EPJ Web of Conferences **265**, 00048 (2022) https://doi.org/10.1051/epjconf/202226500048
Multi-line Diagnostics of the Interstellar Medium

Revealing which Combinations of Molecular Lines are Sensitive to the Gas Physical Parameters of Molecular Clouds

Astrophysics Meet Data Science towards the Orion B Cloud

Jérôme Pety[1,2,*], *Maryvonne* Gerin[2], *Emeric* Bron[2], *Pierre* Gratier[3], *Jan H.* Orkisz[4], *Pierre* Palud[5,2], *Antoine* Roueff[6], *Lucas* Einig[1,7], *Miriam G.* Santa-Maria[7], *Victor* de Souza Magalhaes[1], *Sébastien* Bardeau[1], *Jocelyn* Chanussot[7], *Pierre* Chainais[5], *Javier R.* Goicoechea[7], *Viviana V.* Guzman[7], *Annie* Hughes[7], *Jouni* Kainulainen[4], *David* Languignon[2], *François* Levrier[7], *Darek* Lis[7], *Harvey S.* Liszt[7], *Jacques* Le Bourlot[2], *Franck* Le Petit[2], *Karin* Oberg[7], *Nicolas* Peretto[7], *Evelyne* Roueff[2], *Albrecht* Sievers[1], *Pierre-Antoine* Thouvenin[5], and *Pascal* Tremblin[7]

[1]IRAM, 300 rue de la Piscine, 38406 Saint Martin d'Hères, France
[2]LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, France
[3]Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, France
[4]Chalmers University of Technology, Department of Space, Earth and Environment, Sweden
[5]Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, 59651 Villeneuve d'Ascq, France
[6]Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel,Marseille, France
[7]See https://www.iram.fr/~pety/ORION-B/team.html for the other affiliations

Abstract. Atoms and molecules have long been thought to be versatile tracers of the cold neutral gas in the universe, from high-redshift galaxies to star forming regions and proto-planetary disks, because their internal degrees of freedom bear the signature of the physical conditions where these species reside. However, the promise that molecular emission has a strong diagnostic power of the underlying physical and chemical state is still hampered by the difficulty to combine sophisticated chemical codes with gas dynamics. It is therefore important 1) to acquire self-consistent data sets that can be used as templates for this theoretical work, and 2) to reveal the diagnostic capabilities of molecular lines accurately. The advent of sensitive wideband spectrometers in the (sub)-millimeter domain (e.g., IRAM-30m/EMIR, NOEMA, ...) during the 2010s has allowed us to image a significant fraction of a Giant Molecular Cloud with enough sensitivity to detect tens of molecular lines in the $70-116$ GHz frequency range. Machine learning techniques applied to these data start to deliver the next generation of molecular line diagnostics of mass, density, temperature, and radiation field.

1 The ORION-B program

The ORION-B project (Outstanding Radio-Imaging of OrioN-B, PI: Jérôme Pety & Maryvonne Gerin) is a Large Program of the IRAM 30 meter telescope that has imaged 13×18 pc of the Orion B massive star-forming cloud, located at 400 pc, over 40 GHz of

*e-mail: pety@iram.fr

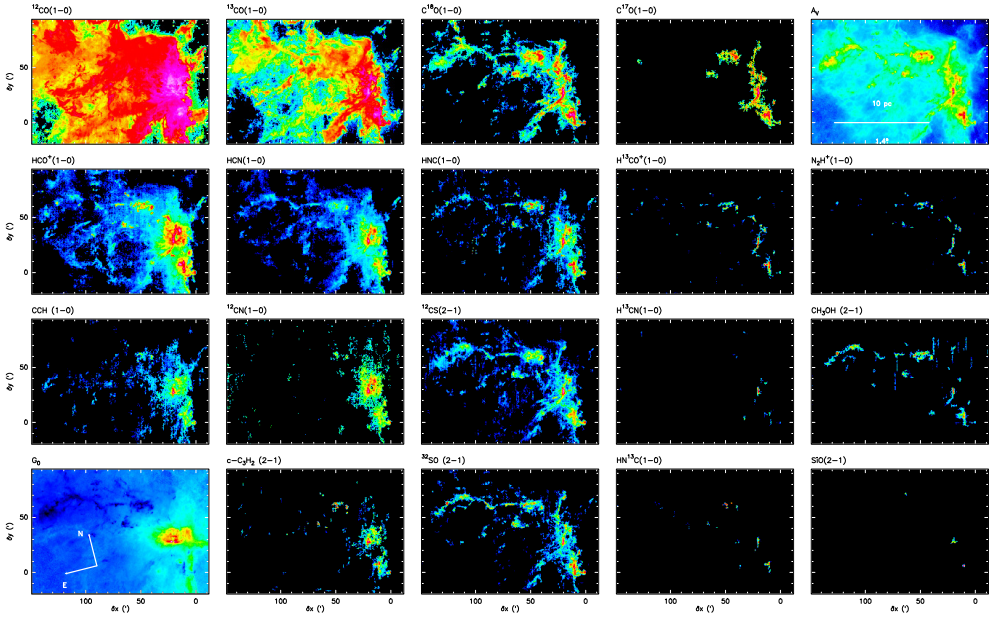

Figure 1. Spatial distribution of the line peak intensity of selected lines in the 3 mm band, plus the dust temperature **(bottom left panel)** and visual extinction **(top right panel)**. Continuum data comes from the publicly available analysis of the Herschel Gould Belt Survey data (PI: P. André) [1]. Only reliably detected pixels are shown. The color-scales are logarithmic to reveal the distribution of faint signal.

bandwidth at a typical resolution of 50 mpc and 0.6 km s^{-1}. An integration time of 850 hours yields about 200 000 images of 800 000 pixels at a typical sensitivity of 0.1 K.

Figure 1 shows the spatial distribution of the line peak brightness for a subset of 18 different molecular transitions. One of the ORION-B main goals is to study how it is possible to constrain the physical conditions (e.g., temperature, column and volume density, thermal pressure, ...) based on such observations. In a pilot study towards the first observed square degree [2], we showed how tracers of different optical depths like the CO isotopologues allow one to fully trace the molecular medium, from the diffuse envelope to the dense cores, while various chemical tracers can be used to reveal different environments. A large fraction of the flux of previously thought "dense gas tracers", as the HCN (1 − 0) line, is coming from translucent gas. Unambiguous detection of the gas whose volume density is larger than 10^4 cm^{-3} requires to detect the $N_2H^+(1-0)$ line.

A clustering algorithm was then applied to the intensities of selected molecular lines, and revealed spatially continuous regions with similar molecular emission properties, corresponding to different regimes of volume density or far-UV illumination [4]. We showed that the (1 − 0) lines of the three main isotopologues of CO are sufficient to cluster the molecular gas in three distinct density regimes ($n_H \sim 100$, 500, and > 1000 cm^{-3}). Adding the HCO^+ and CN (1 − 0) lines reveals clusters 1) of higher volume density ($n_H \sim 7 \times 10^3$ and 4×10^4 cm^{-3}) because the effective critical densities of these lines are between one and two order of magnitude higher than those of the CO isotopologues, and 2) clusters that match the GMC interfaces with the NGC 2023, NGC 2024, and IC 342 H II regions.

A principal component analysis of the twelve brightest lines in the 3 mm wavelength range showed that the first component is well correlated with the column density [5]. The second and third components are correlated at various levels with the volume density of dense cores and the dust temperature, respectively. These results suggest that the wide field

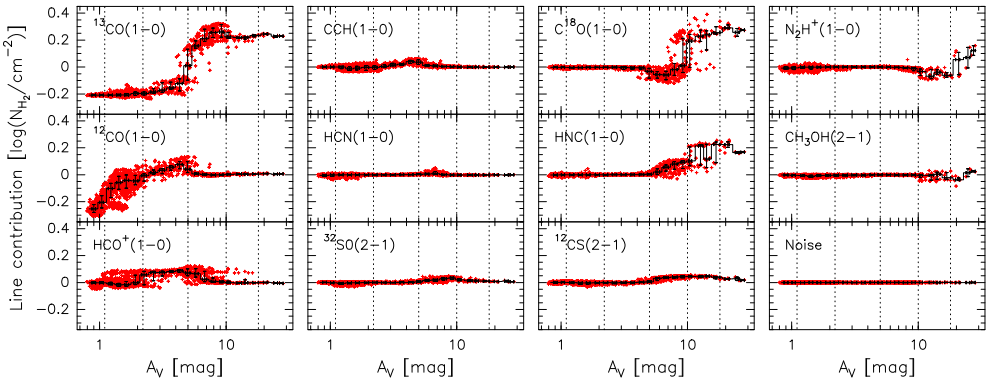

Figure 2. Contribution of the studied lines to the logarithm of the H_2 column density around the mean column density as a function of the visual extinction [3]. The black points show the median values of all data points falling in a regularly sampled interval of the logarithm of the visual extinction. The black error bars show the range of values where 50% of the points in the current bin are located. The vertical dotted lines show visual extinctions of 1.1, 2.2., 5.0, 10.0, and 18.0. The lines are sorted from top to bottom and then from left to right by increasing values of the minimum A_v at which they start to contribute.

emission of molecular lines is sensitive to the column and volume gas densities on one hand and the gas temperature on the other hand. In a subsequent study [3], we quantitatively characterized the correlations between the line integrated intensities and the dust-derived column density using a machine learning regression tool called random forest. We first showed that this correlation can be used to quantitatively infer the molecular column density from the molecular line emissions, even on data that were not used to train the regression. We also showed that the most important contributors to this new kind of column density estimators are in decreasing order the ^{13}CO, ^{12}CO, C^{18}O, and HCO$^+$ $(1-0)$ lines for the bulk of the molecular cloud, while other lines as N_2H^+ $(1-0)$ prove to be important in the densest environments (dense cores). More quantitatively, the logarithm of the gas column density for each line of sight can be to first order analyzed as the sum

$$\log N = \log N_0 + \Sigma_l \log N_l, \tag{1}$$

where $\log N_0$ is the mean column density of the training data set and $\log N_l$ is the contribution of each considered line to the final estimation. Figure 2 shows the evolution of the contribution of each line as a function of the visual extinction. Except for the ^{13}CO $(1-0)$ line that contributes at all A_v and the noise sample that does not contribute at any A_v, the other lines contribute inside a given A_v range. Once generalized this new tool should allow us to accurately estimate the mass of the different (potentially velocity separated) components of a giant molecular cloud, in particular the mass of its filamentary structure [6].

These results hints that the relationships between line intensities and environment in Giant Molecular Clouds are more complicated than often assumed. Sensitivity, excitation, and, chemistry certainly contribute to the observed line intensity distributions. In order to develop the next generation of molecular line diagnostics of mass, density, temperature, and radiation field, we need to understand these different contributions quantitatively. We started this effort by studying the abundances and excitation temperature of the three main CO isotopologues within the framework of the local thermodynamic equilibrium (LTE) for the excitation and radiative transfer [7]. Using the $(1-0)$ and $(2-1)$ transitions of the ^{13}CO and C^{18}O species and the $(1-0)$ line of ^{12}CO, estimates and associated confidence intervals of the column density, excitation temperature and velocity dispersion of these species could be obtained

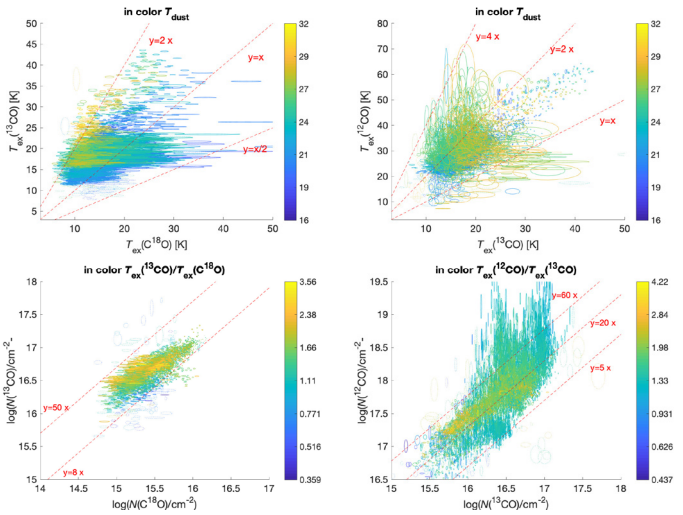

Figure 3. Scatter plots between the column densities **(top)** and excitation temperatures **(bottom)** of the main CO isotopologues [7]. The ellipses represent the interval of confidence for each estimation. The color scale encodes the dust temperature $T_{\rm dust}$ **(top)** and the ratio of the excitation temperatures of the considered species **(bottom)**. Dashed red lines show the loci of ratios: 1/2, 1, 2, and 4 **(top)**, 8 and 50 **(bottom left)**, and 5, 20, and 60 **(bottom right)**.

towards the Horsehead nebula. While the excitation temperature of the ^{13}CO and C^{18}O lines are on average similar, the excitation temperature of ^{12}CO is typically twice as large as the ^{13}CO and C^{18}O ones (top row of Fig. 3). This suggests that the CO isotopologue lines can be mostly excited in more or less dense or warm regions along the same line of sight. Moreover, the ratio of inferred column densities of the CO isotopologues are significantly different from the values given by the ratio of the carbon and oxygen elemental abundances in the solar neighborhood (bottom row of Fig. 3). Assuming they follow the elemental abundances, the [^{12}CO]/[^{13}CO] and [^{13}CO]/[C^{18}O] ratios should be in the ranges $57-67$ and $7.5-9.8$, respectively, while we found $5 \leq {}^{12}\mathrm{CO}/{}^{13}\mathrm{CO} \leq 70$ and $8 \leq {}^{13}\mathrm{CO}/\mathrm{C}^{18}\mathrm{O} \leq 50$ with a clear correlation between these values and the dust-traced far UV illumination. These results can be qualitatively understood as the result of the fractionation reaction of C^+ that enriches diffuse and translucent gas in ^{13}CO as compared to ^{12}CO, and of selective photodissociation that destroys C^{18}O much faster than ^{13}CO in the diffuse envelopes of dense cores and filaments.

2 Challenges for the future

In order to go one step further to infer the physical parameters without human supervision, we will need to invert not only radiative transfer models, but also physics and chemistry models of photo-dissociation regions (PDRs) like the Meudon PDR code. Several challenges need to be handled to succeed.

First, datasets like the ORION-B one imply large brightness dynamical ranges between the brightest and faintest detected lines (a factor ≥ 2500 for the lines integrated over the field of view). This implies that an accurate noise model must be used. In particular, it is necessary to include a multiplicative component in addition to the usual additive noise because the calibration uncertainty (of the order of $5-10\%$) has an important impact at high signal-to-noise ratios. Moreover, the notion of censored information needs to be introduced in the fit to take into account potential upper limits due to the limited instrument sensitivity. Second, a sophisticated PDR code must include many physical and chemical processes, which may have subtle but important impacts on the line strengths. The micro-physical parameters as, eg, the chemical rates, are still often uncertain because physicists and chemists can only study a small fraction of the thousands of reactions that occur in molecular clouds. Finally, there

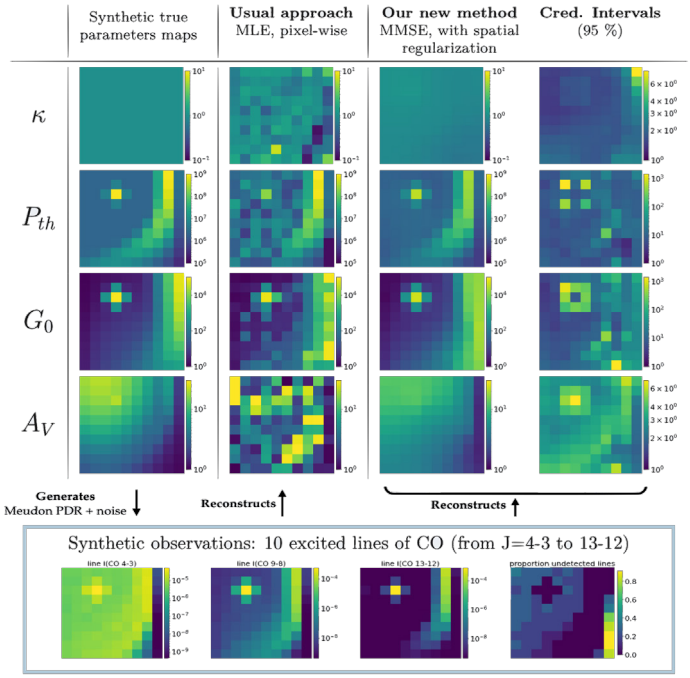

Figure 4. Comparison of the fit of several physical parameters (from top to bottom multiplicative geometry factor, thermal pressure, far UV illumination, visual extinction) between a standard maximum likelihood estimator (MLE), and a new Bayesian approach (MMSE) [8]. The first column shows the input parameters used to produce the synthetic observations of ten different J-lines of CO (bottom row). The other columns show the MLE fit, MMSE fit and its associated credibility interval.

are several hidden variables in the fit process. The most obvious one is the geometry of the cloud that is probably fractal with sharp changes of properties at several spatial scales as a molecular cloud is composed of a diffuse gas around a filamentary structure that surrounds cold cores where star formation happens. Another hidden information is the mass of the bulk of the molecular clouds as their main components (molecular hydrogen and helium) are mostly invisible to us. We thus need to rely on different minor tracers of the mass as the rotational line emission of species with a permanent dipole moment or the continuum emission of the dust.

To overcome these difficulties, we introduced a new, fast Bayesian framework that takes into account 1) a continuous approximation of large grids of PDR models, 2) additive and multiplicative noise, 3) censored information, and 4) a spatial regularization [8]. Figure 4 compares the performances of a standard maximum likelihood estimator (MLE) to the new proposed Bayesian estimator on a mock dataset. Not only the new solution delivers a much less noisy estimation, but it also includes a credibility interval associated to the best estimated set of parameters.

In less than 1 000 hours, the ORION-B large program imaged 5 square degrees over the 3 mm atmospheric window at a sensitivity of $\sim$ 0.1 K inside a typical resolution element of $27'' \times 0.6\,\mathrm{km\,s^{-1}}$ with a single beam receiver. IRAM currently develops a new receiver that will contain 25 beams, each covering the same instantaneous frequency bandwidth of 16 GHz times two polarizations as the EMIR receiver used in the ORION-B project. This tremendous observing power could be used to observe 125 square degrees of a fair fraction of the Galactic plane at a sensitivity of 0.1 K in 1 000 hours. Alternatively this added power could be used to increase the sensitivity over the same field of view as the ORION-B project to 20 mK. Figure 5 gives an idea of the evolution of the number of detected lines when the sensitivity increases by an order of magnitude. For NOEMA, the first step towards using multi-beam receivers is to develop the On-The-Fly observing mode in the coming years. In

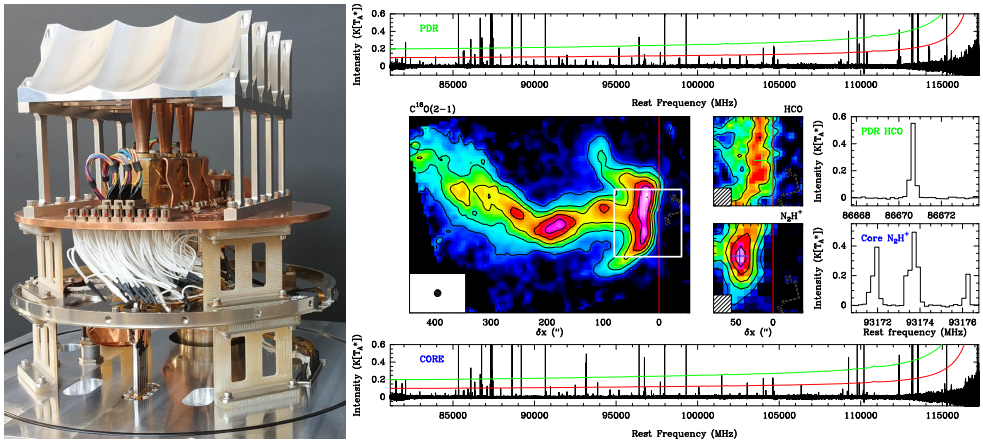

Figure 5. Left: 3 mm HEMT 3-beam prototype (Courtesy of P. Serres, O. Garnier, & the IRAM front-end group). **Right:** Spectra (smoothed to 195 kHz resolution) of the 3 mm band for two different positions in the Horsehead Nebula, obtained by the WHISPER survey [9–11] at the IRAM-30m. The red and green curves show the 1 and 2σ brightness levels achieved in the ORION-B field of view. The photo-dissociation region (PDR) is marked by the green cross on the HCO emission map (middle, top image). The dense core is marked by the blue cross on the $N_2H^+(1-0)$ emission map (middle, bottom image). These fields are reported as a white square on the $C^{18}O$ (2 – 1) integrated emission map displayed on the left panel. The middle right panels display two lines zoomed from the surveys.

other words, the results obtained by the ORION-B project is just a glimpse into what will be routine science of the interstellar medium in the coming decade.

Acknowledgements

This work is based on observations carried out under project numbers 019-13, 022-14, 145-14, 122-15, 018-16, and the large program 124-16 with the IRAM 30m telescope. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain). This work was supported by the French Agence Nationale de la Recherche through the DAOISM grant ANR-21-CE31-0010. It has received financial support from the CNRS through the MITI interdisciplinary programs, and from the Programme National "Physique et Chimie du Milieu Interstellaire" (PCMI) of CNRS/INSU with INC/INP, co-funded by CEA and CNES. J.R.G. and M.G.S.M. thank the Spanish MCINN for funding support under grant PID2019-106110GB-I00.

References

[1] M. Lombardi et al., A&A**566**, A45 (2014), 1404.0032
[2] J. Pety et al., A&A**599**, A98 (2017), 1611.04037
[3] P. Gratier et al., A&A**645**, A27 (2021), 2008.13417
[4] E. Bron et al., A&A**610**, A12 (2018), 1710.07288
[5] P. Gratier et al., A&A**599**, A100 (2017), 1701.04205
[6] J.H. Orkisz et al., A&A**624**, A113 (2019), 1902.02077
[7] A. Roueff et al., A&A**645**, A26 (2021), 2005.08317
[8] P. Palud et al., Subm. (2022)
[9] J. Pety et al., A&A**548**, A68 (2012), 1210.8178
[10] P. Gratier et al., A&A**557**, A101 (2013), 1305.2371
[11] V.V. Guzmán et al., Faraday Discussions **168**, 103 (2014), 1404.7798