 112
5.3.2 Some examples of results by UCLCHEM for the collapse fase. 113
5.3.3 Ice compositions predicted by UCLCHEM for NIR38. 114
5.3.4 Ice compositions predicted by UCLCHEM for J110621. 114
5.3.5 Errors obtained with the UCLCHEM models for NIR38. 115
5.3.6 Errors obtained with the UCLCHEM models for J110621. 116
5.3.7 Ice compositions predicted by UCLCHEM for NIR38 at 10 K. 117
5.3.8 Ice compositions predicted by UCLCHEM for J110621 at 10 K. 117

A.1.1 Screenshots of the Automated GILDAS/CLASS Pipeline. 158

A.2.1 Graph representing an example chemical network. 161

B.1.1 Distribution of losses for RADEX 165
B.1.2 Loss versus the parameters of the CH_3OH E RADEX calculations 165
B.1.3 Loss versus the parameters of the HC_3N E RADEX calculations 166
B.2.1 PDFs for normal and bounded normal distributions 168

C.1.1 Original spectrum taken by JWST of NIR38. 172
C.1.2 Baseline fitting for the spectrum of NIR38. 173
C.1.3 Silicate fitting for the spectrum of NIR38. 173
C.3.1 Example of spectra visualization with the AICE Interactive Toolkit. 177
C.3.2 Example of baseline fitting with the AICE Interactive Toolkit. 178
C.3.3 Example of interpolating a spectrum with the AICE Interactive Toolkit. . . . 179
C.5.1 Evolution of the training of AICE. 184
C.6.1 Histogram of the generated coefficients for our synthetic dataset. 186
C.7.1 Probability density function of bounded normal distributions. 189
C.7.2 Example of the probability density function of a positive variable. 189
C.8.1 Example of synthetic spectrum with simulated saturation. 191
C.8.2 Performance of AICE-s in the validation dataset. 192
C.9.1 Performance of AICE-lite in the validation dataset. 192
C.10.1 Example of automatic band integration using AICE-BI. 193

Agradecimientos

En primer lugar, quiero agradecer enormemente a mi supervisora, Izaskun Jiménez Serra, por su buena dirección durante estos cuatro años. En mi opinión, has sido una buena jefa, amable y muy profesional. Espero haber estado a la altura. En segundo lugar, quiero agradecer a Jesús Martín-Pintado, quien me dirigió unos meses durante la baja por maternidad de mi directora, de forma tan buena como ella. También quiero agradecer a mis compañeros del grupo de astroquímica observacional del Centro de Astrobiología (CAB) –que oficialmente se llama *Grupo de complejidad química en el medio interestelar y formación estelar*– por su compañía y por el ambiente agradable a la par que profesional.

Continúo agradeciendo supervisor de mi estancia doctoral en el laboratorio LERMA (ahora LIRA) de la Universidad de Cergy-París, en Neuville-sur-Oise (cerca de París, Francia), François Dulieu, y a los miembros del grupo, que hicieron que mi estancia fuera muy agradable. Agradezco también el trabajo de los colaboradores de los artículos científicos cuyo contenido se ha presentado en esta tesis (aparte de Izaskun, Jesús y François): Anton, Silvia, Paola, Giuliana y Serena, por mi primer artículo como primer autor sobre L1517B; Julie, Belén, David, Will, Marcos y Jacobo, por mi segundo artículo como primer autor sobre la herramienta AICE; y los miembros del proyecto Ice Age, es especial a mis compañeros de modelaje astroquímico Joseph, Angèle y Miwha –aunque en esta tesis me he centrado en los resultados que obtuve con UCLCHEM. También aprecio los comentarios de un revisor anónimo del primer artículo mencionado, que mejoraron esa parte del trabajo presentado. De forma similar, agradezco el trabajo de los evaluadores externos de esta tesis, Samantha Scibelli y Joan Enrique Romero, cuyos comentarios y sugerencias han mejorado esta obra. Por parte de la universidad, quiero agradecer a mi tutor académico, David Montes –quien también fue el supervisor de mi trabajo de fin de máster, sobre exoplanetas–, y a la coordinadora del programa de doctorado, María Ángeles Gómez, por la ayuda con los trámites y el papeleo de la universidad.

También quiero acordarme de mis antiguos compañeros de la Universidad de Extremadura, del grupo de física estadística SPhinX, donde estuve trabajando 8 meses antes de empezar la tesis. Algunas de las cosas que practiqué en aquel tiempo me han acabado sirviendo para esta tesis, como el tiempo que empleé en hacer mis códigos más estructurados y ordenados y el tiempo en el que estuve practicando programando redes neuronales. Curiosamente, en algún congreso he oído nombres y conceptos que conocía de aquel entonces.

Llegados a este punto, quiero reconocer y agradecer la financiación para realizar esta tesis, proveniente del Ministerio de Ciencia, Innovación y Universidades (antes Ministerio de Ciencia e Innovación) y de la Agenca Estatal de Investigación (MCIU/AEI/10.13039/501100-011033) españolas, así como del Fondo Social Europeo (*El FSE invierte en tu futuro*). Dicha financiación corresponde a las ayudas predoctorales para la formación de doctores de 2019, número PRE2019-91471. También agradezco la financiación de los dos primeros organismos a través de los proyectos PID2019-105552RB-C41 y PID2022-136814NB-I00, con financiación también del Fondo Europeo de Desarrollo Regional (FEDER) para el último proyecto. Esta tesis ha sido posible gracias a la financiación pública, por lo cual estoy muy agradecido. Relacionado con esto, también quiero mencionar el valor de los observatorios y telescopios cuyas observaciones se han usado en esta tesis: el telescopio de IRAM de 30 m (CNRS / IGN / MPG) y el telescopio espacial James Webb (NASA / ESA / CSA). Igualmente, los programas GILDAS y MADCUBA han sido muy útiles para el análisis de datos en esta tesis.

Relacionado con el párrafo anterior, quiero destacar –y agradecer a sus creadores y desarrolladores– el uso de varios programas de código libre que me han sido muy útiles en esta tesis: GIMP, Inkscape, LibreOffice, Lyx y, por supuesto, el lenguaje de programación Python –incluyendo los paquetes NumPy, SciPy, Pandas y Matplotlib. Creo que el código libre y abierto es algo muy valioso para todos. Igualmente, quiero agradecer a decenas de personas anónimas cuyas contribuciones en la página web de Stack Overflow (y páginas similares) me han resultado de ayuda para resolver dudas de programación –o incluso dudas sueltas de aprendizaje automático.

Por último, quiero agradecer a mis padres, Valentín y Maribel, por haberme apoyado siempre y haberme animado a seguir este camino científico. Curiosamente, mi padre es químico y mi madre, bióloga, y he acabado haciendo una tesis sobre astroquímica en el Centro de Astrobiología... Eso sí, yo soy físico de formación, para fastidio de mi padre (es broma).

Aknowledgements

First of all, I want to greatly thank my supervisor, Izaskun Jiménez Serra, for her nice direction during these four years. In my opinion, you have been a good boss, kind and very professional. I hope I have been up to the task. Secondly, I want to thank Jesús Martín-Pintado, who guided me for a few months during my director's maternity leave, being as good as her. I also want to thank my colleagues from the observational astrochemistry group at the Centre for Astrobiology (CAB) – which is officially called *Chemical complexity in the interstellar medium and star formation group*– for their company and for the kind yet professional atmosphere.

I continue thanking the supervisor of my doctoral stay at the LERMA laboratory (now LIRA) of the University of Cergy-Paris, in Neuville-sur-Oise (near Paris, France), François Dulieu, and the members of the group, who made my stay very pleasant. I also appreciate the work of the collaborators of the scientific papers whose content has been presented in this thesis (apart from Izaskun, Jesús and François): Anton, Silvia, Paola, Giuliana and Serena, for my first paper as first author, about L1517B; Julie, Belén, David, Will, Marcos and Jacobo, for my second paper as a first author, about the tool AICE; and the members of the Ice Age project, specially my fellow astrochemical modellers Joseph, Angèle and Miwha – although in this thesis I have focused on the results I obtained with UCLCHEM. I also appreciate the comments of an anonymous referee for the first mentioned paper, which improved that part of the presented work. Similarly, I thank the work of the external evaluators of this thesis, Samantha Scibelli and Joan Enrique Romero, whose comments and suggestions improved the quality of this work. Regarding the university, I want to thank my academic supervisor, David Montes – who also was the supervisor of my master's thesis, about exoplanets –, and the director of the doctoral programme, María Ángeles Gómez, for the help with the procedures and paperwork from the university.

I also want to remember about my former colleagues from the University of Extremadura, from the group SPhinX on statistical physics, where I worked for 8 months before starting the thesis. Some of the things I practiced at that time have ended up being useful for this thesis, such as the time I spent making my codes more structured and clean and the time I was practicing programming neural networks. Interestingly, at some conferences I heard

names and concepts that I knew from that time.

At this point, I want to acknowledge and thank the funding of this thesis, coming from the Spanish Ministry of Science, Innovation and Universities (formerly the Ministry of Science and Innovation) and the Spanish State Research Agency (MCIU/AEI/10.13039/5011000110-33), as well as the European Social Fund (*ESF, Investing in your future*). This funding corresponds to the predoctoral contracts for doctors training of 2019, number PRE2019-91471. I also aknowledge the funding of the first two organizations through the projects PID2019-105552RB-C41 and PID2022-136814NB-I00, with additional funding from the European Regional Development Fund (ERDF) for the last project. This thesis has been possible thanks to public funding, for which I am very grateful. Related with this, I would also like highlight the observatories and telescopes whose observations have been used in this thesis: the IRAM 30 m telescope (CNRS / IGN / MPG) and James Webb space telescope (NASA / ESA / CSA). In a similar way, the software GILDAS and MADCUBA have been very useful for the data analysis in this thesis.

Related to the last paragraph, I want to highlight – and thank its creators and developers – the use of several free software that have been very useful for this thesis: GIMP, Inkscape, LibreOffice and, of course, Python programming language – including the libraries NumPy, SciPy, Pandas and Matplotlib. I believe that free and open software is something very valuable for everyone. Similarly, I want to thank dozens of anonymous people whose contributions on the Stack Overflow website (and similar pages) have been helpful, for solving coding questions – or even sporadic questions about machine learning.

Finally, I want to thank my parents, Valentín and Maribel, for always supporting me and encouraging me to follow this scientific path. Interestingly, my father is a chemist and my mother is a biologist, and I ended up doing a thesis on astrochemistry at the Centre for Astrobiology... However, I am a physicist by training, much to my father's annoyance (just kidding).