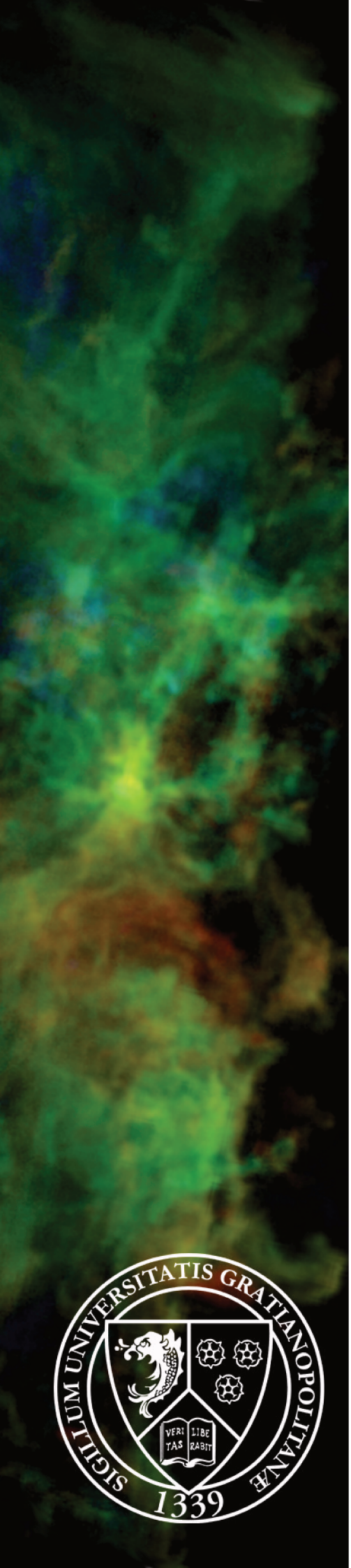
Understanding the structure of molecular clouds : multi-line wide-field imaging of Orion B

Jan Orkisz



Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Physique et Astrophysique**

Arrêté ministériel du 25 mai 2016

présentée par

Jan Orkisz

Thèse dirigée par **Jérôme Pety, astronome (IRAM)**
codirigée par **Maryvonne Gerin, directrice de recherche (LRA-ENS)**

préparée au sein du **Laboratoire de Radio-Astronomie (ENS Paris)**
et de l'**Institut de Radio-Astronomie Millimétrique (IRAM)**
dans l'École Doctorale de Physique de l'Université Grenoble-Alpes

Understanding the structure of molecular clouds: Multi-line wide-field imaging of Orion B

Comprendre la structure des nuages moléculaires : imagerie hyperspectrale d'Orion B

Thèse soutenue publiquement le **22 octobre 2018**,
devant le jury composé de :

M. Alain Abergel
Professeur (IAS), Président du jury
M. Jouni Kainulainen
Professeur assistant (Chalmers University), Examinateur
M. Pierre Lesaffre
Chargé de recherche (LRA-ENS), Examinateur
Mme Frédérique Motte
Directrice de recherche (IPAG, UGA), Examinatrice
M. Mario Tafalla
Professeur (OAN), Rapporteur
Mme Annie Zavagno
Maître de conférences (LAM), Rapporteuse

Remerciements

Je remercie en tout premier lieu mes directeurs de thèse, Jérôme Pety et Maryvonne Gerin. On dit souvent que la relation d'un thésard avec son directeur est un des éléments les plus importants pour la réussite d'une thèse, et, après ces trois ans, je peux témoigner dans ce sens. Si cette thèse s'est si bien passée pour moi, c'est notamment parce que Jérôme et Maryvonne m'ont fourni un excellent équilibre d'encadrement et d'autonomie, tantôt me guidant, tantôt m'aiguillonnant, tantôt me laissant faire mes armes tout seul, toujours avec bienveillance. Leur aide et leur disponibilité ont été précieuses, et d'autant plus appréciées que j'ai conscience de leurs innombrables autres responsabilités. Je me dois aussi de saluer leur patience à mon égard, surtout en ce qui concerne ma tendance à leur soumettre mes travaux à la dernière minute avant les échéances – cette thèse n'a d'ailleurs pas fait exception.

J'ai eu la chance de faire ma thèse au sein d'une équipe nombreuse et riche, rassemblée autour d'un projet ambitieux. Je remercie tous les membres de l'équipe ORION-B de m'avoir si bien intégré et pour toutes les discussions que nous avons eues. Ces échanges m'ont non seulement permis d'avancer dans les recherches présentées dans cette thèse, mais aussi de concevoir assez d'idées et de projets pour avoir du travail pour plusieurs années encore. J'espère que l'avenir me réserve encore bien des collaborations avec cette équipe. Le consortium ORION-B n'étant pas vraiment localisé géographiquement, ce sont le LRA-ENS à Paris, puis l'IRAM à Grenoble, qui m'ont accueilli au quotidien. Merci à tous mes collègues pour leur accueil et pour les conditions de travail chaleureuses dont j'ai bénéficié.

Merci aussi à mes rapporteurs, Annie Zavagno et Mario Tafalla, et à tous les membres du jury, Frédérique Motte, Pierre Lesaffre, Alain Abergel et Jouni Kainulainen, de m'avoir fait l'honneur d'accepter d'évaluer mon travail de thèse, et ce malgré leurs emplois du temps particulièrement chargés.

C'est aussi le moment de remercier tous les enseignants qui m'ont donné le goût des sciences, de la physique, de l'astrophysique en particulier, et de la recherche en général. Je ne peux pas tous les citer car ils ont été nombreux, surtout pendant mes études supérieures, mais je pense notamment à M. Paire, Mme Guilianotto, et M. Lièvre pendant ma scolarité, et, par la suite, MM. Raymond et Le Bourlot. Si je dois enseigner un jour (plus que ce que je n'ai été amené à faire durant cette thèse), j'espère que je saurai me montrer à la hauteur de ce qu'ils m'ont apporté. Je remercie aussi particulièrement François Levrier qui a été pour moi un "fil rouge" du début de mes études à la fin de ma thèse : il m'a donné mes premiers cours d'astrophysique, puis il a été mon tuteur à l'ENS, de nouveau mon enseignant à l'Observatoire de Paris, mon parrain de thèse durant l'année passée à Paris, et enfin mon collègue au sein du projet ORION-B. À chaque étape il s'est montré très disponible et m'a permis de progresser scientifiquement.

Je remercie bien sûr mes parents (on ne le fait jamais assez), qui, bien plus encore que mes professeurs, m'ont permis de devenir qui je suis – un astrophysicien, apparemment – en me guidant, me poussant, m'encourageant et me soutenant quand il le fallait. Merci notamment mon professeur de père, qui, usant de son expérience, m'a servi de directeur de thèse auxiliaire, m'aidant à mener au mieux mes recherches et la rédaction de cette thèse (en allant jusqu'à la relire).

Je tiens aussi à remercier de tout cœur tous ceux qui m'ont permis de traverser l'épreuve qu'a été la rédaction de cette thèse dans les meilleures conditions. Merci donc à Riquet pour m'avoir fourni un cadre de travail exceptionnel ; merci à Françoise pour sa patience et son accueil à toute heure ; merci aux copains de Flos Montis pour les bagarres cordiales, les excellentes bières et la chaleureuse amitié ; merci à Mireille, Thierry, Colette, Laura, Orphée et tous les autres pour m'avoir permis de m'aérer régulièrement le corps et l'esprit. Enfin, à l'évidence, merci à Chloé, qui m'a tellement apporté pendant les trois années de cette thèse, qui a tant donné pour que je tienne le coup pendant les mois de rédaction... Sans elle, je ne sais pas à quoi nous ressemblerions, ma thèse et moi – mais je n'ai pas envie de le savoir.

Charnay-lès-Mâcon & Grenoble, Avril – Juillet 2018

Acknowledgements

First and foremost, my thanks go to my PhD supervisors, Jérôme Pety and Maryvonne Gerin. It is often said that a good relationship between a PhD student and his supervisors is one of the key ingredients for a successful PhD, and, after these three years, I can testify that it is the case. If this PhD went so well for me, it is in particular because Jérôme and Maryvonne provided me with an excellent balance of supervision and autonomy: they advised me, spurred me, or let me try things on my own, always for my own good. Despite their countless other responsibilities, they have been helpful and available when I needed it, which I appreciate greatly. I have to pay tribute to their patience, particularly regarding my tendency to submit my works to their evaluation at the very last moment before deadlines – this thesis was actually no exception.

I was lucky to spend my PhD years as a member of a large and diverse team, gathered around an ambitious project. I am grateful to everyone in the ORION-B team for their welcome and for all the discussions we have had. These interactions not only helped me with the research presented in this dissertation, but also generated enough ideas and projects to keep me busy for many coming years. I hope that my future will see many more collaborations with this team. Given that the ORION-B consortium is rather scattered geographically speaking, my daily work was spent at the LRA-ENS in Paris, and later at IRAM in Grenoble. I thank all my colleagues in both institutions for the friendly atmosphere and the nice working conditions I had there.

My thanks also go to my referees, Annie Zavagno and Mario Tafalla, as well as all the members of the jury, Frédérique Motte, Pierre Lesaffre, Alain Abergel and Jouni Kainulainen, for doing me the honour of accepting to judge the work I did during this PhD, despite their particularly busy schedules.

Now is also a good time to thank all the teachers who have awoken my taste for science, for physics, for astrophysics in particular, and for research in general. I cannot name all of them, are they have been numerous, especially during my university years, but I particularly have in mind Mr Paire, Mrs Guilianotto and Mr Lièvre during my school years, and, later on, Mr Raymond and Mr Le Bourlot. If I am to teach one day (more that the few teaching duties I had during this PhD, that is), I hope that I will live up to what they have taught me. I am particularly grateful to François Levrier, who has always been around from my first years after school to the end of my PhD: he was my first astrophysics teacher, then my tutor at ENS, then my teacher again at the Paris Observatory, my mentor during the first year of PhD spent in Paris, and finally my colleague in the ORION-B project. At each stage, he was open to questions, helpful, and made me a better scientist.

Of course, I wish to thank my parents (one never thanks them enough). Much more than my teachers, they have helped me becoming who I am – an astronomer, it seams – by guiding me, pushing me forward, encouraging me and supporting me when necessary. In particular, I wish to thank my professor dad, who used his experience to act as an additional supervisor, advising me on how to conduct my research and the writing of this thesis (and going as far as proof-reading it).

I also wish to heartily thank all people who helped me going in the best possible conditions through the ordeal that the writing of this dissertation was. Thus, my thanks go to Riquet, for providing me with an amazing working environment; to Françoise, for her patience and for welcoming me at any time; to my comrades at Flos Montis, for the merry fights, the tasty beers and the warm friendship; to Mireille, Thierry, Colette, Laura, Orphée and all the others, for letting me refresh my body and mind regularly. And, most obviously, I wish to thank Chloé, who did so much for me during the three years of this PhD, who put so much at stake to help me make it through the writing months... Without her, I don't know how we would be like, this PhD thesis and me – but I don't want to know what "without her" would mean.

Charnay-lès-Mâcon & Grenoble, April – July 2018

Contents

Résumé

La dernière génération de récepteurs radio, dotés à la fois d'une grande bande passante et d'une haute résolution, fait de toute observation radio-astronomique une étude spectroscopique. Dans le cas de l'imagerie à grand champ du milieu interstellaire, cela génère de très gros jeux de données en grande dimension. Une telle abondance de données fournit de nouveaux outils de diagnostic, mais pose simultanément de nouveaux défis en termes de traitement et d'analyse des données - il est donc nécessaire de développer de nouvelles méthodes. C'est l'objectif du projet ORION-B, qui observe 5 degrés carrés du nuage moléculaire Orion B, soit près de la moitié de la surface du nuage, dans toute la bande à 3 mm. Les observations ont produit des cartes résolues en vitesse de l'émission de dizaines de traceurs moléculaires, qui sont analysées avec des approches statistiques d'apprentissage automatique.

L'accès à des cartes résolues spatialement pour de nombreuses espèces chimiques nous permet d'identifier les traceurs les plus caractéristiques de telles ou telles régions, qui peuvent avoir des densités ou éclairements UV différents. Réciproquement, étudier les données dans l'espace "chimique" multi-dimensionnel permet de segmenter le nuage moléculaire en régions caractérisées par une émission moléculaire similaire, et de quantifier les corrélations les plus importantes entre différents traceurs moléculaires, et entre les traceurs et des quantités physiques telles que la densité ou la température des poussières.

La grande plage dynamique spatiale et spectrale des données fournit également une vue détaillée de la cinématique et de la dynamique du nuage moléculaire, en particulier en ce qui concerne son activité de formation stellaire. Dans cette thèse, je caractérise statistiquement la nature des mouvements du nuages, et je quantifie la fraction de quantité de mouvement contenue dans les modes compressifs et solénoïdaux (rotatifs) de la turbulence. Le nuage est dominé par des mouvements solénoïdaux, tandis que les modes compressifs sont concentrés dans deux régions de formation stellaire. J'aborde aussi la formation stellaire d'un point de vue morphologique, en identifiant et caractérisant les nombreux filaments présents dans le nuage. Ceux-ci s'avèrent avoir de faibles densités, et être très stables face à l'effondrement gravitationnel - mais ils montrent cependant des signes de fragmentation radiale et longitudinale. Tous ces résultats soulignent le rôle du forçage compressif et des filaments denses dans le processus de formation stellaire. Ils sont également cohérents avec la très faible efficacité de formation stellaire d'Orion B, mais suggèrent aussi que le nuage pourrait devenir plus actif à l'avenir.

Liste d'abréviations

ALMA	Atacama Large (sub)Millimetre Array
CMF	Distribution de masse des cœurs (Core Mass Function)
CNM	Milieu froid neutre (Cold Neutral Medium)
FIR	Infra-rouge lointain (Far Infra-Red)
FUV	Ultra-violet lointain (Far Ultra-Violet)
FWHM	Largeur à mi-hauteur (Full Width at Half Maximum)
GMC	Nuage moléculaire géant (Giant Molecular Cloud)
GMF	Filament moléculaire géant (Giant Molecular Filament)
HIM	Milieu chaud ionisé (Hot Ionized Medium)
HGBS	Herschel Gould Belt Survey
HST	Hubble Space Telescope
IMF	Distribution de masse initiale (Initial Mass Function)
IR	Infra-Rouge
IRDC	Nuage sombre infra-rouge (Infra-Red Dark Cloud)
ISM	Milieu interstellaire (Inter-Stellar Medium)
LTE	Équilibre thermodynamique local (Local Thermodynamic Equilibrium)
MC	Nuage moléculaire (Molecular Cloud)
MHD	Magnéto-Hydrodynamique
MIR	Infra-rouge moyen (Mid Infra-Red)
MIS	Milieu Inter-Stellaire
NIR	Infra-rouge proche (Near Infra-Red)
NOEMA	NOrthern Extended Millimetre Array
ORION-B	Outstanding Radio-Imaging of OrioN-B
PCA	Analyse de composantes principales (Principal Component Analysis)
PdBI	Pic de Bure Interferometer
PDF	Fonction de distribution de probabilité (Probability Distribution Function)
PDR	Région photo-dominée (Photon-Dominated Region) *ou* région de photo-dissociation (Photo-Dissociation Region)
PPP	Coordonnées en Position-Position-Position (x, y, z)
PPV	Coordonnées en Position-Position-Vitesse (x, y, v_z)
SF	Formation stellaire (Star Formation)
SFE	Efficacité de formation stellaire (Star Formation Efficiency)
SFR	Taux de formation stellaire (Star Formation Rate) *ou* région de formation stellaire (Star-Forming Region)
UV	Ultra-Violet
VLT	Very Large Telescope
WIM	Milieu tiède ionisé (Warm Ionized Medium)
WNM	Milieu tiède neutre (Warm Neutral Medium)
YSO	Objet stellaire jeune (Young Stellar Object)

Comprendre la structure des nuages moléculaires : imagerie hyperspectrale d'Orion B

1 Introduction

1.1 Le milieu interstellaire

Pendant des millénaires, les hommes ont vu le ciel nocturne comme un fond vide constellé de points brillants, fixes (les étoiles) ou mobiles (les planètes et comètes). L'astronomie était alors la science des cartes du ciel et des éphémérides. Ce n'est qu'au 18^{e} siècle que des observateurs comme Charles Messier ou William Herschel ont commencé à s'intéresser à des objets différents dans le ciel, les "nébuleuses". Et il a fallu attendre le début du 20^{e} siècle pour que la nature des nébuleuses brillantes cataloguées par Messier et des nébuleuses sombres qui intriguaient Herschel soit identifiée : il s'agissait de nuages de gaz et de poussières qui peuvent tantôt émettre, diffuser ou absorber la lumière.

La présence d'un milieu interstellaire (MIS) diffus, non confiné dans les nébuleuses, a ensuite été révélée entre autres par les travaux de Hartmann, qui a remarqué en 1904 la présence de raies d'absorption interstellaires dans des spectres d'étoiles, et de Trumpler, qui a été le premier à quantifier, en 1930, l'*extinction* systématique de la lumière des étoiles par les poussières présentes tout au long de la ligne de visée reliant la source à l'observateur.

Les progrès technologiques dans le domaine radio, initiés dans les années 40, ont permis la détection de la raie à 21 cm de l'hydrogène atomique en 1951, puis, à partir des années 1969-1970, de raies spectrales de molécules telles que NH_3, H_2CO, HCO^+, et, principalement, CO, ce qui a mené à la découverte des nuages interstellaires. Ainsi, notre compréhension de la nature et de la structure du MIS progresse au rythme de la conquête du spectre électromagnétique, et de l'amélioration de la sensibilité, de la résolution angulaire et de la résolution spectrale des observatoires terrestres et spatiaux.

Le milieu interstellaire occupe l'essentiel du volume de la Galaxie, sous la forme d'un disque de rayon $\sim$10 kpc et avec une épaisseur typique de $\sim$ 500 pc. C'est cependant un milieu extrêmement dilué (les nuages moléculaires, qui représentent les régions plus denses du MIS, sont 10^{20} fois moins denses que l'atmopshère terrestre au niveau de la mer, et 10^3 fois moins denses que les meilleurs vides obtenus en laboratoire), si bien que le MIS ne représente que 10% de la masse baryonique de la Galaxie, qui est largement dominée, en termes de matière et de luminosité, par les étoiles. Dans ce contexte, l'importance du milieu interstellaire vient de son lien intime avec la formation des étoiles : la matière galactique circule en permanence dans un cycle qui va de la mort des étoiles à leur naissance en passant par toutes les phases du milieu interstellaire.

Au cours de leur vie, les étoiles éjectent en permanence de la matière dans le milieu environnant, sous forme de vents stellaires. Ces éjections s'intensifient énormément lorsque la fin de vie des étoiles approche, durant la phase dite de géante rouge. Pour les étoiles les plus massives, la fin de vie survient sous la forme d'une explosion de supernova, qui éjecte elle aussi des quantités colossales de gaz, de poussières et d'énergie dans son environnement. La matière ainsi éjectée par les étoiles atteint des températures de plusieurs dizaines de milliers de Kelvins, et se disperse rapidement dans le milieu diffus. Progressivement, le gaz diffus se refroidit radiativement, formant des structures de plus en plus denses et de plus en plus froides. Les plus denses et froides de ces structures, les cœurs denses, peuvent s'effondrer

et atteindre des pressions et des températures telles qu'ils donnent naissance à de nouvelles étoiles.

Au cours de ce cycle, la matière interstellaire peut se retrouver dans différentes phases, qui se mélangent et interagissent en permanence. Ces phases sont approximativement en équilibre de pression les unes avec les autres, et se distinguent notamment par l'état chimique du constituant principal du MIS – l'hydrogène – qui peut être ionisé (HII, ou H^+), neutre (HI, hydrogène atomique) ou moléculaire (H_2). Ainsi, le MIS se compose principalement d'un milieu tiède, qui peut être neutre ou ionisé (*warm ionised medium*, WIM, et *warm neutral medium*, WNM), et d'un milieu froid, qui est neutre. Dans les régions les plus froides et denses du CNM, l'hydrogène se trouve sous forme moléculaire, on parle donc de nuages moléculaires. En plus de cela, des bulles de milieu très chaud, ionisé et très dilué (*hot ionized medium*, HIM) sont générées par les explosions de supernovae.

1.2 Les nuages moléculaires

Cette thèse se concentre sur la phase la plus froide et la plus dense du milieu interstellaire, les nuages moléculaires (et en particulier le nuage Orion B). Les plus opaques d'entre eux peuvent atteindre plusieurs dizaines de magnitudes d'extinction, et les plus grands d'entre eux, les GMCs (*giant molecular clouds*) ont des tailles de plusieurs dizaines de parsecs et des masses de l'ordre de $10^4 - 10^5 M_\odot$. La densité typique d'un GMC est de l'ordre quelques 10^2 cm^{-3}, mais de fortes fluctuations existent autour de cette moyenne, car les nuages sont fortement structurés et contrastés. Cette structure peut être décrite de façon hiérarchique, en terme de sous-nuages, de filaments, de cœurs de plus en plus denses, mais on peut aussi la voir comme une fractale avec un continuum d'échelles spatiales et de densités. Cette approche fractale est en particulier liée à la turbulence magnéto-hydrodynamique (MHD) qui agite le MIS.

Les nuages moléculaires sont au centre des études sur le milieu interstellaire, car c'est en leur sein que les étoiles se forment : leur effondrement gravitationel mène à la formation des cœurs denses et des étoiles. On constate cependant que le taux de formation d'étoiles représente à peine quelques pourcents de ce qu'il serait si cet effondrement avait lieu en chute libre : la contraction du gaz est donc freinée par divers phénomènes, en particulier la pression thermique, la turbulence, et le champ magnétique. Par ailleurs, les premières étoiles formées dans un nuage rétro-agissent sur celui-ci, contribuant à réduire l'efficacité de formation stellaire : elles chauffent le gaz, émettent des vents et des jets qui fournissent de l'énergie cinétique à la turbulence, et, pour les plus massives, explosent rapidement sous forme de supernovae qui peuvent disloquer un nuage de l'intérieur en 20 – 30 millions d'années.

Les nuages moléculaires ont été d'abord identifiés observationnellement en temps que nuages sombres grâce à leur extinction dans le domaine visible et infra-rouge proche. L'émission thermique des poussières, elle, peut être observée en infra-rouge lointain et sub-millimétrique. L'émission des raies spectrales des molécules présentes dans les nuages moléculaires, en général à des longueurs d'onde (sub-)millimétriques, est aussi un traceur observationnel crucial. Cette émission provient de la désexcitation radiative entre niveaux quantiques (en général des niveaux rotationnels) de molécules excitées par collisions. Le dihydrogène, qui est de très loin la molécule la plus abondante dans le MIS, n'a pas de moment dipolaire électrique, et ses transitions rotationnelles sont donc interdites. Il faut donc recourir à l'émission d'autres molécules, plus rares, pour tracer le gaz. La plus couramment utilisée est CO, qui est environ $10^4 - 10^5$ fois moins abondante que H_2. Les raies de ^{12}CO sont assez brillantes pour permettre de cartographier amplement le milieu moléculaire, les isotopologues plus rares (^{13}CO, C^{18}O, C^{17}O...) sont utiles pour observer le gaz à plus forte densité, lorsque les raies de ^{12}CO deviennent optiquement épaisses.

De nombreuses autres raies moléculaires sont également utilisées, comme par exemple NH_3, dont la configuration de niveaux quantiques fait une excellente sonde de température et de densité, ou N_2H^+, qui trace spécifiquement le milieu opaque et froid, comme les cœurs préstellaires. L'observation de nombreux traceurs moléculaires permet d'étudier les réseaux chimiques en œuvre dans le MIS ; réciproquement, la connaissance de la chimie et des conditions d'excitation des molécules permettent d'utiliser ces nombreux traceurs pour contraindre les conditions physiques dans les nuages moléculaires.

Un autre avantage majeur des observations de raies moléculaires est d'avoir accès au profil des raies, qui est essentiellement dû à l'effet Doppler, et traduit donc la cinématique du gaz. Les largeurs de raies

sont souvent supérieures à l'élargissement thermique, et donnent donc accès aux mouvements du gaz le long de la ligne de visée.

Les questions concernant la physique des nuages moléculaires sont complexes et particulièrement importantes, puisqu'elles font le lien entre l'évolution galactique et la formation stellaire. Un des problèmes-clés est la compréhension des rôles respectifs des différents phénomènes qui gouvernent l'évolution des nuages: l'auto-gravité, la turbulence, les champs magnétiques, ainsi que les influences galactiques comme la pression extérieure ou le cisaillement des bras spiraux. La rétro-action stellaire dans les nuages évolués est aussi une source majeure d'incertitude. La description de la structure des nuages sous forme fractale ou hiérarchisée est également problématique, notamment parce que toutes les sous-structures identifiées dans les observations ne correspondent pas toujours à des structures physiques tri-dimensionnelles. En effet, les effets de projection le long de la ligne de visée sont une limitation systématique pour les observateurs, et la question de l'attribution de structures identifiées en position-position-vitesse (PPV) à des objets physiques dans l'espace position-position-position (PPP) est loin d'être évidente. Comprendre la dissipation de la turbulence MHD et son influence sur la condensation et la structuration du gaz peut aussi apporter un éclairage aux questions ci-dessus. Enfin, la question spécifique de la chimie se pose dans les nuages moléculaires, d'une part car nombre de réactions et de réseaux chimiques restent à étudier, d'autre part car l'utilisation des raies moléculaires comme outil de diagnostic pour le milieu interstellaire nécessite encore un étalonnage détaillé.

Le projet ORION-B, dont cette thèse fait partie, tente de répondre à certaines de ces questions. Le chapitre 2 présente en détail le nuage étudié, les observations réalisées, et résume les études statistiques qui se sont concentrées sur différents diagnostics fournis par l'émission moléculaire. Le chapitre 3 étudie la turbulence dans les nuages moléculaires, en particulier la façon d'utiliser des informations PPV pour déterminer les propriétés 3D de la turbulence, et le lien de celles-ci avec la formation stellaire. Le chapitre 4 s'attache aussi à la formation stellaire, mais du point de vue des sous-structures du nuage, en l'occurence les filaments. Les conclusions et perspectives sont présentées dans le chapitre 5.

2 Le projet ORION-B

2.1 Observations et objectifs

Le projet ORION-B (Outstanding Radio-Imaging of OrioN B), comme son nom l'indique, est consacré à l'observation du nuage moléculaire géant Orion B dans le domaine radio, dans le but d'obtenir un jeu de données avec une couverture spatiale et spectrale inégalée.

Orion B fait partie du complexe d'Orion, la région de formation stellaire massive en activité la plus proche du Soleil. Ce complexe couvre l'essentiel de la constellation d'Orion, et contient plusieurs générations d'étoiles de type O et B, qui, pour les plus anciennes, se sont formées il y a 8 à 12 millions d'années. Les plus jeunes d'entre elles, le groupe OB1d, a moins de 2 millions d'années, et se trouve au cœur des nuages Orion A et Orion B, qui sont les deux nuages moléculaires géants les plus massifs de ce complexe.

Orion B a une masse totale de $7 \times 10^4 M_\odot$, et se trouve à une distance d'environ 400 pc du Soleil. Sa taille sur le ciel correspond à environ 30×15 pc. Le nuage peut être approximativement subdivisé en deux moitiés, la moitié nord-est et la moitié sud-ouest. C'est la moitié sud-ouest du nuage qui est la cible du projet ORION-B. À l'extrémité ouest du nuage se trouve notamment la célèbre nébuleuse de la Tête de Cheval, un pilier dense érodé par la photo-évaporation due au rayonnement UV de σ Ori, le système d'étoiles brillantes à l'origine de la région HII IC 434. Le bord sud-ouest d'Orion B abrite également les régions de formation stellaire NGC 2023 et NGC 2024, qui sont de jeunes régions HII respectivement éclairées par des étoiles de types B et O. NGC 2024 est aussi connue sous le nom de nébuleuse de la Flamme, à cause de la présence d'un grand filament sombre et ramifié en avant-plan de la région HII.

Le sud-ouest d'Orion B contient aussi bien des régions diffuses que très opaques, des régions chaudes et illuminées que froides et obscurcies ; de plus, il abrite de la formation stellaire aussi bien spontanée que

déclenchée. C'est donc un environnement très riche, particulièrement adapté pour étudier les conditions physiques et chimiques dans un nuage moléculaire et leur lien avec la formation stellaire.

Les observations d'Orion B, intégralement réalisées avec le télescope de 30 m de l'IRAM, situé à Pico Veleta dans la Sierra Nevada espagnole, ont commencé par deux pointages de grande sensibilité dans la Tête de Cheval, couvrant la totalité des bandes spectrales à 3, 2 et 1 mm, dans le cadre du projet WHISPER de Jérôme Pety et Viviana Guzmán (Guzmán 2013, thèse de doctorat). La volonté de placer ces observations dans un contexte plus large a mené à l'acquisition en 2013 d'un champ d'environ 0,9 $\deg^2$ dans la bande à 3 mm (entre 84 et 116 GHz), ce qui correspond à la phase pilote du projet ORION-B. Au cours des années, l'étendue spatiale des observations a augmenté, ainsi que leur étendue spectrale grâce aux progrès du récepteur du télescope à basses fréquences (72 – 80 GHz). Enfin, en 2016, le projet est devenu un *Large Programme*, avec 550 heures d'observations supplémentaires allouées (pour un total d'environ 850 depuis le début du projet), avec comme objectif la couverture de 5 degrés carrés du ciel, soit environ la moitié de la surface du nuage.

Les observations ORION-B couvrent la quasi-totalité de la bande radio à 3 mm, de 72 à 80 et de 84 à 116 GHz, couverture qui est obtenue en trois réglages du récepteur. La résolution fournie par le télescope de 30 m varie entre 36″ à 72 GHz et 22.5″ à 116 GHz. Le récepteur EMIR est utilisé avec le spectromètre FTS (spectromètre à transformée de Fourier), qui fournit une largeur de bande d'environ 16 GHz à une résolution de 195 kHz. Quelques raies sont par ailleurs observées avec l'autocorrélateur VESPA, qui fournit une résolution de 40 kHz sur une faible largeur de bande. Après calibration et réduction, tous les cubes de données des raies moléculaires détectées sont reprojetés sur une grille PPV régulière de $9'' \times 9'' \times 0.5\,\mathrm{km\,s^{-1}}$ (ou $9'' \times 9'' \times 0.1\,\mathrm{km\,s^{-1}}$ pour les cubes VESPA).

La richesse du jeu de données obtenu dans le cadre du projet ORION-B, avec plus de 25 raies moléculaires cartographiées, permet une grande variété d'études. Les différentes raies permettent de tracer des environnements variés en terme de densité, de température, de conditions chimiques, tandis que la haute résolution spectrale donne un accès détaillé à la dynamique du gaz. Cependant, la couverture spatiale, le volume des données et leur diversité ouvrent des possibilités qui vont au-delà de la simple étude d'un nuage moléculaire.

L'étude de la formation stellaire ne saurait se limiter à l'échelle des nuages moléculaires. En effet, ces derniers sont immergés dans un environnement galactique plus vaste qui peut influer sur leur propriétés chimiques et physiques (via la pression externe, les explosions de supernovae, etc.) ce qui peut contribuer à déterminer la structure de densité et la dynamique des nuages, et ainsi modifier les conditions dans lesquelles les étoiles s'y forment. Cependant, ces effets à grandes échelles sont difficiles à isoler au sein de la Voie Lactée, notamment à cause des limitations des observations qui subissent une forte confusion le long de la ligne de visée. Dans cette situation, les observations extragalactiques sont très précieuses. Cependant, ces mesures souffrent d'incertitudes importantes, d'une part parce que les relations qui permettent de relier les grandeurs observationnelles aux grandeurs physiques (par exemple relier la brillance d'une raie de CO à la densité de colonne totale de gaz moléculaire) peuvent dépendre fortement des conditions physiques (température, illumination,degré d'ionisation...), qui sont mal connues pour les sources extragalactiques ; et d'autre part parce que la résolution spatiale, forcément limitée par la distance des galaxies externes, ne permet pas d'échantillonner avec précision les variations locales dans les conditions du gaz, et donc dans les relations entre observables et grandeurs physiques. Le projet ORION-B peut fournir une référence : l'observation détaillée d'un nuage moléculaire bien connu permet d'étalonner ces relations, et de comparer ensuite les spectres moyens obtenus pour telle ou telle région du nuage Orion B avec des spectres du MIS extragalactique. Par ailleurs, la couverture spatiale des observations ORION-B permet de faire la jonction entre les études de la dynamique du MIS local, qui se concentrent habituellement sur des objets de taille $\lesssim 1$ pc comme des filaments ou des cœurs denses, et les études extragalactiques où les mouvements considérés concernent l'échelle des bras spiraux des galaxies.

Par ailleurs, la taille du jeu de données généré par les observations ORION-B est un défi en elle-même. En effet, les $\sim$ 10 TB de données attendus contiendront 200 000 canaux spectraux pour chacun des

738 000 pixels. Pour l'analyse des données, il n'est donc pas possible de traiter en détail et manuellement chaque spectre, ni chaque carte. Il est donc nécessaire de faire appel à l'*apprentissage automatique* pour extraire des informations de ces données. L'apprentissage automatique est une discipline de l'intelligence artificielle, et consiste à concevoir des algorithmes capables d'améliorer leurs performances dans la réalisation d'une tâche au fur et à mesure qu'ils acquièrent de l'expérience. En astrophysique, les méthodes de classification sont le plus souvent appliquées à des objets individuels, comme des étoiles, des galaxies, des supernovae... plutôt que des milieux continus comme le MIS. L'utilisation de l'apprentissage automatique pour étudier des données hyper-spectrales sur des cartes étendues appartient en fait surtout au domaine de la télédétection terrestre, qui cherche à caractériser des terrains observés depuis l'espace par des satellites. Les observations du projet ORION-B fournissent des cartes d'un grand nombre de traceurs, qui tracent une grande diversité d'environnements, environnements qu'on peut tenter d'identifier et de caractériser par ces méthodes statistiques d'apprentissage automatique.

2.2 Études statistiques de l'émission moléculaire

Les premières études statistiques ont été réalisées sur les données du projet pilote de 0,9 degré carré, avec une couverture spectrale de 84 à 116 GHz.

La première étape d'analyse des observations ORION-B a été le calcul des propritétés moyennes du nuage et de l'émission des raies moléculaires, ainsi que l'estimation des corrélations entre les traceurs moléculaires, l'extinction et la température des poussières (Pety et al. 2017).

La première quantité mesurée pour le nuage est sa masse, qui peut être déterminée de plusieurs façons : en estimant la densité de colonne totale du gaz à partir de l'émission des poussières, en la déduisant de l'émission totale de la raie de ^{12}CO ($J = 1 - 0$), et en supposant un équilibre viriel entre la pression turbulente dans le nuage (déduite de la dispersion de vitesse mesurée pour ^{12}CO) et son auto-gravité. Étant donné l'existence de gaz moléculaire qui n'émet pas en CO, on s'attend à ce que la masse tracée par les poussières soit plus grande que celle tracée par ^{12}COAu contraire, on observe dans Orion B que la masse tracée par CO est de $11000 M_\odot$, celle tracée par les poussières de $3900 M_\odot$, et la masse virielle est de l'ordre de $6200–9500 M_\odot$. Ces écarts peuvent être expliqués par la forte illumination UV subie par Orion B, ce qui chauffe le gaz : CO est alors plus lumineux que ce que prévoit le facteur de conversion standard $X_{\rm CO}$, tandis que l'émission des poussières est dominée par les poussières les plus chaudes, ce qui de fait "cache" l'émission de poussières froides.

Afin de déterminer quelles molécules tracent tel ou tel environnement, le champ observé a été subdivisé en quatre masques d'extinction croissante, correspondant à des milieux diffus, translucide, filamentaire et opaque. La fraction du flux de chaque raie venant de tel ou tel masque a été mesurée, ce qui a révélé que ^{12}CO, C_2H, c-C_3H_2 et HCO^+ sont les meilleurs traceurs du milieu diffu, tandis que N_2H^+, $H^{13}CO^+$ et CH_3OH émettent surtout dans les régions les plus opaques. Ceci montre que les "traceurs de gaz dense" couramment utilisés, tels HCO^+, HCN ou HNC, ne sont pas forcément aussi adaptés pour caractériser ce type d'environnement qu'on le suppose souvent.

Le même procédé a été appliqué à quatre masques correspondant à des températures de poussières croissantes, pour mesurer la sensibilité des molécules à l'éclairement UV lointain. C_2H, c-C_3H_2, CN et HCN ressortent comme émettant surtout dans les régions chaudes et illuminées, tandis que les régions les plus froides sont tracées par N_2H^+, $C^{18}O$, CH_3OH et $C^{17}O$.

Pour rendre ces résultats plus robustes face aux possibles incertitudes de calibrations d'une observation à l'autre, la même étude a été menée sur des rapports de raies, en plus des simples intensités. Il en ressort que $N_2H^+/^{12}$CO est le meilleur traceur du gaz dense, tandis que $C_2H/^{13}$CO trace le mieux le gaz illuminé en UV lointain. La comparaison des rapports de raies dans Orion B avec ceux issus de divers environnements extragalactiques montre que Orion B est une région assez similaire à ce qui peut être observé dans un bras spiral de la galaxie M51 – avec cependant une plus faible fraction de gaz dense dans Orion B.

Les corrélations entre les différents traceurs moléculaires et les grandeurs physiques caractéristiques du gaz comme sa densité ou sa température ont aussi été étudiées à l'aide d'une méthode d'analyse multivariée, à savoir l'Analyse de Composante Principale (PCA, Gratier et al. 2017). Cette méthode permet de déterminer les axes principaux de corrélation dans un jeu de données multi-dimensionnel, ce qui revient à trouver les principaux axes d'inertie du nuage de points correpondant, et ainsi obtenir une nouvelle base de l'espace, plus adaptée à la représentation des données.

Dans notre cas, les cartes (comptant 132 200 pixels) de 12 raies moléculaires ont été étudiées, ce qui donne un nuage de 132 200 points dans un espace à 12 dimensions. Pour que l'analyse ne soit pas totalement dominée par l'amplitude des variations du signal pour les raies les plus brillantes, une renormalisation des données était nécessaire. Une fonction asinh a donc été appliquée aux données : aux grandes valeurs, son comportement est logarithmique, ce qui atténue les variations des données, et aux faibles valeurs, son comportement est linéaire, ce qui évite une divergence du logarithme dans le régime du bruit.

La PCA appliquée aux données renormalisées donne 12 composantes principales (PC). Alors que l'information spatiale n'était pas fournie en entrée à l'algorithme, les PC montrent clairement des structures spatiales cohérentes. Les PC les plus importantes, qui contribuent le plus aux corrélations du jeu de données, ont été comparées avec des grandeurs physiques déterminées indépendamment.

Toutes les raies moléculaires ont des contributions quasiment égales à la PC1, qui à elle seule compte pour 60% des corrélations. Cette composante s'avère aussi extrêmement bien corrélée à la densité de colonne du gaz. En effet, plus il y a de matière le long de la ligne de visée, plus il y a de molécules qui émettent et donc de flux détecté dans les raies. La PC2, qui contient 11% des corrélations, a une forte contribution positive de N_2H^+ et CH_3OH, anti-corrélés avec ^{12}CO et ^{13}CO. Cela correspond donc à un contraste entre les traceurs du milieu dense et ceux du milieu diffus ou transluscide. La comparaison entre cette PC2 et des estimations de densité volumique issues d'un catalogue de cœurs denses montre en effet une corrélation entre ces deux grandeurs. Enfin, la PC3, qui contient 7% des corrélations du jeu de données, présente un contraste entre C_2H et CN d'une part, et ^{12}CO, ^{13}CO, $C^{18}O$, N_2H^+ et CH_3OH d'autre part. Vu que N_2H^+ et CH_3OH tracent les régions opaques et C_2H les régions chaudes et illuminées, PC3 semble donc représenter l'illumination UV – ce qui est confirmé par l'étude de la corrélation entre PC3 et le flux UV déduit de la luminosité des PAH (molécules polycycliques aromatiques hydrocarbonées) dans Orion B.

La PCA permet donc de reconstruire approximativement des grandeurs physiques caractéristiques du gaz du nuage à partir des raies moléculaires, tout en effectuant une compression des données, puisque près de 80% des corrélations du jeu de données peuvent être exprimées avec seulement 3 dimensions (au lieu des 12 initiales), fournissant ainsi une vue synthétique de l'émission moléculaire du nuage.

Une autre approche, complémentaire de la PCA, a consisté à identifier des régions caractéristiques du nuage à partir de leur émission moléculaire. Pour cela, l'algorithme de partionnement (*clustering*) MeanShift a été utilisé (Bron et al. 2018). Le partitionnement consiste à regrouper les points du jeu de données en fonction de leur proximité. Un nombre prédéfini ou non de groupes de points (partitions) peut ainsi émerger. Dans le cas de l'algorithme MeanShift, le groupement se fait autour des maxima locaux de la fonction de densité de probabilité du nuage de points. La densité des points dans l'espace N-dimensionnel sert d'abord à estimer cette densité de probabilité, donc les maxima servent ensuite d'attracteurs pour les points environnants.

Un premier partitionnement a été réalisé avec les seules raies $^{12}CO\,(J = 1 - 0)$, $^{13}CO\,(J = 1 - 0)$ et $C^{18}O\,(J = 1-0)$. Encore une fois, aucune information spatiale n'est fournie a priori, mais le partionnement révèle 8 régions spatialement cohérentes et à peu près concentriques. On peut donc ainsi distinguer des régions de densité croissante. Les deux partitions correspondant au milieu le plus dense, CO-7 et CO-8, ne sont pas emboîtées l'une dans l'autre, et se distinguent surtout par la température du gaz, ce qui se traduit par un ^{12}CO plus brillant dans le gaz plus chaud.

Un second partionnement est réalisé en ajoutant les raies de CN (qui est sensible à l'illumination) et HCO^+ (qui a une densité critique plus élevée que CO). Cette fois, 19 partitions sont identifiées. Elles

sont toujours spatialement cohérentes, mais mais leur grand nombre les rend plus difficiles à interpréter. Les partitions sont donc groupées selon leur intensité moyenne en CN ou HCO^+. Les groupes classés en fonction de CN ne donnent que peu d'information sur le milieu diffus ou translucide, mais proposent une vue intéressante du milieu plus dense : une distinction très nette apparait entre le milieu dense peu illuminé (froid) et le milieu plus chaud et soumis au rayonnement UV, puisque les régions HII autour de NGC 2023 et NGC 2024 ainsi que la PDR au bord de la région HII IC 434 forment un groupe bien distinct. Les groupes triés en fonction de HCO^+ forment une structure plus concentrique, comme CO, mais révèlent des sub-divisions à des densités plus élevées que ce que ne pouvait distinguer les isotopologues de CO. Ceci est probablement dû au fait qu'à l'approche de la densité critique de HCO^+, son mode d'émission change.

L'identification de telles régions caractéristiques à partir des données permet par la suite de raffiner les analyses de spectres moyens ou de corrélations entre les traceurs (à l'aide de la PCA par exemple) en ciblant spécifiquement tel ou tel environnement.

3 Modes de turbulence dans le nuage Orion B

3.1 La turbulence dans les nuages moléculaires géants

La turbulence est "l'agitation d'un fluide qui s'écoule en tourbillons". Cet aspect de l'hydrodynamique a longtemps été délaissé par les physiciens, à cause de son aspect aléatoire qui semblait aller à l'encontre du déterminisme scientifique. Ce n'est qu'à la fin du 19^{e} siècle qu'Osborne Reynolds étudia de près la transition entre les écoulements laminaires et turbulents, et montra qu'elle dépendait de la vitesse v de l'écoulement, de sa taille caractéristique L, de la densité ρ du fluide, et de sa viscosité η. Ces grandeurs se combinent pour former le nombre de Reynolds $R_e = \rho v L/\eta$, qui détermine la nature de l'écoulement. La transition turbulente a lieu pour des valeurs de R_e de l'ordre de 2000 – 13000 en fonction de la géométrie de l'écoulement. En 1941, Kolmogorov fournit une première description statistique des écoulements turbulents, caractérisés par une cascade d'énergie cinétique qui commence à l'échelle d'injection et, via des tourbillons de plus en plus petits, transporte cette énergie jusqu'à l'échelle de dissipation où la viscosité vient freiner le mouvement du fluide.

Dans le milieu interstellaire, et notamment dans les nuages moléculaires, les écoulements de gaz atteignent des nombres de Reynolds de 10^7 et plus – et sont donc largement turbulents. Par ailleurs, ces écoulements sont la plupart du temps supersoniques, ce qui est incompatible avec l'hypothèse de fluide incompressible de la théorie de Kolmogorov. La compressibilité du fluide ajoute donc un degré de liberté supplémentaire à l'écoulement. Enfin, le MIS étant ionisé, les écoulements sont magnétisés, ce qui nécessite de les décrire à l'aide de la magnéto-hydrodynamique (MHD). Bien que la turbulence MHD compressible du milieu-interstellaire soit bien plus complexe que le modèle hydrodynamique incompressible décrit par Kolmogorov, le principe de cascade d'énergie jusqu'à dissipation reste valable. On s'attendrait donc à ce que, une fois la turbulence dissipée (en quelques millions d'années pour un nuage moléculaire) le milieu soit calme. Or il n'en est rien, ce qui implique une injection continuelle d'énergie cinétique, par des mécanismes externes (rotation galactique, collisions de nuages...) ou internes aux nuages (jets, explosions de supernovae, auto-gravité...). Cet état continuellement turbulent des nuages moléculaires joue un rôle clé dans la formation stellaire, puisque d'une part les chocs peuvent créer des sur-densités et déclencher des effondrements gravitationnels, tandis que d'autre part la pression turbulente s'oppose à la gravité.

Il faut en effet distinguer deux types de mouvements dans un écoulement, turbulent ou non : les mouvements solénoïdaux (tourbillonnaires), qui sont caractérisés par une divergence nulle et sont donc les seuls mouvements présents dans les fluides incompressibles, et les mouvements compressifs, qui sont caractérisés par un rotationnel nul. Dans la turbulence, les deux types de mouvement contribuent à la pression turbulente et à la dissipation de l'énergie cinétique, mais les mouvements compressifs contribuent activement à créer des sur-densités, tandis que les mouvement solénoïdaux s'opposent efficacement à l'effondrement gravitationnel via une force centrifuge. L'étude de la contribution relative des mouvements

solénoïdaux et compressifs à la turbulence des nuages moléculaires est donc un point important pour comprendre leur taux de formation stellaire.

3.2 Rôles des modes de turbulence dans la formation stellaire

Une des caractéristiques des écoulements turbulents est l'*intermittence* du champ de vitesse, c'est-à-dire l'existence de régions très localisées dans l'espace et le temps où la vorticité et les gradients de vitesse atteignent des valeurs extrêmement élevées. Cette intermittence a été identifiée dans le MIS, et a pour conséquence une distribution de probabilité non-gaussienne des vitesses et une structure fractale du champ de vitesse. Dans les écoulements compressibles, on observe aussi une intermittence du champ de densité, qui est lui aussi fractal. Cette intermittence est caractérisée dans le MIS par une augmentation de la dispersion de la densité de colonne liée à l'augmentation de sa valeur moyenne (les milieux les plus denses sont donc les plus contrastés). La forme de la relation entre dispersion et moyenne suggère que la distribution de probabilité sous-jacente des fluctuations de densité est log-normale (c'est-à-dire que la distribution de probabilité du logarithme de ces fluctuations est gaussienne).

La largeur de cette distribution de densité est croissante avec le nombre de Mach sonique de l'écoulement : $\sigma_{\rho/\langle\rho\rangle} = b\mathcal{M}_s$ dans le cas hydrodynamique[1]. Le paramètre b dépend en fait du forçage solénoïdal ou compressif de la turbulence ; $b \approx 1$ dans le cas purement compressif et $b \approx 1/3$ dans le cas purement solénoïdal. Ainsi, la distribution des fluctuations de densité est plus large (et donc présente plus de valeurs extrêmes) dans le cas compressif que dans le cas solénoïdal : les chocs compressifs créent plus de sur-densités (et réciproquement, de creux) que les mouvements tourbillonnaires, donnant au fluide une structure en feuillets, avec une dimension fractale proche de 2.

Nombre d'observations sont venues confirmer la forme log-normale de la distribution de densité du MIS, et le lien entre la largeur de cette distribution et la présence de mouvements compressifs, par exemple autour de régions HII en expansion. Cependant, quelques écarts à la simple distribution log-normale ont été observés, comme la possibilité d'avoir deux composantes log-normales (une plutôt solénoïdale et une autre, plus dense, plutôt compressive) ou encore la présence d'une queue de distribution en loi de puissance, signature de l'auto-gravité.

Le lien entre mouvements compressifs et présence de sur-densités devrait avoir un impact en terme de formation stellaire. En effet, celle-ci ne se déclenche qu'au-delà d'un certain seuil de densité à partir duquel l'auto-gravité devient dominante. De plus, la formation stellaire alimente aussi ces modes compressifs via les vents stellaires, les jets, ou l'expansion de jeunes régions HII. Diverses simulations ont confirmé ce résultat intuitif, à l'échelle du nuage moléculaire et même à l'échelle galactique : l'activité de formation stellaire est systématiquement plus intense, plus efficace, et démarre plus rapidement lorsque le forçage de la turbulence est compressif que lorsqu'il est solénoïdal. Cependant, ces résultats montrent aussi implicitement que les nuages moléculaires ne peuvent pas être dominés par des mouvements compressifs, alors même que la plupart des phénomènes qui injectent de l'énergie cinétique dans les nuages moléculaires (auto-gravité, collisions de nuages, explosions de supernovae, jets stellaires...) sont de nature compressive. En effet, les simulations de turbulence purement compressive donnent des taux de formation stellaire nettement supérieurs à ce qui est observé dans le MIS. Cela implique que des instabilités MHD peuvent efficacement mélanger les modes, convertissant des mouvements compressifs en tourbillons et réciproquement. Des simulations ont ainsi montré que, pour des écoulements turbulents hypersoniques ($\mathcal{M}_s > 5$), la fraction de quantité de mouvement dans les modes compressifs et solénoïdaux ne dépend plus des modes dans lesquels l'injection a lieu, mais se retrouve systématiquement dans un état d'équipartition, avec 2/3 de la quantité de mouvement dans les modes solénoïdaux et 1/3 dans les modes compressifs.

[1] Dans le cas magnéto-hydrodynamique, la relation devient $\sigma_{\rho/\langle\rho\rangle} = b\mathcal{M}_s\sqrt{\beta/(\beta+1)}$, où $\beta = 2(\mathcal{M}_A/\mathcal{M}_s)^2$ et $\mathcal{M}_A$ est le nombre de Mach alfvénique.

3.3 Mesure de la fraction de forçage compressif dans Orion B

La détermination de la contribution relative des mouvements solénoïdaux et compressifs dans la turbulence qui agite les nuages moléculaires est donc un élément très important pour la compréhension du taux et de l'efficacité de formation stellaire des nuages moléculaires. Dans le cas de simulations, où la totalité de l'information tridimensionnelle (PPP) du champ de densité et du champ de vitesse est connue, cela peut être directement mesuré en décomposant le champ de vitesse $\vec{v}$ ou le champ de quantité de mouvement $\rho\vec{v}$ en leurs composantes à divergence nulle ou à rotationel nul grâce au théorème de décomposition de Helmholtz. Dans le cas d'observations, en revanche, en l'absence d'une information tridimensionnelle complète, les mesures sont majoritairement limitées à des comparaisons avec les résultats de simulations. Ces comparaisons sont confrontées à de fortes incertitudes, notamment à cause de fortes dégénérescences avec des paramètres tels que le nombre de Mach ou l'intensité du champ magnétique. Cependant, il est possible dans certains cas de mesurer directement la fraction de quantité de mouvement dans les différents modes. En effet, Brunt & Federrath (2014) ont proposé une méthode qui permet, sous conditions, de reconstruire la fraction de quantité de mouvement dans les modes solénoïdaux ou compressifs uniquement à partir de données observationnelles. Cette méthode a été appliquée aux données ORION-B (Orkisz et al. 2017) afin de caractériser les mouvements dans 1,5 degrés carrés de ce nuage.

L'idée directrice de la méthode en question est de remarquer que pour un champ F, la transformée de Fourier de la projection bidimensionnelle du champ $\tilde{F}_{2D}$ est proportionnelle à la coupe à $k_z = 0$ (z étant l'axe de projection) de la transformée de Fourier du champ tridimensionnel $\tilde{F}_{3D}(k_z = 0)$. Si le champ F est isotrope, on peut donc reconstruire statistiquement les propriétés de $\tilde{F}_{3D}$, et donc de F, à partir des informations sur $\tilde{F}_{2D}$. C'est en effet ce que l'on cherche à faire lorsqu'on étudie un objet tridimensionnel (le champ de vitesse d'un nuage par exemple) à partir de données observationnelles bidimensionnelles. En combinant cette propriété de la transformée de Fourier avec les propriétés de projection des modes solénoidaux et compressifs de la décomposition de Helmholtz, on peut remonter au rapport de la variance des modes solénoïdaux du champ de quantité de mouvement à la variance de la quantité de mouvement totale, $R = \sigma^2_{\rho_\perp}/\sigma^2_\rho$, et ce en ne faisant appel qu'à des grandeurs observationnelles, qui peuvent être mesurées à partir d'un cube PPV. Ces grandeurs sont essentiellement les moments pondérés en vitesse du cube PPV, ainsi que les spectres de puissance de ces moments.

Dans le cas d'Orion B, le cube PPV de ^{13}CO ($J = 1 - 0$) a été choisi, car il offre un bon rapport signal-à-bruit tout en évitant les fortes saturations de ^{12}CO ($J = 1 - 0$), qui est optiquement épais sur la majorité du champ observé. Pour réduire encore la sensibilité des résultats au bruit, le cube PPV est débruité en éliminant les voxels individuels en dessous d'un seuil de signal-à-bruit et les "îlots" de voxels contigus dont le signal total est en dessous d'un flux fixée. De plus, on soustrait des spectres de puissance un modèle du spectre du bruit obtenu à partir des transformées de Fourier des canaux spectraux dont on sait qu'ils ne contiennent pas de signal.

La valeur finale du ratio R obtenue pour le champ étudié est de $0{,}72^{+0{,}09}_{-0{,}09} < R < 1^{+0{,}0}_{-0{,}09}$. Comme mentionné plus haut, on s'attend à avoir un mélange des modes qui mène à une équipartition pour des nombres de Mach $\mathcal{M}_s > 5$. Or, dans Orion B, on mesure un nombre de Mach moyen d'environ 6, et la fraction solénoïdale est plus élevée que le $R = 2/3$ attendu à équipartition. Nous n'avons donc pas affaire à de la turbulence "pure", mais à des mouvements d'ensemble solénoïdaux ordonnés qui se superposent à la turbulence. À l'opposé, les zooms ciblés sur les deux régions de formation stellaire NGC 2023 et NGC 2024 (où la température augmente et donc le nombre de Mach diminue) montrent des fractions solénoïdales qui descendent jusqu'à 0,32 et 0,25 respectivement. On constate donc que le nuage dans son ensemble est dominé par des mouvements solénoïdaux, mais qu'il contient des régions de formation stellaire fortement compressives. Ceci est à comparer avec l'activité de formation stellaire dans Orion B. On remarque que ce nuage moléculaire a une efficacité de formation stellaire particulièrement basse : environ 4 fois plus basse qu'Orion A et la plus basse de tous les nuages moléculaires à moins de 500 pc du Soleil. Par ailleurs, il est aussi atypique que la formation stellaire soit totalement concentrée dans des amas, sans objets stellaires jeunes plus isolés. Cela correspond bien à ce que nous montre l'analyse des mouvements du nuage Orion B : l'ensemble du nuage est dominé par des mouvements qui

s'opposent globalement à la formation stellaire, tandis que les régions de formation stellaire, témoins de l'effondrement gravitationnel du gaz et/ou de l'expansion des région HII jeunes, sont fortement compressives.

Cette analyse mériterait d'être placée dans un contexte plus large, qui étudierait le rapport entre la fraction de variance dans les modes solénoïdaux du champ de quantité de mouvement et l'activité de formation stellaire dans plusieurs nuages moléculaires, et qui comparerait les résultats de cette méthode avec d'autres indicateurs de la nature compressive ou solénoïdale des mouvements du gaz, mais il est à l'heure actuelle difficile de générer plusieurs jeux de données de la qualité nécessaire pour réaliser ce type d'analyse. Cependant, l'étude d'observables capables de comparer modèles numériques et observations en terme de turbulence et plus généralement en termes de nature des mouvements qui agitent le gaz est une voie très prometteuse pour améliorer notre compréhension de la formation stellaire.

4 Les filaments d'Orion B

4.1 Les filaments dans les nuages moléculaires géants

De part leur forme allongée, les "filaments" de toutes sortes attirent l'œil des astronomes. Bien que le terme puisse selon les contextes décrire des objets très différents en terme de tailles (de quelques milliparsecs à plusieurs centaines de parsecs) et de densités (du milieu extragalactique au milieu moléculaire dense), les filaments qui ont le plus capté l'attention ces dernières décennies sont les filaments denses présents dans les nuages moléculaires. Ces filaments ont été remarqués dans le voisinage de régions de formation stellaire, ce qui suggère qu'ils jouent un rôle important dans la concentration de la matière depuis l'échelle des nuages moléculaires (plusieurs dizaines de parsecs) à l'échelle de la formation stellaire (de l'ordre du milliparsec). Leur présence dans tous les nuages a été largement confirmée par les observations du télescope *Herschel*.

Les filaments moléculaires présentent une grande diversité de formes (linéaires, tortueux ou ramifiés), de tailles et de densités. Leur longueur varie d'un ordre de grandeur, de même que leur diamètre, mesuré entre 0.03 et 0.4 pc (avec une épaisseur typique de l'ordre de 0.1 pc). Les densités de colonne au centre des filaments sont de l'ordre de 10^{21} à 10^{22} cm^{-2}, ce qui correspond à des densités volumiques de 10^4 à 10^5 cm^{-3}. Les masses linéiques des filaments vont de quelques unités à quelques centaines de $M_\odot$/pc, tandis que les masses totales vont de $\sim 1 M_\odot$ à $\sim 10^3 M_\odot$ pour les plus massifs. Les filaments sont par ailleurs sous-structurés, avec des séries de cœurs denses le long des filaments, comme des perles sur un collier, ainsi que la possible présence de "fibres", qui sont des structures longitudinales, plus fines, qui s'entremêlent pour former le filament.

L'origine des filaments moléculaires semble directement liée au processus de formation stellaire, même si les mécanismes exacts sont encore débattus. Pour former une étoile, il faut accumuler assez de matière dans un petit volume. Pour y arriver à partir de l'échelle du nuage moléculaire, on peut notamment diminuer le nombre de dimensions en augmentant progressivement la densité, en formant des feuillets de gaz, puis des filaments plus denses, puis enfin des cœurs préstellaires très denses. Plusieurs phénomènes peuvent contribuer à réduire progressivement la dimensionnalité : la gravité, qui tend à accentuer toute anisotropie, les champs magnétiques, qui peuvent guider la contraction ou l'expansion de la matière le long des lignes de champ, la turbulence, qui peut générer des surfaces de chocs ou des lignes de tourbillons, et la compression due à des phénomènes "ordonnés" comme la pression du gaz, la pression radiative, ou les vents stellaires. À peu près toutes les combinaisons de ces phénomènes ont été proposées et étudiées pour passer de trois dimensions à deux puis de deux à une, formant ainsi les filaments. Le point de vue le plus fréquent (mais pas universel) est celui où le gaz est accumulé sous forme d'un feuillet, sous l'action de mouvements compressifs turbulents ou ordonnés, qui ensuite se fragmente en filaments sous l'effet de l'auto-gravité, le champ magnétique contribuant plus ou moins à accentuer les anisotropies à chacune des deux étapes.

Le mécanisme de formation des filaments peut avoir un impact sur leurs propriétés internes, en

particulier sur leur diamètre, qui peut être fixé par un équilibre de pressions ou par l'échelle dissipative de la turbulence, mais les simulations ne s'accordent pas encore sur des observables qui permettraient d'identifier clairement le ou les mécanismes de formations des filaments.

Une fois les filaments unidimensionnels formés, il reste à créer des sur-densités ponctuelles pour arriver aux cœurs préstellaires. La grande majorité des cœurs denses dans les nuages sont localisés dans des filaments, et leur distribution de masse préfigure bien celle des étoiles, ce qui confirme que les filaments sont à l'origine des cœurs préstellaires. On peut former ces "points" en fragmentant les "lignes" (les filaments) ou en les intersectant.

Le premier cas est bien identifié : les cœurs peuvent être issus d'une fragmentation longitudinale due à une instabilité gravitationnelle des filaments. La fragmentation peut se produire à plusieurs échelles en fonction de mécanismes d'instabilité différents, qui se développent à partir de masses linéiques de l'ordre de 15 – 16 $M_\odot$/pc. Cette fragmentation peut parfois se faire de façon hiérarchisée, avec plusieurs groupes de cœurs se formant le long d'un filament, avec des distances typiques entre les cœurs de 0,1 à 0,5 pc environ. Cependant, si la gravité était la seule cause de fragmentation des filaments, les distributions de taille, masse et espacement des cœurs devraient être très piquées. Or il n'en est rien : les instabilités sont initiées par des perturbations turbulentes dont l'amplitude peut être importante et qui, de part leur nature aléatoire, élargissent ces distributions.

D'un autre côté, la convergence de plusieurs filaments en un nœud central semble être un mécanisme privilégié pour la formation d'étoiles massives. L'aspect de ces nœuds indique qu'ils pourraient se former à partir d'une zone de gaz dense compressée en un feuillet et évoluant sous l'effet de son auto-gravité : le nœud central et les filaments qui l'entourent, approximativement parallèles, se forment donc simultanément. La matière environnante est accrétée sur les filaments, puis s'écoule le long de ceux-ci et vient alimenter un ou plusieurs cœurs très massifs au centre du nœud, pouvant donner naissance à des étoiles de type O et B.

Ces deux mécanismes ne sont pas mutuellement exclusifs : la fragmentation longitudinale peut notamment être observée dans des filaments qui par ailleurs convergent vers un nœud massif.

Outre les questions sur la formation des filaments et leur évolution pour former des cœurs denses, quelques autres problèmes se posent lors de l'étude des filaments moléculaires. Tout d'abord, il y a la question de savoir si les objets observés correspondent toujours dans l'espace à des structures unidimensionnelles et continues. En effet, dans certains cas on peut avoir des "filaments" qui sont en fait des objets 2D vus par la tranche. Par ailleurs, des simulations ont montré qu'il pouvait y avoir confusion le long de la ligne de visée, avec des structures apparemment continues qui sont faites de plusieurs fragments indépendants à des distances différentes. À propos de sous-structures, l'apparition des fibres dans les filaments est mal comprise, et deux visions s'opposent : une où les filaments s'effilochent par fragmentation radiale et forment progressivement des fibres qui se fragmentent ensuite longitudinalement pour former des cœurs, et une autre où les fibres se forment d'abord via la turbulence et se regroupent sous l'effet de mouvements à grande échelle pour former les filaments. Enfin, le diamètre des filaments est source de controverse : les largeurs des filaments observés ou simulés varient environ d'un facteur 4 autour de la valeur de 0,1 pc, mais les méthodes de mesure ne sont pas toujours les mêmes, et la question de savoir si cette échelle de 0,1 pc résulte d'un mécanisme fondamental et donc possède une certaine universalité est loin de faire l'unanimité.

4.2 Caractérisation des filaments d'Orion B

Dans Orion B, nous avons cherché à étudier l'ensemble du réseau filamentaire présent dans 1,9 degrés carrés dans la partie sud-ouest du nuage (Orkisz et al. 2018).

Pour cela, nous avons utilisé le cube PPV de $C^{18}O$ ($J = 1 - 0$), qui a servi à construire une carte de densité de colonne du gaz dense dans le nuage. Les filaments ont été détectés à l'aide de deux méthodes d'analyse morphologique différentes, toutes deux faisant usage des propriétés de la matrice hessienne, qui quantifie les dérivées partielles secondes du champ, et permet d'identifier les "sommets" qui correspondent

aux cœurs denses et les “lignes de crête” qui correspondent aux filaments. Les deux réseaux de filaments ainsi identifiés contiennent plus de 100 filaments chacun.

Chaque filament est ensuite étudié en termes de diamètre, de densité de colonne, de masse linéique et volumique. Ceci permet ensuite de déterminer leur stabilité vis-à-vis de l’effondrement gravitationnel, en fonction de leur pression interne. Les résultats montrent qu’on a affaire à des filaments de diamètre proche de 0,11 pc avec un écart-type assez faible (0,03 pc) indépendant de la masse linéique des filaments. Les masses linéiques présentent une distribution log-normale, similaire à la distribution des densités de colonne du milieu turbulent environnant, et sont globalement plus faibles que ce qui a été observé dans d’autres nuages. Ceci a comme conséquence que très peu de filaments sont instables gravitationnellement – seuls quelques uns présents au voisinage immédiat des régions de formation stellaire approchent ou dépassent le seuil d’instabilité. À noter que la définition du seuil de stabilité dépend fortement de la pression interne considérée : en considérant seulement la pression thermodynamique, on obtient environ 1% de filaments instables et 8% de filaments marginalement stables, mais si on prend aussi en compte la pression turbulente, ces fractions tombent à 0,1 et 0,6% environ. Cette faible quantité de filaments instables (et donc susceptibles de former des cœurs préstellaires) est cohérente avec les autres résultats sur la faible activité de formation stellaire d’Orion B, et sa concentration au sein des régions NGC 2023 et NGC 2024. Cependant, au moins un cas de fragmentation longitudinale a été identifié, dans le filament dit du Colibri, et ce malgré une masse linéique sensiblement sous-critique du filament.

Nous avons également étudié l’alignement des filaments d’Orion B avec le champ magnétique mesuré dans le nuage. On s’attend à ce que les filaments soient plutôt parallèles au champ magnétique lorsqu’ils sont sous-critiques, et perpendiculaires quand ils sont sur-critiques. En pratique, on observe une orientation relative dominante vers $40 - 60°$, qui ne peut être interprétée de façon satisfaisante à cause de problèmes de projection. En revanche, l’étude du champ de vitesse au voisinage des filaments révèle des résultats intéressants. Il apparait que la dispersion de vitesse le long de la ligne de visée augmente au centre des filaments (ce qui est encore plus visible avec $C^{18}O$ ($J = 1 - 0$) qu’avec ^{13}CO ($J = 1 - 0$), qui trace plutôt l’enveloppe des filaments), alors que la norme du gradient de centroïde de vitesse, qui donne accès à la dispersion de vitesse dans le plan du ciel, est stable voire diminue vers le centre des filaments. Ceci peut s’expliquer par des chocs d’accrétions sur les filaments ou par l’éventuelle présence de fibres.

Les propriétés atypiques des filaments d’Orion B, en particulier en terme de masse linéique, peuvent être en autres dues à la façon dont les filaments ont été identifiés. En effet, nos méthodes de détection étaient assez “laxistes”, ce qui a permis de sélectionner autant de filaments de faible masse. Cependant, on peut noter qu’il n’existe pas de consensus actuellement sur la définition précise de ce qu’est un filament, ni sur la façon de les identifier.

Cette étude pourrait être étendue en améliorant la détection des filaments grâce à la structure du cube PPV (la détection actuelle se fait en deux dimensions), et en approfondissant l’étude du champ de vitesse longitudinal le long des filaments, par opposition à l’étude transversale faite pour l’instant. Par ailleurs, le filament du Colibri va bénéficier d’observations interférométriques à haute résolution grâce à NOEMA (et, probablement, ALMA) qui permettront d’étudier en détail la façon dont il se fragmente.

5 Conclusions

5.1 Une vue inédite du nuage moléculaire géant Orion B

L’étude des nuages moléculaires géants est un élément clé pour comprendre la formation des étoiles. Particulièrement complexes en terme de physique et de chimie, ils posent un défi observationnel fort, qui a été relevé par le projet ORION-B. Le jeu de données acquis est inédit et inégalé : il couvre une grande surface (12×18 pc, soit près de la moitié du nuage) à une résolution de 60 mpc, une bande passante de 40 GHz à une résolution de 195 kHz, et à une sensibilité de 0,1 K, ce qui donne un rapport signal-à-bruit maximal de 400 pour les raies moléculaires les plus brillantes. Ceci permet donc de détecter un grand nombre de molécules, d’étudier leur distribution spatiale en détail, sur une grande portion du nuage

moléculaire, et avec la possibilité d'accéder aux mouvements du gaz grâce à la haute résolution spectrale. Cela fournit des données à haute dynamique et avec un grand nombre de dimensions d'information.

L'ampleur de ce jeu de données rend possible (et exige) un traitement statistique plus ou moins automatisé de l'information. Des études statistiques ont ainsi permis de mettre en lumière des variations dans le facteur X_{CO} en fonction de l'illumination du gaz, et d'identifier les traceurs moléculaires les plus typiques du gaz diffus ou dense, et des régions obscurcies ou illuminées. L'émission moléculaire de différents traceurs a aussi pu être quantitativement corrélée à des grandeurs physiques caractéristiques du nuage comme la densité de colonne, la densité volumique ou l'illumination UV du gaz. Enfin, différentes régions typiques ont pu être identifiées sur la base de leur signature moléculaire.

La résolution spectrale des données couplée à leur grande dynamique spatiale a permis d'utiliser la raie de ^{13}CO ($J = 1 - 0$) pour caractériser les modes de la turbulence dans Orion B. Cette étude a montré que le nuage est principalement animé de mouvements solénoïdaux, ce qui est cohérent avec sa très faible activité de formation stellaire, mais que les deux principales régions de formation stellaire ont, elles, des mouvements fortement compressifs.

Par ailleurs, la raie de C^{18}O ($J = 1 - 0$) a été utilisée pour identifier et étudier les filaments présents dans le nuage moléculaire. Cette étude souligne elle aussi la faible activité de formation stellaire du nuage, qui pourrait être due à un état évolutif peu avancé. En effet, les nombreux filaments détectés (plus de 100) sont généralement peu denses et assez agités, et donc stables face à l'effondrement gravitationnel. Cependant, quelques filaments instables sont identifiés dans les régions de formation stellaire, et de la fragmentation longitudinale est observée.

Bien que ces études aient fourni des réponses importantes, elles ont aussi posé de nouvelles questions, et l'avenir du projet ORION-B est bien rempli. Tout d'abord, il est nécessaire d'étendre les études déjà réalisées sur des champs restreints aux 5 degrés carrés complets des observations, entre autres pour identifier d'autres environnements typiques et pour caractériser les variations spatiales du facteur X_{CO}. De plus d'autres méthodes d'analyse statistiques sont en développement, en particulier en considérant plus l'aspect astrochimique des données. Les études statistiques à grande échelle vont aussi de plus en plus bénéficier d'études approfondies de régions ciblées, notamment à plus haute résolution ou à d'autres longueurs d'onde, ce qui apporte une vision complémentaire et des points de référence pour les méthodes statistiques. Sur le plus long terme, on peut espérer une convergence progressive des diverses études, pour prendre simultanément en compte les statistiques de l'émission, l'information chimique, spectrale (cinématique) et aboutir à une vision très complète du nuage Orion B.

5.2 Perspectives

Au-delà du projet ORION-B, l'avenir de la radio-astronomie est fait de jeux de données de plus en plus massifs, qui permettront des observations plus détaillées et/ou l'étude d'échantillons statistiques conséquents du milieu interstellaire. Ainsi, le récepteur à 25 pixels qui devrait être installé sur le télescope de 30 m de l'IRAM dans les prochaines années pourrait transformer un Large Program comme ORION-B en un projet observationnel de routine, avec 30 à 40 h de temps de télescope. Réciproquement, avec autant d'heures, on pourrait soit observer le même champ avec une sensibilité 5 fois supérieure, soit observer 25 champs de cette taille dans divers nuages moléculaires, ce qui fournirait un échantillon conséquent d'environnements diversifiés. Ceci pourrait permettre à terme de développer une véritable classification des nuages moléculaires qui donnerait une vision claire de leur évolution.

Cependant, ces progrès ne seront possibles qu'en combinant observations, théorie et simulations numériques. En effet, même si des efforts sont réalisés pour surmonter cette limitation, les observateurs resteront toujours tributaires des "dimensions manquantes" dans leurs données, et ont besoin de modèles numériques tridimensionnels à comparer avec les observations. Réciproquement, les simulations numériques tentent d'incorporer de plus en plus d'éléments et d'échelles, mais il reste impossible de tout simuler depuis l'échelle galactique jusqu'au disque protoplanétaire avec une chimie et une rétro-action stellaire réaliste, il est donc nécessaire d'utiliser des modèles théoriques simples mais efficaces pour prendre en compte certains phénomènes.

La synergie entre des simulations réalistes, capables d'identifier les observables clés qui peuvent faire la distinction entre des scénarios de formation stellaire et d'évolution de nuage moléculaire différents, et des vastes quantités d'observations variées, capables d'apporter des statistiques solides sur les nuages moléculaires de la Voie Lactée (et, dans une certaine mesure, sur les nuages extragalactiques aussi), pourra révéler comment ces incroyables pouponnières d'étoiles se forment dans les galaxies, comment elles évoluent sous l'effet d'influences multiples et comment elles sont finalement détruites par leur progéniture stellaire.

Abstract

The new generation of wide-bandwidth high-resolution receivers turns almost any radio observation into a spectral survey. In the case of wide-field imaging of the interstellar medium, large, highly multi-dimensional datasets are generated. Such a wealth of data simultaneously provides new diagnostic tools and creates new challenges in terms of data processing and analysis - new methods therefore need to be developed. This is the goal of the ORION-B project, which is observing 5 square degrees of the Orion B molecular cloud, or about half of the cloud's surface, over the entire 3 mm band. The observations have provided velocity-resolved maps of the emission of tens of molecular tracers, which are analysed using statistical, machine learning approaches.

Having access to spatially resolved maps from many molecular species enables us to identify the tracers most typical of given regions, with different gas densities and illuminations. Conversely, the multi-dimensional data in chemical space allow us to segment the molecular cloud into typical regions based on their molecular emission, and to quantify the most meaningful correlations of different molecular tracers with each other and with physical quantities such as density or dust temperature.

The large spatial and spectral dynamical range of the dataset also provides a detailed view of the kinematics and dynamics of the molecular cloud, in particular in relation with its star formation activity. In this thesis, I statistically characterize the nature of the motions in the cloud, and quantify the amount of momentum in the compressive and solenoidal (rotational) modes of turbulence. The cloud is dominated by solenoidal motions, with the compressive modes being concentrated in two star-forming regions. I also approach star formation from a morphological point of view, by identifying and characterizing the numerous filaments present in the cloud. These prove to have rather low densities, and to be very stable against gravitational collapse - and yet they show signs of longitudinal and radial fragmentation. All these results highlight the role of compressive forcing and dense filamentary structures in the star formation process. They are also consistent with the overall very low star formation efficiency of the Orion B cloud, while suggesting that it might host more triggered star formation in the future.

List of abbreviations

ALMA	Atacama Large (sub)Millimetre Array
CMF	Core Mass Function
CNM	Cold Neutral Medium
FIR	Far Infra-Red
FUV	Far Ultra-Violet
FWHM	Full Width at Half Maximum
GMC	Giant Molecular Cloud
GMF	Giant Molecular Filament
HIM	Hot Ionized Medium
HGBS	Herschel Gould Belt Survey
HST	Hubble Space Telescope
IMF	Initial Mass Function
IR	Infra-Red
IRDC	Infra-Red Dark Cloud
ISM	Inter-Stellar Medium
LTE	Local Thermodynamic Equilibrium
MC	Molecular Cloud
MHD	Magneto-Hydrodynamics
MIR	Mid Infra-Red
NIR	Near Infra-Red
NOEMA	NOrthern Extended Millimetre Array
ORION-B	Outstanding Radio-Imaging of OrioN-B
PCA	Principal Component Analysis
PdBI	Pic de Bure Interferometer
PDF	Probability Distribution Function
PDR	Photon-Dominated Region, or Photo-Dissociation Region
PPP	Position-Position-Position (x, y, z) coordinates
PPV	Position-Position-Velocity (x, y, v_z) coordinates
SF	Star Formation
SFE	Star Formation Efficiency
SFR	Star Formation Rate (or Star-Forming Region)
UV	Ultra-Violet
VLT	Very Large Telescope
WIM	Warm Ionized Medium
WNM	Warm Neutral Medium
YSO	Young Stellar Object

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Interstellar Medium

Many lectures, books, or dissertations contain detailed, comprehensive and well-written explanations about the Interstellar Medium (ISM), the history of its discovery, its structure, properties and characteristic processes (Lequeux et al. 2004; Draine 2011; Schulz 2012; Ryden & Pogge 2015, to name a few). The aim of this chapter is therefore not to vainly attempt to surpass or even emulate the many authors who have covered the subject, but merely to provide the reader with the necessary amount of context to understand the challenges and questions that ISM studies face today, and how this thesis is trying to answer some of them.

After a brief presentation of the origins of ISM studies and a general description on the ISM, the focus will be set on Giant Molecular Clouds, the processes that govern their evolution, and in particular those that control their star formation activity.

1.1.1 A brief history of the interstellar medium

For millenia, the night sky was for people a dark background on which bright spots shine, most of them immobile, a few of them wandering across the sky over months or years. At that early stage in the history of science, astronomy was all about drawing charts of the sky, defining constellations and computing ephemerides to predict the relative motions of the Sun, Moon and planets.

Only in the 18th century, after Newton's theories had turned the focus of astronomers on celestial mechanics rather than the purely descriptive approach they had to that point, several observers started to notice hints that there was more in outer space than the Solar System and stars. Charles Messier and William Herschel can be credited, in the second half on this 18th century, for their pioneering interest in "nebulae". While hunting for comets, Messier (1781) identified tens of bright, cloudy, and, very importantly, immobile objects in the sky, which turned out, for a large fraction, to be swaths of bright gas in our Galactic neighbourhood (Fig. 1.1, left). Herschel (1785), on the other hand, observed dark patches on the sky, and wondered if they were "holes in the sky", regions of stellar fields where, for some reason, stars were absent. Over a century after Herschel's initial remark, Barnard (1913, 1919), who compiled a catalog of such dark objects (Fig. 1.1, right), understood that these were the dark counterparts of the bright clouds observed by Messier: dark nebulae are clouds of gas and dust obscuring the view and thus hiding the background stars.

The discovery of the diffuse ISM started with Hartmann (1904), who noticed the presence of an absorption line of Ca^+ in stellar spectra, at a radial velocity and linewidth different from that of the star. He concluded that this line must come from a cloud containing calcium vapour somewhere along the line of sight. The ubiquity of the diffuse ISM was revealed by Trumpler (1930), who noticed that the brightness of stellar clusters decreases faster than their apparent diameter. There is therefore an *extinction* of starlight along the line of sight, which is caused by absorption and scattering of light by dust present between the stars.

Figure 1.1: *Left:* The Messier 42 object, also know as the Orion Nebula, one of the first identified regions of the local ISM (figure by NASA, ESA and the HST Orion Treasury Project Team). *Right:* The Barnard 72 object ("the Snake"), a typical dark cloud (figure adapted from Stéphane Guisard, Los Cielos de América).

Our understanding of the nature and structure of the ISM has progressed with the conquest of the electromagnetic spectrum. In particular, technological progresses in the radio domain in the 1940s led to the detection of the 21 cm line of atomic hydrogen in 1951 (Ewen & Purcell 1951; Muller & Oort 1951), which revealed the existence of a diffuse, atomic component in the Galaxy, visible both in emission and absorption, and thus present in both a warm and a cold phase. Later, by the end of the 1960s, the detection of molecules such as NH_3 (Cheung et al. 1968), H_2CO (Snyder et al. 1969), HCO^+(Buhl & Snyder 1970; Klemperer 1970) and, most importantly, CO (Wilson et al. 1970) led to the discovery of molecular clouds.

The sensitivity, angular resolution, and spectral coverage of ground-based and space observatories have kept increasing ever since, giving us access to new ranges of spatial scales, temperatures, energies, and improving our knowledge of the spatial structure and chemical composition of the ISM.

1.1.2 The cycle and structure of the interstellar medium

The interstellar medium makes up most of the volume of our Galaxy, it forms a disk in approximate hydrostatic equilibrium with a radius of $\sim 10\,\mathrm{kpc}$ and a typical scale height of $\sim 500\,\mathrm{pc}$. It is also an excellent laboratory to study extreme physical conditions: the range of temperatures and, even more, densities (molecular clouds are 10^{20} times less dense than the Earth's atmosphere at sea level, and 10^3 times less dense than the best vacuum achieved in a laboratory) allow the existence of a whole variety of processes, and in particular exotic chemical reactions, which occur in the ISM in ways that are very difficult to study experimentally. But, in a Galaxy vastly dominated in terms of mass fraction and luminosity by stars, the importance of the ISM, which accounts for about 10 % of the baryonic mass of the Milky Way, comes from its intimate link with the stellar population, and in particular the process of star formation: galactic matter constantly circulates in a cycle that goes from stellar death to stellar birth, and passes through all the phases of the interstellar medium.

During their lifetime, stars expel matter into the surrounding medium via stellar winds, which are more intense for more massive stars. When the end of their lifetime draws near, medium-mass stars ($0.5 - 8\,M_\odot$) enter their red giant phase, when the stellar winds intensify and a significant fraction of the stellar mass is returned to the ISM. For more massive stars ($> 8\,M_\odot$), death occurs in a violent way, through a supernova explosion, which again ejects large amounts of gas and dust into the environment,

together with a colossal energy in the form of a blast wave and radiations. All the gas ejected from stars at all stages of their evolution reaches tens of thousands of Kelvins, and disperses quickly in a diffuse interstellar medium. Over time, the diffuse gas cools down by radiating its energy through continuum or line emission, and accumulates into colder and colder, and denser and denser clouds. Some parts of the coldest and densest clouds condensate even further, and form dense cores, which can collapse and, by reaching extremely high central densities and pressures, give birth to new stars. Once the gas of the parent cloud has been either accreted onto young stars or dissipated by stellar feedback, the loop has been closed and the situation is back to evolved stars expelling matter into the surrounding interstellar medium.

During this cycle, on its way from a dying star to a newly forming one, interstellar matter passes through different types of environment, or phases, that are spatially intermingled and constantly interacting with one another. These phases can be distinguished by the chemical state of the main constituent of the ISM – hydrogen, which can be in its ionized (HII, or H^+), neutral (HI, i.e., atomic H) or molecular (H_2) form. Distinctions also arise from pressure equilibrium considerations[1].

Given that the heating and cooling curve of the interstellar medium presents a bi-stable configuration, Field et al. (1969) proposed an isobaric model where two (actually three) phases where in pressure equilibrium with each other: a warm phase, with an ionized and a neutral medium (the WIM and the WNM), and a cold, neutral medium (CNM). McKee & Ostriker (1977) expanded this model by proposing the presence of a additional phase, which is created by bubbles of hot and diffuse ionized gas (HIM) originating in supernova explosions. These overlapping bubbles actually fill the majority of the ISM volume (Fig. 1.2).

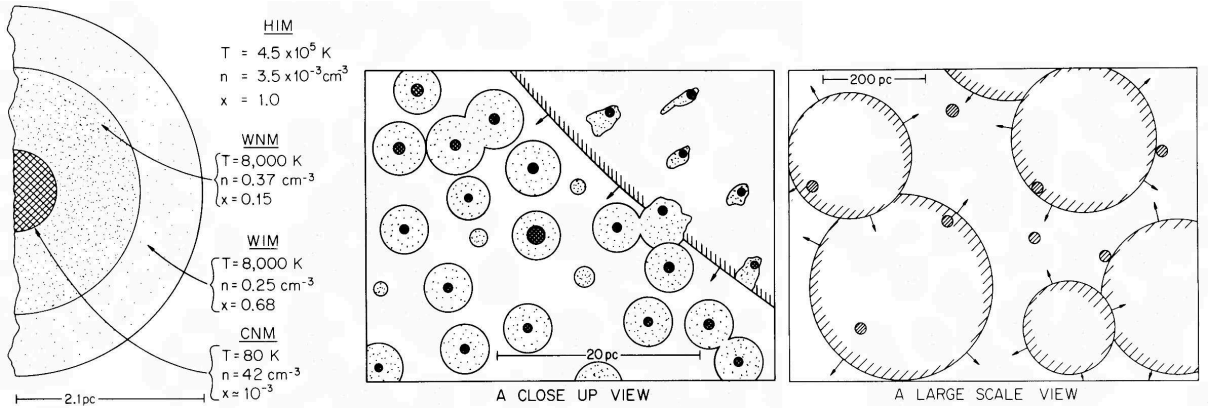


Figure 1.2: The phase structure of the ISM as proposed by McKee & Ostriker (1977). While the physical properties of the different phases (temperature T, volume density n and ionization fraction x) are generally correct, the spatial representation proposed by McKee & Ostriker (1977) is rather wrong. The proposed spatial scales of clouds rather correspond to globules than to molecular or atomic clouds that can reach hundreds of parsecs across, and the spatial distribution of clouds is a continuous, complex and turbulent structure, rather than a lattice of small, layered clouds. However, the idea of the cold atomic clouds being embedded in regions of warm atomic and ionized medium, all these clouds being impacted by expanding supernova bubbles, agrees well with the modern picture.

While the McKee & Ostriker (1977) picture of the ISM has been refined since, the general principles, as understood today, are roughly the same: the interstellar medium is divided in three phases defined by pressure equilibrium, the hot phase generated by supernova explosions, the warm phase and the cold phase. The latter two are further subdivided based on the chemical state of hydrogen, distinguishing between a warm ionized phase and a warm neutral phase, and between a cold neutral medium and a cold molecular medium, referred to as molecular clouds. All these phases are turbulent and show significant

[1]Pressure has to be understood here in a general sense, with thermal, turbulent, radiative and magnetic contributions. Earlier theories of the general ISM structure only considered thermal pressure, but it has become evident since that the other contributions are non-negligible, and actually probably all contribute to the phase equilibrium with similar orders of magnitude.

density and temperature fluctuations, and are generally embedded in each other by order of decreasing temperature. The physical properties of each phase are summarized in Table 1.1.

Phase	Temperature	Density	Ionization	Mass	Volume fraction
	K	cm^{-3}	%	$10^9\ M_\odot$	%
HIM	10^6	0.003	100	0.01	50
WIM	10^4	0.1	70	1	25
WNM	8000	0.5	10	0.3	25
CNM	100	50	10^{-2}	3	3
MC	20	10^3	10^{-3}	1	0.1

Table 1.1: Typical physical properties of the main phases of the interstellar medium: the hot ionized medium (HIM), the warm ionized medium (WIM), the warm neutral medium (WNM), and cold neutral medium (WNM) and molecular clouds (MC). The table only gives a schematic view of the ISM structure, and the values should only be considered as orders of magnitude, given the large scatter of observed conditions, and the large uncertainties, in particular regarding the total mass of each phase and the fraction of the ISM volume it occupies (adapted from Lequeux et al. 2004; Draine 2011; Ryden & Pogge 2015).

1.2 Giant Molecular Clouds

1.2.1 What are (giant) molecular clouds?

This thesis focuses on the coldest and densest phase of the interstellar medium, namely molecular clouds (and in particular the Orion B molecular cloud, Chap. 2). Such clouds are categorized based on their visual aspect and their total mass. Diffuse and translucent clouds have visual extinctions A_V of less than 1 or 5 magnitudes respectively, while dark clouds are, as the name suggests, very opaque at visible wavelengths ($A_V \sim 10$ mag) and infra-red dark clouds (IRDC) are even darker, being opaque even in the mid-IR ($A_V > 20$ mag). Except for very small globules with masses of the order of 1 $M_\odot$, molecular clouds have masses of tens or hundreds of solar masses, and span several parsecs. Larger structures, called Giant Molecular Clouds (GMC), can span tens of parsecs and weight up to $10^5\ M_\odot$. The typical gas density of a GMC is of the order of several $10^2\ cm^{-3}$. It is however only an average, as these clouds are highly sub-structured and display large density fluctuations. These sub-structures can be described in a hierarchical way, with denser and denser clumps, filaments, and cores embedded in each other (see Chap. 4), with mean densities increasing by an order of magnitude at each step. A more statistical approach, in which molecular cloud structure is an almost scale-free fractal from cloud to cores, is favoured nowadays. Among the smallest, densest and coldest objects studied in molecular clouds are prestellar cores. They have sizes of the order of 0.1 pc, and mean densities of several $10^5\ cm^{-3}$, with even higher central densities, which can lead to star formation. On the other hand, low-density bubbles and cavities can be hollowed out in Giant Molecular Clouds by the star formation activity, in the form of expanding HII regions ionized by young massive stars, or under the effect of stellar outflows and jets. With the exception of embedded HII regions, the dense and opaque molecular clouds have generally a very low ionization fraction (about 10^{-3}% on average, down to 10^{-6}% in cold cores), which is nonetheless sufficient to provide a coupling of the gas to magnetic fields, so that the motions of the gas cannot be described by simple hydrodynamics, but require a magneto-hydrodynamic (MHD) description (see Chap. 3).

(Giant) Molecular Clouds are at the heart of ISM studies, since they are the place where star formation occurs. Arguably, most of the evolution and dynamics of molecular clouds is either a cause or a consequence of star formation. Molecular clouds are gravitationally bound, and undergo a global gravitational collapse that eventually leads to the formation of dense cores and stars. However, were this global contraction to happen in an unhindered way under the action of gravity alone (i.e., in free-fall), one would expect a star formation rate of $10^2 - 10^3\ M_\odot$/year in the Galaxy, whereas the observed star formation rate is rather of the order of a few $M_\odot$/year. The star formation efficiency of molecular clouds

is thus of a few percent (up to $\sim 10\%$ in some GMCs), which implies that the collapse is slowed down by several counteracting processes – mostly thermal pressure, turbulence, and magnetic fields. Thermal energy needs to be radiated away so that thermal pressure can be overcome, which is why stars can only be born in the very coldest cores. Turbulence can either enhance or inhibit cloud collapse, but in general it acts like an effective pressure term, and needs to be dissipated by viscous forces (see Chap. 3 and 4, also Falgarone & Phillips 1990; Goodman et al. 1998). The coupling of the gas with the magnetic field can be weakened by ambipolar diffusion or magnetic reconnection (Mouschovias 1991; Lazarian 2005).

In addition to that, stellar feedback from the first stars formed in a cloud participates in reducing the star formation rate and eventually destroys the molecular cloud: young stars re-heat the gas, jets and outflows re-inject energy into the turbulent cascade, and, if this is not sufficient to evaporate and/or disperse the cloud, the most massive among young stars can quickly lead to supernova explosions which will tear the cloud apart from the inside, dispersing it over time scales of $\sim 10 - 30\,\mathrm{Myr}$ (McKee & Ostriker 2007; Iffrig & Hennebelle 2015).

1.2.2 Molecular observations

As mentioned in Sect. 1.1.1 and 1.2.1, molecular clouds were initially discovered – and are still currently studied – via their extinction properties in the visible and near-IR domain. They can also be observed at far-IR and sub-millimetre wavelengths thanks to the thermal continuum emission of dust. Another very useful category of observational tracers is the spectral line emission of molecules present in the clouds, mostly observed in the sub-millimetre and millimetre range.

Line emission is caused by the radiative de-excitation of a molecule from an upper quantum state to a lower one. In the millimetre domain, the excitation is usually collisional, and the lines are given by transitions between rotational levels (mostly), which correspond to the (quantified) amount of angular momentum carried by a molecule spinning around its axis. The most common molecule in molecular clouds is, by far, H_2. It is unfortunately a symmetric molecule, with no electric dipole moment, and therefore its rotational transitions are forbidden. It is thus necessary to observe other molecules, and use them as proxies to characterize the properties of the molecular gas, although they only represent a very small fraction of it. The most common tracer of molecular gas is CO (and its rarer isotopologues, such as ^{13}CO, C^{18}O, C^{17}O, etc.). The abundance of ^{12}CO with respect to H_2 is of the order of $10^{-5} - 10^{-4}$, but it is however relatively easily detected, and has been widely used to map molecular gas in the Milky Way and in external galaxies. Its strong emission makes it a good molecule to detect molecular gas, but not necessarily to locally measure its amount with accuracy, as ^{12}CO lines, and in particular the commonly observed ^{12}CO ($J = 1 - 0$) transition at 115 GHz, quickly become optically thick. This is where the rarer isotopologues are useful, as they enable the observers to measure the ^{12}CO opacity, by comparing the various line ratios with the known isotopic ratios.

Many other molecular tracers are also used, although not as extensively as CO. One can for example cite NH_3, which, due to its quantum level configuration, is a very good temperature and density probe, H_2O, which is often observed in the form of masers, or N_2H^+, which is a specific tracer of obscured and cold environment such as prestellar cores. The observation of several molecular tracers can give access to the relative abundances of these molecules, which are all part of complex chemical networks that involve photo-chemistry, gas phase chemistry, reactions on dust grains via adsorption and desorption of molecules... By comparing the observed molecular abundances with chemical models, one can thus infer constraints on the physical conditions in the cloud (spatial distribution of density, temperature, radiation field...) and on its evolution. However, one must be cautious when using molecular emission for physico-chemical diagnosis of the molecular environment, as the detected emission can depend on the chemistry of the cloud, on the excitation of the molecules, and the sensitivity of the observations (see Sect. 2.2.1).

Another advantage of line observations over continuum ones is that, provided a good enough spectral resolution, one can have access to the line profile, which, in the absence of opacity effects, is mostly shaped by the Doppler effect. Thermal motions of the gas produce a velocity dispersion of the emitting molecules, which widens the spectral lines, but very often the line profiles are much broader and more

complex than simply the thermal linewidth, in which case they provide information on the kinematics of the gas along the line of sight (see in particular Chap. 3).

1.2.3 What next?

The questions on GMC physics are complex and particularly important, as these objects are the bridge between galactic evolution and star formation. Questions such as matter exchange rates between the different phases of the Galactic ISM, or the relative contribution of spiralling jets and magnetic braking in the angular momentum problem of star-forming cores, are therefore actually relevant for the understanding of Giant Molecular Clouds, but other unsettled issues are more specific to the molecular cloud scale.

One of the key issues is the understanding of the relative roles of different phenomena in driving the cloud's evolution, with contributions from self-gravity, turbulence, magnetic field, as well as galactic influences in the form of external pressure or spiral arm compression waves etc. Stellar feedback is also a major source of uncertainty, maybe even the most important one when dealing with evolved clouds harbouring active star formation. In particular, the attribution of supersonic motions of the gas to turbulence, gravitational collapse or the relative motions of independent (fragments of) clouds is far from fully agreed on, and the same goes for the hierarchical density structure of clouds which can be described by gravitational fragmentation or by a turbulent fractal. In that context, the definition and even the very existence of sub-structures in molecular clouds is a challenge, since not all the clumps, filaments etc. identified by the observer's eye correspond to objects defined by actual physical properties. From an observational point of view, the difficulty comes from the systematic issue of line-of-sight confusion and averaging, which, in particular, raises the question of the possibility to map structures from the position-position-velocity (PPV) space sampled by Doppler-resolved observations to the physical, three-dimensional position-position-position (PPP) space. The way (magneto-)hydrodynamic turbulence dissipates and the influence of that dissipation on the way the gas condensates can bring answers to the question of cloud sub-structures. In addition to all the above issues, the specific question of chemistry also arises in molecular clouds, which is intimately linked with the physical processes in the cloud. Not only many reaction paths and networks still need to be understood, but from the observational point of view, the diagnosis provided by molecular observations still require detailed benchmarking.

This thesis, and the ORION-B project it is part of, try to answer some of these questions, one step at a time. Chapter 2 presents in detail the studied molecular cloud, the project in its entirety, and statistical studies focused on the molecular diagnosis and their ability to trace physical conditions or identify characteristic regions in a molecular cloud. Chapter 3 focuses on turbulence in molecular clouds, in particular on using PPV information to constrain the 3D properties of turbulence, and on the way these motions of the gas are linked to star formation. Chapter 4 also analyses star formation, but from the point of view of the sub-structures in the cloud, and in particular filaments. Conclusions and perspectives are given in Chapter 5.

Chapter 2

The ORION-B project

2.1 Observations and objectives

This thesis was entirely dedicated to the ORION-B (Outstanding Radio-Imaging of OrioN-B) project. As the name suggests, the project is focused on observations of the Orion B Giant Molecular Cloud in the radio domain, and aims at acquiring an unprecedented dataset on this molecular cloud, in terms of both spectral and spatial coverage. The following section presents the observational target as well as the scientific and technical objectives of the project.

2.1.1 A large radio-astronomical programme

The Orion B Giant Molecular Cloud

The Orion complex is the region of ongoing star formation of both massive and low-mass stars closest to the Sun, and is thus one of the most studied regions in the sky. The entire complex spans most of the Orion constellation (Fig. 2.1, left), and extends into the Monoceros, Taurus, Perseus and Eridanus constellations. At the heart of this massive structure lies the Orion OB association, which is a large group of young massive stars. These stars have formed gradually, starting 8 – 12 Myr ago (OB1a) and continuing till present, with the OB 1d group (Orion Nebula Cluster and NGC 2024 cluster, in the Orion A and Orion B molecular clouds respectively) being less than 2 Myr old (Bally 2008). The strong stellar winds of the massive stars, and the explosion as supernovae of the 10 to 20 most massive among them have carved the Orion-Eridanus superbubble, which contains the Orion A and Orion B molecular clouds, and whose most obvious manifestation is Barnard's Loop (Fig. 2.1, left).

Orion A and Orion B are the two most massive Giant Molecular Clouds in the Orion complex, with masses of the order of $10^5 M_\odot$. While both clouds have their denser edge towards the centroid of the OB association, Orion A has an elongated, "cometary" shape while Orion B's structure is more complex (Fig. 2.1, right). Although both clouds were shaped by the energy released by the OB association (Maddalena et al. 1986; Bally et al. 1987; Bouy & Alves 2015), it is unclear whether they are the remainder of a pre-existing cloud or if they formed from gas swept up and accumulated as the bubble expanded. In Orion A, this influence has taken the form of triggered star formation (Bally 2008; Bouy & Alves 2015). In Orion B, the clearest influence is that of the σ Ori multiple system (belonging to the OB 1c population of stars, like the oldest stars formed in Orion A) onto the edge of the cloud.

The Orion B Giant Molecular Cloud has a total mass of $7 \times 10^4 M_\odot$ (Lombardi et al. 2014), and lies at a distance of about 400 pc (Menten et al. 2007; Lombardi et al. 2011; Schlafly et al. 2014). Its projected size on the sky corresponds to about 30×15 pc. It can be roughly divided into two halves, the north-eastern one and the south-western one (Fig. 2.1).

It is this south-western half (marked by the green rectangle in Fig. 2.1), and in particular its western edge known as the Lynds 1630 dark nebula, which is the main focus of the ORION-B project. Its most famous feature is the Horsehead nebula (or Barnard 33), a dark pillar emerging from the dense

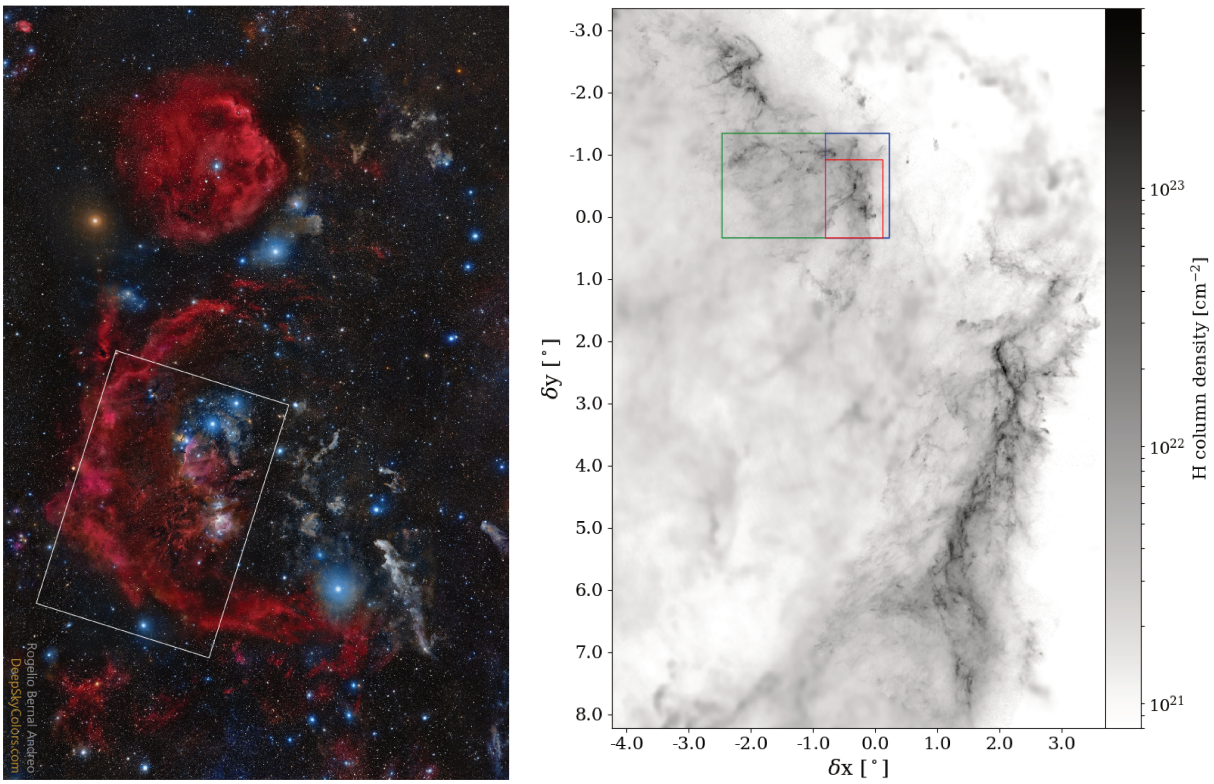


Figure 2.1: *Left:* The Orion constellation seen in visible light. The white frame corresponds to the field seen in the right panel. The large red arch is Barnard's Loop. Near the centre of the image are the three stars of the Orion Belt: from left to right Alnitak, Alnilam and Mintaka. Next to Alnitak, the yellowish glow of NGC 2024 can be seen. Just south is the tiny silhouette of the Horsehead Nebula. Further south, Orion's sword is actually M 42, the Orion Nebula (photograph by R. Bernal Andreo, DeepSkyColors.com) *Right:* Total hydrogen column density of the larger Orion A Giant Molecular Cloud to the south and smaller Orion B Giant Molecular Cloud to the north, derived from a combination of *Herschel* and *Planck* telescope FIR data. The coordinates are centred on the photo-dissociation region (PDR) at the edge of the Horsehead nebula. Overlaid are the successive stages of observation of the ORION-B project: pilot study in red, extension in blue, large programme in green (based on data from Lombardi et al. 2014).

molecular cloud and silhouetting against the bright HII region IC 434 (Fig. 2.2). The pillar was carved from the molecular cloud by the photo-evaporation induced by FUV radiation from σ Ori: the neck of the Horsehead stayed in the shadow of the dense ridge while less dense material was eroded away from the surface of the cloud (Reipurth & Bouchet 1984; Tremblin et al. 2012). The Horsehead has been identified as being rotating around its axis, a characteristic feature of such pillars (Hily-Blant et al. 2005), and it hosts several young stellar objects, whose formation was triggered by the compression induced by the HII region (Reipurth & Bouchet 1984; Lefloch & Lazareff 1994). This south-western part of the cloud also hosts two of the four important star-forming regions of Orion B: NGC 2023 (also known as LBN 954) and NGC 2024 (LBN 953, the Flame nebula) – the other two, NGC 2068 and NGC 2071, lie in the north-eastern part of the GMC. NGC 2023 is a reflection nebula, illuminated by the B1.5V star HD 37903. The Flame Nebula (NGC 2024) is a bright nebula, a young HII region owing its name to the dark, branching filament that obscures it (Fig. 2.2, left). The filament reaches tens of magnitudes of extinction, and thus hides at visible wavelengths the cluster of young stars at the centre of the HII region (Fig. 2.2, right). The brightest and hottest of them all, NGC 2024/IRS 2b, is of type O8 or B2, and has been identified as the source that ionizes the gas in the nebula (Bik et al. 2003).

The south-western part of the Orion B Giant Molecular Cloud thus contains both diffuse ($A_V \sim 1$, $n_H \sim 10^2\,cm^{-3}$) and dense ($A_V \sim 10^2$, $n_H \sim 10^6\,cm^{-3}$) gas (Bron et al. 2018), warm and FUV-

Figure 2.2: *Left:* The south-western edge of the Orion B GMC seen in visible light. The bright star in the foreground is Alnitak, which belongs to Orion's Belt. In the top right corner, the dark pillar of the Horsehead Nebula emerges against the bright pink, Hα glow of the IC 434 PDR. At the base of the Horsehead, the reflection nebula NGC 2023 shines blue. Further down is the small IC 435 nebula. In the left, the large Flame nebula (NGC 2024) lights the gas brightly, while being obscured by a dark dusty filament in its foreground (photograph by Warren Keller, Cherry Mountain Observatory). *Right:* The same objects seen in the near infra-red (composite of J, H and K bands). IC 434 is now dark, the faint bluish glow of the edge of the Horsehead and the PDR comes from scattering on dust grains. The reddish glow of the cloud comes from non-equilibrium emission of dust. In NGC 2023 and 2024, we can now see through the dust that obscured them, which reveals the many young stars forming in NGC 2024 (image by VISTA, ESO).

illuminated ($T \sim 100\,\mathrm{K}$, $G_0 \sim 10^4$) as well as cold and obscured ($T \sim 10\,\mathrm{K}$, $G_0 \sim 1$) regions (Pety et al. 2017), and hosts both triggered and spontaneous star formation. It is therefore an extremely varied environment, particularly well suited to study the physical and chemical conditions in molecular clouds and their link with star formation.

IRAM-30m observations

Everything started with two deep pointed observations. Starting in 2011, the WHISPER project (Wideband High-resolution Iram-30m Surveys at two Positions with Emir Receivers), led by Jérôme Pety and Viviana Guzmán, took advantage of the upgrade of the heterodyne spectrometers at the IRAM-30m telescope (Fig. 2.3), to acquire detailed spectra of two positions in the Horsehead Nebula (the edge of the PDR and the B33-MM1 dense core) over the entire 3 mm, 2 mm and 1 mm bands, detecting hundreds of spectral lines of about sixty molecules (Guzmán et al. 2011, 2012a,b; Pety et al. 2012; Gratier et al. 2013; Guzmán et al. 2013; Guzmán 2013; Guzmán et al. 2014, 2015; Fuente et al. 2017).

In 2013, the next natural step was to put the detailed studies of these two positions into the broader context of their environment. The goal was therefore to map in the 3 mm band the entire Horsehead nebula and the region where it anchors itself into the molecular cloud. This mapping project changed its aim in an unexpected way due to bad weather during the observations. Clouds rendered the atmospheric transparency too low to map faint lines, and pointing variations due to anomalous refraction too large compared to the limited field of view – it was thus deemed more useful to make a shallower, but much wider map (~0.9 deg^2, i.e., 3000 arcmin^2 instead of the 61 arcmin^2 planned initially) of the south-western edge of Orion B.

This was the starting point of the ORION-B project: the new goal was not to detect hundreds of lines to perform detailed astrochemical diagnosis, but to observe in many relatively bright molecular tracers as possible over a statistically meaningful fraction of an entire molecular cloud. The observational objective

Figure 2.3: The antenna and station of the IRAM-30m telescope on Pico Veleta, in Sierra Nevada, Spain.

Stage	Observing time	Observed area	Frequency range
	hours	arcmin^2	GHz
Pilot project	133	3185	84–116
Northern extension	101.5	2635	84–116
Spectral extension	65.5	5820	72–80 ; 88–96
Large programme	550	10795	72–80 ; 84–116
Total	850	16615	72–80 ; 84–116

Table 2.1: Spatial and spectral coverage of the successive stages of the ORION-B project, with the corresponding observing time.

in 2014 was to complete the spatial and spectral coverage of the pilot project: the goal was to observe the Horsehead, NGC 2023 and NGC 2024 over the entire 3 mm spectral window offered by the EMIR receiver, i.e., from 84 to 116 GHz (which could be covered in just two tunings).

In the following years the ambitions of the project increased. First, the spatial coverage was extended to the North, in order to map entirely a large filament that flows from NGC 2024 to a diffuse region. Second, taking advantage of the new capabilities the EMIR receiver at low frequencies, the entire region was observed again with a third tuning which covered the 72–80 GHz frequency range, and allowed to re-observe the 88-96 GHz range. At this stage, in 2016, it became clear that the project would not grow any more if the increases were to be done bit by bit, so a big leap was taken and the ORION-B project turned into a Large Programme, with 550 hours of observations granted over the course of 6 semesters, and the goal to cover about 5 square degrees (about 12×18 pc), i.e., almost half of the Orion B Giant Molecular Cloud. These 550 hours of observations are scheduled to be finished by the end of 2019. The successive stages of the ORION-B project are summarized in Table 2.1, and their spatial coverage can be seen in Fig. 2.1 (right).

The IRAM-30m observations use an on-the-fly scanning strategy. It means than rather than pointing the single-pixel receiver at a given position on the sky and tracking it during the integration time, before moving to the next position, the telescope is continuously scanning the sky, and the integrated signal is dumped at regularly spaced intervals during the scanning motion. For the ORION-B project the dump rate of the IRAM-30m telescope was pushed to its limit, which results in a scanning speed of $\sim 20''$/s. The main scanning direction corresponds to the vertical axis of Fig. 2.1 (right), except for the low-frequency tuning, which scans along the horizontal axis. Thus, the spectral range which is covered

twice (88–96 GHz) benefits of a criss-crossed observation pattern. The spatial resolution provided by the IRAM-30m telescope varies with the observation frequency, and ranges from 36″ at 72 GHz to 22.5″ at 116 GHz.

The calibration of the observations, which requires to subtract the contribution of the atmosphere to the radio emission measured by the telescope, is done by means of position switching, i.e., periodically the telescope is moved to a position close to the target (∼ 12′ from the Horsehead for the pilot study), but devoid of molecular line emission, and the signal recorded at this reference position is assumed to only contain the atmospheric contribution.

All the observations are carried out using the EMIR receiver (Eight MIxer Receiver, Carter et al. 2012), using the FTS (Fast Fourier-Transform Spectrometer, Klein et al. 2012), which yields a total bandwidth of ∼16 GHz at a 195 kHz channel spacing for each tuning. The higher spectral resolution but narrower bandwidth VESPA auto-correlator (VErsatile SPectrometer Array, Paubert 2000) is used simultaneously to observe several molecular lines, at a spectral channel spacing of 40 kHz.

The data reduction is done using the GILDAS/MRTCAL (Marka et al. 2017) and GILDAS/CLASS (Pety 2005) software. This process includes instrumental and atmospheric calibration, baseline correction, frequency-to-velocity conversion (using Doppler shift), and re-gridding of the position-position-velocity (PPV) cubes of all detected molecules onto a common grid of $9' \times 9' \times 0.5\,\mathrm{km\,s^{-1}}$ (except for the VESPA cubes which are re-gridded onto a $9' \times 9' \times 0.1\,\mathrm{km\,s^{-1}}$ grid).

The entire ORION-B project is expected to yield a total of ∼ 10 TB of raw data. A more detailed description of the observation strategy and the data reduction can be found in Pety et al. (2017).

Despite a relatively modest sensitivity (typically between 0.1 and 0.2 K– to be compared with the 0.008 K of WHISPER), over 25 molecules were detected in the cloud, among which ^{12}CO, ^{13}CO, $C^{18}O$, $C^{17}O$, HCO^+, HCN, HNC, CN, N_2H^+, SO, C_2H, CH_3OH, SiO, DCO^+, N_2D^+, etc. (Fig. 2.4, left), and the spectral resolution enables us to characterize in great detail the kinematics of the gas (Fig. 2.4, right). This combination of wide bandwidth and high spectral resolution, i.e., a large chemical diversity of tracers and accurate velocity measurements, provides us with an extremely rich dataset which is a powerful tool to understand the physics of Giant Molecular Clouds.

2.1.2 Orion B and beyond

The richness of the dataset acquired by the ORION-B project enables a large variety of studies to be performed on the Orion B Giant Molecular Cloud. The diversity of molecular tracers allows us to observe different types of environments (the translucent bulk of the cloud, the filaments, dense cores, PDRs), and study their dynamics thanks to the high spectral resolution. The availability of lines emitted by both molecular ions and neutrals gives access to the ionization fraction of the gas, and thus to the way it couples with the magnetic field. Other astrochemical processes can be characterized, such as the link between the deuteration of molecules, or the HCN to HNC ratio, and the gas temperature. All these approaches can be combined with the a priori knowledge of this well-studied GMC, for which independent temperature, column density, magnetic field measurements are available, along with catalogues of dense cores and YSOs.

These studies can constrain the relation of the chemical and physical properties of the cloud with its star-formation activity and the influence of the cloud's environment, thus helping answer some of the main questions on molecular clouds (Sect. 1.2.3). However, the spatial extent, data volume and sample diversity of the ORION-B observations also opens possibilities that go beyond the study of an individual molecular cloud.

A template for the extragalactic ISM

The study of the ISM and the star formation process cannot stop at the scale of molecular clouds, as these clouds are embedded in a galactic environment which has a significant impact on their physical and chemical properties. The density and pressure of the surrounding atomic gas, motions imposed by local

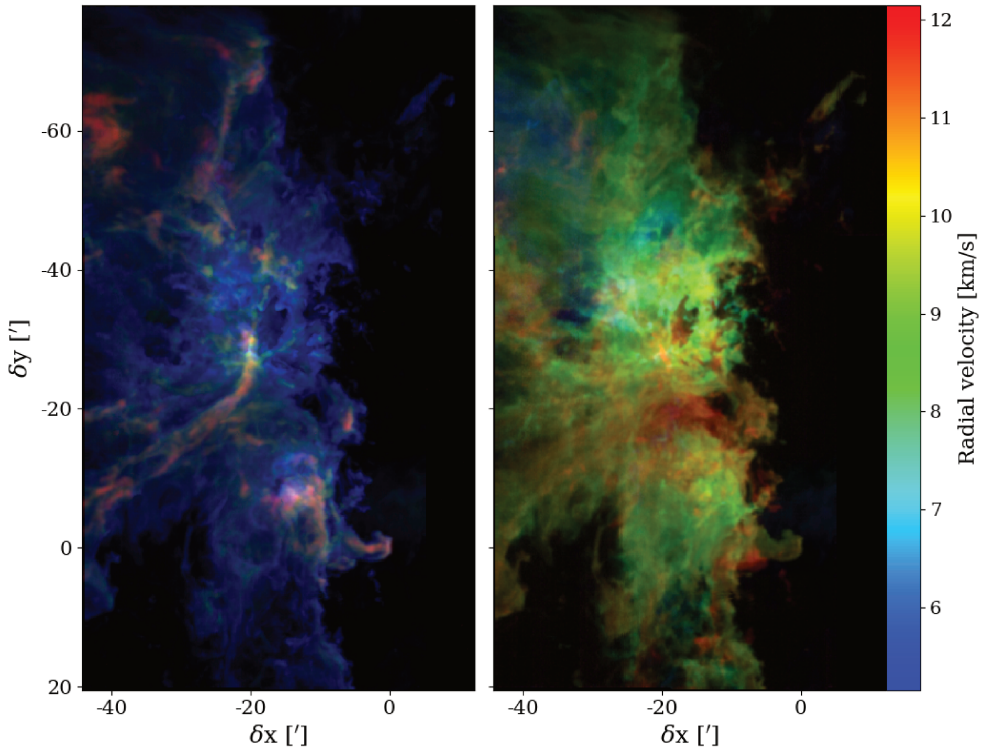


Figure 2.4: Chemistry and kinematics of the Orion B Giant Molecular Cloud revealed by IRAM-30m observations. *Left:* The ^{12}CO, ^{13}CO and C^{18}O ($J = 1 - 0$) peak temperatures are superimposed in blue, green and red respectively. The variations in spatial distribution of these CO isotopologues illustrate just a fraction of the molecular diversity in Orion B. *Right:* The emission of ^{12}CO ($J = 1 - 0$) in each velocity channel is encoded by a corresponding colour. The resulting composite image shows how complex the velocity field is in the molecular cloud.

or global processes, such as supernova explosions or galactic shear, can shape the clouds, determine their density distribution and control their internal dynamics, thus setting the conditions for star formation. However, the effects of these galactic processes on star formation are difficult to disentangle in the Milky Way, where our ability to observe the Galaxy as a whole is intrinsically limited. Besides, it is also important to probe a large variety of galactic environments and geometries, which can only be found in other galaxies than our own. This is why, to study star formation on the largest scales and in all possible conditions, astronomers have turned to observations of the ISM in nearby galaxies.

Over the last decades, a wide variety of surveys have observed significant samples of nearby galaxies, using different spectral tracers. In addition to datasets resulting from all-sky surveys, one can for example cite the FCRAO Extragalactic CO Survey (Young et al. 1995) which targeted 300 galaxies in ^{12}CO ($J = 1 - 0$), SINGS (*Spitzer* Infrared Nearby Galaxies Survey, Kennicutt et al. 2003, 75 galaxies observed in infrared, from 3.6 to 160 µm), THINGS (The HI Nearby Galaxy Survey, Walter et al. 2008, 34 galaxies observed in HI at 21 cm), HERACLES (HERA CO-Line Extragalactic Survey, Leroy et al. 2009, 18 galaxies observed in ^{12}CO ($J = 2 - 1$)), COLD GASS (CO Legacy Database for the GALEX Arecibo SDSS Survey, Saintonge et al. 2011a, 350 galaxies observed in UV, optical, HI and ^{12}CO ($J = 1 - 0$)), PAWS (PdBI Arcsecond Whirpool Survey, Schinnerer et al. 2013, high resolution mapping of the M51 galaxy in ^{12}CO ($J = 1 - 0$)) or the EMPIRE survey (EMIR Multiline Probe of the ISM Regulating Galaxy Evolution, Bigiel et al. 2016, 9 galaxies observed in HCN, HNC and HCO$^+$ ($J = 1 - 0$) and their ^{13}C isotopologues). The various tracers used in these surveys serve as a proxy to measure the physical properties of the studied galaxies, in particular the amount of atomic gas (HI emission), molecular gas

(CO lines), dense gas (HCN, HNC, HCO^+)[1], or stellar mass and star formation rates (UV, IR or Hα emission). Velocity-resolved observations of spectral lines also provide access to the large-scale dynamics of the galaxies, such as shear or streaming motions.

Correlations between these tracers enable to study major questions related to the largest scales of star formation. The most obvious is the measure of the fraction of the mass of galaxies contained in stars (M_*), in atomic gas ($M_{\rm atom}$) or in molecular gas ($M_{\rm mol}$). Leroy et al. (2009) estimate from HERACLES data that $M_{\rm mol} \approx 0.2 - 0.6 M_{\rm atom}$ and $M_{\rm mol} \approx 0.03 - 0.1 M_*$, while Saintonge et al. (2011a) propose $M_{\rm mol} \approx 0.3 M_{\rm atom}$ and $M_{\rm mol} \approx 0.07 M_*$ based on COLD GASS. However the stellar and gas content of galaxies evolves over time (Saintonge et al. 2016), with the gas being depleted and more stars being formed, on time scales of the order of a few Gyr (Bigiel et al. 2011; Saintonge et al. 2011b; Leroy et al. 2013; Kreckel et al. 2016).

Extragalactic surveys mostly approach star formation via the correlations between the mass reservoirs in atomic or (dense) molecular gas and the star formation rate. For example, Bigiel et al. (2010b,c) compare HI data from THINGS and UV observations from *GALEX* and conclude that HI is the main mass reservoir of star formation, so that the SFR is actually limited by the extremely inefficient HI to H_2 conversion, which happens on time-scales of the order of 100 Gyr. Bigiel et al. (2011, 2016) also find linear correlations between the molecular gas surface density of galaxies (measured by HERACLES), or the amount of dense gas ($> 10^5$ cm^{-3}, derived from the HCN brightness observed by EMPIRE), and the SFR (Fig. 2.5).

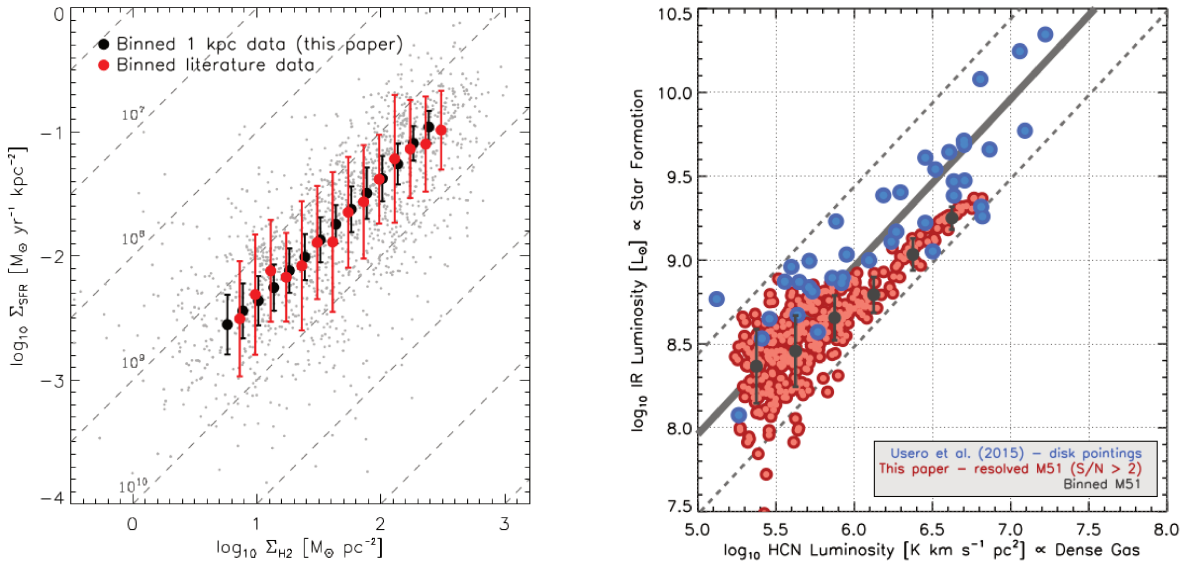


Figure 2.5: *Left:* The surface density of molecular gas, traced by CO, in 30 nearby galaxies, correlates linearly with the star formation rate deduced from their infrared luminosities, albeit with a significant scatter. *Right:* In the M51 galaxy, the spatial distributions of HCN ($J = 1 - 0$) emission, which is assumed to trace dense gas ($> 10^5$ cm^{-3}), and infrared luminosity, also correlate linearly (figures from Bigiel et al. 2011, 2016).

The reliability of these correlations is subject to caution, as the relations used to convert the observables to physical quantities (e.g., a CO line brightness to a total column density of molecular gas) are not always well-constrained and not necessarily universal. They can depend on physical and chemical parameters such as temperature, UV illumination, metallicity, chemical processes... Bigiel et al. (2010a) notice variations of the $X_{\rm CO}$ conversion factor from CO brightness to H_2 column density, under the influence of massive star formation, Groves et al. (2015) estimates variations of the dust to gas ratio in galaxies using

[1]These molecules are commonly assumed to be tracers of dense gas, although things might actually be more complex than that, see Sect. 2.2.1 and A.1

CO, HI and infrared data. Jiménez-Donaire et al. (2017b,a) show that HCN, HNC and HCO^+ actually emit in gas with lower densities than expected, and that the $^{13}CO/C^{18}O$ ratio varies in galaxies, increasing with galacto-centric radius but decreasing with local SFR...

The local variations revealed by such studies are an important element, but they are highly uncertain given that extragalactic observations are fundamentally limited by their spatial resolution. For example, one of the highest resolution maps of a galaxy in $^{12}CO\,(J = 1 - 0)$ to date has been obtained by the PAWS programme, for the M51 nearby galaxy, using the PdBI and the IRAM-30m telescope. A single resolution element of that map corresponds to a spatial scale of about 40 pc, which is more than twice the entire spatial extent of the ORION-B project (Fig. 2.6). This means that observations of the extragalactic ISM can barely resolve individual molecular clouds, not to mention different environments in a single cloud. It is thus very difficult to constrain the conditions in which the gas is emitting in detail, and therefore to obtain accurate estimations of the gas properties based on atomic and/or molecular emission alone: Leroy et al. (2017) have tried to simulate the unresolved emission of several molecular tracers in nearby galaxies, and have shown that the results vary significantly depending on the density distribution of the gas and the excitation conditions of the chemical tracers.

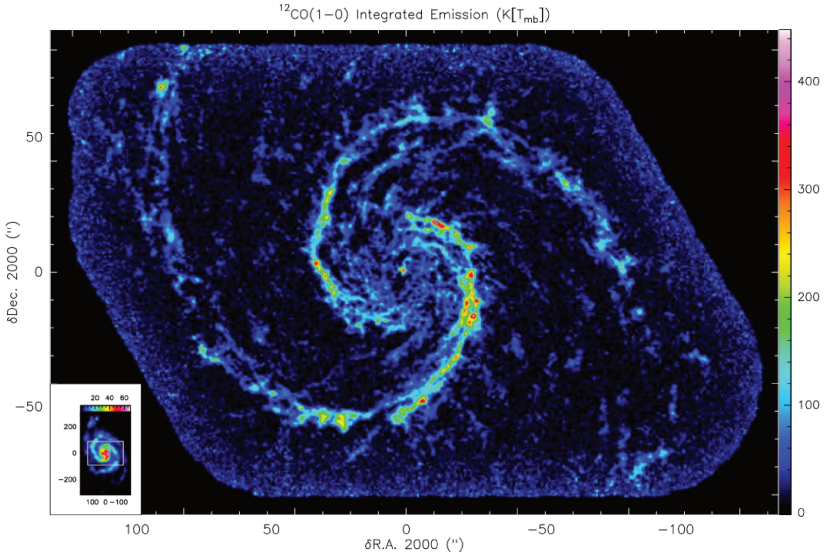


Figure 2.6: Interferometric mosaic of the M 51 galaxy observed in $^{12}CO\,(J = 1 - 0)$ with the PdBI as part of the PAWS large programme (PI: Eva Schinnerer). At a distance of 7.6 ± 1 Mpc and with a resolution of $\sim 1.1''$, each element of resolution corresponds to ~ 40 pc. In contrast, the entire ORION-B field of view covers 12×18 pc (figure from Pety et al. 2013).

This is where a local ISM survey such as the ORION-B project becomes important. By providing a detailed view of a well-known molecular cloud in many molecular tracers, such observations can constrain the conditions in which the tracers are emitting. Average spectra of the entire cloud or of its typical sub-regions can be compared to extragalactic observations where a spectrum corresponds, at best, to an entire molecular cloud (Sect. 2.2.1). The contribution of the various molecular tracers to the emission of dense, diffuse, dark, illuminated, warm or cold regions can be characterized (Sect. 2.2.2), thus providing benchmarks for the use of molecular tracers to characterize properties of the extragalactic ISM.

On the other hand, the spatial extent of the ORION-B project bridges the gap from sub-cloud-scale to galactic-scale dynamics. Most velocity-resolved studies of the ISM focus on objects the size of which are of the order of the parsec or less, such as infra-red dark clouds, filaments, hubs, etc. (see Chap. 4), and yet the importance of large-scale motions of the gas on the ability of a molecular cloud to form stars is crucial (e.g., Herrera et al. 2012; Henshaw et al. 2016). However, these large scales are very difficult to constrain in the Milky Way, which can only be seen from the inside with substantial line-of-sight confusion. Observational studies therefore have to focus on nearby galaxies to characterize the interplay

between galactic dynamics, stellar feedback and star formation. For example, the PAWS project has revealed the complexity of the ISM on galactic scales, with large variations of the star formation activity (Schinnerer et al. 2013, 2017): the mass distribution of GMCs varies throughout the galaxy (Colombo et al. 2014), and their star formation efficiency can be e.g., lowered by streaming motions (Meidt et al. 2013), or enhanced at the crossings of families of galactic orbits, or at the interaction of the bar and the spiral arms (Beuther et al. 2017). By observing the dynamics of a molecular cloud on 12×18 pc, while achieving a resolution of the order of 0.06 pc, the ORION-B project can quantitatively relate the motions imposed on the cloud by its galactic environment with its internal kinematics directly linked with star formation (Chap. 3 and 4).

Restating ISM questions in statistical terms

The size of the ORION-B dataset is a challenge in itself. The expected $\sim$ 10 TB of raw data will contain 200 000 spectral channels over 5 square degrees, which corresponds to 24 times the area of the sky. The final maps will contain $\sim$ 738 000 pixels, and $\sim$ 67 000 independent spectra at the typical resolution of $\sim 25''$. This is far more spectra than spectral surveys with a limited number of pointings (e.g., WHISPER) usually analyse, and far more molecular tracers than what is usually characterized by mapping observations. Such an amount of data both requires and enables new statistical analysis methods. One of the goals of the ORION-B project is thus to pave the way for the science that becomes possible thanks to the new generation of wide-bandwidth receivers coupled with high-resolution spectrometers in radio-astronomy. This is all the more necessary that the data-rates delivered by the major radio-telescopes will keep increasing in the coming years, delivering far more data that radio-astronomers usually analyse (Fig. 2.7).

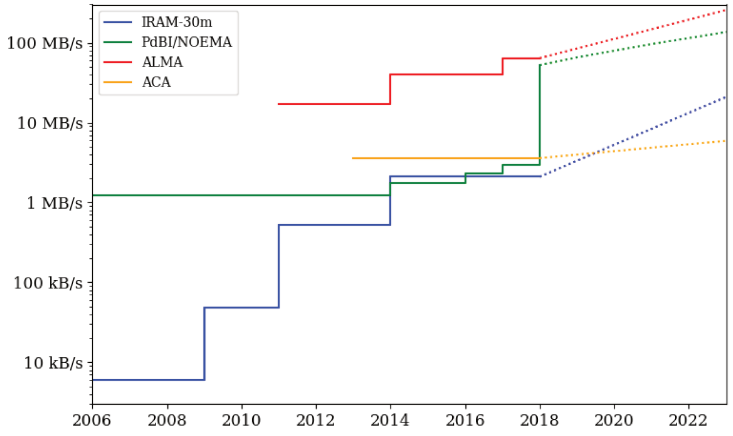


Figure 2.7: Evolution of the raw data-rates delivered by several major radio-astronomical observatories: the IRAM 30-metre single dish telescope, as well as the NOEMA (formerly PdBI) and ALMA (main array and ACA) interferometers. New receivers, antennas and correlators are foreseen, leading to a further increase in data rate.

The automated, statistical analysis of large datasets belongs to the field of *machine learning*. Machine learning is a domain of artificial intelligence which aims at designing algorithms which are able to improve their performances (in terms of data analysis, problem solving, etc.) thanks to the experience they gather, rather than explicitly programmed static algorithm which always execute their task in a predetermined way. As the "learning" process can require large amounts of data, this approach is particularly suitable to analyse extremely large datasets.

A machine learning process can be supervised (if the algorithm is provided with a training set) or unsupervised (if the algorithm has to find patterns in the data by itself). Typical machine learning tasks include classification, which attributes user-specified labels to data, clustering, which corresponds to unsupervised classification since the algorithm has to find categories by itself, regression, i.e., model

fitting, and dimensionality reduction, which can be used for better visualization or for data compression. While supervised methods yield results that are easy to understand, given that the interpretation is generally pre-existent in the training set provided to the algorithm, they can potentially miss some information present in the dataset, if it does not match the learned patterns. Conversely, unsupervised methods make the best of the structure of the dataset, but their results can be difficult to interpret. The use of machine learning tools with scientific goals indeed requires that these tools provide some clear results in terms of data analysis, but also that they are interpretable, in order to relate the patterns in data to physical parameters and to infer causalities.

One of the best known classification methods in astrophysics is the Hertzsprung-Russell diagram (Russell 1914, Fig. 2.8), in which the stars are sorted depending on their spectral types (or temperatures) and absolute magnitudes (or luminosities). Without the need for any statistical data processing, the diagram naturally reveals several categories of stars (the main sequence, the asymptotic giant branch, and the white dwarf group). However, not all datasets can be as easily classified by hand, especially when the number of dimensions is larger than two. Over the course of time, various methods such as Principal Component Analysis, Random Forest, Self-Organizing Maps, Neural Networks, etc. have been developed and applied to the classification of galaxy morphologies (Jimenez et al. 2008; Beck et al. 2018; Domínguez Sánchez et al. 2018), stellar variability classes (Wyrzykowski & Belokurov 2008; Moretti et al. 2018), supernovae (Ishida & de Souza 2013), stellar spectra (Plewa 2018)...

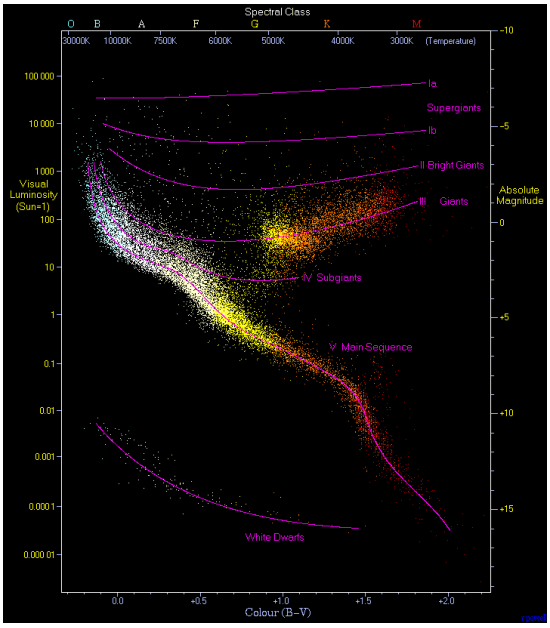


Figure 2.8: This Hertzsprung-Russell diagram sorts the stars of the Hipparcos catalogue based on their color or spectral type (horizontally) and their luminosity or absolute magnitude (vertically). It reveals several clusters of stellar types, most importantly the main sequence of "normal" stars, the giant branch, and the group of the white dwarfs (figure adapted from Richard Powell).

In astronomy, automated classification methods are thus mostly applied to individual objects, not to the analysis of continuous medium such as the ISM. In molecular clouds, machine learning methods have rather been used as cataloguing tools, identifying hierarchies of spatially contiguous clouds, clumps, or cores (Williams et al. 1994; Colombo et al. 2015), generally making use of a single tracer. The domain of space science in which hyperspectral information is used to identify areas with different spectral and physical properties is actually Earth remote sensing (e.g. Akbari et al. 2014; Inglada et al. 2017), where data from aerial and satellite imaging are used to distinguish between different environments (forests, fields, deserts...) on Earth and characterize their properties (tree height, water salinity...).

ORION-B observations provide maps of a great variety of velocity-resolved molecular lines. This diversity enables probing a wide range of different environments, in terms of physical and chemical

properties. Diversity also implies complexity, from the perspective of the dataset, as these environments, and these properties, need to be disentangled. Several important astrophysical questions can be asked and tentatively answered with a machine learning approach (Sect. 2.2.2 and 2.2.3). What are the characteristic regions in the Orion B molecular cloud, in terms of chemical composition, physical properties? What are the molecular lines and line ratios that trace each of these regions? Which physical properties can we reliably infer from the spectral signature of each line of sight?

2.2 Statistical studies of the molecular emission

The first statistical analysis of the molecular emission was performed on the pilot project (Table 2.1, Fig. 2.1 right), therefore using a field of view of approximately 0.9 square degree, and a spectral coverage from 84 to 116 GHz, which excludes all deuterated species detected at lower frequencies.

2.2.1 A first look at the spatial distribution of the molecular emission

The first step of the analysis of ORION-B observations was the computation of average properties of the cloud, of the emission, and the qualitative estimation of the correlations present between the different molecular tracers, the extinction and the dust temperature (Pety et al. 2017, Sect. A.1).

The studied tracers are the $(1-0)$ transitions of ^{12}CO, ^{13}CO, HCO$^+$, HCN, CN, C^{18}O,C$_2$H, HNC, C^{17}O, N$_2$H$^+$, H^{13}CN, H^{13}CO$^+$ and HN^{13}C, and the $(2-1)$ transitions of CS, SO, c-C$_3$H$_2$, CH$_3$OH and SiO. In addition to data from IRAM-30m observations, the column density (or A_V extinction) and dust temperature maps obtained by Lombardi et al. (2014) from *Herschel Gould Belt Survey* (André et al. 2010; Schneider et al. 2013) and *Planck* (Planck Collaboration XIX 2011) far infra-red data are also used as references.

One of most basic quantities derived in observations of the galactic or extragalactic ISM is the mass of clouds. It is typically done by deriving a column density of gas from either continuum or line emission, then summing it over the cloud's entire surface. When using the continuum emission of interstellar dust, the typical procedure is to fit the FIR spectral energy distribution (SED) of the dust at different wavelength with a modified black body, which yields an effective temperature and a total column density for the dust (Lombardi et al. 2014). When using spectral lines, either a column density of the emitting atom or molecule is computed under certain assumptions on the physical state of the gas and its optical thickness (Mangum & Shirley 2015), or, most commonly, a conversion factor from the line's total intensity to a total column density of hydrogen is used, like the X_{CO} factor widely used for ^{12}CO $(J = 1-0)$ observations of the ISM (Bolatto et al. 2013). All these methods suffer from limitations. Dust emission depends on the properties of the dust grains, which are far from being well-constrained (Ysard et al. 2018), and molecular emission depends on the excitation of the molecules (Mangum & Shirley 2015; Liszt & Pety 2016). Using a molecule such as ^{12}CO to trace the mass of molecular gas does not take into account the variations in the relative amount of so-called CO-dark gas, which is molecular but does not contain any (detectable) CO emission (Smith et al. 2014a). Both measurements also depend on the gas metallicity, via variations in the dust-to-gas ratio and the CO-to-H ratio (Bigiel et al. 2010b; Groves et al. 2015). Besides, all measurements suffer from layering effects, i.e., the physical and chemical state of the gas can vary along the line of sight, and the averaged emission that is observed can give a biased measurement of the cloud's properties. This is particularly true in the case of temperature layering for continuum observations, since dust emission increases rapidly with its temperature (Shetty et al. 2009). The layering issues can to some extent be mitigated by computing differential column densities by temperature bin (e.g., using the PPMAP method, Marsh et al. 2015, 2017) instead of a single effective temperature and column density per line of sight. Even though it is a computationally expensive approach and it requires many wavelengths to avoid degeneracies, it is a promising prospect to disentangle different regions along a line of sight.

In Orion B, the mass of the cloud was estimated using three methods: 1) the dust-traced mass was derived from the FIR emission fitted by Lombardi et al. (2014), 2) the CO-traced mass was derived from

the ^{12}CO ($J = 1 - 0$) integrated intensity converted by the standard X_{CO}, and 3) the virial mass was computed assuming equilibrium between the turbulent pressure and gravity, using the velocity dispersion derived from the average spectrum of ^{12}CO ($J = 1 - 0$).

Given the presence of CO-dark gas, one would expect the dust-traced mass to give the highest estimate, followed by the virial mass and the CO-traced mass. However, in the 0.9 square degree of Orion B, the dust-traced mass amounts to $\sim 3900 M_\odot$, the virial mass lies in the $6200 - 9500 M_\odot$ range, while the CO-traced mass amounts to $\sim 11000 M_\odot$. The CO-traced mass is thus 2.7 times higher than the dust-traced mass, even though there is probably some amount of CO-dark gas in the field. This can be explained by the fact that, compared to an "average" GMC, the studied field of view in Orion B is permeated by a strong UV illumination, coming from four HII regions (IC 434, IC 435, NGC 2023 and NGC 2024), which results in a higher excitation temperature of CO and higher dust temperatures. The excited CO produces a brighter emission, and thus biases the CO-traced mass towards higher values, whereas the SED of warm dust is completely dominated by the highest temperature, effectively hiding the emission of cold dust – which in turn lowers the estimation of the dust-traced mass. This dependency of the X_{CO} factor on the environment, and in particular the presence of FUV illumination coming from young massive stars, is consistent with extragalactic results by Bigiel et al. (2010a), and shows that "standard" ways to derive the properties of an "average" cloud should be used with caution[2].

An intuitive way to characterize which molecular lines are tracers of what kind of medium is to divide the cloud into typical regions and measure the fraction of emission that comes from each region for each molecular line. In Orion B this approach has been used with the A_V extinction and the dust temperature. These two quantities, relatively easily derived from FIR continuum observations (Lombardi et al. 2014), are good proxies for the column density of the gas (and therefore, to some extent, for the volume density as well) and for the FUV illumination (since FUV radiation significantly contributes to heating the dust grains). Thus, if a line has for example most of its emission coming from the region of low A_V but equivalent amounts of emission in all the regions defined by temperature, it can be considered as a good tracer of diffuse gas, being quite insensitive to temperature or illumination effects.

The observed field was divided into four regions of increasing A_V, representing diffuse, translucent, filamentary and dense gas, with the thresholds set at 2, 6 and 15 magnitudes of A_V (Fig. 2.9, left). The fractions of each line's emission in each of these regions are shown in Fig. 2.9 (right). The horizontal lines define families of lines which predominantly emit in the low ($A_V < 6$), low and intermediate ($A_V < 15$), intermediate and high ($A_V > 6$), or high ($A_V > 15$) extinction regimes. ^{12}CO, CCH, c-C_3H_2 and HCO^+ appear as the best tracers of diffuse and translucent gas, while the best tracers of well-shielded gas are N_2H^+, $H^{13}CO^+$ and CH_3OH.

The same procedure is repeated with T_{dust} (which is used as a proxy for G_0, i.e., the incident FUV radiation field, Habing 1968). The cold, lukewarm, warm and hot regions of the gas are delimited by thresholds at 19.5 K, 23.5 K and 32 K. These regions and the fraction of each molecular line's emission they contain are presented in Fig. 2.10.

While the trends are less clear than in the case of A_V masks, it is still possible to classify the lines into families, depending of the fraction of their flux which comes from regions of low or high FUV illumination. C_2H, c-C_3H_2, CN and HCN appear as the best tracers of warm, FUV-illuminated gas, while N_2H^+, $C^{18}O$, CH_3OH and $C^{17}O$ are the best tracers of cold, obscured regions. The presence of the rarest isotopologues of CO in the latter category can be explained by selective photo-dissociation and self-shielding: while ^{12}CO can quickly reach column densities high enough to self-shield from FUV radiation (which also benefits ^{13}CO via fractionation reactions with C^+), the rarest isotopologues can be photo-dissociated much deeper into the cloud before they start to self-shield efficiently (see Sect. 2.2.3).

[2]The entire field of view of the ORION-B large programme, which covers 5 square degrees, includes a large fraction of cold gas, both diffuse and dense. One can expect the average W_{CO}/A_V to get closer to the standard X_{CO}. Given the results of Pety et al. (2017) in the 0.9 square south-western degree field, the measure of the spatial variations of the local X_{CO} factor is one of the main objectives of the Large Programme.

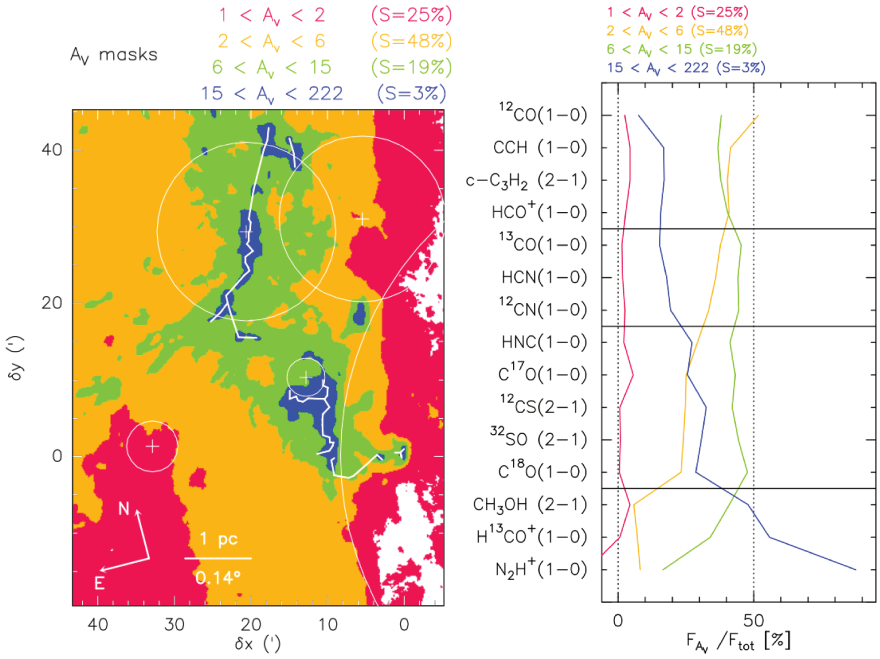


Figure 2.9: *Left:* A_V extinction masks defined based on *Herschel* data. *Right:* fraction of the flux of each molecular line coming from these masks (figures from Pety et al. 2017).

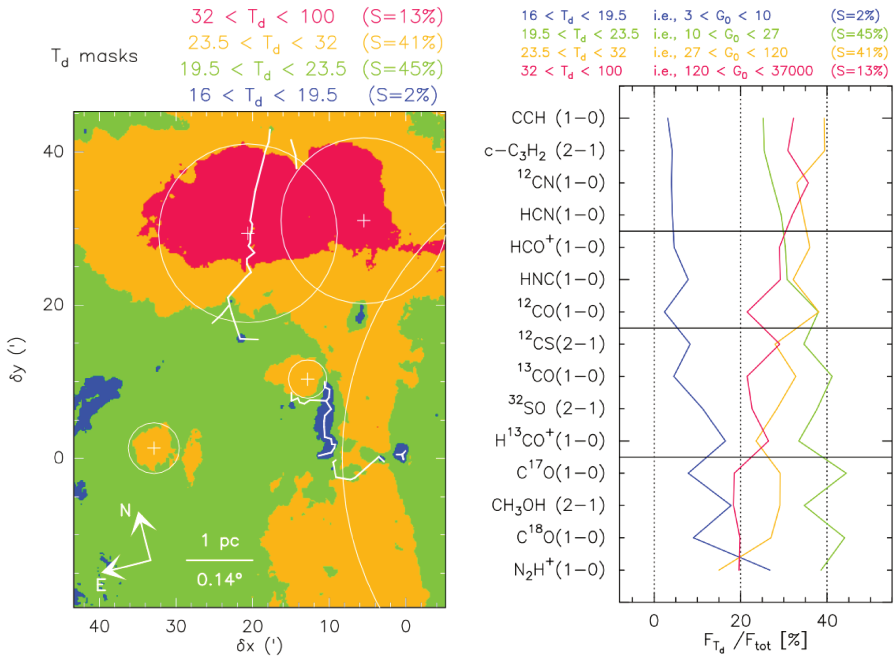


Figure 2.10: Same as Fig. 2.9, but for masks of increasing FUV illumination (or, equivalently, increasing dust temperature). (figures from Pety et al. 2017).

The results of these classifications, and in particular those based on the A_V extinction, raise the question of the reliability of the "dense gas tracers" commonly used when measuring the fraction of dense gas in the extragalactic ISM. Usually, molecules such as HCO^+, HCN or HNC are considered to be tracers of dense gas (e.g. Bigiel et al. 2016) because of their high critical density – which corresponds to the density at which collisional de-excitation equals the radiative emission (Shirley 2015). Molecules are often considered to be efficiently excited only above their critical densities, although they can actually be

excited and thus detected in much more diffuse gas (Shirley 2015; Liszt & Pety 2016). Fig. 2.9 shows that less than 30% of the HCO^+, HCN and HNC flux comes from regions with densities larger than 10^4 cm^{-3}. Remarkably, the HCO^+ ($J = 1 - 0$) emission actually mostly comes from diffuse and translucent gas.

To understand which molecules can be tracers of a given medium, one has to look at several complementary aspects: 1) chemistry, which defines how abundant the molecule is in the medium, 2) excitation, which determines if the molecule can emit photons in the observed transition, and 3) detection, which simply measures whether the molecule's emission reaches the detector with a strong enough flux to lie above the noise level. All these three aspects are linked by radiative transfer. Fig. 2.11 illustrates how over-simple assumptions can lead to a biased selection of tracers of dense gas. The left column shows the emission of two isotopologues of CO, ^{13}CO and C^{17}O, which have the same critical density (2×10^3 cm^{-3}), and the right column shows two molecules, HCO^+ and N_2H^+, with critical densities of 2×10^5 cm^{-3}. ^{13}CO and HCO^+ have large filling factors, while the emission of C^{17}O and N_2H^+ is present only in a small fraction of the field. This does not mean that C^{17}O is a molecule only found in dense gas. Chemical fractionation and photo-dissociation mean that this isotopologue is probably not present at densities as low as ^{12}CO and ^{13}CO, but the main reason why it is only visible in such a limited portion of the field is sensitivity: C^{17}O is about 1800 times less abundant that ^{12}CO (neglecting any chemical effects) – its emission is thus so faint that it is only detected at the highest column density. However, would the sensitivity be better, it could probably be observed over most of the field of view, since stacking spectra allows us to detect its presence in regions of lower A_V ($2 < A_V < 6$) (see Fig. 6 in Pety et al. 2017).

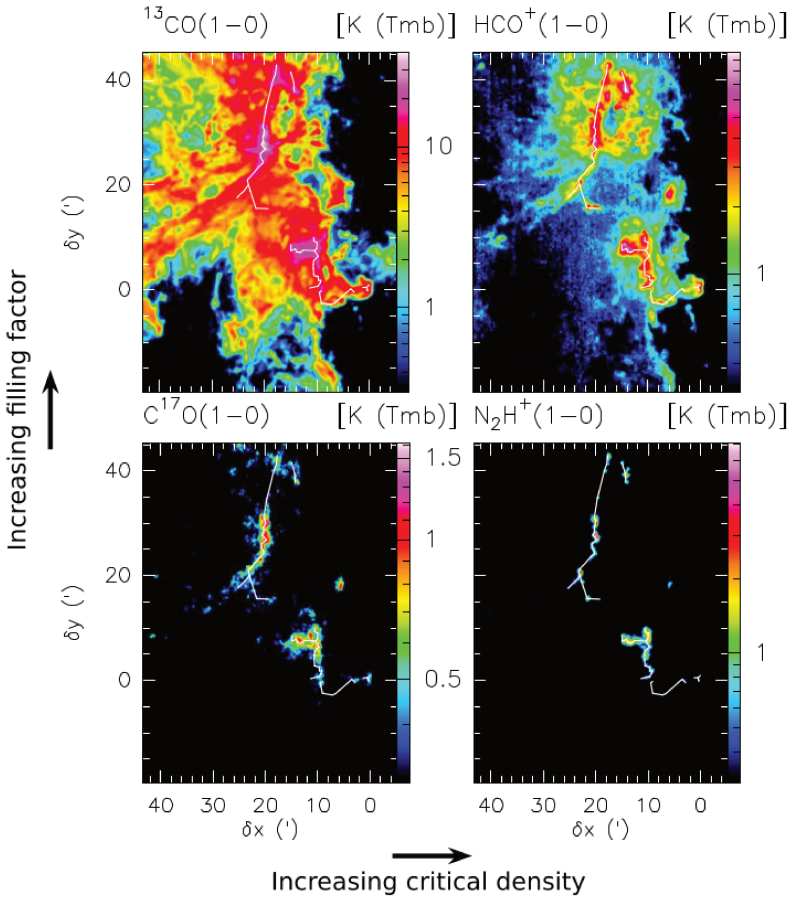


Figure 2.11: These peak temperature maps of four different molecular lines illustrate the pitfalls that one can encounter when neglecting some aspects of the chemistry/excitation/detection trio that governs the observed brightness of molecular lines. C^{17}O is not only present in dense gas, despite its low filling factor in the observed field of view, and HCO^+, despite its high critical density, does not trace only dense gas, contrarily to what is commonly assumed (figure adapted from Pety et al. 2017).

HCO^+ and N_2H^+ have the same critical densities, and even similar peak temperatures in the observed field (~ 10 K), which shows that the difference between their detected emissions is not a matter of excitation or sensitivity: it needs to be explained in terms of chemistry. HCO^+ can be formed by protonation of CO, or by various reactions involving carbonated cations, such as CH_2^+, and it is mostly destroyed by dissociative recombination when encountering a free electron, a destruction that re-creates CO. It can thus be formed by abundant species in environments with a variety of densities, ionization fractions... N_2H^+, on the other hand, can only be formed by protonation of N_2, and is destroyed either by dissociative recombination, or by reacting with CO (thus forming HCO^+). Therefore, N_2H^+ can only be present in regions where N_2 is sufficiently abundant and CO is (almost) absent: this happens only in cold, dense cores where CO has been depleted by freeze-out on dust grains. N_2H^+ is thus intrinsically a molecule exclusively associated with dense and cold gas.

The use of line intensities as tracers of physical conditions of the gas can be compromised in the case of calibration uncertainties. In that situation, the use of line ratios is advantageous, as it easily enables comparing different sources with one another. Besides, as most line intensities predominantly correlate with the column density of the gas (Pety et al. 2017; Gratier et al. 2017), line ratios are more sensitive to physical and chemical effects.

The study of line ratios involving the two brightest transitions (^{12}CO ($J = 1-0$) and ^{13}CO ($J = 1-0$)) in Orion B yields results consistent with the study of line intensities: the $N_2H^+/^{12}$CO ratio is a much better tracer of the dense gas ($> 10^4$ cm^{-3}) than the ratios involving HCO^+, HCN or HNC, and UV illumination is best traced by the $C_2H/^{13}$CO ratio. It becomes even more evident when comparing the average line ratios in Orion B with extragalactic values obtained from the literature. Compared to the ratios commonly used to trace the fraction of dense gas, such as $HCO^+/^{12}$CO, HCN/^{12}CO, or HCO^+/HCN, the $N_2H^+/^{12}$CO ratio shows the largest variations between a quiet environment such as Orion B, and starburst galaxies, ULIRGS (ultra-luminous infra-red galaxies) or AGNs (active galactic nuclei), which form large amounts of stars in a large amounts of dense cores. Similarly, the $C_2H/^{13}$CO ratio reaches its highest values for ULIRGs and AGNs, which are strongly illuminated, and is almost an order of magnitude lower in Orion B, in M51's spiral arm or in the Large Magellanic Cloud, which do not harbour massive star formation.

In general, the line ratios in Orion B are most similar to the ones observed in a spiral arm of the M51 galaxy (Fig. 2.6) – which suggests that it is similar to the GMCs that this spiral arm contains. The main difference is the lower $N_2H^+/^{12}$CO ratio, which indicates a particularly low amount of dense material in Orion B. This can be related with the low star formation activity of Orion B (see Chap. 3 and 4), which implies a low number of dense, star-forming cores in the cloud.

This preliminary study, which identified the molecular lines which are the best tracers of given physical quantities, and separated the cloud into characteristic regions, can be performed in a more data-driven way, using multivariate analysis to measure correlations (Sect. 2.2.2) and clustering to segment the field of view based on its molecular emission (Sect. 2.2.3).

2.2.2 Measuring the correlations between the molecular tracers

In Gratier et al. (2017, see Sect. A.2), the correlations between the emissions of the twelve brightest molecular lines presented in Pety et al. (2017) were analysed using one of the simplest multivariate analysis tools, namely Principal Component Analysis (PCA, Hotelling 1933; Jolliffe 2002).

The principle of PCA is illustrated in Fig. 2.12. Applied to a dataset composed on X samples with N features each (i.e., X points in an N-dimensional space), this linear method aims at constructing an orthogonal basis of the N-dimensional space that is best suited to describe the axes of variation of the dataset, which are called principal components (PCs). It can be equivalently seen as finding the main axes of inertia of the cloud of points around its barycentre, as finding the hyperplanes of increasing dimension that minimize the loss of variance of the dataset by projection (these two approaches are self-evident in Fig. 2.12), or as a diagonalization of the covariance matrix defined by the dataset.

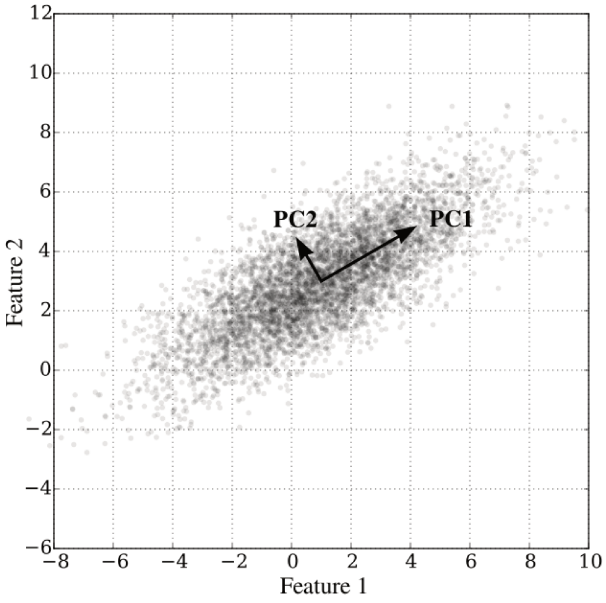


Figure 2.12: The concept of Principal Component Analysis illustrated in a simple two-dimensional case. The set of samples (scattered points) with two features each is better described using the new basis defined by PC1 and PC2. To first order, the data points can even be described in terms of PC1 alone (figure from Wikipedia).

In our case, the samples are 132 200 pixels, and their features are the integrated intensities of spectral lines of 12 molecular tracers: ^{12}CO, ^{13}CO, CS, HCN, HCO$^+$, SO, CN, HNC, C_2H, C^{18}O, N_2H^+ and CH_3OH. Spatial information is not included in the analysis, only the spectral information is considered. Since the maximal intensities over the field can vary by more than an order of magnitude between the different lines, the variance of the 12-dimensional cloud of points would simply be dominated by the lines with the largest intensities: PC1 would roughly correspond to the ^{12}CO axis, PC2 to ^{13}CO, etc. To avoid this behaviour, a variant of PCA is used where the variations of each feature are normalized by their standard deviation (Hotelling 1933) – the PCA thus corresponds to the diagonalization of the correlation matrix instead of the covariance matrix.

Normalization does not change the fact that the intensities of some tracers have a much broader dynamic range over the noise level than others. To avoid biasing the PCA towards the extreme values which can happen in the line intensity maps with large dynamic ranges, the data need to be further re-parametrized to suppress these extreme values. The most common way to achieve this is to apply a logarithm transform to the data. However, the data is noisy, which can bring the intensities close to or even below zero. A logarithmic transform would thus give infinite weight to regions of the data set where no signal is detected – which is clearly not desirable. The solution is to use the asinh function, which has a linear behaviour near zero and shifts to a logarithmic behaviour at high values. Thus, the extreme values are suppressed without unnecessarily enhancing the noisy regime.

Using an asinh transform is also desirable from the data interpretation point of view. Since PCA is a method which uses linear algebra, the operations performed on the samples (i.e., N-dimensional vectors) are additions and subtractions. In the noisy regime, if we assume that the noise in each feature is Gaussian and independent, it is a good thing to have an almost linear transform, and apply additions and subtractions to the noise, since a (weighted) sum of Gaussian, independent random variables is also a Gaussian random variable – the properties of noise of the PCs can thus easily be expressed in terms of the noise properties of the initial features. In the regime of high signal, where the behaviour of the asinh transform is logarithmic, additions and subtractions of the transformed data corresponds to products and ratios of the original data – the correlations revealed by PCA can thus directly be compared to the line ratios studied in Sect. 2.2.1.

PCA applied to the normalized and re-parametrized dataset made of 132 200 samples with 12 molecular line intensity features each yields 12 PCs, which are linear combinations of the 12 features, and provide a new basis in which each of the 132 200 samples can be re-expressed. The 12 line intensity maps can thus be equivalently replaced by 12 PC maps. As is can be seen in Fig. 2.13, these maps clearly display spatially coherent structures, despite the lack of spatial information provided to the algorithm. Since PCA is an unsupervised method, its results can be difficult to interpret, this is why we mostly focus on PC1, PC2 and PC3, which contain the highest fraction of the correlation of the dataset, and therefore have the best signal-to-noise ratio and the most striking spatial features. However, all PCs down to PC12 display spatially coherent structures – their lower importance is either linked to the fact that their values are relevant in a very small portion of the field of view, or that they represent very subtle behaviour of the data (e.g., PC11 has non-zero values only in a few dense cores, and PC12 clearly presents extended structures which correlate ^{12}CO, ^{13}CO and C^{18}O, maybe tracing opacity or isotopic fractionation effects).

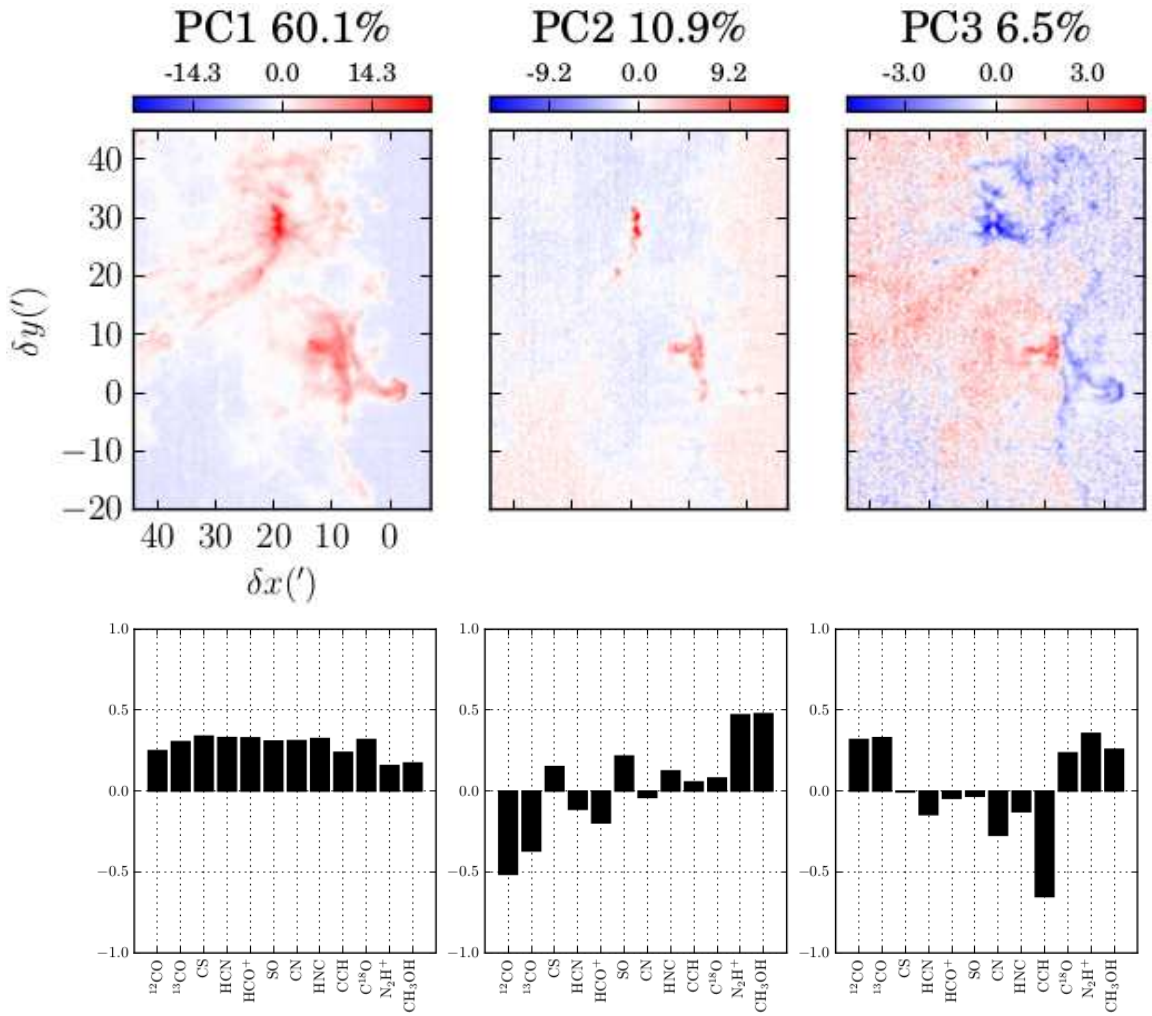


Figure 2.13: The first three Principal Components (out of 12) obtained from integrated intensities of molecular lines. *Top:* Fraction of correlation of the dataset that is contained in each PC, and spatial variations of the said PC, clearly showing spatially coherent structures. *Bottom:* Relative contribution of each original feature (line intensity) to a given PC – in terms of linear algebra these are the coordinates of the vectors of the new PC basis expressed in the initial feature basis.

We try to push the interpretation of the obtained PCs further than just considering the correlations between features, and try to relate each PC qualitatively and quantitatively to physical quantities (which is rarely done in studies of PCA in the ISM, with the notable exception of Ungerechts et al. 1997). Such a

quantitative interpretation of the PCs can provide an understanding of the physical conditions in a large field of view without detailed chemical, physical and radiative modelling, when dealing with massive datasets.

PC1 is relatively easy to interpret. As it can be seen on Fig. 2.13 (bottom left), it is a linear combination of all lines with positive coefficients of almost equal amplitude. It thus represent the overall intensity of the molecular emission, and can be related to the total column density: the more matter there is along the line of sight, the brighter all lines are. Several exceptions to this simple relation can be noticed, with a few lines contributing less than the average to PC1. It is the case of ^{12}CO, which is optically thick and thus has a saturated emission over a large portion of the field – more matter therefore does not imply brighter emission; it is also the case of N_2H^+ and CH_3OH, which are only formed in/around dense cores (Pety et al. 2017), and thus do not emit more brightly when the overall column density increases, but only when the amount of matter in cold, dense gas is larger. The correlation between PC1 and the column density N_{H_2} inferred from Lombardi et al. (2014) was measured using Spearman's rank correlation coefficient, which takes into account the relative ordering of the values of the compared variables. The obtained correlation is excellent (90%) with an almost linear relation between $\log(N_{H_2})$ and PC1. This is consistent with the results of the preliminary analysis in Pety et al. (2017), which showed that to first order all line emissions are linearly correlated with A_V (or equivalently N_{H_2}).

PC2 reaches large values only in the dense cores harboured by the Horsehead pillar, the NGC 2023 complex and the large filament in NGC 2024. PC2 mostly shows the anti-correlation opposing N_2H^+ and CH_3OH with ^{12}CO and ^{13}CO (Fig. 2.13 middle, top and bottom). It thus seems obvious that it is associated with the presence of cold, dense gas, where CO has been depleted by freeze-out. As the re-parametrization is in its logarithmic regime in these regions of high signal for all four tracers, this anti-correlation corresponds to a ratio for the original data – just like in Pety et al. (2017), it thus appears that the most important ratio (here, anti-correlation) to trace dense gas is $N_2H^+/^{12}$CO (together with $CH_3OH/^{12}$CO, and the same ratios with ^{13}CO). The quantitative correlation with the volume density n_H was done using volume densities obtained from a census of dense cores by Kirk et al. (2016). The results are less convincing than with PC1 and N_{H_2}: using Spearman's rank correlation coefficients, PC2 (anti-)correlates best with N_{H_2}, and n_H correlates best with PC1. However, since PC1 and N_{H_2} are already attributed, the second best option is linking PC2 and n_H (correlation coefficient of 22%). The lesser quality of this relation can be explained by the reliability of the volume density measurements by Kirk et al. (2016), who make strong assumptions on the properties of the dense cores they observe, but also by the fact that volume density might not be the best quantity to consider. Given the way N_2H^+ is formed and destroyed, it is not expected to trace a wide range of volume densities, but rather to specifically trace dense and cold gas, above a certain density threshold and below a certain temperature threshold. The physical quantity actually traced by PC2 could therefore rather be the amount (or the fraction) of dense and/or cold gas along the line of sight. This is however particularly difficult to measure, especially if the density and temperature thresholds are not precisely defined.

PC3 shows a strong East-West asymmetry, with strongly negative values being concentrated in the PDR at the edge of the IC 434 HIIregion and in the western edge of the cloud, including the Horsehead nebula, as well as in a roughly circular region around NGC 2024, and positive values in the inner part of the cloud, in particular near NGC 2023 (Fig. 2.13 top right). It thus seems that PC3 traces UV illumination of the gas, since the contrast is between HII regions and PDRs on the one hand, and more obscured regions on the other hand. PC3 has positive contributions from CO isotopologues as well as from N_2H^+ and CH_3OH, which are known to trace obscured gas, and notable negative contributions from C_2H, CN, HNC and HCN (Fig. 2.13 bottom right). We thus find the same tracers of obscured and illuminated regions respectively as in Pety et al. (2017) and Fig. 2.10. The anti-correlations of re-parametrized data can again be translated to ratios of the original data. The best ratio to trace UV illumination is thus C_2H/N_2H^+, however it is of limited use, given that N_2H^+ is only detected in a very limited fraction of the field. The next most useful ratios, involving widely detected tracers, are thus the ratios of C_2H and CN with ^{12}CO or ^{13}CO, here again in agreement with Pety et al. (2017). The quantitative correlation with UV illumination was estimated using PAH emission at 12 µm (Meisner & Finkbeiner 2014) normalized by the column

density as a proxy. PC3 correlates better with UV illumination (42%) than with N_{H_2} or n_H. The estimation of the incident UV field is however very approximate – a detailed radiative transfer model applied to a realistic volume density model of the cloud combined with measurement of transmitted and scattered visible and UV light could probably give a better correlation with PC3, but is again near infeasible in practice.

Provided that the interpretation of the results is done properly, PCA can thus provide estimates of the column density, volume density (or amount of dense gas) and UV illumination of a molecular cloud, using the integrated intensity of a limited number of molecular lines (here, twelve, but PC1 to PC3 could have probably been obtained similarly with only 9 or even 7 well-chosen lines) without need for further modelling.

In addition to the difficulties of interpretation in terms of physical quantities and proper correlation measurement mentioned above, one of the main limitations of PCA is that it is a purely linear method, whereas the relations of the molecular line intensities with each other and with the physical parameters are mostly non-linear. PCA thus captures the main trends with the first PCs, and the next PCs are corrections of the non-linearity of the relations – rather than having each PC representing a different kind of correlation in the data. Different physical and chemical regimes can also exist in the GMC, each with its own set of relationship between the various parameters, but PCA treats the dataset globally, and can thus be biased towards certain regimes, or trace the shift from one regime to another rather than the physical relations within each given regime. Piecewise-linear and non-linear variants of PCA do exist, and could overcome these limitations, but they tend to be more difficult to interpret. On the other hand, the use of linear PCA yields PCs which seem to contain significant information down to the last dimensions, which is also difficult to interpret: the "right" representation of the dataset in a new parameter space should ideally only have a limited number of relevant dimensions that would encompass all the correlations of the dataset.

There is thus room for more methodological research as far as multivariate analysis of the ORION-B dataset is concerned. However, regarding the identification of different physical regimes in order to characterize each of them individually, *clustering* can provide some degree of answer.

2.2.3 Identifying typical regions based on their molecular emission

As mentioned, classification methods, such as clustering, applied to the ISM, are most often used to identify physically contiguous regions, either solely based on their plane-of-the-sky coordinates (Williams et al. 1994; Colombo et al. 2015), or including also information on their line-of-sight velocity (Rosolowsky & Leroy 2006; Henshaw et al. 2016). In contrast, what was attempted in Orion B by Bron et al. (2018, see Sect. A.3) is to cluster the data points using only their molecular line intensities. It is thus a clustering done in *chemical* coordinates, rather than spatial coordinates.

Clustering aims at separating components in the multi-dimensional probability distribution function (PDF) of a dataset, and identifying the groups of points (clusters) belonging to each component. Some properties of the components can be set a priori, or, alternatively, the clustering algorithm can have more freedom to determine the clusters in a purely data-driven way. In Bron et al. (2018), we have used the Meanshift clustering algorithm, the principle of which is illustrated on Fig. 2.14. The algorithm estimates the underlying PDF from the distribution of the data points, using a smoothing kernel, the width of which can either be fixed, or be dynamically adapted to the density of data points. The maxima of the PDF are thus identified, each of them being associated with a "potential well" that define the clusters.

The dataset consisted in 141 050 samples with, initially, only 3 features, namely the line integrated intensities of ^{12}CO, ^{13}CO and C^{18}O, which are among the brightest observed tracers. Assuming constant isotopic abundance and neglecting (photo-)chemical effects, the relative abundances of these CO isotopologues should be uniform, so that the study of these line intensities should mostly reveal radiative transfer (opacity) effects.

The 8 clusters (named CO-1 to CO-8) obtained using these three tracers are presented in the left panel of Fig. 2.15. As in the case of PCA, despite the lack of spatial information provided to the algorithm, the resulting structures are spatially coherent. They show a clearly nested pattern which is expected

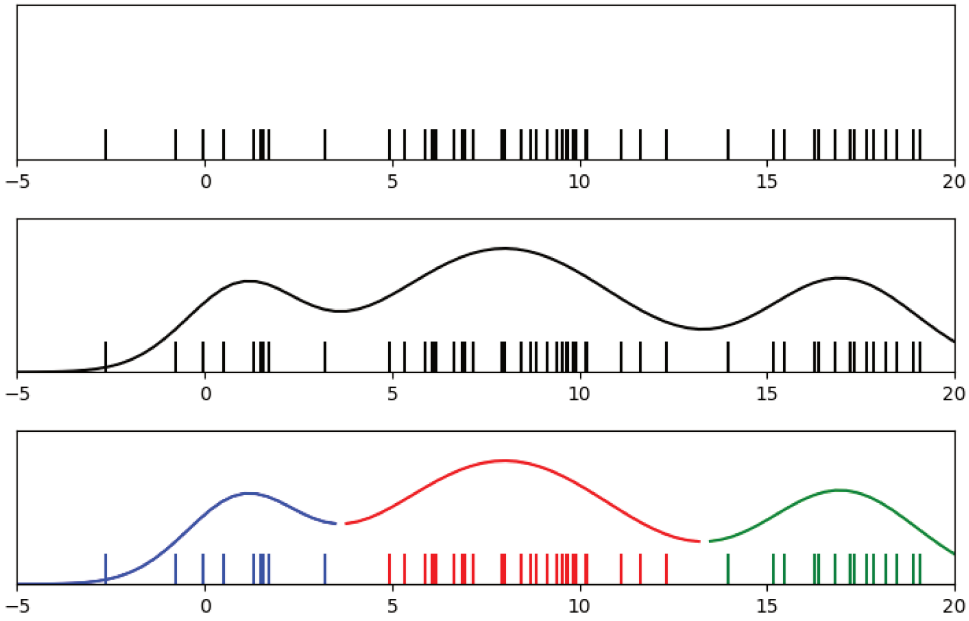


Figure 2.14: The basic principle of the Meanshift clustering algorithm, illustrated in one dimension. *Top:* The data points (samples) are scattered with a given density distribution. *Middle:* using a smoothing kernel, the PDF is estimated from the data distribution. *Bottom:* The peaks of the PDF are identified, and the corresponding clusters of points are defined.

to correspond to increasing (column) density of the gas. CO-8 diverges a bit from this pattern, as it is embedded in CO-6, not CO-7, and surrounds the star-forming regions NGC 2023 and NGC 2024. An East-West asymmetry is also visible: the transition from cluster 1 (almost no CO emission) to cluster 6 (relatively high column density gas) is much steeper on the western edge of the cloud, near the IC 434 PDR, than on the eastern, less illuminated side of the field.

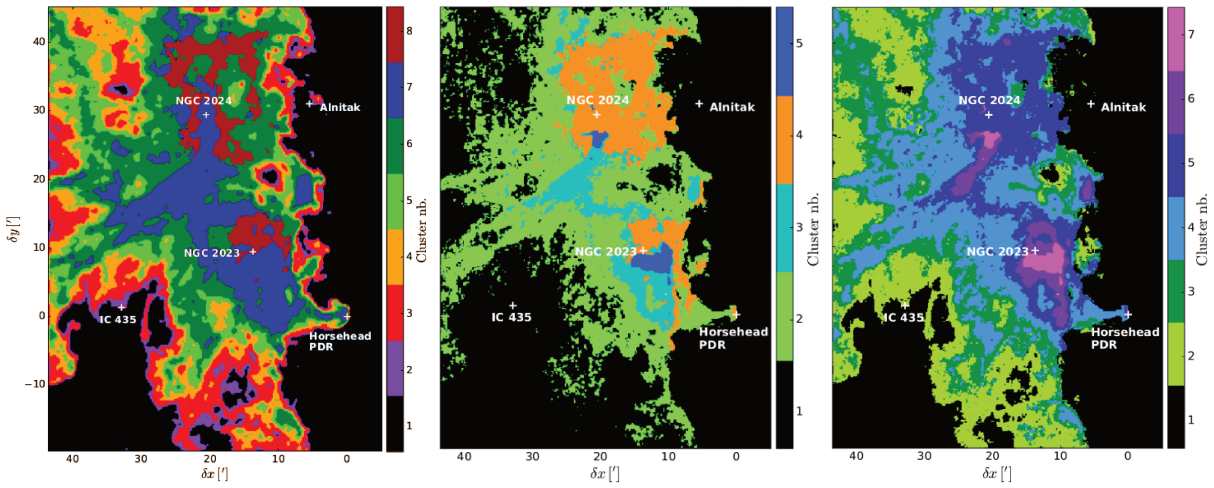


Figure 2.15: *Left:* Clusters obtained by applying the Meanshift algorithm to the line intensity maps of ^{12}CO, ^{13}CO and C^{18}O only, showing a nested pattern of increasing (column) densities. *Middle and right:* Adding CN and HCO$^+$to the features of the dataset yields a total of 19 clusters. To ease the interpretation of the effects of the new tracers, the clusters are fused into larger groups based on their most probable CN (*middle*) and HCO$^+$(*right*) intensities respectively. These groupings reveal FUV illumination effects and higher densities than CO isotopologues (figures from Bron et al. 2018).

The clusters can be easily related to the different density regimes visible in the volume density PDF

of the studied field. CO-1, in which only ^{12}CO is detected, corresponds to a first peak of the density PDF, at $\sim 100\,\mathrm{cm}^{-3}$, and thus represents diffuse gas. Clusters CO-2 to CO-5 correspond to a second peak of the PDF, at $\sim 500\,\mathrm{cm}^{-3}$. Only ^{12}CO (which quickly becomes optically thick) and ^{13}CO are detected in these clusters, which correspond to translucent gas. Clusters CO-6 to CO-8, in which all three CO isotopologues are detected, correspond to a last peak and to the extended tail of the PDF at high densities ($\gtrsim 1000\,\mathrm{cm}^{-3}$).

Other distinctions between the clusters can be attributed to dust temperature (and thus, FUV illumination) regimes. Clusters CO-1 to CO-6 all have similar temperature ranges, with mean values of 24 – 25 K. In contrast, the temperature in CO-7 is typically lower, with an average of 22 K, and CO-8 presents an average temperature of 34 K, with a temperature distribution peaked around 27 and 38 K.

These interpretations are confirmed by the results of LTE radiative transfer modelling. They show that ^{12}CO is optically thick almost everywhere, so that its line intensity mostly depends on its excitation temperature, unlike ^{13}CO, which reaches opacities larger than 1 at the transition between CO-6 and CO-7. Higher ^{12}CO intensities in cluster CO-8 are therefore directly linked to a higher gas temperature. This modelling also shows that the ^{13}CO/C^{18}O abundance ratio is closer to the elemental abundance ratio ~ 8 in CO-7 than in the other clusters, and in particular CO-8, where ^{13}CO/C^{18}O = 28. This is consistent with a well-shielded medium in CO-7, and strongly FUV-illuminated gas in CO-8. CO isotopologues are photo-dissociated by FUV radiation, and less abundant isotopologues self-shield less efficiently than ^{12}CO. Besides, ^{13}CO can be formed efficiently by isotopic fractionation reactions of ^{13}C$^+$ with ^{12}CO even in medium where ^{13}CO is not self-shielded, unlike C^{18}O – the ^{13}CO/C^{18}O ratio is thus even larger in PDRs.

Clustering based only on CO isotopologues has therefore the ability to separate different density regimes within the cloud, as well as to provide a first distinction between cold dense gas and FUV-illuminated, warm dense gas.

The next step of the clustering analysis was to add two dimensions to the feature space, namely the emission of CN and HCO$^+$. Both tracers are well-detected (better that C^{18}O, on average), have larger critical densities than CO ($2 \times 10^6\,\mathrm{cm}^{-3}$ for CN, $2 \times 10^5\,\mathrm{cm}^{-3}$ for HCO$^+$, $2 \times 10^3\,\mathrm{cm}^{-3}$ for CO), and are sensitive to FUV illumination – especially CN has been shown to be a good tracer of FUV-illuminated regions (Sect. 2.2.1 and Pety et al. 2017).

Clustering yields a total of 19 clusters. The lower signal-to-noise ratio of the CN and HCO$^+$ emission in diffuse and translucent gas, compared to ^{12}CO and ^{13}CO, results in more ragged boundaries compared to the CO clustering in these density regimes, which makes the transition from diffuse to dense gas look more noisy, with a loss of details. In particular, the steep transition from CO-1 to CO-6 observed on the western side of the cloud has been reduced to a single border between the IC 434 PDR and the rest of the cloud – even cropping the shape of the Horsehead nebula. The nested pattern of clusters is also not as obvious as before. These elements, combined with the large number of clusters, make them more difficult to interpret in terms of physical and chemical conditions. Therefore, in order to ease the interpretation of the clusters, as well as to focus mostly on the specific information brought in by CN and HCO$^+$, the clusters were fused into groups of clusters based on their most probable CN and HCO$^+$ intensities, yielding the groups CN-1 to CN-5 and HCO$^+$-1 to HCO$^+$-7 respectively (Fig. 2.15, middle and right).

The CN-based groups of clusters do not bring much information on diffuse, translucent and even moderately dense gas: most of clusters CO-1 to CO-6 are covered by groups CN-1 and CN-2, where CN is either invisible or weakly detected. However, groups CN-3, CN-4 and CN-5 show a striking pattern. CN-4 surrounds the star-forming regions NGC 2023 and NGC 2024 in a fashion similar to CO-8, but it is also present at the border of the IC 434 PDR, including at the tip of the Horsehead pillar. On the other hand, CN-5 is nested in CN-3, which mostly corresponds to the dense and colder environment traced by CO-7. It is thus tempting to interpret these results as the fact that CN-4 traces FUV-illuminated regions, both inside the cloud around the star-forming region and at the edge of the molecular cloud, while CN-3 and CN-5 trace different regimes of cold, dense gas. Indeed, when comparing the average properties of the different groups, we find that CN-1 and CN-2 have low densities ($< 1 \times 10^3\,\mathrm{cm}^{-3}$) and CN-5 as

the highest density, at 4×10^4 cm^{-3} (more than any CO-based cluster). CN-3 and CN-4 have similar densities ($\sim 3 - 4 \times 10^3$ cm^{-3}), but they are strongly separated in terms of temperatures (and thus FUV illuminations): CN-4 has a median temperature of 36 K, against 22 K in CN-3 (and 21 K in CN-5). The addition of CN to the clustering therefore reveals the distinction between dense and illuminated gas in PDRs (CN-4), dense and cold gas in the filamentary medium (CN-3) and very dense and very cold gas in dense cores (CN-5).

The HCO^+-based groups are more reminiscent of the CO clusters, with a clearly nested pattern. Groups HCO^+-1 to HCO^+-4 trace the transition from diffuse to moderately dense gas, similarly to clusters CO-1 to CO-6. The HCO^+-5 group mostly corresponds to a fusion of the CN-3 and CN-4 groups, indicating that the distinction between obscured and illuminated dense gas is generally lost. The temperature distribution of the pixels in that group clearly shows two peaks, which correspond to the cold and hot medium (CN-3 and CN-4) respectively. However, groups HCO^+-6 and HCO^+-7 (which corresponds to CN-5) give a more detailed view of the cold and dense gas. In particular, the group HCO^+-6 characterizes intermediate densities ($\sim 7 \times 10^3$ cm^{-3}) between the filamentary medium (HCO^+-5, 4×10^3 cm^{-3}) and the dense cores (HCO^+-7, 4×10^4 cm^{-3}). The high densities in groups HCO^+-6 and HCO^+-7 come close to the critical density of HCO^+. It is thus possible that, despite the fact that most of the HCO^+ emission comes from diffuse and translucent gas (Pety et al. 2017), the transition between the weak excitation regime (Liszt & Pety 2016) of HCO^+ and its normal thermalized excitation regime, which occurs around the critical density of the tracer, can actually characterize the presence of dense gas.

Additional chemical modelling has shown that it is indeed the presence of CN and HCO^+ in the dataset that enables the clear identification of FUV-illuminated regions. By using the non-LTE radiative transfer modelling code RADEX (van der Tak et al. 2007), we have shown that in PDRs, the relative abundances of CN and HCO^+ compared to $C^{18}O$ can be increased by up to two orders of magnitude, compared to more obscured regions.

Although the choice of tracers is crucial both in terms of signal-to-noise ratio and of the physical effects that are to be analysed, clustering a molecular cloud using only its molecular emission has proven to be able to distinguish regions of different densities, temperatures and FUV-illuminations based on opacity, photo-chemistry or excitation effects. Multi-variate analysis methods such as PCA are able to suggest which tracers should be used to reveal such and such physical effect. Conversely, clustering can segment the observed field into regions with given physical regimes, which can to some extent solve the inability of PCA to cope with non-linear relations between variables and multiple regimes.

Chapter 3

Turbulence modes in the Orion B molecular cloud

3.1 Turbulence in Giant Molecular Clouds

3.1.1 A brief introduction to turbulence

Turbulence is usually described as the "disordered" flow of an agitated fluid, displaying eddies and unpredictable motions.

For centuries, this complex and seemingly random phenomenon has kept at bay physicists in search of universal order and determinism. In the early sixteenth century, Leonardo da Vinci can be credited with detailed studies of turbulent flows, but his goal was mostly descriptive and artistic, as he tried to capture with precision the patterns of waves and vortices in water (Fig. 3.1, left). It is only in the late nineteenth century that turbulence entered the realm of natural science, with attempts of mathematical descriptions of the phenomenon, thanks to the efforts of Osborne Reynolds (1883).

Reynolds focused on the determination of the transition between the laminar regime (in which fluids "calmly" flow following the deterministic equations obtained in the previous centuries by Lagrange, Gauss, Navier, Stokes, etc.) and the turbulent regime in which instabilities disrupt the deterministic flow predicted from the initial conditions (Fig. 3.1, right). He found that this transition depends on the velocity of the flow v, its characteristic size L (set by an obstacle or boundaries), and the nature of the fluid via its density ρ and its viscosity η. These quantities combine into a dimensionless number, the so-called Reynolds number R_e

$$R_e = \frac{\rho v L}{\eta}. \tag{3.1}$$

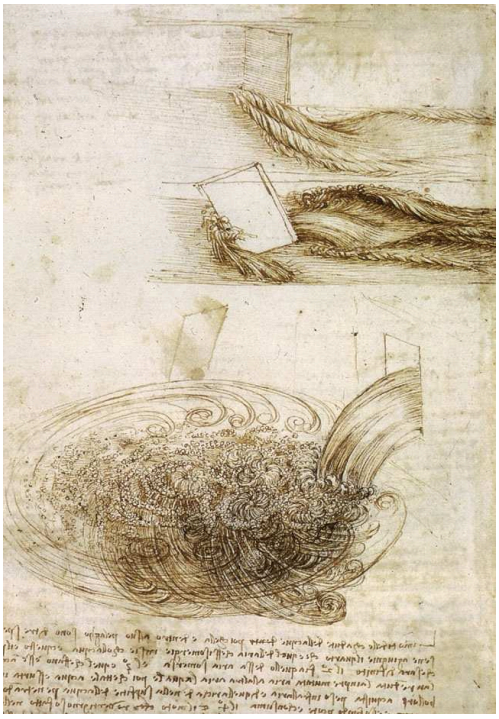
This number can be interpreted as the ratio of the inertial and the viscous (dissipative forces) by having a look at the Navier-Stokes equation of hydrodynamics:

$$\rho\left(\frac{\partial \vec{v}}{\partial t} + \vec{v}\cdot\nabla\vec{v}\right) = -\vec{\nabla}P + \vec{F} + \eta\nabla^2\vec{v} \tag{3.2}$$

where P and $\vec{F}$ are the pressure and the external forces. Leaving aside the pressure and forcing terms, the two contributions to the evolution of the velocity field are the non-linear advection term $\rho\vec{v}\cdot\nabla\vec{v}$, which tends to increase the inhomogeneities, and the diffusive (viscous) term $\eta\nabla^2\vec{v}$. Approximating the spatial derivatives by a division by the typical spatial scale L, the ratio of these two contributions is thus

$$\left|\frac{\rho\vec{v}\cdot\nabla\vec{v}}{\eta\nabla^2\vec{v}}\right| \approx \frac{\rho v^2/L}{\eta v/L^2} = \frac{\rho v L}{\eta} = R_e. \tag{3.3}$$

In other words, turbulence arises in a hydrodynamic flow when the viscous forces are too weak to damp the inhomogeneities enhanced by the non-linear advection term, leading to the development of instabilities.



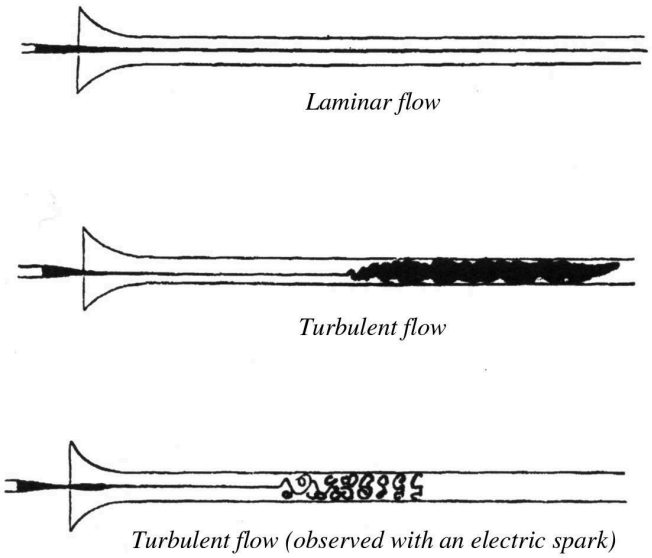


Figure 3.1: *Left:* Sketches of turbulent flows made by Leonardo da Vinci, who ca. 1508 was trying to find patterns in these seemingly random motions of water. *Right:* Sketches made by Osborne Reynolds when studying the transition between laminar and turbulent flow (Reynolds 1883). The top one depicts the injection of ink into a laminar flow of water, the middle one a turbulent flow, and the bottom one a snapshot of a similar turbulent flow, which reveals the eddies and vortices in the turbulence.

Reynolds found that the transition to turbulence occurs for R_e in the range 2000 – 13000, depending on the geometry of the flow.

When the flow reaches a very large Reynolds number, it cannot remain stable on the initial scale L, so it breaks down into smaller structures (eddies) on smaller and smaller scales, until the eddies reach a final scale L_D on which the viscous force is strong enough to dissipate the inertial force and the Reynolds number $R_{e\mathrm{D}}$ is not turbulent any more – hence the name of dissipation scale. Half a century after Reynolds' research, the seminal work of Kolmogorov (1941) described in statistical terms this "cascade" of kinetic energy from the initial scale (the injection scale) down to the dissipation scale in incompressible turbulence, characterized by an energy spectrum $E(k) \propto k^{-5/3}$, where k is the wavenumber $2\pi/L$ for each scale.

3.1.2 Turbulence in the interstellar medium

The study of turbulence is at the heart of the attempts of understanding the dynamics of the flows of interstellar gas and dust, given the extremely large Reynolds number that characterizes the interstellar medium (e.g. von Weizsäcker 1951). A simplistic view of a molecular cloud can describe it as a structure of pure molecular hydrogen about 10 pc or 3.10^{17} m across, with a density of $\sim 10^3$ H cm^{-3}, agitated by flows with velocities ~ 5 km s^{-1}, at a temperature of ~ 20 K. The Reynolds number can be re-written as $R_e = \frac{vL}{u\lambda}$ where u is the thermal speed of the particles and λ their mean free path. For molecular hydrogen at ~ 20 K, $u \sim 0.5$ km s^{-1}, and with collisional cross-sections of $\sim 10^{-16}$ cm^2, $\lambda = 1/(\rho\sigma) \sim 10^{11}$ m, thus $R_e \sim 5 \times 10^7$, far larger than the turbulent limit of a few 10^3.

The above calculation also brings to light the fact that $v > u$ – here by a factor of ~ 10 resulting from the assumed velocity of ~ 5 km s^{-1}, while for example in Orion B gas velocities span a range of ~ 20 km s^{-1}, with a typical FWHM velocity dispersion of ~ 4 km s^{-1} (Pety et al. 2017). In other words, the motions of the gas in the ISM are highly supersonic (with sonic Mach numbers $\mathcal{M}_s = v/c_s \sim v/u$ of

$$u = \sqrt{\frac{8k_{\rm B}T}{\pi\mu m_{\rm H}}} \tag{3.4}$$

$$c_s = \sqrt{\frac{\gamma k_{\rm B}T}{\mu m_{\rm H}}} \tag{3.5}$$

$$v_A = \frac{B}{\sqrt{4\pi\rho_{\rm c}}} \tag{3.6}$$

where
- $k_{\rm B}$ is Boltzmann's constant
- T is the gas temperature
- μ ≈ 2.37 is the molecular mass
- $m_{\rm H}$ is the mass of a hydrogen atom
- γ $\approx 5/3$ is the adiabatic index (ratio of specific heats)
- B is the magnetic field strength
- $\rho_{\rm c}$ is the mass density of charged particles

Table 3.1: Relevant velocities in a gas: mean particle velocity u (according to the Maxwell-Boltzmann distribution), as well as sound velocity c_s and Alfvén velocity v_A in magneto-hydrodynamics.

the order of 10, where c_s is the speed of sound, see Table 3.1). Since flows of gas travel faster than sound waves do, the medium cannot be approximated as being incompressible. Therefore, the main assumption of Kolmogorov's theory, that of an incompressible flow, does not hold any more: without the condition of a divergence-free velocity field, an extra degree of freedom is available, which can deeply modify the behaviour of the flow (the complexity of compressible turbulent flows is discussed in more details in Sect. 3.2).

The interstellar medium in not only highly turbulent and highly compressive – it is also a magnetized medium with an ionization fraction $n_e/n_{\rm H}$ ranging from about 10^{-6} in molecular clouds to $\lesssim 1$, as seen in Sect. 1.1.2. At the microscopic level, it means that each charged particle of the gas is not only subject to collisions with other particles, but also to the Lorenz force exerted on it by the interstellar magnetic field. At the astronomic level, it means that flows of charged particles, i.e., electric currents, create magnetic fields. This interplay of electromagnetism and hydrodynamics is known as magneto-hydrodynamics (MHD). Supersonic (and sometimes super-Alfvénic, see Table 3.1) MHD turbulence have proven to be a problem so complex that, after 40 years of research, no convincing analytical description of its energy spectrum has been obtained (Hennebelle & Falgarone 2012), and numerical simulations remain the best way to gain insight into its properties.

No matter how complex the turbulent motions in the ISM might be, the turbulent cascade brings the kinetic energy from the injection scale to the dissipation scale. Shocks are expected to dissipate the turbulence entirely in a few crossing times, i.e., in a few Myr for a cloud with a size of a few pc and a velocity dispersion of a few $\mathrm{km\,s^{-1}}$ (e.g. Mac Low 1999). Therefore, once this dissipation has elapsed, the gas flow should come to a quiet, laminar state. This is however not the case, as all molecular clouds are turbulent: it is thus a permanent, not a transient state. For this to be possible, continuous injection of kinetic energy is needed. This injection of kinetic energy can be internal or external, and the relative importance of the various contributing phenomena is unknown: galactic shear, colliding flows, supernova explosions near or in the cloud, stellar jets and outflows, and, at all scales, gravity (Hennebelle & Falgarone 2012, and references therein).

Turbulence plays a major role in the way a molecular cloud forms stars. In a scenario where only gravity would play a role, all of a cloud's mass would collapse and condensate into stars in a few free-fall times $\tau_{\rm ff}$. One could even expect that the shocks and density fluctuations created by the turbulence would

hasten this process. However, the star formation rate per free-fall time $\mathrm{SFR_{ff}}$, which is the related to the star formation rate $\frac{\mathrm{d}M_\star}{\mathrm{d}t}$, the mass of the cloud M_c and the free-fall time by $\mathrm{SFR_{ff}} = \frac{\mathrm{d}M_\star/\mathrm{d}t}{M_\mathrm{c}}\tau_\mathrm{ff}$ is of the order of 1–2% (Krumholz & McKee 2005). Star formation therefore appears as fundamentally slow and inefficient.

3.1.3 Compressive and solenoidal modes

While the magnetic fields also play an important role in slowing down the gravitational collapse of molecular clouds (which is beyond the scope of this thesis), turbulence is the main phenomenon that limits star formation. Shocks and condensations are typical of *compressible* turbulence, but we also mentioned vortices and eddies as the manifestation of *incompressible* turbulence. These motions are still largely present in compressible turbulence, and they tend oppose gravitational collapse rather than favour it. Intuitively, one can see that gravitational collapse results from centripetal forces, while eddies, being rotational motions, tend to impose centrifugal forces on the gas, impeding the collapse. On the other hand, Giant Molecular Clouds are subject to stellar feedback due to the young stars they formed. This feedback, in the form of radiation, outflows, jets, drives expansion motions that decay quite rapidly (~ 1 Myr) due to turbulence (Mac Low 1999). Even the most violent feedback, the supernova explosions of the most massive and short-lived stars, rarely disrupt the parent cloud efficiently, but rather accumulates shells of dense gas in a "snowplough" effect (Walch & Naab 2015).

As we can see, we have to distinguish between two types of motions in a turbulent flow: the motions of compression and expansion, which are exclusive to compressible turbulence, and the eddies, which are common to incompressible and compressible turbulence. They correspond to what is known as the compressive and solenoidal modes of the velocity field respectively. These modes are defined by the Helmholtz decomposition theorem (Helmholtz 1858), which states that any three-dimensional vector field $\vec{F}(\vec{x})$ can be described as the sum a purely longitudinal (compressive), curl-free field $\vec{F}_\parallel(\vec{x})$, and a purely transverse (solenoidal), divergence-free field $\vec{F}_\perp(\vec{x})$,

$$\begin{aligned} \vec{F}(\vec{x}) &= \vec{F}_\perp(\vec{x}) + \vec{F}_\parallel(\vec{x}) \qquad (3.7)\\ &= \vec{F}_\mathrm{sol}(\vec{x}) + \vec{F}_\mathrm{comp}(\vec{x}) \\ \vec{\nabla}\cdot\vec{F}_\mathrm{sol} &= 0 \qquad (3.8)\\ \vec{\nabla}\times\vec{F}_\mathrm{comp} &= 0 \qquad (3.9) \end{aligned}$$

The effect of supersonic turbulence on self-gravitating gas can be compared to an effective pressure, in which the turbulent velocity dispersion replaces the sound velocity (e.g. Peretto et al. 2014). This pressure has to be overcome by gravity in order to collapse and star formation to take place. Thus, star formation can only occur in "islands of calm in a turbulent sea", regions where the turbulent kinetic energy has been dissipated and motions have become subsonic (Goodman et al. 1998). Such regions have a characteristic size of about 0.1 pc (Goodman et al. 1998; Hennebelle 2013).

Both compressive and solenoidal modes contribute to the effective turbulent pressure, and to the dissipation of the turbulent kinetic energy, via shocks or in regions of extreme vorticity respectively. However, on the one hand, compressive modes of turbulence contribute to the formation of local condensations of matter, and on the other hand, solenoidal modes provide a strong support against gravitational collapse. Therefore, in the study of interstellar turbulence, among other major points of study such as the power spectrum of compressible MHD turbulence or the magnetic, hydrodynamic and radiative mechanisms of energy dissipation in protostellar cores, the understanding of the nature and relative contribution of solenoidal and compressive modes is particularly important to understand of the link between motions on the scale of Giant Molecular Clouds and star formation.

3.2 The role of turbulence modes in star formation

3.2.1 Turbulence modes and density PDFs

One of the characteristics of turbulent flows is the *intermittency* of the velocity field – i.e., the existence of very localized regions in time and space with very enhanced vorticity and velocity gradients. Intermittency results in a non-Gaussian probability distribution of velocities, and a fractal structure of the velocity field. Intermittency of the ISM turbulence has been brought to light by Falgarone & Phillips (1990), who showed the presence of self-similar non-Gaussian spectral line wings over a wide range of spatial scales. Further studies (Falgarone & Puget 1995; Lis et al. 1996) revealed that the spatial structure of the intermittent regions is indeed fractal (filamentary), and showed that the regions of increased vorticity can efficiently dissipate the kinetic energy into heat.

The specificity of supersonic (compressible) turbulent flows such as the ones encountered in the ISM compare to incompressible ones is that their *density* field is also intermittent and also has a fractal structure. The combination of local over-densities and dissipation of kinetic energy can result in excellent conditions to facilitate gravitational collapse, and thus lead to the formation of very dense gas and to star formation.

This intermittency of the density field in the ISM has been brought to light in observational data by Padoan et al. (e.g. 1997), who have shown that the dispersion of extinction (and thus column density) increases with its mean value. This appeared to be consistent with the accumulation of mass in a very small fraction of the volume observed in simulations of supersonic turbulence. In these simulations, a scaling relation appeared between the mean and the dispersion of the logarithm of the volume density:

$$\sigma^2_{\log(x)} = -2\,\langle \log(x) \rangle \quad \text{where} \quad x = \rho / \langle \rho \rangle \tag{3.10}$$

where ρ is the volume density of the gas and the brackets denote a spatial average. This scaling relation matches also very well the measurement made on observations of interstellar extinction. It is characteristic of a log-normal PDF for the fluctuations x of the volume density ρ (i.e., the PDF of $\log(x)$ is a Gaussian distribution):

$$P\left(\log(x)\right) = \frac{1}{\sqrt{2\pi\sigma^2_{\log(x)}}} \exp\left(\frac{-\left(\log(x) - \langle \log(x) \rangle\right)^2}{2\sigma^2_{\log(x)}}\right). \tag{3.11}$$

The dispersion of the volume density also increased with the mean velocity of the gas, quantified by its sonic Mach number $\mathcal{M}_s$:

$$\sigma_x = b\mathcal{M}_s \quad \text{and} \quad \sigma^2_{\log(x)} = \log\left(1 + b^2\mathcal{M}_s^2\right). \tag{3.12}$$

Padoan et al. (1997) found the parameter b to be of the order of 0.5. Later studies of supersonic hydrodynamic turbulence, using various simulation methods and at higher resolution (e.g. Kritsuk et al. 2007; Federrath et al. 2008) have found different various of b, roughly in the $\sim 0.2 - 1$ range. As we will see later, this parameter is of prime interest as far as the star formation rate is concerned.

Although these simulations were performed in the hydrodynamic case (without any magnetic field), the shape of the density PDF remains log-normal in the magneto-hydrodynamic case. The difference is only in the width σ of this distribution, which depends on the coupling of the magnetic field with the gas ($B \propto \rho^\alpha$) and on the Alfvénic Mach number $\mathcal{M}_A$. The extreme cases, with $\alpha = 0$ or $\alpha = 1$, correspond to a magnetic field decoupled from the gas (and thus to the hydrodynamic case), and to a magnetic field frozen in the gas, in what is known as ideal MHD, respectively. Observational results by Crutcher et al. (1999) or Padoan & Nordlund (1999) suggest that $\alpha \sim 0.5$, which yields in the supersonic *and* super-Alfvénic case (Molina et al. 2012) a PDF of the density fluctuations x characterized by:

$$\sigma^2_{\log(x)} = \log\left(1 + b^2\mathcal{M}_s^2 \frac{\beta}{\beta+1}\right) \quad \text{where} \quad \beta = 2\left(\frac{\mathcal{M}_A}{\mathcal{M}_s}\right)^2. \tag{3.13}$$

Here again, the last parameter that needs to be determined is b.

Federrath et al. (2008) have suggested that the value of b could be linked to the type of forcing that drives the turbulence. They performed simulations of supersonic turbulence in one, two or three dimensions, with continuous forcing of purely solenoidal or purely compressive motions, and found that $b \approx 1$ in the purely compressive case, and $b \approx 1/D$ (where D is the number of dimensions) in the purely solenoidal case. For interstellar, three-dimensional turbulence ($D = 3$), b thus lies in the $[1/3\ ;\ 1]$ range. Applying this to Eq. 3.12 or 3.13 yields a broader distribution of densities in the compressive case than is the solenoidal one, as it can be seen of Fig. 3.2.

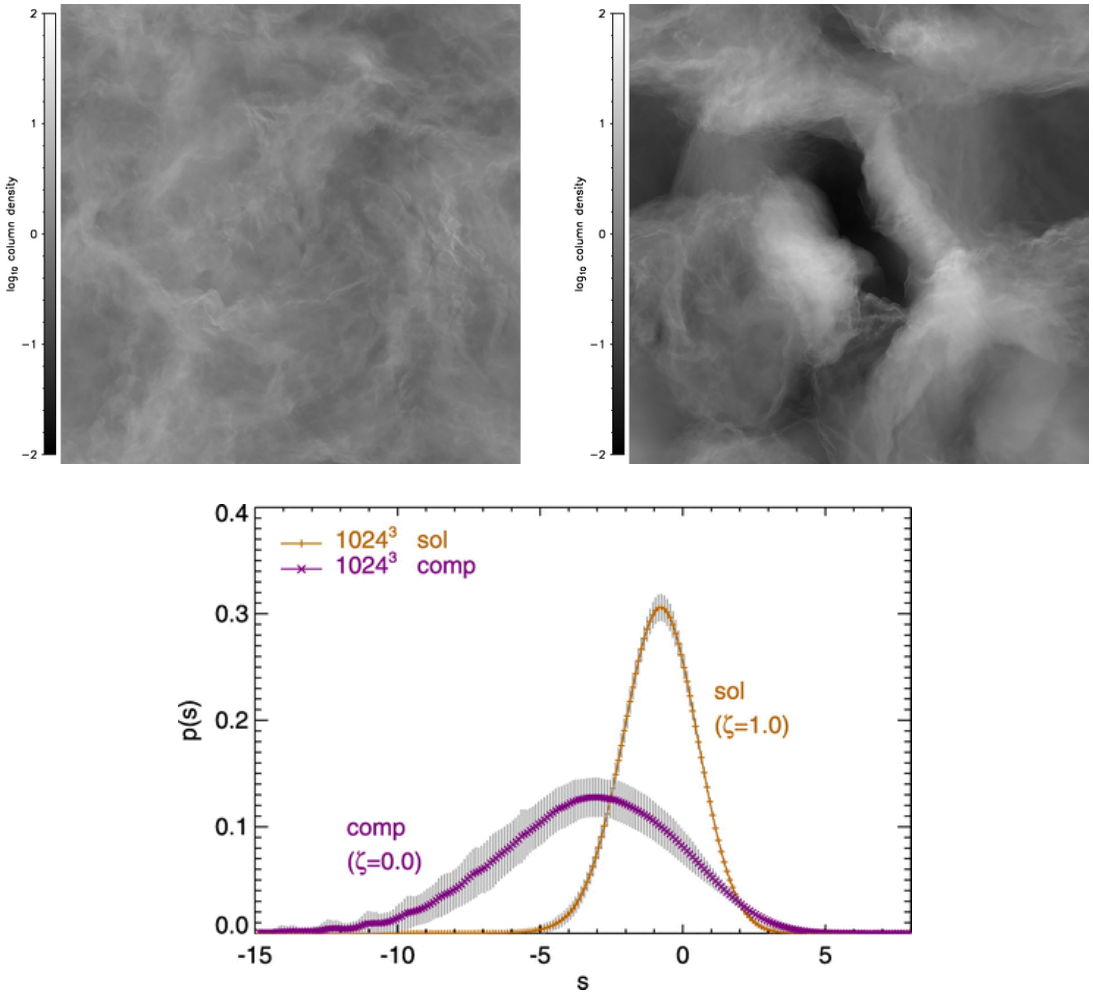


Figure 3.2: *Top:* Random snapshots of the normalized column density obtained in turbulence simulations with continuous forcing, in the purely solenoidal (*left*) and purely compressive (*right*) cases. *Bottom:* Density probability distributions of solenoidal and compressive simulations, where $s = \log(x)$, clearly displaying the Gaussian shape of the PDF of $\log(x)$ (figures from Federrath et al. 2008).

To quote Federrath et al. (2010), "it is evident that compressive forcing produces higher density contrasts, resulting in higher density peaks and bigger voids compared to solenoidal forcing". The difference between the two distributions can be explained by the fact that compressive forcing generates shocks and concentrates the mass of the (simulated) cloud into sheet-like structures, with a fractal dimension close to 2 (Kritsuk et al. 2007), while the solenoidal turbulence merely "stirs" the gas.

The link between b and the dominant modes in the turbulence was further highlighted by Federrath et al. (2010), who showed that b is directly proportional to the fraction of energy in the compressive modes of the velocity field, and that is can be simply related to the fraction of the forcing power injected into the compressive modes:

$$b = \frac{1}{3} + \frac{2}{3}\left(\frac{F_{\rm comp}}{F_{\rm tot}}\right)^3 \tag{3.14}$$

where F is the power injected by the continuous forcing, in the compressive modes or in total. However, they also note that the PDFs show departure from the ideal log-normal case. The main cause of this feature is intermittency, but the non-ideal behaviour is expected to increase in the case of non-isothermal, magnetized or self-gravitating turbulence – all elements that are left out of purely hydrodynamical simulations, but that are key in shaping the actual ISM turbulence.

Observational confirmations of these theoretical or numerical predictions have come in great numbers over the years. The log-normal shape of the density distribution has for example been tested in various regions of the Perseus molecular cloud by Goodman et al. (2009). Fig. 3.3 shows the very good agreement of the observational data with the log-normal prediction, over a large range of column densities. Federrath et al. (2010) note that the Shell region (open squares, bottom left of Fig. 3.3) displays the largest column density dispersion, and corresponds to an expanding region around a young massive star, which is associated with compressive forcing.

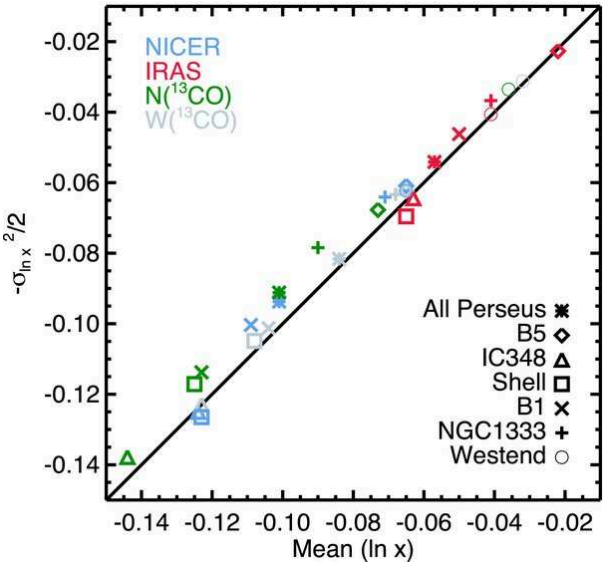


Figure 3.3: Various regions of the Perseus molecular cloud were observed with different tracers (dust traced by infra-red emission or extinction, gas traced by ^{13}CO emission) and characterized in terms of their column density distribution, showing a very good agreement with Eq. 3.10 (figure from Goodman et al. 2009).

Federrath et al. (2010) report several other studies the results of which can be compared with simulations to obtain an insight into the solenoidal or compressive nature of the turbulence in various other regions of the sky. For example, PCA measurements of the scaling between column density and velocity by Heyer et al. (2006) suggest that the interior of the HII region in the Rosette molecular cloud is mostly compressive, while the part of this cloud outside the HII region is mostly solenoidal, a result consistent with the Perseus MC. The Δ-variance technique applied to Fourier spectra of the velocity field, when compared with the results of Ossenkopf & Mac Low (2002) on the Polaris Flare, suggest that its turbulence is injected on large scales with a mixture of solenoidal and compressive modes, with a majority of compressive ones, whereas the analysis of Centroid Velocity Increments (CVIs) by Hily-Blant et al. (2008) shows that on intermediate scales, the turbulence in the Polaris Flare is mostly solenoidal – the latter can be linked to the absence of star formation in the molecular cloud (see below).

Kainulainen et al. (2009) have shown that the column density PDF of quiet, non-star-forming molecular clouds is consistent with a log-normal distribution, but that in star-forming, self-gravitating ones, a strong power-law tail can be seen in the PDFs at high column densities. Tremblin et al. (2014a) (Fig. 3.4) also report PDFs that stray away from the ideal log-normal model due to star-formation activity and stellar feedback. They obtain good results when fitting the PDFs with the sum of two log-normal distributions and adding a power-law tail at high densities. They associate the lower-density, narrower log-

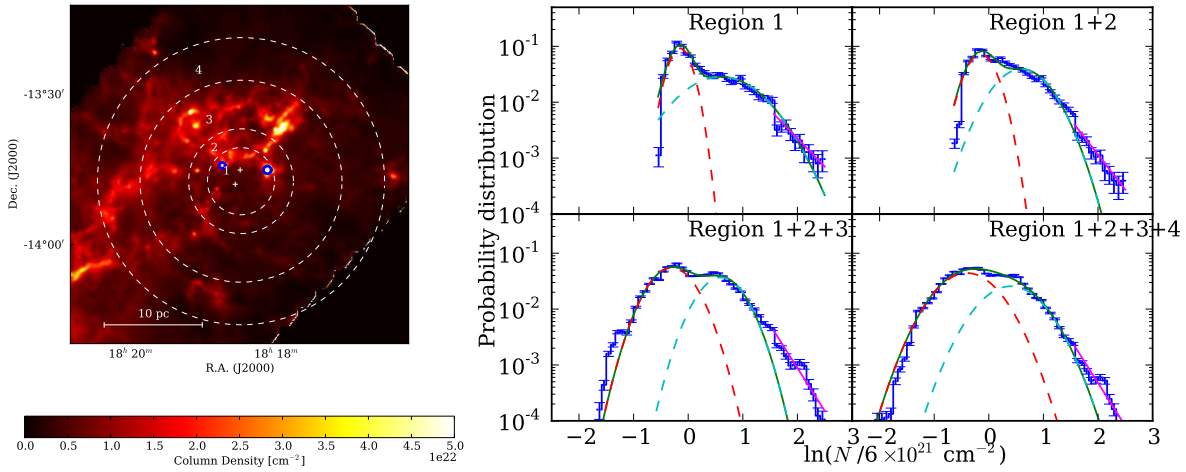


Figure 3.4: *Left:* Column density map obtain from *Herschel* observations of the M 16 molecular cloud, with four concentric regions defined around the young O-type stars (marked by the white crosses) which ionize the surrounding molecular gas. *Right:* Column density PDFs of these regions, fitted with the sum of two log-normal distribution and a power-law tail. (figure from Tremblin et al. 2014a).

normal distribution with the turbulence of the bulk of the gas, and the higher-density, broader component with the compression induced by the expansion of the young HII region created by young ionizing stars. The power-law tail is identified as corresponding to self-gravitating structures.

3.2.2 Turbulence modes and star-formation efficiency

Given that the star formation rate and star formation efficiency depend on the fraction of dense gas above certain threshold (Lada et al. 2010) or even on the entire density PDF (Krumholz & McKee 2005), they are also expected to depend on the fractions of solenoidal and compressive modes in the turbulence, since this ratio contributes to shaping the density PDF. Intuitively, one expects compressive modes to enhance the SFR and SFE, as they lead to the formation of denser regions through shocks, which can facilitate gravitational collapse (which is also a compressive motion). In turn, star formation feeds the compressive modes, since young stars drive outflows or the expansion of young HII regions that are also compressive in nature (Goodman et al. 2009; Federrath et al. 2010; Tremblin et al. 2014a).

Numerical simulations have come to confirm this intuitive view. Girichidis et al. (2011) have run simulations of star formation in a self-gravitating, turbulent, supersonic molecular cloud, with a varying turbulent injection. They have shown very clearly that star formation occurs much earlier when the forcing is purely compressive, compared to cases with a mixed, or a purely solenoidal turbulence forcing (Fig. 3.5) – provided that the initial density profile of the cloud is not so steep that turbulence has a negligible effect on its gravitational collapse. However, interestingly, the injection of turbulence in different modes does not seem to have an impact on the initial mass function of the stars.

Federrath & Klessen (2012) have also compared a range of analytical models of star formation (Fig. 3.6, left), and run numerical simulations (Fig. 3.6, right), with compressive, mixed or solenoidal forcing of the turbulence. All studies agree with the fact that the star-formation rate and the star-formation efficiency increase with the fraction of compressive motions in the turbulence.

When venturing on scales larger that the molecular cloud, simulations by Renaud et al. (2014) also provide a striking example of the link between star formation and compressive turbulence, in their simulation of starbursts (intense episodes of star formation) in a galactic merger, triggered by the intergalactic tides. Fig. 3.7 shows how, during the merging event, the fraction of compressive modes increases greatly, simultaneously increasing the star formation rate of the merging galaxies – the maxima of the energy in compressive turbulence modes match peaks in star formation activity.

In summary, it appears that the compressive or solenoidal modes of turbulence play a key role that

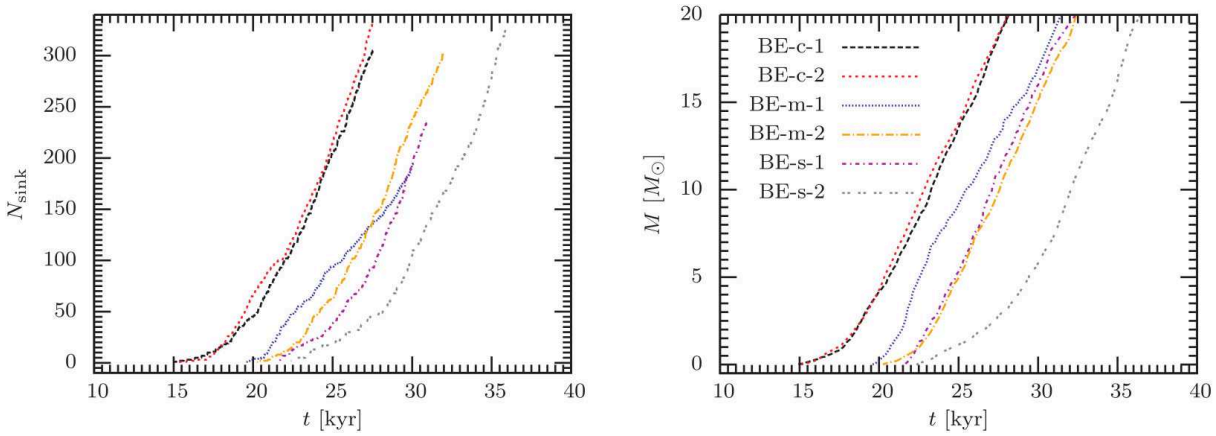


Figure 3.5: *Left:* Number of sink particles (and thus stars) formed in a simulated molecular cloud as a function of time, for simulations runs with turbulence injected in the compressive (c-1, c-2) or solenoidal (s-1, s-2) modes, or in a mixture of those (m-1, m-2). *Right:* Total mass accumulated in the sink particles as a function of time, for the different simulation runs (figure adapted from Girichidis et al. 2011).

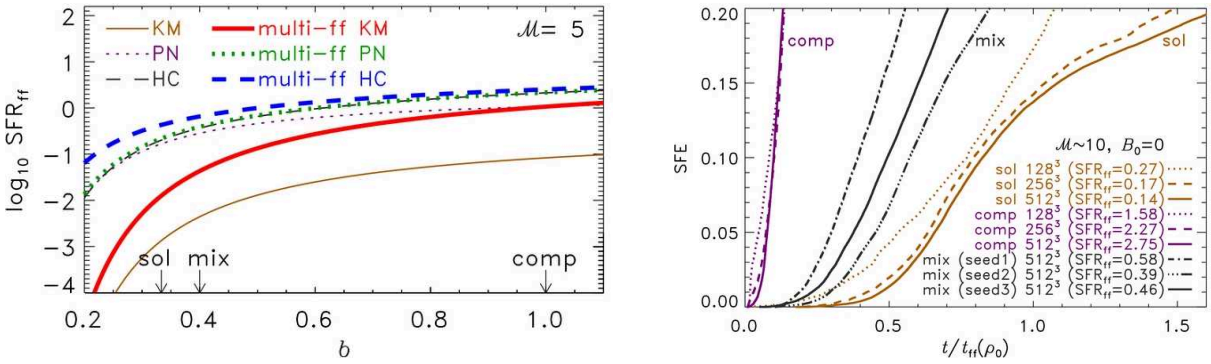


Figure 3.6: *Left:* The star formation rate deduced from various analytical model, using a single or multiple (multi-ff) free-fall time, consistently increases with the fraction of compressive modes in the turbulence forcing. *Right:* Simulations show that the star formation efficiency in not only higher, but increases much more quickly when the turbulence forcing is purely compressive, rather than mixed or purely solenoidal (figure adapted from Federrath & Klessen 2012).

controls star formation in molecular clouds. All things being kept equal, a cloud subject to compressive turbulence will contain very dense sub-regions compressed by shocks, and likely to collapse and form stars under the combined effect of hydrodynamics and self-gravity, while a cloud subject to solenoidal turbulence will be supported against self-gravity by the centrifugal force of the eddies and stirring motions generated by the turbulence, thus impeding its star formation activity. The observational characterization of the fraction of turbulence in solenoidal or compressive modes in a molecular cloud is therefore crucial to understand its evolution in terms of star formation and stellar feedback.

3.3 Measuring the fraction of compressive forcing in a GMC

3.3.1 Indirect vs. direct measurements of the compressive fraction

The studies of the turbulence forcing in observational data almost exclusively rely on comparison with simulation results, where the turbulence parameters are either known or accurately retrieved. Such studies include analysis of the properties of the density PDF (Goodman et al. 2009; Federrath et al. 2010; Kainulainen et al. 2013a, etc.), correlations between the SFR, SFE and/or column density (e.g. Federrath & Klessen 2012) or other statistical analysis methods such as CVIs (Hily-Blant et al. 2008; Federrath

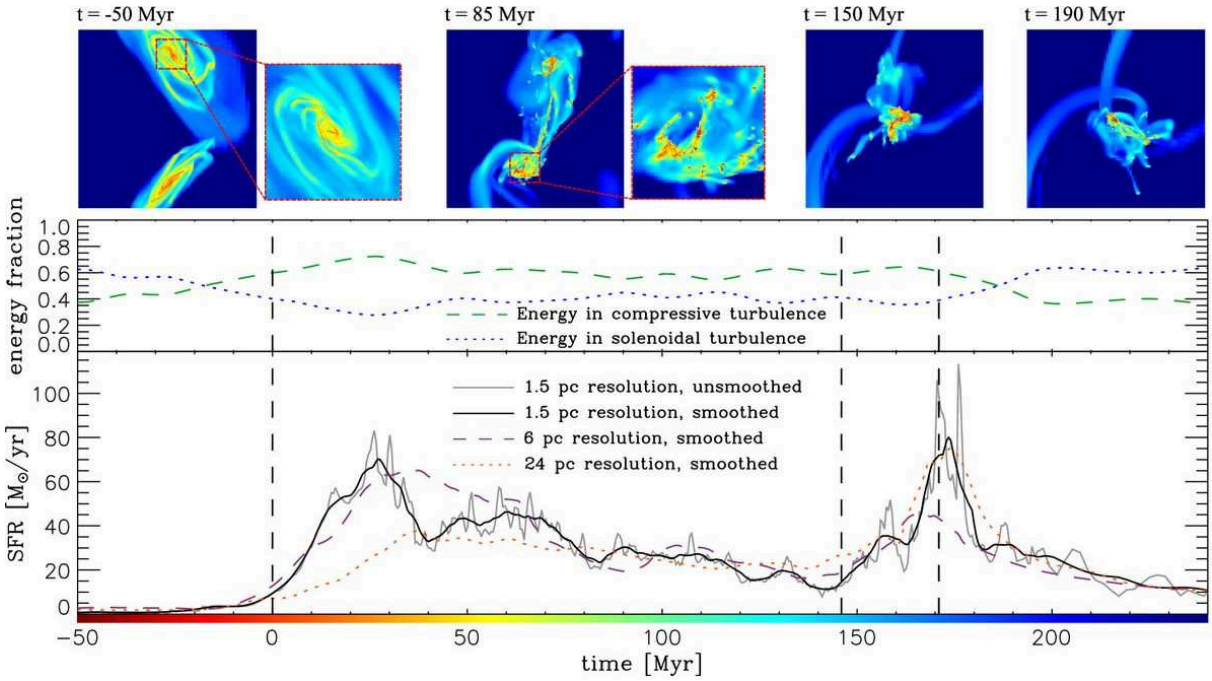


Figure 3.7: *Top row:* Snapshots of the simulation of the merging of two spiral galaxies. *Middle:* Evolution of the fraction of energy in compressive and solenoidal modes of turbulence. The compressive fraction increases significantly during the merger event, in particular during pericentre passages. *Bottom:* Star formation rate in the two galaxies during the merger event. The increases of the SFR closely follow the increases of the fraction of energy in compressive modes of turbulence (figure adapted from Renaud et al. 2014).

et al. 2010)... In all cases, the fraction of turbulence in the solenoidal or compressive modes is determined by indirect means, and this determination in faced by a number of degeneracies due to e.g., the Mach number or the magnetic field strength.

Despite these limitations, such studies have not only confirmed the star formation picture drawn by numerical simulations (as seen in Sect. 3.2), but they have also brought new constraints and a better understanding of the actual physical processes, for example by showing that the star formation efficiency of molecular clouds is not compatible with turbulence with a dominant compressive forcing (Kainulainen et al. 2013a), despite the fact that many of the processes that drive the turbulence, such as gravitational collapse, colliding flows, supernova explosions or stellar jets and outflows are likely to inject momentum directly into the compressive modes of turbulence (Federrath et al. 2008; Hennebelle & Falgarone 2012). This can be explained by (magneto-)hydrodynamical instabilities which contribute to mixing the modes, turning eddies into compressive shocks or compressive motions into vortices. One can for example mention the Rayleigh-Taylor instability which creates swirls and vertices around a compressive pillar, and is believed to contribute to the shape of the tendrils of the famous Crab nebula (Hester et al. 1996; Porth et al. 2014). This mixing of modes was confirmed in numerical simulations by Brunt & Federrath (2014), who have shown that at hypersonic Mach numbers, the fraction of momentum density in the compressive or solenoidal modes converges for hypersonic Mach number ($\mathcal{M}_s > 5$) to a fully mixed equipartition situation that is independent on the turbulent forcing (Fig. 3.8). The equipartition ratio (2/3 of solenoidal modes, 1/3 of compressive ones) can simply be explained in terms of the number of degrees of freedom needed to build such modes, and it agrees well with the general picture of supersonic or hypersonic molecular clouds with a turbulence that is more solenoidal than compressive.

The reason why observational studies have to find ways around to indirectly retrieve information on the turbulence forcing is that observations suffer from a number of limitations compared to simulations. Angular resolution is a first issue, but the main problem comes from the projection effect that is inherent

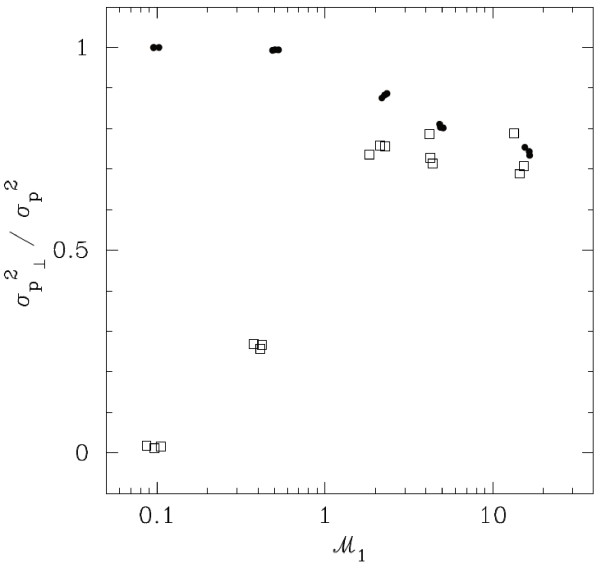


Figure 3.8: Fraction of power in the solenoidal modes of the momentum density field $\rho\vec{v}$ obtained in simulations with purely solenoidal (black squares) or purely compressive (open squares) forcing, as a function of the sonic Mach number (figure adapted from Brunt & Federrath 2014).

to astronomical observations. It comes from the fact the the studied objects (in our case, molecular clouds) are three-dimensional objects, while the observations are necessarily limited to the plane of the sky, and are thus merely projections of the observed objects along the line of sight. This implies a lack of knowledge along the third dimension (line of sight confusion), only made worse by the non-linearities of radiative transfer. The projection affects not only the measurement of density fields, but of velocity fields as well: only the line-of-sight component of the velocity can be retrieved thanks to the Doppler effect in the case of line observations at high spectral resolution, the plane-of-the-sky velocity remains inaccessible.

In simulations, where the entire three-dimensional (PPP) density scalar field $\rho(x, y, z)$ and the entire three-dimension velocity vector field $\vec{v}(x, y, z)$ are known, the most direct may to retrieve the fraction of velocity $\vec{v}$ or momentum density $\rho\vec{v}$ in solenoidal or compressive modes of turbulence is to explicitly perform the Helmholtz decomposition of these vector fields into their divergence-free and curl-free components – an approach that is impossible in the case of observational data.

However, Brunt & Federrath (2014) have devised a method that enables to reconstruct statistical properties of a vector field using the properties of its two-dimensional PPV projection. Building on the experience of Brunt et al. (2010), who reconstructed the variance of a 3D scalar field based on the variance of its 2D projection, they proposed to reconstruct the fraction of power in the solenoidal modes of the momentum density field. This method, despite having a number of requirements and assumptions, is quite a breakthrough, in so far as it gives directly access, from purely observational data, to the ratio between the compressive and solenoidal modes of turbulence, rather than to an indirect consequence of this ratio.

3.3.2 The case of Orion B, as seen in ^{13}CO($J = 1 - 0$)

In Orkisz et al. (2017), we have applied the method of Brunt & Federrath (2014) for the first time to actual observational data – as it only had been tested on simulations so far. The ORION-B dataset, which by then could provide a field of view of 1.5 square degrees in ^{13}CO ($J = 1 - 0$), was perfectly suited to test this method, as if offered a large spatial dynamic range, a relatively isotropic and isolated object and a good velocity resolution, as the method requires.

On the one hand, the analysis has shown that the Mach number in Orion B is on average hypersonic, so one expects the turbulence modes to be in equipartition (Fig. 3.8). However, the solenoidal fraction in the entire field is higher than the expected 2/3, showing that, on average, the cloud significantly lacks

compressive motions – this result can be related to the particularly low SFE in Orion B. On the other hand, zooms into the star-forming regions NGC 2023 and NGC 2024 (Fig. 3.9) have revealed a lower Mach number and much lower solenoidal fraction, which correlates with the star-formation activity highlighted by the density of YSOs in these two regions. In the case of NGC 2024, the dynamics of the gas (in particular the filaments traced by $C^{18}O$ ($J = 1 - 0$), see Chap. 4) show the presence of infalling gas as well as expansion of the young HII region created by the bright stars – thus confirming the tight link between star formation efficiency and compressive turbulence.

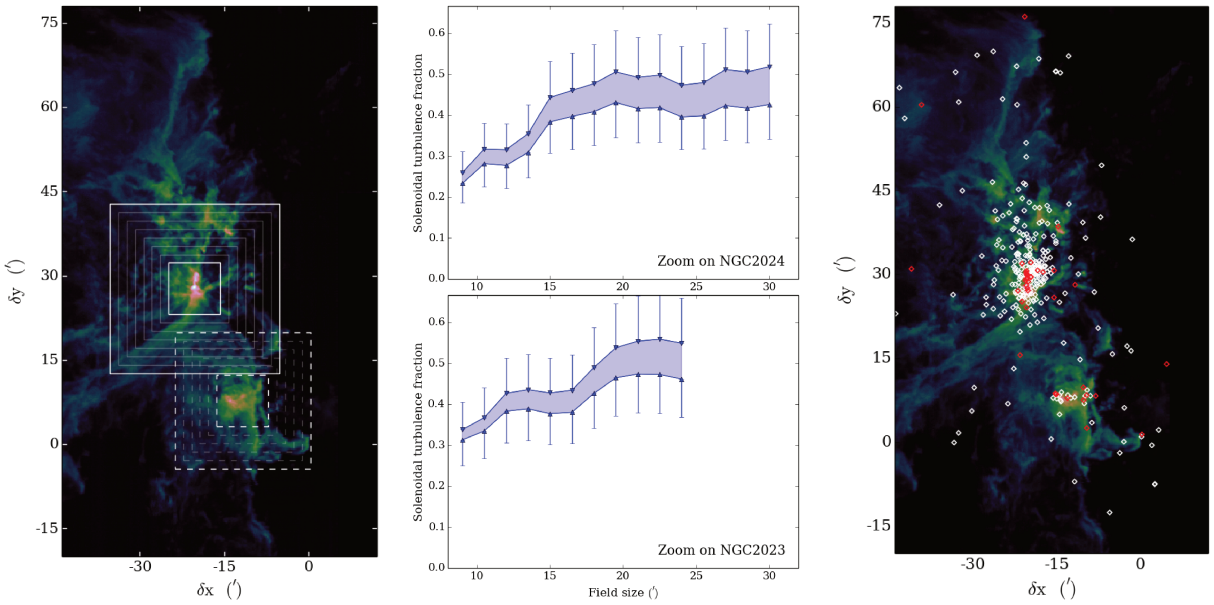


Figure 3.9: *Left:* Zooms on the star-forming regions NGC 2023 (bottom, dashed frames) and NGC 2024 (top, solide frames), overlayed on the peak temperature map of the ^{13}CO ($J = 1 - 0$) line. *Centre:* Fraction of solenoidal motions in the frames defined on the left panel, showing that the inner part of the star-forming region in much more compressive than the surrounding medium. *Right:* Position of protostars (red) and young stars (white) in the studied field – star formation is concentrated in the highly compressive regions (figures adapted from Orkisz et al. 2017, YSOs from Megeath et al. 2016).

A&A 599, A99 (2017)
DOI: 10.1051/0004-6361/201629220

Astronomy & Astrophysics

Turbulence and star formation efficiency in molecular clouds: solenoidal versus compressive motions in Orion B⋆

Jan H. Orkisz[1,2,3], Jérôme Pety[2,3], Maryvonne Gerin[3], Emeric Bron[4,8], Viviana V. Guzmán[6,7], Sébastien Bardeau[2], Javier R. Goicoechea[8], Pierre Gratier[5], Franck Le Petit[4], François Levrier[3], Harvey Liszt[9], Karin Öberg[6], Nicolas Peretto[10], Evelyne Roueff[4], Albrecht Sievers[11], and Pascal Tremblin[12]

[1] Univ. Grenoble Alpes, IRAM, 38000 Grenoble, France
e-mail: orkisz@iram.fr
[2] IRAM, 300 rue de la Piscine, 38406 Saint-Martin-d'Hères, France
[3] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, École normale supérieure, 75005 Paris, France
[4] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, 92190 Meudon, France
[5] Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, Allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[6] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, 02138, USA
[7] Joint ALMA Observatory (JAO), Alonso de Cordova 3107 Vitacura, Santiago de Chile, Chile
[8] ICMM, Consejo Superior de Investigaciones Cientificas (CSIC), 28049 Madrid, Spain
[9] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA, 22903, USA
[10] School of Physics and Astronomy, Cardiff University, Queen's buildings, Cardiff CF24 3AA, UK
[11] IRAM, Avenida Divina Pastora, 7, Núcleo Central, 18012 Granada, España
[12] Maison de la Simulation, CEA-CNRS-INRIA-UPS-UVSQ, USR 3441, Centre d'Étude de Saclay, 91191 Gif-Sur-Yvette, France

Received 30 June 2016 / Accepted 2 January 2017

ABSTRACT

Context. The nature of turbulence in molecular clouds is one of the key parameters that control star formation efficiency: compressive motions, as opposed to solenoidal motions, can trigger the collapse of cores, or mark the expansion of HII regions.
Aims. We try to observationally derive the fractions of momentum density (ρv) contained in the solenoidal and compressive modes of turbulence in the Orion B molecular cloud and relate these fractions to the star formation efficiency in the cloud.
Methods. The implementation of a statistical method applied to a ^{13}CO(J = 1−0) datacube obtained with the IRAM-30 m telescope, enables us to retrieve 3-dimensional quantities from the projected quantities provided by the observations, which yields an estimate of the compressive versus solenoidal ratio in various regions of the cloud.
Results. Despite the Orion B molecular cloud being highly supersonic (mean Mach number ∼6), the fractions of motion in each mode diverge significantly from equipartition. The cloud's motions are, on average, mostly solenoidal (excess >8% with respect to equipartition), which is consistent with its low star formation rate. On the other hand, the motions around the main star forming regions (NGC 2023 and NGC 2024) prove to be strongly compressive.
Conclusions. We have successfully applied to observational data a method that has so far only been tested on simulations, and we have shown that there can be a strong intra-cloud variability of the compressive and solenoidal fractions, these fractions being in turn related to the star formation efficiency. This opens a new possibility for star formation diagnostics in galactic molecular clouds.

Key words. turbulence – methods: statistical – ISM: clouds – ISM: kinematics and dynamics – radio lines: ISM – ISM: individual objects: Orion B

1. Introduction

The evolution of molecular clouds is controlled by a complex interplay of large-scale phenomena and microphysics: chemistry and interaction of the matter with the surrounding far-UV and cosmic-ray radiation control the thermodynamic state of the gas and its coupling to the magnetic field. The medium is highly turbulent, with Reynolds numbers reaching 10^7 and magnetic Reynolds numbers reaching 10^4 (Draine 2011). Magneto-hydrodynamic (MHD) turbulence is one of the main counter-actions to gravity (Hennebelle & Falgarone 2012; Federrath & Klessen 2012; Padoan et al. 2014), as well as the major mechanism that shapes the clouds: their fractal geometry is related to the properties of their turbulent velocity field (Pety & Falgarone 2000; Hily-Blant et al. 2008; Federrath et al. 2009). The dissipation timescale for the turbulent energy of a molecular cloud is shorter (∼1 Myr, Mac Low 1999) than the age of these clouds (∼20−30 Myr, Larson 1981). Hence, a continuous energy injection must exist (Hennebelle & Falgarone 2012, and references therein). The proposed injection mechanisms may be either external, for instance Galactic shear or nearby supernovae explosions (Kim & Ostriker 2015), or internal, like the expansion of HII regions and molecular outflows of the recently

⋆ Based on observations carried out at the IRAM-30 m single-dish telescope. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain).

formed stars (Hennebelle & Falgarone 2012). The nature of turbulence, and notably its solenoidal or compressive forcing, also plays a key role in the star formation efficiency (SFE) of molecular clouds (Federrath & Klessen 2012, 2013). In particular, compressive motions bear the mark of various phenomena related more or less directly to the formation of stars: infall on filaments, collapsing dense cores, expansion around young stars, etc. As a result, we can expect that a more compressive cloud is a more active cloud, probably more likely to be forming stars, as proposed by Federrath & Klessen (2012).

In this work, we propose to measure for the first time (to our knowledge) the relative fractions of momentum in the solenoidal and compressible modes of turbulence in a molecular cloud, following a method devised and tested on numerical simulations by Brunt et al. (2010) and Brunt & Federrath (2014). Our goal is to obtain a quantitative estimation of these fractions from observational data, and to compare their ratio with the star formation efficiency that was derived from independent data. We investigate whether different fractions of compressive and solenoidal motions might provide a diagnostic for the variation of the star formation efficiency among molecular clouds. A different proportion of compressive forcing might be the reason why some molecular clouds form stars at a high rate while others do not (Federrath et al. 2010; Federrath & Klessen 2012; Renaud et al. 2014).

Our object of study is a large region of a nearby giant molecular cloud (GMC), namely the south-western edge of the Orion B cloud (Barnard 33 or Lynds 1630). Orion B is relatively close to us, at a typical distance of ~400 pc (Menten et al. 2007; Schlafly et al. 2014), so that a spatial resolution of 25″ corresponds to 0.05 pc or 10^4 AU in the cloud. The total mass of Orion B is estimated to be $7 \times 10^4\,M_\odot$ (Lombardi et al. 2014), and the average incident FUV radiation field is $G_0 \sim 45$ (Pety et al. 2017). Orion B is located in the Orion GMC complex (Kramer et al. 1996; Ripple et al. 2013), east of the famous Orion Belt. Alnitak, the eastmost of the three belt stars shines in the foreground of the cloud. The south-western edge of the cloud represents an ideal laboratory to study star formation, and features several remarkable regions. First, the cloud is illuminated by the massive star σ Ori that creates an HII region, the emission nebula IC 434, bounded on its eastern side by an ionization front. Silhouetted against this bright background, a dark cloud can be seen: the famous Horsehead nebula. HD 38087 also creates a small HII region, IC 435. Still embedded in Orion B, the star forming region NGC 2024, known as the Flame Nebula, hosts several massive O-type young stellar objects, which have created compact HII regions inside the cloud. NGC 2024 lies just east of the Alnitak star, and is crossed by a filament that is seen in absorption in visible light, and in emission in the radio range. The reflection nebula NGC 2023 is a quieter counterpart of NGC 2024, hosting young B-type stars. It lies north-east of the Horsehead nebula. The rest of the cloud contains extended and quieter areas with strong filamentary structures (Fig. 1). This area has been extensively observed in the 3 mm range with the IRAM-30 m telescope (PI: J. Pety). This survey has led to a series of papers (Liszt & Pety 2016; Pety et al. 2017; Gratier et al. 2017) to which this article belongs.

The paper is organised as follows. In Sect. 2, we briefly describe the observations by the ORION-B (Outstanding Radio Imaging of OrioN B) collaboration, and the data we use here. In Sect. 3, we present the concepts and equations of the statistical method and the details of its implementation, from noise filtering to the computation of power spectra and the estimation of velocity-density correlations. The results are described in Sect. 4, and discussed in Sect. 5 with a special emphasis on the relation of the turbulence properties with the star formation efficiency in Orion B. In the appendix, we present our computation of the Mach number map in the cloud.

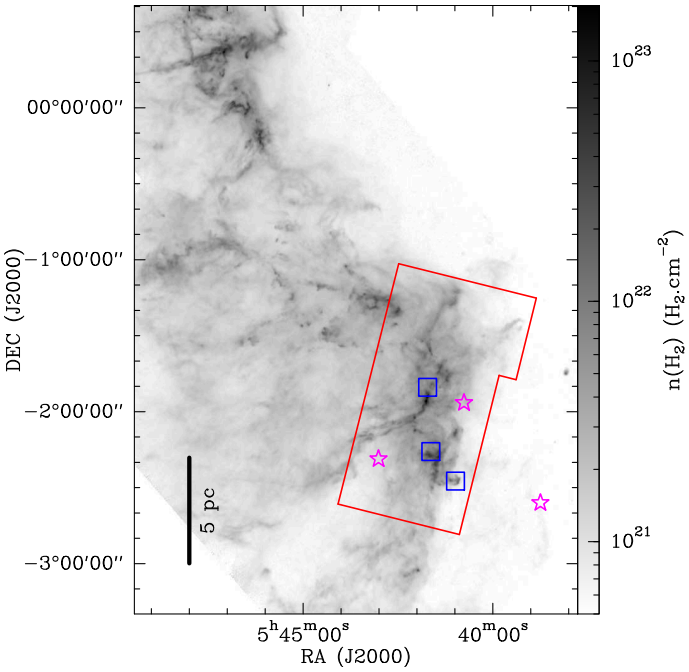


Fig. 1. H_2 column density map of the south-western part of the Orion B giant molecular cloud, derived from *Herschel* Gould Belt Survey observations (André et al. 2010; Schneider et al. 2013). The field observed by the Orion-B collaboration and used for this work is overlaid in red. The blue squares mark the nebulae NGC 2024, NGC 2023 and the Horsehead, and the pink star symbols mark the stars Alnitak, HD 38087 and σ Ori (north to south).

2. Observations

2.1. The Orion B project dataset

The Orion B project (PI: J. Pety) has already mapped the south-western edge of the Orion B molecular cloud with the IRAM-30 m telescope over a field of view of 1.5 square degrees in the full frequency range from 84 to 116 GHz at 200 kHz spectral resolution (Pety et al. 2017). The red rectangle in Fig. 1 shows the field of view observed up to now, over an H_2 column density map produced by the *Herschel* Gould Belt Survey consortium (André et al. 2010; Schneider et al. 2013). The mapped area covers 56×98 arcmin (about 6×11 parsecs at the assumed distance of Orion B, 400 pc) in size.

So far, the observations provided about 250 000 spectra over a 32 GHz bandwidth, yielding a position-position-frequency cube of $370 \times 650 \times 160\,000$ pixels, each pixel covering $9'' \times 9'' \times 0.5\,\mathrm{km\,s^{-1}}$ (at Nyquist sampling). The reduced dataset amounts to 100 GB of data. The data reduction is described in details in Pety et al. (2017).

This dataset is unique for its extensive coverage, both spatially and spectrally, at a typical resolution of 25″ and a median noise of 0.1 K [T_{mb}] per channel over a 32 GHz spectral range. From the spatial point of view, the survey gives us access to a large range of scales in the molecular cloud, from 50 mpc to 10 pc. For the analysis of turbulence properties (see Sect. 4.3), it means that we are able to study a large fraction of the inertial range of the turbulence, with a potential view on

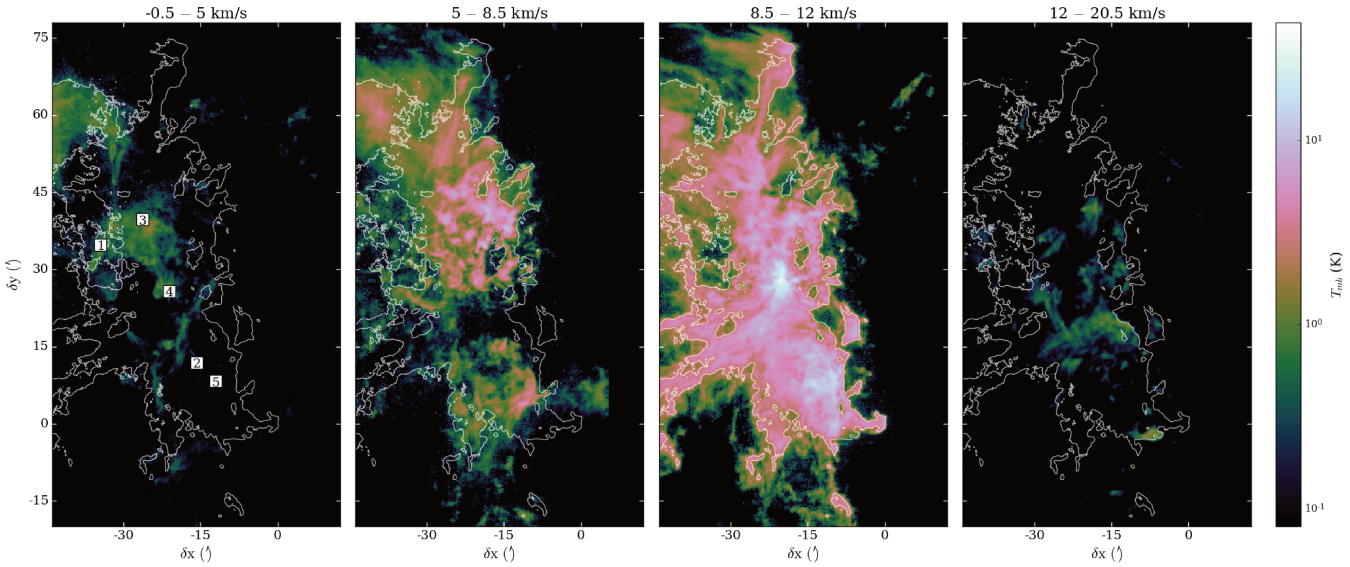


Fig. 2. Maps of the average brightness temperature of the ^{13}CO(J = 1–0) line in four contiguous velocity ranges. The main-beam temperature scale is indicated by the color bar on the right. The contour shows the value of 8.9 K km s^{-1} in the W_0 map, corresponding to 0.43 K in the mean temperature map integrated over the −0.5–20.5 km s^{-1} velocity range. The set of coordinates used for the observational campaign takes the Horsehead PDR as a reference point, and aligns the IC 434 PDR along the vertical axis (14° counter-clockwise rotation with respect to equatorial coordinates). The numbered squares in the *first panel* show the positions of the spectra presented in Fig. 3, from left to right.

the injection scale. The dissipation scale, on the other hand, is of the order of a milli-parsec (Hennebelle & Falgarone 2012; Miville-Deschênes et al. 2016), and, at a distance of 400 pc, is only accessible using millimetre interferometers, and out of reach for the IRAM-30 m telescope.

From the spectral point of view, having such a large bandwidth observed in one go enabled us to image over 20 chemical species (Pety et al. 2017), including those listed in Table 1. As opposed to several small bandwidth mappings, the spectral lines in this survey are observed in the same conditions and are well inter-calibrated, which gives an unprecedented spectral accuracy for such a large field of view.

2.2. The ^{13}CO spectral data cube

Most of the work presented here was performed on the ^{13}CO(J = 1–0) datacube[1], which covers a velocity range of 40 km s^{-1} centred around the source systemic velocity of 10.5 km s^{-1} and a rest frequency of 110.201 354 GHz. The datacube presents an root mean square (rms) noise of σ = 0.17 K, and a median signal-to-noise ratio of $T_{\rm peak}/\sigma$ = 7.9.

The ^{13}CO(J = 1–0) line was chosen because it offers one of the highest signal-to-noise ratios over the whole map, but it does not feature as much saturation as the ^{12}CO(J = 1–0) line. Signal is present at a S/N greater than 5 in the whole map, except in the HII regions around σ Ori and HD 38087, where molecular gas is photodissociated. The brightest regions are the NGC 2023 nebula, the center of the NGC 2024 nebula, and its northern edge (Fig. 4, left panel).

^{13}CO is a good tracer of molecular gas, from moderately diffuse and translucent regions ($A_{\rm V}$ = 1–5 mag) up to moderately dense and shielded gas (10^4 cm^{-3}, $A_{\rm V}$ = 10 mag). CO has a small dipole moment (0.11 D), hence the rotational lines have low Einstein coefficients (e.g., Mangum & Shirley 2015). This leads to relatively easy collisional excitation, excitation temperatures approaching the kinetic temperature, and moderate line opacities except for the most abundant species, ^{12}CO. The abundance ratio ^{13}CO/^{12}CO is equal to ^{13}C/^{12}C or about 1/60 when chemical fractionation reactions, which are limited to the most diffuse regions, are inefficient (Wilson & Rood 1994). Therefore, the ^{13}CO abundance relative to H_2 remains approximately constant at a level of $\sim 2 \times 10^{-6}$ (Dickman 1978) across most of the cloud volume. ^{13}CO starts to be depleted on dust grains in cold cores, but these represent only a small fraction of the mass and a negligible fraction of the volume of the Orion B molecular cloud (Kirk et al. 2016).

Figure 2 shows the ^{13}CO(J = 1–0) signal integrated over four contiguous velocity ranges. It showcases the complexity of the spectral structure of the cloud, with prominent variations of the line profile with the position, which is a consequence of the strong turbulence at play in the molecular cloud.

Up to four spectral components appear along each line of sight within the field (Fig. 3). A main component is visible around 10 km s^{-1}, and a secondary component at lower velocity (about 5 km s^{-1}). Sometimes an extra component at higher velocity (about 14 km s^{-1}), or secondary peaks around 5 and 10 km s^{-1} appear too. The first two components are the most significant ones at the scale of the whole cloud, being the only ones visible in the mean spectrum of ^{13}CO(J = 1–0), and have average velocities of 9.7 km s^{-1} and 4.9 km s^{-1} respectively. All the components, despite being quite distinct on the spectral axis are, however, thought to be part of the Orion B cloud (see discussion in Pety et al. 2017).

3. Deriving the relative fraction of solenoidal motions from a position-position-velocity cube

To measure the fraction of the solenoidal and compressive turbulence modes, we apply the method developed by Brunt et al. (2010) and Brunt & Federrath (2014). In this section, we first recall the method, its assumptions, and the way we implemented it.

[1] The data products associated to this paper are available at http://www.iram.fr/~pety/ORION-B

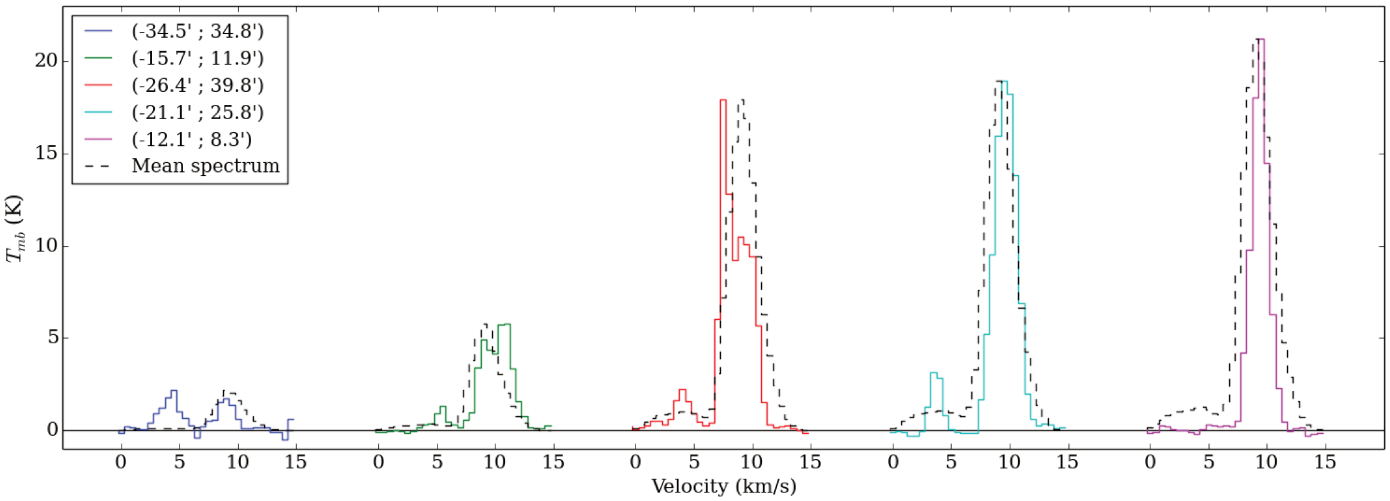


Fig. 3. ^{13}CO(J = 1−0) spectra along selected lines of sights in the Orion B cloud, showing the diversity of velocity components (up to four per spectrum). The coordinates of the various lines of sights are given in arc-minutes in our custom set of coordinates (δx, δy). The average spectrum for the whole field, normalized to have the same peak temperature, is superimposed with a dashed line for comparison. The positions of the spectra on the map are shown by white squares in the first panel of Fig. 2.

3.1. Description of the method

3.1.1. Principles and assumptions

The key point of this method is the fact that the objects we observe are fundamentally 3-dimensional (e.g., a molecular cloud), but the observer only has access to a 2-dimensional projection along the line of sight of that object. Brunt et al. (2010) developed a method to retrieve properties of the 3-dimensional object that we are interested in, and which corresponds to a 3-dimensional field F_{3D}, via the properties of the 2-dimensional observational F_{2D}, which is a projection of F_{3D} along the z axis. To achieve this, they use the fact that the Fourier transform $\tilde{F}_{2D}$ of the 2-dimensional field is proportional to the $k_z = 0$ cut through the Fourier transform $\tilde{F}_{3D}$ of the 3-dimensional field. In short, $\tilde{F}_{2D}(k_x, k_y) \propto \tilde{F}_{3D}(k_x, k_y, k_z = 0)$. If these fields are isotropic, i.e., if they are functions of $k = |\boldsymbol{k}|$ alone, with $\boldsymbol{k} = (k_x, k_y, k_z)$ the wave vector and k the wave number, the 2-dimensional field enables us to reconstruct average properties of the 3-dimensional field thanks to symmetry arguments.

The Brunt & Federrath (2014) method was developed as an application of the Brunt et al. (2010) method to the case of vector fields. For a vector field, such as a velocity or momentum field, the dimensionality reduction that is due to the projection is made worse by the fact that only one component of the vector (the line-of-sight one) can be measured, thanks to the Doppler effect. In this case, the next main tool to retrieve 3-dimensional properties is the Helmholtz theorem, which enables to decompose any vector field in its divergence-free (solenoidal) and curl-free (compressive) components, $\boldsymbol{F}_\perp$ and $\boldsymbol{F}_\parallel$. These components are related via a local orthogonality in Fourier space, $\tilde{\boldsymbol{F}}_\perp(\boldsymbol{k}) \perp \tilde{\boldsymbol{F}}_\parallel(\boldsymbol{k})$. Solenoidal modes can be pictured as the modes of a turbulent incompressible field, made of vertices and eddies. On the other hand, compressive modes, made of compression and expansion motions, are more likely to be generated by phenomena linked to star formation.

The application of these methods implies several requirements on the studied dataset. As mentioned earlier, the statistical isotropy of the cloud is the first necessary point, and enables the use of 2-dimensional averages as a means to estimate 3-dimensional properties. It means that the method cannot be applied to individual filaments, or to clouds where a strong anisotropy is suspected, e.g., owing to the presence of a strong magnetic field at low Mach numbers.

Second, the field is required to go smoothly to zero on its borders. This property is needed to ensure that the decomposition of the field is unique, since the Helmholtz decomposition is, in theory, defined up to a vector constant. It is also a necessary condition for good behaviours of the Fourier transform, since actual observational data are not periodic fields, unlike hydrodynamical simulations (see discussion in Brunt et al. 2010). This implies that the studied field should be bounded in space like, for example, a gravitationally bound cloud. In the case where the signal extends up to the edge of the observed field, the dataset has to be apodized.

Finally, from a practical viewpoint, Brunt et al. (2010) have shown that their method works best for fields with power spectra that are not too steep. Steep power spectra give measurements that are very sensitive to the low spatial frequencies, which are usually uncertain owing to poor statistics.

The compliance of our dataset with these requirements is discussed in detail in Sect. 5.1.

3.1.2. Equations and notations

The studied quantity is the momentum density field (hereafter momentum), $\boldsymbol{p} = \rho\boldsymbol{v}$, with ρ and $\boldsymbol{v}$ the volume density and the velocity.

The 3-dimensional quantity we infer is

$$R = \sigma^2_{p_\perp}/\sigma^2_p, \tag{1}$$

the ratio of the variance of the transverse (solenoidal) momentum to the variance of the total momentum. For short, R will be referred to as the solenoidal fraction in the rest of this paper.

According to Brunt & Federrath (2014), at hypersonic Mach numbers ($M = v/c_{\rm sound} > 5$) the solenoidal fraction does not depend any more on the type of forcing, but instead converges towards $R \sim 2/3$. Brunt & Federrath (2014) note that this behaviour is different from what is observed by Federrath et al. (2011), where the solenoidal fraction converges to different values depending on the type of forcing, but this is due to the fact

that Brunt & Federrath (2014) consider the *momentum* density field, while Federrath et al. (2011) describe the *energy* density field.

This value of $R \sim 2/3$ can be simply explained in terms of equipartition of momentum between the compressive and solenoidal modes (see, e.g., Federrath et al. 2008). A value of R lower than 2/3 therefore means that there is more momentum in the compressive modes of the flow, and that the cloud is thus more likely to form stars. The influence of the Mach number is further discussed in Sect. 4.1.

The available observables are a position-position-velocity cube, its velocity-weighted moments and the power spectra of these moments. We make two major assumptions about this datacube, namely that the ^{13}CO($J = 1{-}0$) line is optically thin and that its emissivity only depends on the ^{13}CO volume density. These assumptions are true within less than 20%, with the exception of a few lines of sight towards the center of NGC 2024, which are more saturated, but represent about 2% of the whole field. Under these assumptions, the position-position-velocity cube can be seen as a density-weighted velocity field: the spectrum obtained for each line of sight results from the projection of the emission of the matter present along this line of sight, and moving at various velocities.

The useful moments are the zero-th, first, and second order moments of the momentum field, W_0, W_1, and W_2, which are defined as follows in Brunt & Federrath (2014):

$$W_0 = \int I(v)\mathrm{d}v, \quad W_1 = \int vI(v)\mathrm{d}v, \quad W_2 = \int v^2 I(v)\mathrm{d}v. \tag{2}$$

The spectral line intensity, $I(v)$, may have contributions from various positions along the line of sight. Given our assumptions on emissivity, and assuming that the natural linewidth is negligible compared to the overall velocity dispersion, we can describe these moments in an alternative way:

$$W_0 \propto \int \rho(z)\mathrm{d}z, \quad W_1 \propto \int v(z)\rho(z)\mathrm{d}z, \quad W_2 \propto \int v(z)^2\rho(z)\mathrm{d}z. \tag{3}$$

The solenoidal fraction can be written in terms of these observational quantities as follows (see Brunt & Federrath 2014, for details):

$$R \approx \left[\frac{\langle W_1^2\rangle}{\langle W_0^2\rangle}\right]\left[\frac{\langle W_0^2\rangle/\langle W_0\rangle^2}{1 + A(\langle W_0^2\rangle/\langle W_0\rangle^2 - 1)}\right]\left[g_{21}\frac{\langle W_2\rangle}{\langle W_0\rangle}\right]^{-1} B. \tag{4}$$

The A and B factors are functions of the power spectra of W_0 and W_1,

$$A = \frac{\left(\sum_{k_x}\sum_{k_y}\sum_{k_z} f(k)\right) - f(0)}{\left(\sum_{k_x}\sum_{k_y} f(k)\right) - f(0)}, \quad B = \frac{\sum_{k_x}\sum_{k_y}\sum_{k_z} f_\perp(k)\frac{k_x^2+k_y^2}{k^2}}{\sum_{k_x}\sum_{k_y} f_\perp(k)}, \tag{5}$$

where $f(k)$ and $f_\perp(k)$ are the angular averages of the power spectra $\tilde{W}_0(k_x, k_y)$ and $\tilde{W}_1(k_x, k_y)$, respectively.

Brunt & Federrath (2014) introduce a statistical correction factor, g_{21}, of order unity, which measures the correlations between the variations of the density and the velocity fields, and is defined as

$$g_{21} = \frac{\langle \rho^2 v^2\rangle/\langle\rho^2\rangle}{\langle \rho v^2\rangle/\langle\rho\rangle}. \tag{6}$$

Brunt & Federrath (2014) show that this may be written as

$$g_{21} = \left\langle\left(\frac{\rho}{\rho_0}\right)^2\right\rangle^{-\epsilon}. \tag{7}$$

The 3-dimensional variance of the volume density, $\langle(\rho/\rho_0)^2\rangle$, can be derived using the Brunt et al. (2010) method. The exponent ϵ is a small, positive constant, which can be obtained as the exponent of the σ_v^2 vs. ρ power law, i.e., the typical velocity dispersion in the cloud as a function of volume density. If the density and velocity fields were uncorrelated, g_{21} would be equal to 1.

3.2. Implementation

Actual data suffer from several limitations that need to be dealt with to apply the method described previously. The field of view, the angular resolution, and the sensitivity are limited. This section describes how these issues were dealt with.

3.2.1. Noise filtering

Computation of line moments is sensitive to noise in the line wings. It is well known that masking the position-position-velocity cube where the signal stays undetected improves the determination of the centroid velocity and linewidth. To define the mask containing the pixels detected at high significance, these pixels are first grouped into continuous brightness islands that are made of neighbouring pixels in the position-position-velocity space, whose S/N is larger than 2. The noise level, σ, is measured outside the studied velocity range (-0.5, $20.5\,\mathrm{km\,s^{-1}}$). The list of islands is then sorted by decreasing total flux. The first island contains about 97.8% of the total signal. The following ones are small signal clumps that are spatially or spectrally isolated from this main block and, after about a few hundred islands, we are left with single-cell islands that are just noise peaks.

While it is easy to visually assess that the first few islands correspond to genuine signal, it is more complex to determine the transition to pure noise, as a significant fraction of the total line flux could be hidden in pixels of faint brightness, at low S/N. We thus studied the influence of the number of islands used on R, the solenoidal fraction in the studied cube. This influence is modest, mainly because over 97% of the signal is located in the first island. Using up to about 80 islands yields a very stable value of R with less than 0.1% variation. A steep increase of R is observed when we enter the noisy domain. We thus used the 80 brightest islands for all other calculations.

3.2.2. Moment computation

After selecting the signal islands in the position-position-velocity cube, the moments are integrated from -0.5 to $20.5\,\mathrm{km\,s^{-1}}$. The calculations have to be performed in the center-of-mass frame of reference of the cloud, which implies determining the centroid velocity of the cloud in the LSR frame. This center-of-mass velocity is simply given by

$$V_\mathrm{c} = \frac{\langle W_1^\mathrm{obs}\rangle}{\langle W_0\rangle}, \tag{8}$$

where W_1^obs is the first moment field in the observer's frame of reference. W_0, on the other hand, is not velocity-weighted and, therefore, does not depend on the frame of reference. For the observed field of view, we obtain $V_\mathrm{c} = 9.16 \pm 0.90\,\mathrm{km\,s^{-1}}$.

The velocity scale in the observer's frame of reference is shifted by V_c, before computing W_1 and W_2 in the center-of-mass frame of reference:

$$W_1 = \int (v_\mathrm{obs} - V_\mathrm{c})\cdot I(v_\mathrm{obs})\mathrm{d}v_\mathrm{obs}, \quad W_2 = \int (v_\mathrm{obs} - V_\mathrm{c})^2\cdot I(v_\mathrm{obs})\mathrm{d}v_\mathrm{obs}. \tag{9}$$

The resulting fields are shown in Fig. 4.

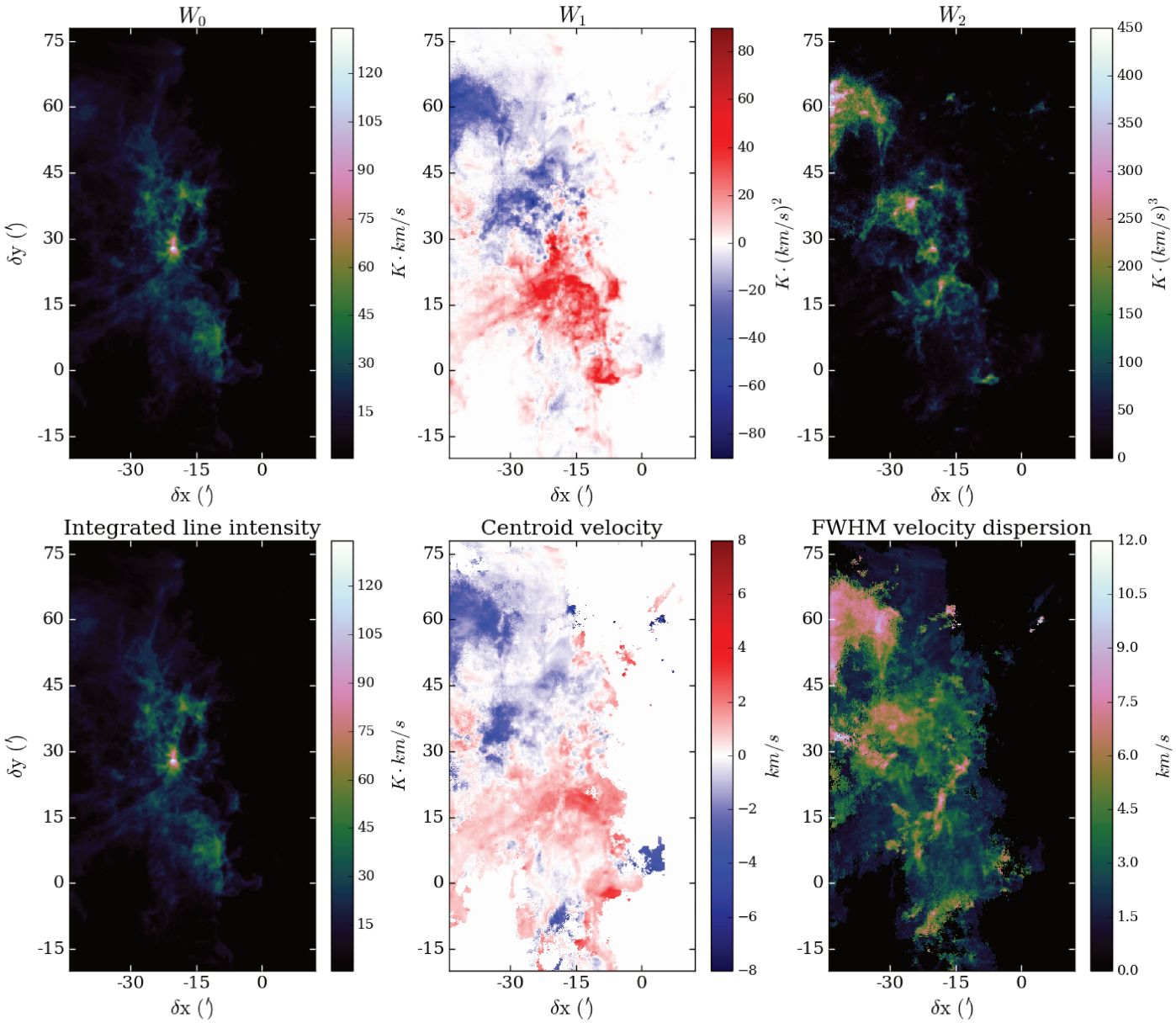


Fig. 4. *Top*: maps of the ^{13}CO(J = 1−0) field moments in the cloud's frame of reference. *Bottom*: maps of the physical quantities directly derived from each field, with the centroid velocity being simply W_1/W_0, and the Full Width at Half Maximum (FWHM) velocity dispersion given by $2\sqrt{2\ln(2)}\sqrt{W_2/W_0}$ – the normalization of the FWHM corresponds to that of a field with purely Gaussian line profiles.

3.2.3. Apodization

Computing the power spectra of the W_0 and W_1 fields requires to take the Fourier transform of these fields. Although calculating the FFT of 2-dimensional fields is an easy task, several numerical artefacts must be taken care of. In particular, the observed area does not reach the edges of the Orion B molecular cloud in all directions, as illustrated in Fig. 1. This sharp truncation of the ^{13}CO emission will create artefacts in the Fourier transform, owing to the convolution of the true Fourier spectrum by a sinus cardinal function that oscillates at high frequencies. Apodizing the field is required to avoid this behaviour. We have chosen to multiply the intensity by $1-\cos(\pi x/w)$, where x is the pixel coordinate and w the apodization width. This function goes from 0 to 1 over w pixels. This apodization function is used in Martin et al. (2015). We keep the region affected by apodization as small as possible to minimize signal alteration. An apodization width of about 5% of the smallest dimension of the field (i.e., roughly 25 pixels) was the smallest value that efficiently smoothed out the high-frequency artefacts. This is consistent as well with the width determined in Martin et al. (2015).

Once the field is apodized, it must be made square to follow the isotropy requirements of the method. The square was built by padding the right side (west) of the observed area with zeros, since this is the location of the HII region associated with σ Ori, i.e., no signal is detected past the western edge of the field of view. After this, the Fourier transform is calculated using the FFT implementation of Numpy 1.8 (Cooley & Tukey 1964).

Apodization is a linear filter of the data, and thus has effects on the power spectrum at all frequencies. While apodization allows us to clean the spectrum at high spatial frequencies, it also alters the spectrum at the lowest frequencies. For example, $f(0)$ – the value of the power spectrum at spatial frequency $k = 0$ – is directly proportional to the spatial integral of W_0^2, therefore losing some signal because of the apodization will reduce the value of

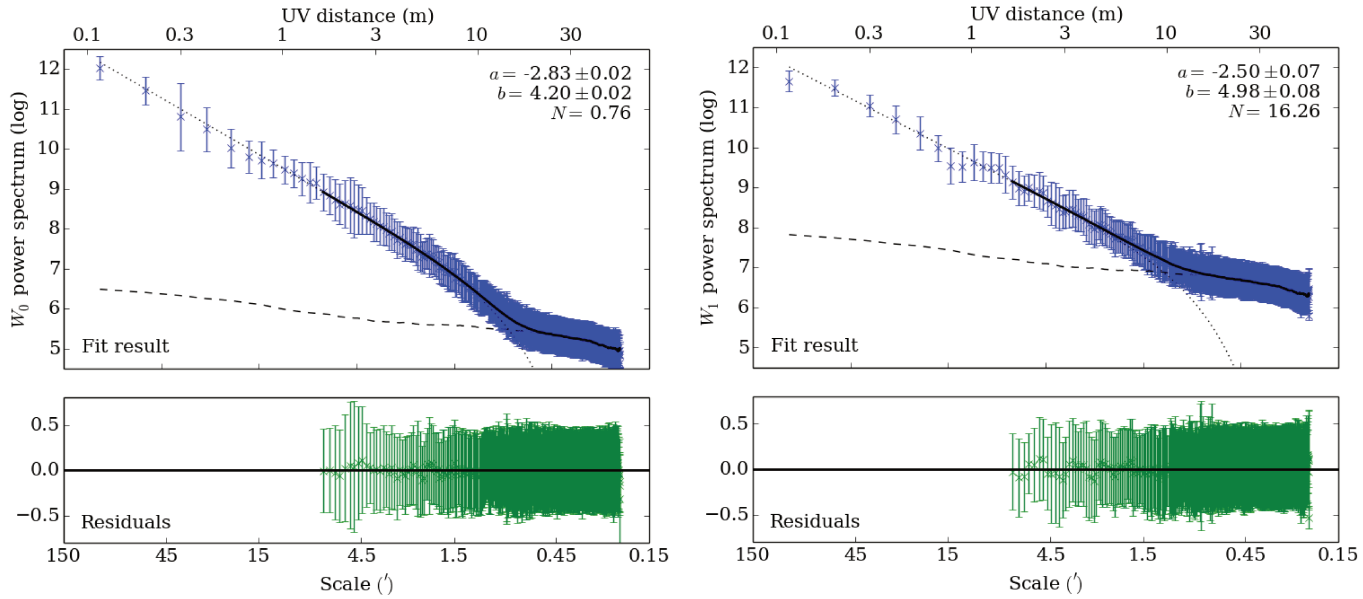


Fig. 5. *Left*: power-law fitting of the W_0 power spectrum. *Upper panel*: data (blue crosses), fit result (thick solid line) plotted over the fitted domain, power law convolved with the Gaussian beam (dotted line) extrapolated to all spatial frequencies, noise model (dashed line). *Lower panel*: residuals. The scale in the Fourier space is given in UV distance as for interferometric observations, which enables us to visually relate the resolution and the telescope diameter. *Right*: same results, except for the W_1 power spectrum.

$f(0)$. The method chosen to keep the good parts of the apodized and non-apodized spectra was the following: we first computed the FFT of both the apodized and non-apodized fields, respectively $\tilde{W}_{\rm ap}$ and $\tilde{W}_{\rm nap}$, then mixed them smoothly around $k = 0$ using a narrow (about 10 Fourier-space pixels) 2-dimensional Gaussian $G_{\rm mix}(k)$. The resulting Fourier field is

$$\tilde{W}_{\rm final} = G_{\rm mix}(k) \cdot \tilde{W}_{\rm nap}(\boldsymbol{k}) + (1 - G_{\rm mix}(k)) \cdot \tilde{W}_{\rm ap}(\boldsymbol{k}). \quad (10)$$

The power spectrum thus behaves like the non-apodized spectrum at low k, keeping the correct value of the field integral $f(0)$, and like the apodized spectrum at high k, free of the spectral parasites created by the sharp edges of the map.

3.2.4. Power spectra computation and fit

The apodized and corrected FFT of the field needs to be transformed into an angle-averaged power spectrum $f(k)$ or $f_\perp(k)$. This is simply done by binning the modulus of the spatial frequencies, and averaging the points found in these radial bins. The resulting discrete function can then be linearly interpolated into a continuous function. A critical element in this exercise resides in the sampling of the spatial frequency axis. On the one hand, the resulting angle-averaged spectrum should be as detailed as possible but, on the other hand, a larger number of bins can lead to empty bins, containing no sampled points at all. As a result, we used a number of bins of $S/1.45$, with S the size in pixels of the square field, so that the size of a bin in the Fourier space corresponds to slightly more than the length of the diagonal of pixels in the Fourier space.

Additional observational constraints (noise, beam shape, etc.) affect the determination of the power spectrum. Following Martin et al. (2015), the power spectra are fitted with a power law, modified to take into account the single-dish beam and the noise. In our case, the beam is modelled as the Fourier transform of a Gaussian of FWHM equal to the cube resolution, i.e., 23.5″. This corresponds to about 2.61 pixels. The convolution in the image space corresponds to a multiplication in the Fourier space that mostly affects the high spatial frequencies.

The noise is not a Gaussian white noise, because of inter-pixel correlations and systematics. We therefore use the power spectra of 30 signal-free channels and average them to obtain a template of the noise power spectrum. In this case, we use the fully noisy data cube, not the 80 first signal islands, to have the same spatial correlations in noise for each channel (with or without signal), which would not be the case with a masked data cube. The noise template intends to reproduce a systematic behaviour, but we can only use a finite number of channels with random noise. The noise template is therefore smoothed to make this systematic pattern stand out more. Given that W_1 is, just like W_0, a linear combination of the channel maps, the same noise template is used for both power spectra.

The fit is performed in the $\log(k)$-$\log(f)$ space, so that the straight line of the power law stands out more. The fitting function is therefore the logarithm of

$$10^{(a \cdot \log(k) + b)} \cdot \tilde{G}_{\rm beam}(k)^2 + N \cdot noise(k) \quad (11)$$

where $G_{\rm beam}$ is the Gaussian beam, $\tilde{G}_{\rm beam}$ is its (Gaussian) Fourier transform, and *noise* is our noise template. The fitted parameters are a, b, and N. During the fit, the data are weighted by the inverse of the variances obtained in each bin when computing the power spectrum.

The choice of fitting the power spectrum with a modified power law implies that the underlying physical processes should produce a power law. This is indeed the case for the inertial range of scales in Kolmogorov turbulence, and can be applied as well to Burgers turbulence (see, e.g., Federrath 2013). However, the power spectrum of turbulence starts to deviate from a power law at scales where the energy is injected (low spatial frequencies) or dissipated (high spatial frequencies). The power spectra computed from our dataset somewhat deviates from a power law at low spatial frequencies (see Fig. 5). The power law can therefore only be fitted and then used above a given spatial frequency. The power-law range starts around ~5.5′. At high spatial frequencies,

no deviations from the power law are detected (the fits render the observations very well). This means that either the dissipation scale happens at lower angular scale than the resolution of the observations or it is hidden in the noise.

Once the fit has been performed, the final version of the power spectra is built using both the fit result and the observational angle-averaged power spectra, and used for further calculations. The final power spectrum is a pure power law (without the beam and noise) above the $1/5.5$ arcmin^{-1} threshold, and is equal to the linearly interpolated angle-averaged power spectrum below this threshold. In particular, we enforce $f_\perp(0) = 0$, since we are working in the cloud's rest frame.

3.3. Density-velocity correlations

We used the information on the mean line profiles to estimate the slope ϵ of the relation between the velocity dispersion and the local density (see Eq. (7)). Among the lines detected in the mean spectrum, we selected lines with different spatial distributions (Pety et al. 2017). To trace the low density gas, we selected ^{12}CO($J = 1$–0) and HCO$^+$($J = 1$–0) since these lines present very extended emission and have moderate excitation requirements (Pety et al. 2017; Liszt & Pety 2016). We included ^{13}CO($J = 1$–0) as our tracer of the bulk of the gas. The somewhat denser and more shielded gas is well traced by C^{18}O($J = 1$–0), while we selected N_2H^+($J = 1$–0) for the dense cores. For these five species, we determined the FWHM by fitting a Gaussian line profile to the mean profile of the whole map. Only the 10 km s^{-1} component, which is present for all five species, was used for this fit. For N_2H^+ we used the HFS fit method in GILDAS/CLASS[2], which makes use of the information on the hyperfine structure.

While these lines are emitted by gas over a wide range of densities, there is a minimum density under which the line is not detected because of lack of excitation or because the molecule is not present in low density gas. It is this minimum density that corresponds to the velocity dispersion of the line. To derive the densities associated with the line emission, we adopted three different methods.

For the low density and extended emission tracers, we derived the volume density by comparing the minimum gas column density where the emission is detected and the resulting size of the emission regions. This leads to gas densities of a few hundred cm^{-3} for ^{12}CO($J = 1$–0), HCO$^+$($J = 1$–0) and ^{13}CO($J = 1$–0). The emission of ^{12}CO($J = 1$–0) is dominated by the low density regions (Pety et al. 2017). The emission of HCO$^+$($J = 1$–0) is also dominated by low to intermediate density regions, and comes from the weak excitation regime (Liszt & Pety 2016). In both cases, opacity broadening is not very significant. We must, however, consider that the line widths for these tracers are upper bounds, owing to the effect of opacity broadening. For these molecules, we are also limited by the sensitivity of the observations, so that the densities should be regarded as upper limits. Hily-Blant et al. (2005) analysed the structure and kinematics of the Horsehead nebula and derived the density of the extended region traced by C^{18}O as 3–5×10^3 cm^{-3}. We kept this value as the typical density traced by C^{18}O. Hily-Blant et al. (2005) show that higher density regions exist with densities significantly larger than 10^4 cm^{-3}. As very few pixels are detected in N_2H^+ in the region of the Horsehead nebula, we used the catalogue of dense cores identified by Kirk et al. (2016) in their SCUBA2 map of the Orion B complex. We found 55 cores associated with N_2H^+($J = 1$–0) emission. The densities were derived using the extracted fluxes and effective radii, a uniform dust temperature of 20 K, and assuming spherical geometry for all cores. The mean density is $10^{4.1}$ cm^{-3} with a scatter of about a factor of two. The temperature of the N_2H^+ cores was not individually checked, but it is likely to be lower than 20 K. In turn, this implies an even higher density of N_2H^+. Therefore, the derived density should be regarded as a lower limit.

[2] See http://www.iram.fr/IRAMFR/GILDAS/ for more details on the GILDAS software.

Table 1. Spectral tracers used in this study, and observed with the IRAM-30 m telescope.

Line ($J = 1$–0)	Frequency (GHz)	*FWHM* (km s^{-1})	log(n(H$_2$)) (cm^{-3})	Ref.
^{12}CO	115.271202	4.08 ± 0.04	2.17 ± 0.3	1
HCO$^+$	89.188525	3.91 ± 0.08	2.34 ± 0.3	1
^{13}CO	110.201354	2.97 ± 0.03	2.65 ± 0.3	1
C^{18}O	109.782173	2.55 ± 0.01	3.60 ± 0.15	2
N_2H^+	93.173764	1.79 ± 0.10	4.10 ± 0.3	3

Notes. The linewidths are derived from these observations, the typical densities from these observations (Ref. 1) or other works (Ref. 2: Hily-Blant et al. 2005, Ref. 3: Kirk et al. 2016).

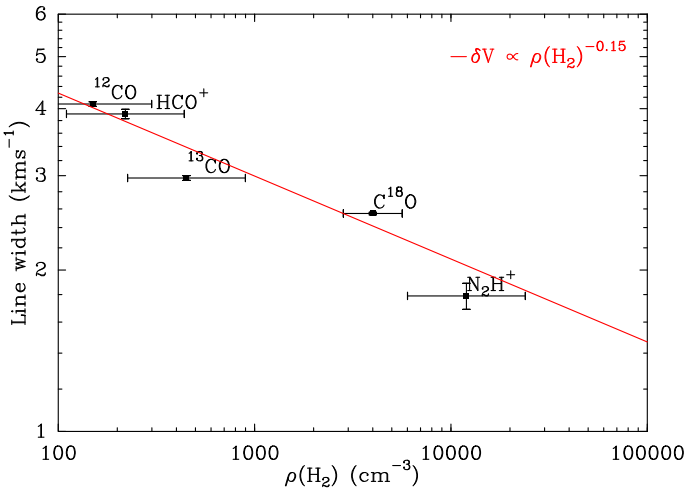


Fig. 6. Variation of the FWHM as a function of the gas density. Each point refers to the ($J = 1$–0) spectral line of a different molecule. The red line shows the least squares fit, with a slope $\alpha = -\epsilon = -0.15$.

Table 1 presents the resulting data, which are illustrated in Fig. 6. The slope $\alpha = -\epsilon$ is derived from a least squares fit of the variation of the FWHM with the density. We derive $\epsilon = 0.15 \pm 0.03$. The possible systematics on the densities traced by ^{12}CO($J = 1$–0), HCO$^+$($J = 1$–0), and N_2H^+($J = 1$–0), as well as on the linewidths of ^{12}CO($J = 1$–0) and HCO$^+$($J = 1$–0), all tend to make the power law steeper. Therefore, we keep this value of ϵ as an upper limit.

3.4. Estimation of the uncertainties

The computation of uncertainties implies computing the uncertainty of each element of Eq. (4). We start from the average rms noise level in the data cube, 0.17 K [$T_{\rm mb}$]. This enables us to compute the noise level for the W_0, W_1, and W_2 maps. The computation is straightforward compared to the computation of the uncertainty of the centroid velocity and linewidth because W_1, and W_2 are not normalized by W_0, i.e., their noise distributions

stay Gaussian, whatever the value of W_0. Due to the velocity weighting, the absolute uncertainty increases significantly with the moment order. However, the relative uncertainties on $\langle W_0 \rangle$, $\langle W_1 \rangle$ and $\langle W_2 \rangle$ are similar, with values ranging from about 10% for the whole field to 5% for the deepest zooms on NGC 2023 and NGC 2024 (see Fig. 8). The uncertainties on the sums A and B were explicitly computed according to the error bars described in Sect. 3.2.4. The relative uncertainty on A ranges from 24% for the full field to 11% for the deepest zooms, and stays around 13% for B.

For the overall relative uncertainty, one must not only take into account the errors of the individual terms, but also the correlation between the different variables. In our case, the different variables are strongly correlated, since they are all by-products of the same data cube. Therefore, we chose to use the average of the various relative errors as a rule-of-thumb estimate of the overall error. This approach yielded an overall relative error of 13% for the whole field, and 8% for the deepest zooms. We kept the highest value to allow for a safety margin.

This 13% relative error $\Delta R/R$ corresponds to the median noise-to-signal ratio in the field, $\sigma/T_{\rm peak}$, which is also of the order of 13% – but testing with simulated noise whether this is a coincidence or not is out of the scope of this article.

4. Results

In this section, we first briefly present the derivation of the Mach number in the cloud, then we compare the obtained power spectra with other results in molecular clouds, and finally, we give the results of our computation of the solenoidal fraction R in the Orion B cloud.

4.1. Mach number

According to Brunt & Federrath (2014) and Federrath et al. (2011), at hypersonic Mach numbers ($M > 5$) the ratio of solenoidal and compressive modes does not depend any more on the type of forcing. It is thus important to also derive the distribution of the Mach number to be able to interpret the results.

Using the maps of the sound speed derived from ^{12}CO($J = 1 - 0$) and dust temperature maps, two different estimations of the Mach number, $M_{\rm max}$ and $M_{\rm exc}$, were computed (see appendix for details). Figures A.2 and A.3 show their spatial distributions and compare their histograms. The shapes of the histograms of Mach numbers computed with $T_{\rm max}$ and $T_{\rm exc}$ are very similar, and both show a large tail at high Mach numbers. Table A.1 lists several characteristic values of both distributions. The most probable value (~3.5) of the Mach number is much smaller than the mean or median values (~6), for both distributions.

Schneider et al. (2013) estimate the average Mach number to be of ~8, with approximately 30–40% error, deriving this value from *Herschel* dust temperature and CO linewidth. They also find that Orion B has the highest Mach number of the set of studied clouds. Our results are compatible with this mean value, and they also provide the spatial and statistical distribution at good angular resolution. In particular, the Mach number is much smaller than the average in the star forming regions NGC 2023 and NGC 2024, below the hypersonic regime.

4.2. Power spectra

When the whole field is considered, the fit yields an exponent $a_0 = -2.83 \pm 0.02$ for the W_0 field, and $a_1 = -2.50 \pm 0.07$ for the W_1 field. When zooming into specific regions of the cloud (NGC 2023 and NGC 2024, see Fig. 8), the values of these exponents range from $a_0 = -2.52 \pm 0.08$ (widest field) to $a_0 = -3.04 \pm 0.05$ (smallest field), and from $a_1 = -2.24 \pm 0.09$ to $a_1 = -2.81 \pm 0.06$. While the indices of W_1 power spectra are rarely reported, there are many values of spectral indices for integrated line intensity maps in the literature, and our a_0 values fall well in the range of spectral indices for observations of CO emission, dust emission, HI emission and absorption compiled by Hennebelle & Falgarone (2012): the values range from −2.5 to −3.2, with most values around −2.7.

To have a meaningful result for the A and B coefficients, the power-law slope of the power spectra must follow the two steepness requirements mentioned by Brunt et al. (2010). On the one hand, the spectra should not be too steep, so that the weight of the low spatial frequencies, for which the available information is scarce, does not become too large in the A and B sums. On the other hand, the slopes should be steep enough to avoid divergence of these sums at large frequencies. A slope of $a = -3$ is at the limit between these two contradicting constraints, since the divergence of the 3-dimensional sum is only logarithmic. In our case, the sums are finite, owing to the finite resolution of the observations, so that with slopes between −2.24 and −3.04, the sums do not grow too quickly and do not give too much weight to the low spatial frequencies.

4.3. Turbulence mode ratio

We determined a relative error of about 13% on the calculation of the ratio R from the position-position-velocity cube. The correction factor g_{21} is determined independently, and we assume a range of possible values for g_{21}, the lower limit being given by our calculations of Sect. 3.3, and the upper limit resulting from the minimum estimate of $\epsilon \simeq 0.05$ according to Brunt & Federrath (2014).

For the entire ^{13}CO field, we obtain the following range of values:

$$0.72^{+0.09}_{-0.09} < R_{^{13}\rm CO} < 1^{+0.0}_{-0.09}. \quad (12)$$

To gain a deeper understanding of the dynamics at stake in the Orion B cloud, the method was also applied to several subregions of the ^{13}CO map.

First, to check the reliability of the method and the homogeneity of the field, we applied a sliding square window, whose side is equal to the smallest dimension of the mapped area (Fig. 7). This avoids zero-padding the studied field. The results show that, even though the values are in general somewhat lower for these sub-fields than for the whole ^{13}CO field, they remain marginally compatible with this result, within the estimated uncertainties.

Second, we searched for systematic variations of the fraction of solenoidal modes when zooming into specific regions of the map. In particular, signs of the solenoidal or compressive forcing are expected to appear mainly in regions of low Mach numbers (see Sect. 4.1). The zooms were thus performed into the NGC 2023 and NGC 2024 star forming regions (Fig. 8), where the Mach number lies between about 3 and 5 (see Fig. A.2), mostly because the speed of sound is higher in regions of higher gas temperature, but also because the velocity dispersion is a bit lower.

Moreover, these regions offer the advantage of presenting a strongly localized emission. One of the requirements of the Brunt & Federrath method is to use isolated fields. This is clearly

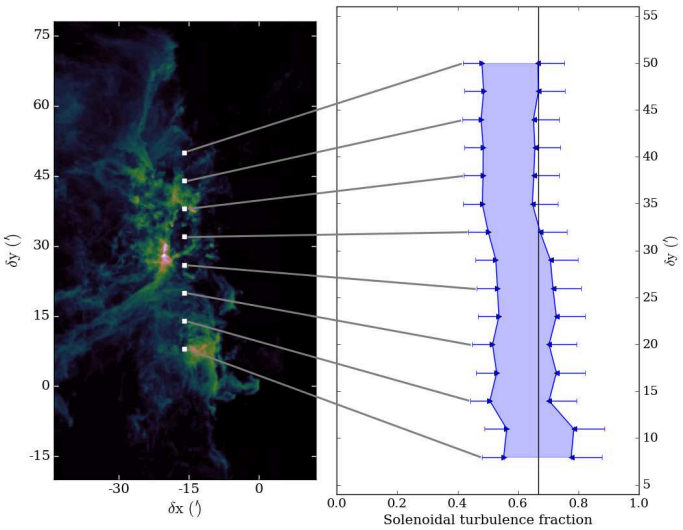


Fig. 7. Solenoidal turbulence fraction for sliding square areas with a width equal to the one of the full map (56 arcmin). The shaded area corresponds to the g_{21} uncertainty, while the error bars show the experimental uncertainties due to observational noise. The vertical line marks the equipartition limit. The map on the *left* presents the centres of the square areas for which the calculations were performed, superimposed on the ^{13}CO($J = 1-0$) $T_{\rm peak}$ map.

not the case any more when zooming into these specific regions, and apodization is more necessary than ever to ensure that the signal falls smoothly to zero on the edges of the field. However, by using fields for which most of the signal is concentrated near the center, the effects of apodization are minimized, allowing us to stay as close as possible to the requirement of an isolated field. The smaller the field, the lower the value of R: this indicates an increasing proportion of compressive forcing.

5. Discussion

5.1. Compliance with the method's assumptions

To apply the Brunt & Federrath method to real observational data, we were able to overcome several difficulties and sources of uncertainty.

First, the compliance of the dataset with the requirements of the method must be checked, namely the isotropy of the studied cloud, and its isolation. The whole field and the zooms present two opposite situations. The isolation criterion is well met in the case of the whole field: we have almost no signal to the west and to the south of the field, and very diffuse regions to the north and to the east (see Fig. 4, first column), so that there is almost no need for apodization to have the signal going down to zero on the edges of the field. For the zooms, on the other hand, we are well into the cloud, so that there is signal all the way to the edges of the field. However, since we study local maxima of the emission, most of the signal is in the center, which minimizes the effects of the necessary apodization on the final results (only 7.9% of lost signal for the deepest NGC 2023 zoom, 7.6% for NGC 2024).

As far as the isotropy criterion is concerned, the 2D projection is quite obviously isotropic in the case of the zooms, since the considered fields are square, and much less in the case of the whole field, in which the region with signal has an aspect ratio of about 2:1. The third dimension is unknown, and in any case cannot match simultaneously the dimensions of the whole field and those of the deepest zooms. However, for the large diffuse regions like for the bright, compact regions, the dimension along the line of sight is supposed to be of the order of the dimensions in the plane of the sky. In the case of a zoom on a bright region embedded in a diffuse one, the signal and, therefore, the dimension on which it is emitted along the line of sight, is dominated by the bright gas. Therefore, since the zooms are centred on a bright region, the corresponding datacube behaves almost as if the bright region was isolated and isotropic. Besides, the non-angle-averaged power spectra of W_0 and W_1 do not show any apparent anisotropy at any scale (except for the windowing effects). If the power spectra are isotropic in two dimensions, then statistically we can expect the third dimension to follow this isotropy as well. Therefore, the isotropy requirement seems to be fulfilled by our dataset and the method can be applied.

Second, as was mentioned by Brunt & Federrath (2014), the method is sensitive to values of the power spectra at large spatial scales (low frequencies of the power spectra) owing to the characteristics of the sums in the parameters A and B. We therefore had to find a way to obtain a smooth and reliable function that would represent an angle-averaged power spectrum at all frequencies. Once the power spectra were binned and fitted, we have two versions of the spectra, each with its flaws. The binned (data only) spectrum suffers from observational effects (noise and beam) but has also larger uncertainties at low spatial frequencies. The fitted spectrum, if extrapolated to all spatial frequencies (fit only), can give unphysical results in the lowest frequencies because they lie outside the power law validity range. For example, if the field had been zero-padded all around to create a very large square, such an extrapolation would give very high values of the spectrum at low frequencies, whereas physically they should be very low, since the field would on average be almost empty. These flaws led us to choose the composite scheme described in Sect. 3.2.4, where the fit result is used only in the power-law domain, and the low frequencies keep the angle-averaged power spectrum as it is. Using a different version of the power spectra leads to quite different final results, as illustrated in Fig. 9. However,we note that, even though the absolute value of the solenoidal fraction R varies, the relative variations are consistent across scales, whatever the power spectra computing scheme. Thus, results such as the unusually high solenoidal fraction at the scale of the whole cloud, or the variations of R when zooming out of the star forming regions, stay valid and are further discussed below.

5.2. Physical interpretation

To our knowledge, this work is the first attempt at applying the Brunt & Federrath method to actual observational data (Brunt & Federrath 2014; Lomax et al. 2015). The results need to be compared with what has been done so far on numerical simulations.

Both Schneider et al. (2013) and our calculations (see Sect. 4.1 and appendix) show that we are in a context of highly supersonic turbulence, with a mean Mach number of about 6. We therefore expect a full mixing of the turbulence modes, so that the momentum equipartition would predict a solenoidal fraction of $R = 2/3$ and a compressive fraction of $1 - R = 1/3$. Deviations from this ratio can either be a sign of a specific forcing for the turbulence in the case of transonic Mach numbers, as shown in the case of simulations (Brunt & Federrath 2014), or indicate that an ordered flow is superimposed on top of the turbulent flow.

The global value of $R > 0.72 \pm 0.09$ and the values in Fig. 7 can agree, within the error bars, with the expected value $R = 2/3$

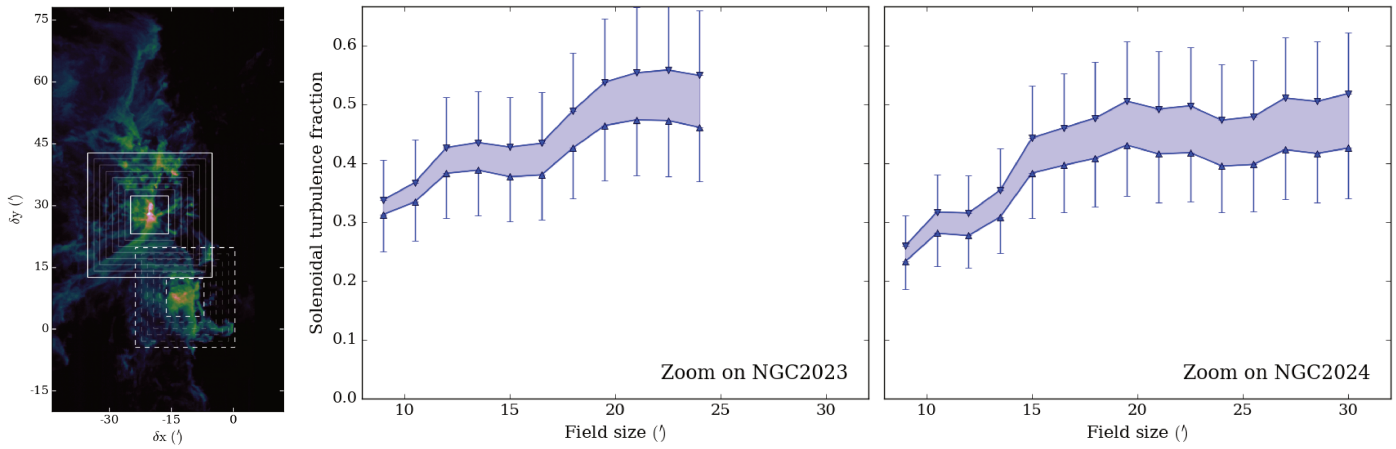


Fig. 8. Solenoidal turbulence fraction for zooms on the NGC 2023 and NGC 2024 star forming regions. The shaded area corresponds to the g_{21} uncertainty, while the error bars show the experimental uncertainties due to observational noise. The upper limit of the plots marks the equipartition limit. The map on the *left* presents the areas for which the calculations were performed (solid squares: NGC 2024, dashed squares: NGC 2023), superimposed on the ^{13}CO(J = 1−0) $T_{\rm peak}$ map.

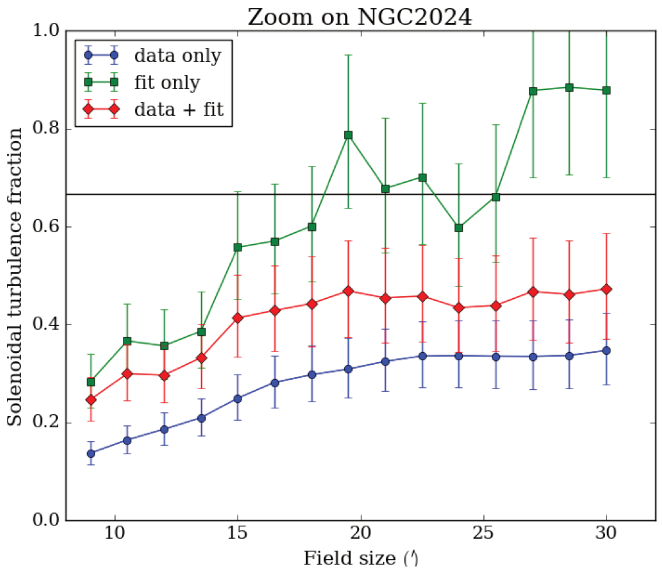


Fig. 9. Comparison of the calculation results in the case of the zoom on NGC 2024 with three different methods of computing the power spectra: either the power spectra resulting from Fourier-transformed data are used directly, with a linear interpolation between the points (data only, blue), or the result of the power law fit at high frequencies is extrapolated to low frequencies to provide a power law throughout the whole range of frequencies (fit only, green), or a composite version of the power spectrum is built using the raw data at low frequencies, and the fit result at high frequencies (data + fit, red). The error bars account for the observational noise as well as for the g_{21} uncertainty. The horizontal line marks the equipartition limit.

in the case of equipartition. However, the value $R > 0.72 \pm 0.09$ is still quite high. This can be the sign of a deviation in favour of solenoidal modes. At high Mach numbers, it would imply that an ordered solenoidal flow is superimposed on top of the turbulence. And Fig. 4 (Col. 2) shows that there is a large-scale differential motion in the cloud with the southern part receding, while the northern part is approaching. This velocity shift could be the sign of a large-scale rotation of the whole cloud, which could dominate the smaller-scale motions and be responsible for the large fraction of solenoidal modes.

The fact that the turbulence in the Orion B molecular cloud is, on large scales, mostly solenoidal, is in agreement with the fact that, for its mass, Orion B has a low star formation rate (Lada et al. 2010). Estimations of its SFE vary, with values ranging from 0.4% to 3% (Lada 1992; Carpenter 2000; Federrath & Klessen 2013; Megeath et al. 2016), but all studies show that the SFE in Orion B is about four times lower than in the neighbouring cloud Orion A. In general, Orion B's SFE is regarded as particularly low, with Megeath et al. (2016) stressing that is has the lowest SFE among all molecular clouds closer than 500 pc. This remarkable feature of the Orion B cloud could be partially explained by the solenoidal flows that drive its velocity field, and hinder collapsing motions that could trigger star formation.

In contrast, Fig. 8 shows a major deviation from the equipartition in favour of compressive motions in two specific regions. When zooming deeply into the star forming regions NGC 2023 and NGC 2024, we obtain solenoidal fractions as low as $R = 0.25$. The high fraction of compressive flow in these two regions most likely results from the infall of matter onto the star forming region, and/or the expansion of the HII regions around the young massive stars (Tremblin et al. 2014a; Geen et al. 2015). The HII regions themselves are not observed in molecular tracers, so that this expansion is detected indirectly through the compression of the molecular gas at the ionization front of the HII regions.

For both regions, R grows when zooming out, and will eventually reach the average values displayed on Fig. 7. R tends to decrease when reaching a field size of about 20′. This behaviour can be due to the fact that the other star forming region is entering the field of view, since the distance between the two cores is about 22′, but it can also be related to the geometry of each region.

In the case of NGC 2023, this size of 20′ also corresponds to the PDR of the Horsehead Nebula coming into the field and, owing to the pressure at the photo-dissociation front, it is expected to be a compressive region (Ward-Thompson et al. 2006), which is proven by the detection of at least one young star and one protostar in this region (Megeath et al. 2012).

In the case of NGC 2024, we see that R starts increasing again after 25′, even though NGC 2023 comes more and more into the field. The variations of R around 20′−25′ might therefore be related to the location of the edge of the HII region around

NGC 2024. This HII bubble exerts a pressure on the surrounding gas (Tremblin et al. 2014a,b), and this edge is therefore a highly compressive region that can be seen in the form of an arc north of NGC 2024 (Megeath et al. 2012). This region has a far lower surface density of stars that the inner part of NGC 2024, but it is likely to be younger that NGC 2024 (since it is a consequence of the expansion of the HII region), and therefore might have formed only very young protostars, poorly detected by *Spitzer* (Megeath et al. 2016), or no stars yet – but it might become a very active region in the future.

The sharp contrast between the large-scale solenoidal flow and the highly compressive flows in the NGC 2023 and NGC 2024 nebulae is in agreement with the spatial distribution of star formation observed in Orion B: Lada (1992) observes that the large-scale SFE in the Orion B cloud is an order of magnitude lower than in the most massive cores, Carpenter (2000) shows that, at his detection level, 100% of young stars in Orion B are located in clusters (NGC 2068 in the north, and NGC 2024 in the south), which is unusual among the studied clouds, and Lada et al. (2013) conclude that Orion B is very ineffective at forming stars at $A_K < 2.0$ mag, compared to other GMCs.

In a broader perspective, not only do these variations of R confirm observationally the intuitive link between compressive motions and star formation, as proposed in simulations (Federrath & Klessen 2012; Padoan et al. 2014), but they also show that there can be an intra-cloud variability of the solenoidal fraction, in addition to the inter-cloud one. This shows that the large-scale environment of the cloud, although it plays a major role in driving the turbulence of the molecular cloud, cannot alone explain the repartition of solenoidal and compressive motions in the cloud: any denser region created by the density fluctuation in the compressible turbulent gas (e.g., Nolan et al. 2015, and references therein) can lead, under the effect of self-gravity or stellar feedback, to the formation of very localized, strongly compressive regions, even in the context of a mostly solenoidal flow. There is therefore no universal solenoidal fraction that can be applied generally to all clouds, and there are even intrinsic variations of R from region to region within a cloud.

6. Conclusion

From a practical point of view, our work has shown that it is possible to apply the numerical method of Brunt & Federrath (2014) to observational data to determine the fraction of solenoidal and compressive motions in a molecular cloud, using molecular lines as tracers of the density and velocity fields. We were able to pinpoint the observational requirements to apply this method to a dataset.

The spatial dynamic range is an important element, mostly to provide good quality power spectra. The field must be large enough to provide good statistics at low spatial frequencies, but it must also have a good spatial resolution, so that the power spectra have enough points to correct properly for the beam and noise effects. We found our minimum field size to be of at least 50 independent pixels.

In addition to the spatial resolution, having many independent spectral channels is of great help when correcting for the beam and noise effects. The spectral resolution must also be sufficient to resolve the studied spectral line.

The S/N also proved to be a key element during the calculations. An average S/N of at least 5 is desirable: our datacube, which has a mean S/N of 7.8, yielded a relative observational uncertainty of 13% on the fraction of momentum in the solenoidal modes, and Figs. 7 and 8 show that this observational uncertainty contributes significantly to the overall uncertainty, and is even dominant at low solenoidal fractions.

The last point of the computation – the density-velocity correlation – which also significantly contributes to the overall uncertainty, requires us to use many spectral tracers of various typical densities. In that case, a spectral survey such as the one of the Orion B project is invaluable, in so far as all the needed tracers are available and inter-calibrated.

From a physical point of view, the measurements have shown that the motions in the Orion B molecular cloud are highly supersonic, with a mean Mach number of ~6. However, the Mach number maps show large variations, with some regions being only moderately supersonic. These variations are due both to the variations of the temperature and to the turbulent velocity distribution in the cloud.

The largest scales of the cloud seem to be dominated by a rotational motion, which can be identified by a high solenoidal fraction in the flow. At smaller scales, we have shown that the motion is largely dominated by compressive (infall and/or outflow) motions in the vicinity of the NGC 2023 and NGC 2024 star forming regions. The northern edge of NGC 2024 and the photo-dissociation front of the Horsehead nebula are also likely to be highly compressive regions, according to our results. This is in agreement with the observations of the star formation efficiency in Orion B, which is unusually low at the scale of the whole cloud, and exclusively concentrated in clusters (NGC 2023 and NGC 2024).

The example of Orion B also shows that the star formation efficiency in a molecular cloud does not only depend on its overall fraction of momentum in the solenoidal modes of turbulence, but also on the local variations of this fraction, which can be driven by internal phenomena such as self-gravity and stellar feedback.

This method could be applied in the future to study the variations of the solenoidal fraction between different molecular clouds, or between different regions or different chemical tracers within a given cloud. In the case of Orion B, we intend to analyse other data cubes for tracers such as $C^{18}O$ or HCO^+, which trace, respectively, more compact and more diffuse regions of the cloud, to probe different environments in terms of density and temperature.

Acknowledgements. We wish to thank Patrick Hennebelle, Edith Falgarone, and Pierre Lesaffre for fruitful conversations on hydrodynamics and magnetohydrodynamics, as well as our referee for enlightening remarks on compressible turbulence in the interstellar medium. This research also made use of data from the *Herschel* Gould Belt survey (HGBS) project (http://gouldbelt-herschel.cea.fr). The HGBS is a *Herschel* Key Programme jointly carried out by SPIRE Specialist Astronomy Group 3 (SAG 3), scientists of several institutes in the PACS Consortium (CEA Saclay, INAF-IFSI Rome and INAF-Arcetri, KU Leuven, MPIA Heidelberg), and scientists of the *Herschel* Science Center (HSC). This work was supported by the CNRS/CNES program Physique et Chimie du Milieu Interstellaire (PCMI). We thank the CIAS for its hospitality during the two workshops devoted to this project. V.V.G. thanks for support from the Chilean Government through the Becas Chile program. P.G.'s postdoctoral position was funded by the INSU/CNRS. P.G. thanks ERC starting grant (3DICE, grant agreement 336474) for funding during this work. NRAO is operated by Associated Universities Inc. under contract with the National Science Foundation.

References

André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
Brunt, C. M., & Federrath, C. 2014, MNRAS, 442, 1451
Brunt, C. M., Federrath, C., & Price, D. J. 2010, MNRAS, 403, 1507
Carpenter, J. M. 2000, AJ, 120, 3139
Cooley, J. W., & Tukey, J. W. 1964, Math. Comp., 19, 297
Dickman, R. L. 1978, ApJS, 37, 407

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton University Press)
Federrath, C. 2013, MNRAS, 436, 1245
Federrath, C., & Klessen, R. S. 2012, ApJ, 761, 156
Federrath, C., & Klessen, R. S. 2013, ApJ, 763, 51
Federrath, C., Klessen, R. S., & Schmidt, W. 2008, ApJ, 688, L79
Federrath, C., Klessen, R. S., & Schmidt, W. 2009, ApJ, 692, 364
Federrath, C., Roman-Duval, J., Klessen, R. S., Schmidt, W., & Mac Low, M.-M. 2010, A&A, 512, A81
Federrath, C., Chabrier, G., Schober, J., et al. 2011, Phys. Rev. Lett., 107, 114504
Geen, S., Hennebelle, P., Tremblin, P., & Rosdahl, J. 2015, MNRAS, 454, 4484
Goldsmith, P. F. 2001, ApJ, 557, 736
Gratier, P., Bron, E., Gerin, M., et al. 2017, A&A, 599, A100
Hennebelle, P., & Falgarone, E. 2012, A&ARv, 20, 55
Hily-Blant, P., Teyssier, D., Philipp, S., & Güsten, R. 2005, A&A, 440, 909
Hily-Blant, P., Falgarone, E., & Pety, J. 2008, A&A, 481, 367
Kim, C.-G., & Ostriker, E. C. 2015, ApJ, 802, 99
Kirk, H., Di Francesco, J., Johnstone, D., et al. 2016, ApJ, 817, 167
Kramer, C., Stutzki, J., & Winnewisser, G. 1996, A&A, 307, 915
Lada, E. A. 1992, ApJ, 393, L25
Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687
Lada, C. J., Lombardi, M., Roman-Zuniga, C., Forbrich, J., & Alves, J. F. 2013, ApJ, 778, 133
Larson, R. B. 1981, MNRAS, 194, 809
Liszt, H., & Pety, J. 2016, ApJ, 823, 124
Lomax, O., Whitworth, A. P., & Hubber, D. A. 2015, MNRAS, 449, 662
Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, A&A, 566, A45
Mac Low, M.-M. 1999, ApJ, 524, 169
Mangum, J. G., & Shirley, Y. L. 2015, PASP, 127, 266
Martin, P. G., Blagrave, K. P. M., Lockman, F. J., et al. 2015, ApJ, 809, 153
Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, AJ, 144, 192
Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, AJ, 151, 5
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Miville-Deschênes, M.-A., Duc, P.-A., Marleau, F., et al. 2016, A&A, 593, A4
Nolan, C. A., Federrath, C., & Sutherland, R. S. 2015, MNRAS, 451, 1380
Padoan, P., Federrath, C., Chabrier, G., et al. 2014, in Protostars and Planets VI, 77
Pety, J., & Falgarone, É. 2000, A&A, 356, 279
Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98
Renaud, F., Bournaud, F., Kraljic, K., & Duc, P.-A. 2014, MNRAS, 442, L33
Ripple, F., Heyer, M. H., Gutermuth, R., Snell, R. L., & Brunt, C. M. 2013, MNRAS, 431, 1296
Rohlfs, K., & Wilson, T. L. 2004, Tools of radio astronomy (Berlin: Springer)
Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, ApJ, 786, 29
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Tremblin, P., Anderson, L. D., Didelon, P., et al. 2014a, A&A, 568, A4
Tremblin, P., Schneider, N., Minier, V., et al. 2014b, A&A, 564, A106
Ward-Thompson, D., Nutter, D., Bontemps, S., Whitworth, A., & Attwood, R. 2006, MNRAS, 369, 1201
Wilson, T. L., & Rood, R. 1994, ARA&A, 32, 191

Appendix A: Mach number estimation

In this appendix, we describe the computations that yielded the Mach number maps that enable us to estimate if the studied regions of the molecular clouds were below or above the hypersonic limit, which is characterized by the equipartition of momentum in the solenoidal and compressive modes of turbulence. We first compute the speed of sound using two different temperature maps, then derive the speed of the flow from the ^{13}CO(J = 1−0) position-position-velocity cube, and finally obtain statistics on the 3-dimensional Mach number.

A.1. Speed of sound

The speed of sound in the gas needs to be derived from a temperature map, by means of $c_{\rm sound} = \sqrt{\gamma \cdot R_{\rm gas} \cdot T / m_{\rm mol}}$, where γ is the adiabatic index, $R_{\rm gas}$ the universal gas constant and $m_{\rm mol}$ the molar mass of the gas.

We have direct access to two temperature maps: the dust temperature, computed by the *Herschel* Gould Belt Survey consortium (Schneider et al. 2013), and the peak temperature of the ^{12}CO map, from our IRAM-30 m observations. The dust temperature is expected to be a lower bound for the kinetic temperature, because dust radiates more efficiently than gas. The gas and dust become coupled only for densities larger than 10^4 cm^{-3} (Goldsmith 2001).

For optically thick gas, the excitation temperature $T_{\rm exc}$ is determined (e.g., in Rohlfs & Wilson 2004) by

$$\exp\left(\frac{h\nu}{k_{\rm B}T_{\rm exc}}\right) - 1 = \left(\frac{T_{\rm peak}}{h\nu/k_{\rm B}} + \frac{1}{\exp\left(\frac{h\nu/k_{\rm B}}{T_{\rm CMB}}\right) - 1}\right)^{-1} \quad \text{(A.1)}$$

where ν = 115.271202 GHz is the frequency of the ^{12}CO(J = 1−0) transition, and $T_{\rm CMB}$ = 2.728 K is the cosmic microwave background temperature. We assume that the excitation temperature is close to the kinetic temperature of the gas (local thermodynamic equilibrium).

We assume that at very low column density (and therefore low $T_{\rm peak}$), the kinetic temperature is underestimated (since the computation is only valid for optically thick gas). The dust temperature is supposed to be a lower limit close to the kinetic temperature of the gas, except when it is not deemed reliable any more: above a threshold of 60 K, we deem that *Herschel* and *Planck* observations do not yield the best temperature estimate due to their limited wavelength ranges (50−600 μm for *Herschel*). At these temperatures, ^{12}CO(J = 1−0) is usually saturated enough to give a good result for the excitation temperature, as can be seen from the ^{12}CO(J = 1−0)/^{13}CO(J = 1−0) line ratio which diverges significantly from the ^{12}C/^{13}C abundance ratio (Pety et al. 2017; Ripple et al. 2013). We can therefore construct a third temperature map $T_{\rm max}$ using the dust temperature and the excitation temperature of ^{12}CO(J = 1−0): we use the excitation temperature whenever it is above the 60 K threshold, and in other regions we use the maximum of the gas excitation temperature and the dust temperature (Fig. A.1, panels 1 to 3).

When computing the speed of sound from the temperature map, the nature of the gas comes into play. We considered a 75−25% mixture in mass of molecular hydrogen and helium, which yields a molecular mass of 2.513 kg mol^{-1}. Since we are not simply dealing with a mono-atomic or diatomic ideal gas, the adiabatic index γ of this mixture has to be determined. To compute γ, we used tabulated NIST values of the calorific capacities C_P and C_V of the two gases at an average temperature of 24 K. The resulting value, γ = 1.66674, is very close to the value 5/3 that would be expected for a mono-atomic ideal gas, which could be expected since, at such low temperatures, the ro-vibrational modes of H_2 are frozen.

A.2. Flow velocity

The turbulent velocity dispersion was computed using the velocity dispersion along each line of sight, which can be derived from the W_2 map. We determined the flow velocity dispersion as $u = \sqrt{W_2/W_0}$ (Fig. A.1, last panel).

This computation implies two assumptions. First, we assume that the natural and thermal widths of the lines are negligible, compared to the turbulent broadening. Second, we neglect the opacity broadening as well. While it is correct that thermal broadening has less than 1% effect on the line width, the assumption for the opacity broadening is less obvious. We determined that the expected correction would be of the order of 10% for the brightest lines of sight, reducing the width of the lines and therefore the Mach number. However, given that the correction would have been difficult to implement for non-Gaussian lines, as is the case in our ^{13}CO(J = 1−0) field, and that the correction would be significant only for a small fraction of the lines of sight, we left the turbulent velocity field uncorrected.

A.3. Results

The ratio of the turbulent velocity to the sound speed gives us the Mach number along the z axis. The total Mach number is $M = \sqrt{3}M_z$, when the turbulence is isotropic. The Mach number is determined twice, using two maps: the excitation (gas) temperature, and the maximum of dust and excitation temperatures, as described above. We then compare the results of the two computations (Figs. A.2 and A.3).

To estimate the typical Mach number at the scale of the whole cloud, we plot the histogram of the obtained maps (Fig. A.3). The shapes of the histograms are very similar for $T_{\rm max}$ and $T_{\rm exc}$. Owing to the shape of both distributions, with a large tail at high Mach number, the most probable value of the Mach number for both distributions (M = 3.5), is significantly smaller than the mean and the median values, as shown in Table A.1.

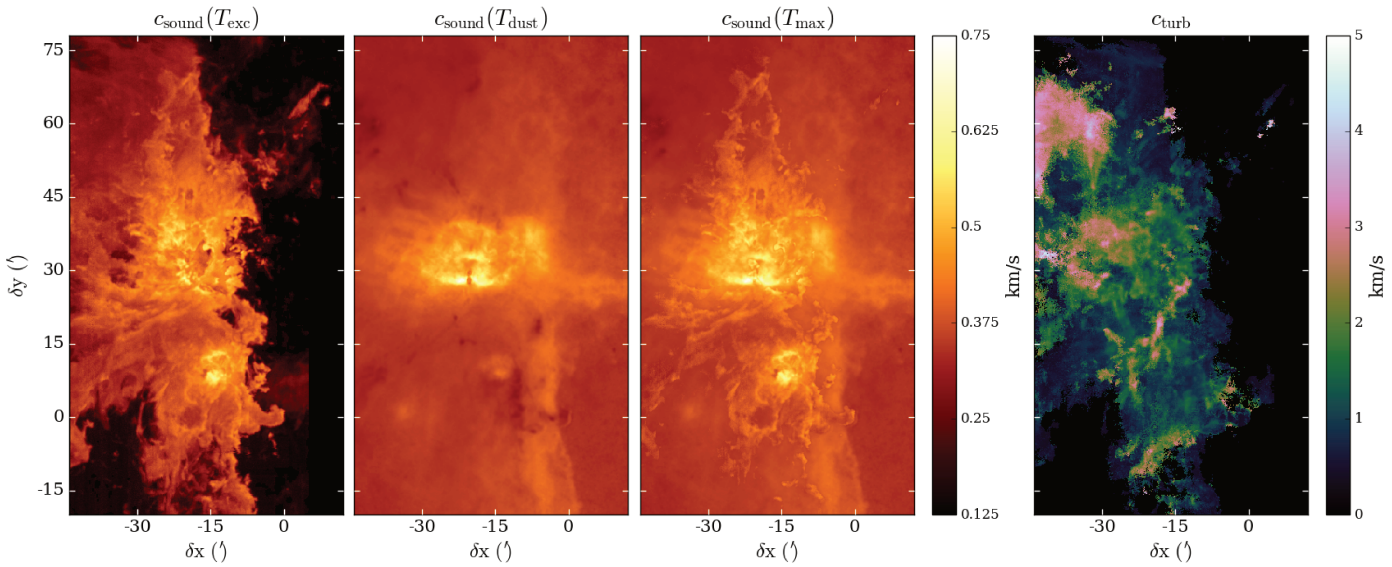


Fig. A.1. *First three panels*: speed of sound computed for the estimated excitation temperature of the gas, the dust temperature, and the maximum of the two previous ones. *Last panel*: turbulent flow velocity dispersion. We assume that having one or several spectral components along the line of sight is irrelevant, since one component gives a turbulent velocity dispersion, and two components show a high-velocity shock.

Table A.1. Typical values of the temperature and the Mach number distribution, for the two computed temperature maps.

Value	$T_{\rm max}$	$T_{\rm exc}$	$M_{\rm max}$	$M_{\rm exc}$
Most probable	21.1	22.3	3.5	3.5
Mean	25.5	15.8	6.1	6.5
Median	23.4	15.6	4.8	5.0

Notes. The $T_{\rm exc}$ distribution is truncated below 10 K, to avoid being biased by the large number of pixels at low $T_{\rm peak}$ in ^{12}CO.

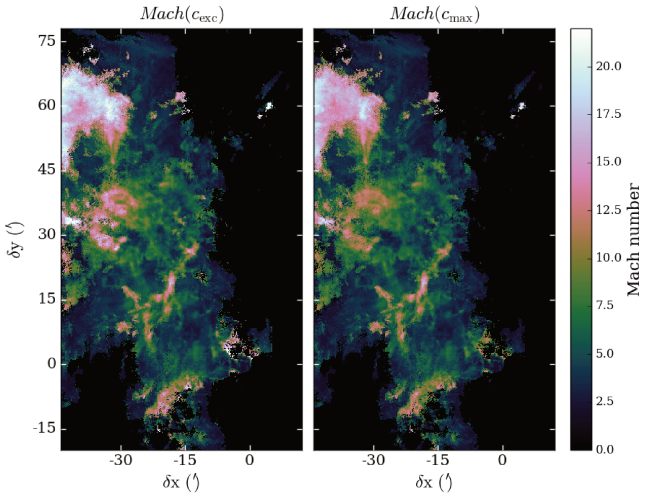


Fig. A.2. Mach number maps computed using the two different temperature maps, *left* using $T_{\rm exc}$, *right* using $T_{\rm max}$.

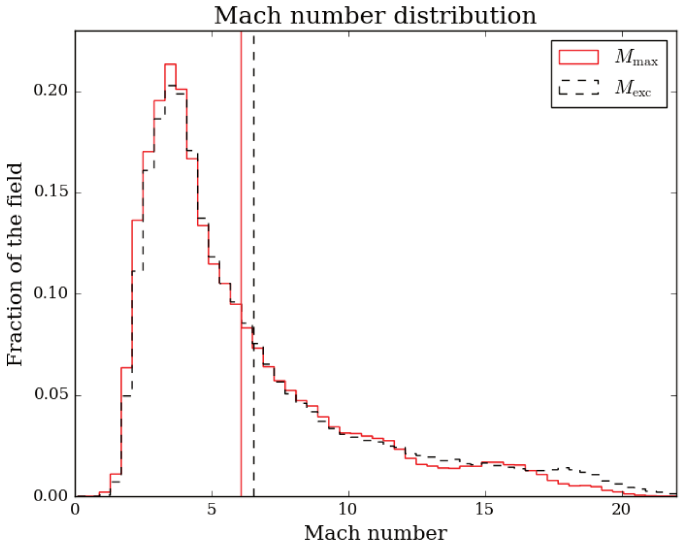


Fig. A.3. Histogram of the Mach number map, for $T_{\rm max}$ ant $T_{\rm exc}$. The vertical bars show the position of the mean value of each distribution.

3.4 Comments and prospectives

The main results of the analysis of the proportion of solenoidal and compressive modes in the momentum density field of Orion B point at the unusually low star formation activity in this Giant Molecular Cloud. This average property is consistent with its particularly low proportion of dense cores traced by N_2H^+ (Sect. 2.2.1), as well as with the average dynamical properties of the filaments present in the cloud (Sect. 4.2). Conversely, the star forming regions NGC 2024 and NGC 2023 feature highly compressive motions, a strong, localized N_2H^+ emission and gravitationally unstable filaments. Whether the low SFE in Orion B is an individual feature of this GMC or if it is the sign of a "young" cloud (i.e., if star formation in Orion B can be expected to become more active over time) is unclear. The detailed analysis of a large number of molecular clouds would be needed to determine if there is a clear trend of increasing SFE with the age of the cloud – which would point at the importance of triggered star formation.

While the focus of this chapter was on turbulence only, it is important to remember that it is far from being the only process in play that defines the PPV structure of a GMC: one of main questions in studies of molecular clouds is the relative contribution of various processes (self-gravity, turbulence, galactic influence, stellar feedback) in shaping a cloud and defining its velocity field. Many studies of ISM turbulence focus on simulations that do not aim at reproducing accurately what is observed in molecular clouds, but rather favour an idealized approach, where turbulence is typically driven in the Fourier space in a periodic box, without gravity, radiation fields, etc. It was in particular the approach that Brunt & Federrath (2014) had when devising the method applied in Orkisz et al. (2017). However, the results show that we are in a situation where the motions of the cloud cannot be described as idealized turbulence.

As discussed in Sect. 3.3.1 and 3.3.2, at hypersonic Mach numbers, the ratio between solenoidal and compressive modes in well-mixed, fully developed turbulence is expected to be 2/3 – 1/3. In other words, if the ratio between modes is different, the turbulence is not fully developed or well-mixed. At moderately supersonic Mach numbers, it might be because instabilities do not develop quickly enough to evenly distribute the motions imposed by the continuous driving between the turbulence modes. At hypersonic Mach number, it implies that at least part of the cloud's motions are *ordered* motions of some kind rather than "pure" turbulence.

It is important to note that the Helmholtz decomposition theorem, which is the backbone of the mode analysis presented here, can be applied to any vector field, not only to a purely turbulent velocity (or momentum density) field. For example, the velocity field of a steadily expanding bubble would only contain a compressive component, and that of a solid-body rotation only a solenoidal one. The results of the analysis should therefore be seen in a broader frame, as as a way to characterize which modes dominate the *motions* of the cloud, rather that the turbulence of the cloud. Being able to identify deviation from the well-mixed, ideal turbulence in molecular clouds thus provides a tool to discriminate between the influence of different processes on molecular clouds. In the case of Orion B, for example, one can conclude that if solenoidal motions dominate on large scales in a hypersonic regime, it is possible that Galactic shear is a dominant influence, since it modifies the signature of turbulence, and gravitational collapse would have led to an excess of compressive motions.

On the other hand, the isotropy assumption of the Brunt & Federrath (2014) method might not hold so well in the case of ordered motions, since e.g. an ordered rotation around a single axis yields a strongly anisotropic velocity field. The method therefore requires farther benchmarking using simulations, not only with mixture of modes in idealized turbulence, but also in the context of a superposition of various compressive or solenoidal ordered motions.

In general, the comparison of the statistics of the density and velocity (or momentum density) fields in simulations and observations is the main path that will lead to a better understanding of the relative contributions of the different processes that shape Giant Molecular Clouds. Nowadays, detailed and "realistic" simulations with very different sets of parameters are able to reproduce fairly well a broad range of observables, such as the cloud masses, density distributions and typical SFR. The need is for observables that will be able to discriminate between the different models, and the measure of the contributions of solenoidal and compressive modes in the velocity, or momentum density field is a good

candidate for that. Of course, the results of the Brunt & Federrath (2014) method also need to be compared with other approaches, such as the characterization of the column density PDF (Kainulainen et al. 2013a; Alves et al. 2017), the "velocity-channel analysis" (VCA, Kandel et al. 2016) or better estimations of the statistics of the volume density distribution than the first approximations made in Pety et al. (2017) and Bron et al. (2018).

From the practical, observational point of view, the application of the Brunt & Federrath (2014) method still requires refinement. Indeed, using the PPV cube of a molecular tracer gives access to the projection of the momentum density field only to first order, as it is limited by the properties of the tracer: its chemical network and therefore its presence in such or such environment of the cloud, its opacity, the sensitivity of the observations... All these elements bias the observations, and mean that the PPV cube is not simply a projection of the gas motions. A way to overcome these limitations could be to try and reconstruct as precisely as possible an accurate PPV cube that would describe the motions of the gas rather that the spectra of a molecule. To achieve this, multiple molecular tracers giving access to different column density regimes (Pety et al. 2017; Bron et al. 2018), coupled with estimations of the line-of-sight structure of the temperature (Marsh et al. 2015, 2017) could give a better view than a single tracer with a single temperature map. The use of measurements of the distances to ISM structures thanks to extinction tomography (Capitanio et al. 2017; Lallement et al. 2018) would also help constrain the three-dimensional structure of the cloud and of its environment – which is needed in particular to confirm that all velocity components arise from the same physical region.

Chapter 4

Filaments in the Orion B molecular cloud

4.1 Filaments in Giant Molecular Clouds

4.1.1 Description of filaments

Any type of elongated structure in astronomy tends to be called a filament, even though the term can designate very different physical objects. Such filaments, regardless on their nature, size and environment, have always drawn interest, as their shape catches the eye and makes them easily recognizable.

Apart from the dark filaments visible on the surface of the Sun, the first mention of interstellar filaments dates back to Edgeworth (1946), who proposed very early on a pioneering theory of star formation occurring in dense filaments of gas, which form as a necessary step during the contraction of clouds of interstellar gas (see Sect. 4.1.2 and 4.1.3). The first observational reference to "filaments" is owed to Gaze & Shajn (1952) to describe two nebulae in Cygnus.

The last decades have seen growing interest in the dense filaments present in GMCs. Several filamentary regions have drawn specific interest early on. The Taurus molecular cloud (Schneider & Elmegreen 1979; Gaida et al. 1984; Heyer et al. 1987; Hartmann 2002; Goldsmith et al. 2008) is a prime example, but one can also cite the northern part of the Orion A Giant Molecular Cloud (Bally et al. 1987; Wiseman & Ho 1998; Johnstone & Bally 1999), the Perseus region (Hatchell et al. 2005), etc. Filaments are also observed in translucent molecular clouds, such as the Polaris Flare (Falgarone et al. 1998; Miville-Deschênes et al. 2010). At even lower densities, elongated condensations in the diffuse, atomic ISM have also been referred to as filaments (or cirrus) (Low et al. 1984; Gibson et al. 2000; Peratt 2015), suggesting that there can be some extent of universality in the formation of such structures in the interstellar gas over a broad range of physical conditions.

More recently, attention has been drawn towards Giant Molecular Filaments (GMF), which are actually extremely elongated Giant Molecular Clouds, reaching several tens (if not hundreds) of parsecs in length. The 80-pc long "Nessie" nebula is an archetypical GMF (Jackson et al. 2010). Such structures have also been observed in nearby galaxies, in the form of "spurs" or "feathers" of gas exiting spiral arms (Koda et al. 2009; Schinnerer et al. 2017). Upon entering the inter-arm region of galaxies, these pressure-confined structures, which initially have moderate aspect ratios, get stretched into giant filaments by the shear (differential rotation) of the gas (Duarte-Cabral & Dobbs 2017). The term "filament" is also sometimes even applied to elongated features of atomic or molecular gas seen inside or around external galaxies, with scales of the order of $\gtrsim 10^3$ pc (e.g., Leroy et al. 2009; Fukui & Kawamura 2010). At the other end of the spectrum in terms of sizes, descriptive terms such at wisps, fibres or, again, filaments, can be used to refer to very small ($\sim 10^{-3}$ pc) elongated dense structures detected in the dense ISM by high angular resolution optical imaging (Fig. 4.1).

Of all these different filamentary objects, it is the dense molecular filaments in molecular clouds that this chapter focuses on. They have received most attention due to their systematic presence in and around star-forming regions, suggesting that they play a key role in channelling matter from molecular cloud scales (up to several tens of parsecs) down to the scales at which star formation occurs (a few thousands

Figure 4.1: Optical view of the dark filament of the Flame Nebula (NGC 2024). The background light provided by massive young stars and the HII region provides a detailed view at wavelengths at which filaments are usually obscured. The entire field covers about 1.7×1.3 pc in size, the "main" filament is ~ 0.2 pc wide, but the smallest visible elongated sub-structures have widths in the $3 - 6$ mpc range (image by Adam Block, Mt. Lemmon SkyCenter, U. Arizona).

of AU, i.e., about a milliparsec). Imaging surveys of the *Herschel* space observatory have revealed that filaments are not only widespread, but actually ubiquitous in molecular clouds and star forming regions (André et al. 2010, 2014). Several such molecular filaments can be easily seen in Orion B (e.g., Fig. 2.4, left). Filaments are not only ubiquitous *among* molecular clouds, but also *within* molecular clouds, which rarely contain a single filament or filamentary region. The filaments can be found in the inner part of molecular clouds (e.g., the Integral-Shaped Filament in Orion A, or the filaments of Orion B) or just as well lie at the outskirts of the cloud (notably the L 1495/B 213 complex in Taurus).

The aspect and dimensions of filaments in molecular clouds are very diverse, as there is no general agreement on what types of structures should be referred to as molecular filaments – the only constants being that filaments are elongated and denser that their environment. They can be quite uniform and rectilinear, like the Musca filament (Fig. 4.2, left), the Hummingbird filament in Orion B (Sect. 4.2, or Orkisz et al. 2018, Fig. 8), or very curved and sub-structured, like the Snake filament in Ophiuchus (Fig. 1.1, right) or the Integral-Shaped Filament in Orion A (Fig. 4.2, right). They can reach several parsecs in length (e.g., the Integral-Shaped Filament) or be less than a parsec long (e.g., Fig. 4.4, left).

The width of filaments is a debated issue (see Sect. 4.1.4). Arzoumanian et al. (2011) have brought up the often mentioned value of 0.1 pc as a "typical" width. Observations of filaments yield a wide range of mean or median values, from 0.03 to 0.4 pc (Henshaw et al. 2017; Panopoulou et al. 2014), with most values found in the 0.06 – 0.25 pc range [1], often with a significant scatter in each study. The same goes with simulations, where filament widths ranging from 0.04 to 0.35 pc are reported, depending on the simulation setup, the measurement method, again with significant scatter (Smith et al. 2014b; Gómez & Vázquez-Semadeni 2014; Kirk et al. 2015; Federrath 2016; Chira et al. 2018).

The mass and the linear density of filaments strongly depend on the definition of their width, since this affects how much matter is attributed to the filament and to its environment. One way to overcome this width issue was used by Stutz & Gould (2016), who provide a diameter-dependent average linear

[1] By order of increasing reported filament width, one can for example cite André et al. (2010); Hacar & Tafalla (2011); Salji et al. (2015); Hatchell et al. (2005); Arzoumanian et al. (2011); Palmeirim et al. (2013); Könyves et al. (2015); Peretto et al. (2014); Hartmann (2002); Juvela et al. (2012); Hennemann et al. (2012).

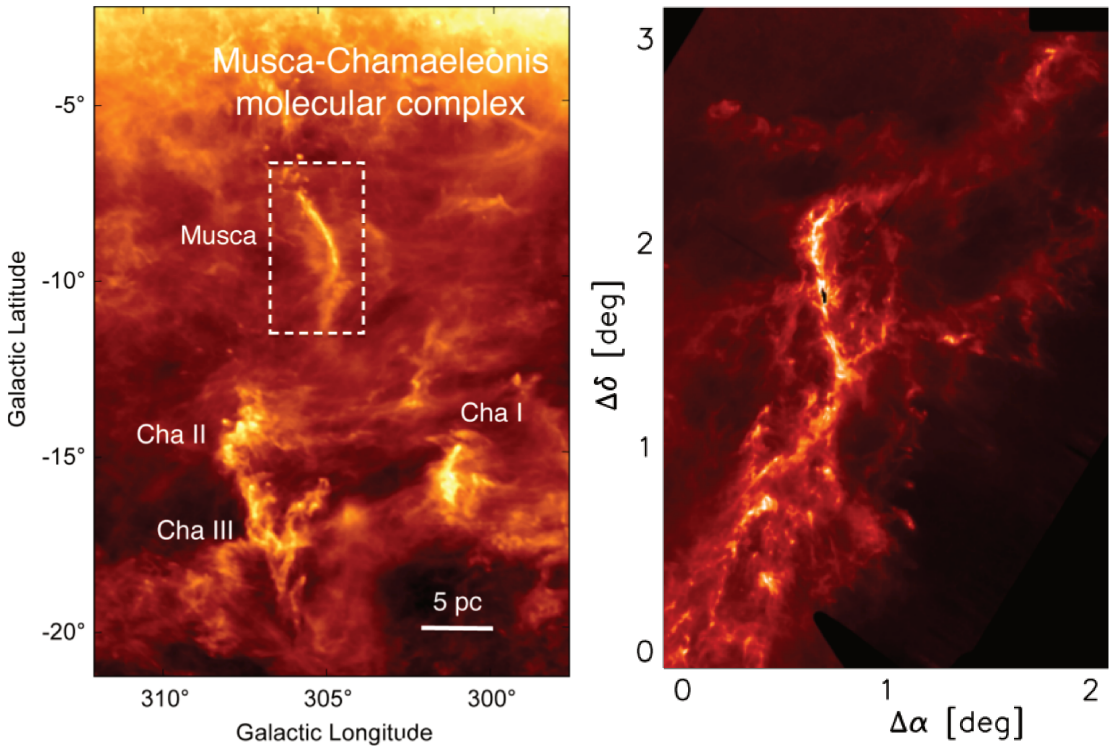


Figure 4.2: *Left:* The very rectilinear, isolated, 6 pc-long Musca filament, as seen at 857 μm by the *Planck* telescope (figure from Hacar et al. 2016). *Right:* Column density map of the 8 pc-long Integral-Shaped Filament in the northern part of Orion A, showing a complex structure around and inside it (figure in relative coordinates, adapted from Stutz & Kainulainen 2015).

density for the Integral-Shaped Filament of $385 \times (d/\mathrm{pc})^{3/8}\ M_\odot\,\mathrm{pc}^{-1}$. Given the order of magnitude of filament widths, this corresponds to a linear density in the $130 - 240\ M_\odot\,\mathrm{pc}^{-1}$ range, and a total mass of the filament of the order of 1000 to 2000 $M_\odot$. The Integral-Shaped Filament is however recognized for being a high-mass filament. The commonly studied molecular filaments have often lower linear densities, of the order of a few $M_\odot$/pc to several tens of $M_\odot$/pc (Arzoumanian et al. 2011; Schisano et al. 2014; Panopoulou et al. 2014). Their total masses vary significantly, given the diversity in filament length, and reach from $\sim$ 1 to a several hundreds of solar masses (Hacar & Tafalla 2011; Busquet et al. 2013; Kainulainen et al. 2016).

Less dependent on the definition of the width are the central column density and volume density of the filaments. Observed central column densities are quite consistently reported as lying between a few 10^{21} and a few $10^{22}\,\mathrm{cm}^{-2}$ (Wiseman & Ho 1998; Goldsmith et al. 2008; Juvela et al. 2012; Williams et al. 2018, etc.). Similar values are obtained in simulations (Myers 2011; Chira et al. 2018). Exceptionally, column density can reach as low as a few $10^{20}\,\mathrm{cm}^{-2}$ (in Polaris Flare, Arzoumanian et al. 2011) or higher than $10^{23}\,\mathrm{cm}^{-2}$ (in Cygnus X, Hennemann et al. 2012). Despite the strong dependency on fore- and background contributions and assumed filament geometry, the derived volume density values are also quite consistent, between a few 10^4 and a few $10^5\,\mathrm{cm}^{-3}$ (Wiseman & Ho 1998; Johnstone & Bally 1999; Hatchell et al. 2005; Peretto et al. 2012; Kirk et al. 2013; Kainulainen et al. 2016), again in good agreement with theoretical or numerical works (Fiege & Pudritz 2000a; Myers 2011; Kirk et al. 2015). With the notable exception of Arzoumanian et al. (2011), observational studies rarely tackle a broad distribution of filaments, but rather a small number of individual objects, so it is difficult to obtain a representative distribution of filament column and volume densities.

It is to be noted that these typical values of filament linear density, column density and volume density not only vary among filaments, but also within a filament – just like the filament width. Filaments are therefore far from being uniform cylinders (however computationally appealing this description might be)

but complex structures with shape and density variations reminiscent of the turbulent medium they are found in (Roy et al. 2015).

The first molecular line observations of filaments have revealed that they are quiescent structures, with low velocity dispersion (along the line of sight, observationally visible as narrow line-widths) and low velocity divergence (i.e., the centroid velocity throughout the entire filament does not vary much). The non-thermal velocity dispersion of filaments cannot be definitively attributed to a single phenomenon, such as rotation, collapse or turbulence, but it is in general close to the sound speed (Hatchell et al. 2005; Peretto et al. 2007; Hacar & Tafalla 2011; Tafalla & Hacar 2015). It seems that the velocity dispersion is lower in the inner part of filaments than in their envelope (Hatchell et al. 2005), it is also higher in massive, gravitationally unstable filaments (Arzoumanian et al. 2013, also Sect. 4.1.3). The filaments can have coherent centroid velocity on parsec scales, in disagreement with Larson (1981) scaling relations. It shows that filaments disconnect to some extent from the turbulent regime of the surrounding cloud (Goldsmith et al. 2008; Panopoulou et al. 2014; Tafalla & Hacar 2015), which is confirmed by simulations (Smith et al. 2014b, 2016).

Different types of large-scale ordered flows of gas in filaments have been observed in the molecular clouds and in numerical simulations. In addition to the gravitational radial contraction which can possibly explain the increased velocity dispersion of massive filaments (Arzoumanian et al. 2013), velocity gradients in the filament's environment also suggest the presence of accretion flows which channel surrounding gas onto the filament (Schneider et al. 2010; Kirk et al. 2015). This accretion flow can sometimes take the shape of striations, which are elongated dense structures perpendicular to the main filament (Palmeirim et al. 2013). It has also often been noted, for filaments that merge into a massive hub (e.g., Fig. 4.4, left), that gas flow *along* the filaments towards this central spot (Kirk et al. 2013; Peretto et al. 2013, 2014).

Filaments are also sub-structured. One of the first features ever noticed in molecular filaments was that they often contain dense clumps along their spine, hence the term of "globular filaments" used by Schneider & Elmegreen (1979). Chains of evenly spaced dense cores, compared to beads on a string, have been widely observed in various filamentary environments (Hartmann 2002; Busquet et al. 2013; Hacar et al. 2013). More recently, it has been shown that these condensations along filaments can follow a two-level hierarchical pattern, with widely spaced groups of closely spaced cores (Takahashi et al. 2013; Teixeira et al. 2016; Kainulainen et al. 2017). This phenomenon has also been observed and studied in numerical simulations (Clarke et al. 2016, 2017). Longitudinal fragments are not the only ones present, however, and it seems that radial fragments are present as well. Hacar et al. (2013) have shown that filaments can be made of so-called "fibres", which are narrower, subsonic, velocity-coherent structures (Fig. 4.3, top). Similar observations were also reported by Panopoulou et al. (2014) and Hacar et al. (2018). The presence of these fibres was also identified in simulations, in the form of intertwined ribbon-like "sub-filaments" which make up the filament (Smith et al. 2014b, 2016; Clarke et al. 2017, and Fig. 4.3, bottom), and the simulations confirm the subsonic and velocity coherent character of these structures.

4.1.2 On the origin of interstellar filaments

The basic process that is needed to achieve star formation in a molecular cloud is to bring enough relatively diffuse gas from a large volume into a very dense, small volume where eventually the pressure will become high enough to trigger nuclear reactions. One way to achieve this is to increase the density while reducing the number of dimensions, and this is where filaments step in. The ubiquity of filaments is a very good argument in favour of a general star formation mechanism in which the gas is gathered from diffuse 3D clouds into moderately dense 2D sheets, then into denser 1D filaments and eventually into very dense, "point-like" cores. This scenario is even more supported by observations that suggest that filaments are embedded in sheets of gas, rather than in a spherical or cylindrical medium (Kirk et al. 2013; Palmeirim et al. 2013; André 2017). The question of the origin of filaments thus boils down to understanding how molecular gas can condensate from three dimensions to two, and then to only one.

Four major categories of phenomena can be distinguished which can contribute to condensation

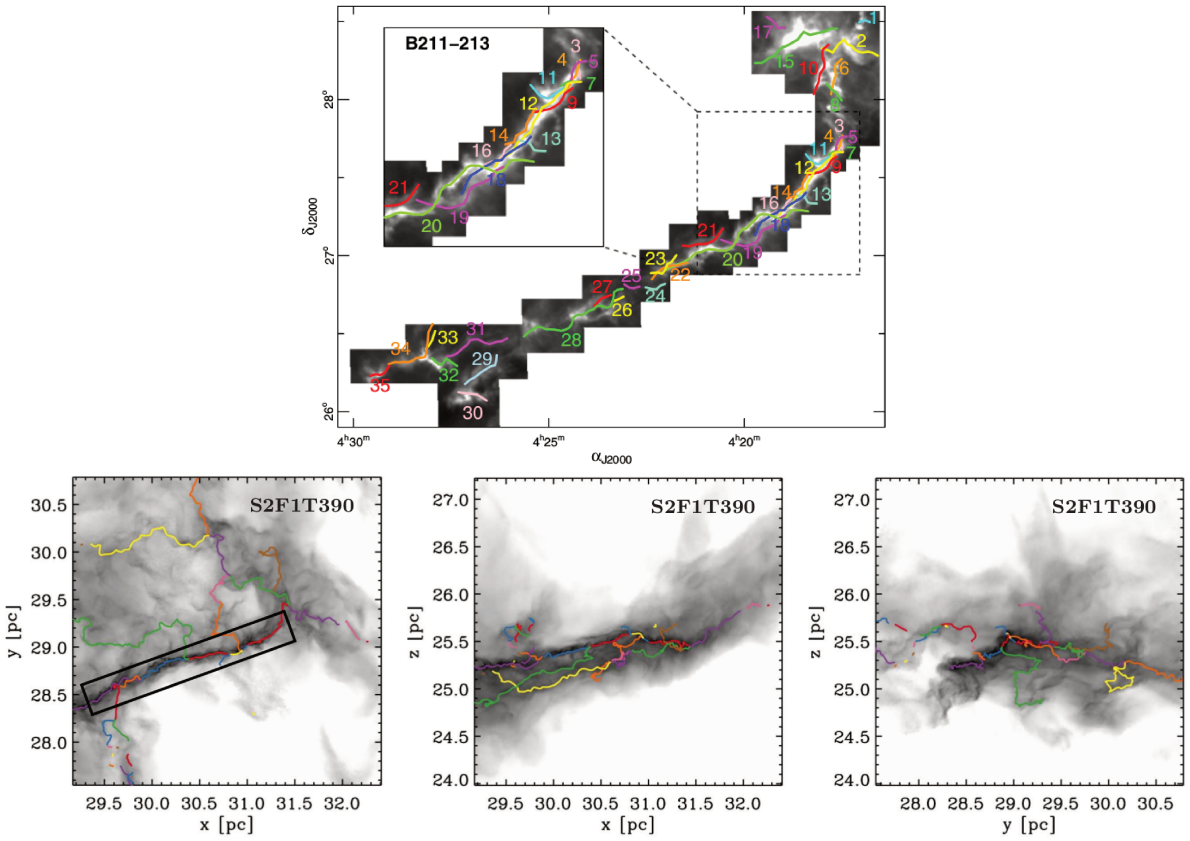


Figure 4.3: *Top:* Velocity-coherent fibres identified in the L1495/B213 Taurus filament using $C^{18}O$ ($J = 1 - 0$) line observations (figure from Hacar et al. 2013). *Bottom:* Three projections of the sub-filaments make up a filament obtained in a simulation. Depending on the viewing angle, the "main" filament visible in 2D can contain more or fewer superimposed sub-filaments (figure from Smith et al. 2016).

and dimensionality reduction: gravity, magnetic fields, turbulence (be it through its compressive or its solenoidal modes), and ordered compression in the form of ambient gas pressure, radiative pressure, expanding shells or winds[2].

Self-gravity tends to increase inhomogeneities, therefore any non-uniform and finite medium subjected to self-gravity only will collapse or fragment into structures of high aspect ratios along its shortest dimension. Large-scale pressure waves created by phenomena such as winds from OB associations, supernova explosions... sweep up the gas and gather it into dense sheets. Two such sheets can cross and thus create a one-dimensional dense structure such as a filament. Turbulence can partly act in the same way, by creating more random shocks which compress the matter into sheets, which then can cross and form filaments. Turbulence can also lead to the formation of elongated tubes of high vorticity, where turbulence dissipation is particularly efficient. Large-scale magnetic fields also have the ability to increase inhomogeneities: in the ideal MHD approximation, where magnetic fields are "frozen" into the gas, compression perpendicular to the field lines is impeded by magnetic pressure. Therefore, matter preferentially flows along the magnetic field lines. All these mechanisms can of course be present at the same time, and have different contributions depending on the environment and initial conditions. It is also not necessarily the same mechanism that is dominant in the formation of sheets and their evolution into filaments: almost any combination of the four mechanisms can *a priori* shape filaments out of a cloud, and in practice most of these combinations have been considered in studies of molecular filaments (Table 4.1).

[2]To some extent, pressure, winds, outflows etc. can be grouped together with compressive turbulence modes, the main difference being the large-scale coherence of these phenomena, both in spatial and time scales.

The purely gravitational point of view was pioneered by Edgeworth (1946). A blob-like cloud collapses fastest along its smallest dimension, leading to the formation of a self-gravitating sheet, which will then collapse into one or several filaments, by enhancing any inhomogeneity present in the sheet. In this picture of hierarchical gravitational fragmentation, filaments are not structures in equilibrium, but are the channels through which matter flows from sheets into dense cores – filaments are long lived because they accrete at a rate similar to the one at which their matter gets accreted onto cores (Gómez & Vázquez-Semadeni 2014).

The gravitational fragmentation of a (self-gravitating) sheet of gas into filaments is commonly mentioned as a likely origin of filaments, but other phenomena are often considered for the formation of the sheet. For example, Nakamura & Li (2008) study the gravitational fragmentation of a sheet that was flattened by flows along the magnetic field lines, and finds that the magnetic field slows down the fragmentation process, whereas Miyama et al. (1987); Myers (2009) or Peretto et al. (2012) propose that gravitational evolution occurs after the gas has been brought to two dimensions by external pressure. Hartmann (2002) rather sees the flattening as being generated by the largest scales of turbulence. Note that both Hartmann (2002) and Peretto et al. (2012) suggest that the direction of the dimensionality reduction (alignment of the sheet or filaments, respectively) can also be due to magnetic fields.

Compressed shells or converging flows are rarely seen as capable of efficiently forming filaments by themselves, but this first compression step is a good candidate to initiate the fragmentation into filaments, either through self-gravity, or through turbulent fragmentation (Hennebelle et al. 2008) or magneto-hydrodynamic instabilities. In particular, Nagai et al. (1998) show that a compressed sheet of gas threaded by a magnetic field can fragment into filaments which can be either perpendicular to the field (if the sheet is dense enough for gravitational contraction to occur, gas flows along the field lines and gathers into filaments) or parallel to it (if the sheet is not gravitationally bound, blobs of gas will dissipate, but do so along the field lines, yielding elongated structures parallel to the field).

Another popular scenario to form filaments goes through the formation of compressed sheets by shocks in a turbulent medium. If the transition from 2D to 1D is also attributed to turbulence, the paradigm is turbulent hierarchical fragmentation (Padoan et al. 2001; Klessen 2004), in which all the coherent spatial structures in the volume are created by the crossing of shocks generated by the largest wavelengths of the turbulent velocity field. In addition to self-gravity, magnetic fields can also play a major role in shaping shocks into filaments. In particular, Inoue & Fukui (2013); Inoue et al. (2018) propose a mechanism by which any dense clump of gas encountering a shock wave in the presence of a magnetic field gets squeezed in all directions except along the field lines, as the impact on denser medium deforms the shock front, and the oblique shock channels the matter into the convex point of the deformed shock wave. Hennebelle (2013) also propose a magneto-turbulent origin for non self-gravitating filaments, where the importance of the magnetic field is brought to light by the fact that non-magnetized turbulence tends to form dense sheets rather than filaments. The filaments in the the magnetized case are prevented from dispersing by Lorentz forces, and are mostly created by the turbulent strain, which tends to elongate any dense clump of gas. This elongation by the turbulent strain is reminiscent of the filamentary shape of the regions of highest intermittency, identified by She & Leveque (1994) and applied to the context of the turbulent ISM, under the name of "vorticity filaments" by Pety & Falgarone (2000). These highly dissipative structures may be responsible for the quiescence of filaments. Interestingly, vorticity filaments recently came back into light, as they were associated with the formation of fibres (or sub-filaments), which are particularly velocity coherent and generally subsonic (Clarke et al. 2017).

In summary, the most frequently mentioned (but neither consensual nor universal) mechanism for the formation of interstellar filaments is that a cloud is flattened by a compression wave (caused by turbulence, winds...) into a sheet, which then fragments into filaments under the effect of self-gravity, with magnetic field playing a more or less important role in enhancing the anisotropies at both stages of collapse. However, no mechanism of filament formation is either completely satisfactory and generalizable, nor can it be completely rejected. Depending on conditions such as density, gas temperature, ionization fraction, initial velocity dispersion... gravity, magnetic fields, turbulence and pressure will play roles of various importance. For example, Federrath (2016) have run simulations with and without gravity, turbulence,

magnetic fields and stellar feedback, and all of them yield filaments with properties reasonably close to observations.

2D → 1D

3D → 2D	Turbulence	Self-gravity	Magnetic field	Pressure, winds
Turb.	Pety & Falgarone (2000) Padoan et al. (2001) Smith et al. (2016)	Hartmann (2002)	Hennebelle (2013) Inoue & Fukui (2013)	Arzoumanian et al. (2013)
Gravity		Edgeworth (1946) Gómez & Vázquez-Semadeni (2014)	Busquet et al. (2013)	
B field		Nakamura & Li (2008)	Goldsmith et al. (2008) Palmeirim et al. (2013)	
Pressure	Hennebelle et al. (2008)	Miyama et al. (1987) Myers (2009) Peretto et al. (2012)	Nagai et al. (1998)	Schneider & Elmegreen (1979) Hacar et al. (2013)

Table 4.1: A summary of the proposed mechanisms that shape molecular gas into filaments. It is only a very simplified view of the topic, as some theories propose directly the formation of 1D objects, or an interplay of several phenomena. The fact that two papers are grouped into the same cell does not imply that they contain identical, or even equivalent theories.

The processes that form filaments also have an impact on their internal properties, and in particular their width and their sub-structures. The issue of width strongly depends on the filament formation scenario. In the case of purely gravitational hierarchical fragmentation, the filament width is set by the balance between the radial pressure gradient inside the filament and the thermal and ram pressure of the accreting gas (Gómez & Vázquez-Semadeni 2014). In MHD (gravo-)turbulent scenarios, however, the filament width is rather set by the scale at which turbulence decays. Arzoumanian et al. (2011) propose that the 0.1 pc filament width roughly corresponds to the sonic scale (the scale at which the turbulent velocity dispersion equals the sound speed), which is confirmed in simulations by Federrath (2016). The fact that this width is constant notwithstanding variation in filament linear densities is then explained by Arzoumanian et al. (2013) as resulting from a balance between accretion and contraction. Hennebelle (2013) suggests that the filament width corresponds in simulations to the turbulence dissipation scale, and compute that the order of magnitude of the dissipation scale induced by ion-neutral fraction in dense gas matches the observed filament widths[3]. Smith et al. (2014b), on the other hand, notices in simulations that the dissipation scale corresponds to smaller objects, namely fibres or sub-filaments, which are then gathered together into larger, thicker filaments.

4.1.3 Filaments and star formation

The previous section discussed the many ways which can bring the molecular gas from three dimensions of a cloud into a roughly one-dimensional filament. This section focuses on the less controversial last stage of condensation, which means creating "point-like" dense cores out of filaments. Geometrically, it can happen in two ways: point-like objects can be fragments of 2D objects, or they can lie at the crossing of two (or more) of these 2D objects. In practice, both situations are observed in the ISM: dense cores can be the result of filament fragmentation, or form in hubs of merging filaments (Fig. 4.4).

Following the observation of globular filaments by Schneider & Elmegreen (1979), studies of the stability of filaments against self-gravity were conducted by Larson (1985), and later by Inutsuka & Miyama (1992, 1997), showing that fragmentation into dense cores was indeed expected. These cores would then lead to star formation, thus setting the Initial Mass Function (IMF), and would be spaced by 4 times the filament's diameter. Fiege & Pudritz (2000b) show that magnetic field can somewhat stabilize

[3] Hennebelle (2013) also notes that this dissipation scale can decrease by orders of magnitude when the ionization fraction increases, e.g., at the outskirts of molecular clouds or near HII regions. This could explain the extremely narrow fibres observed in the NGC 2024 HII region in Fig. 4.1.

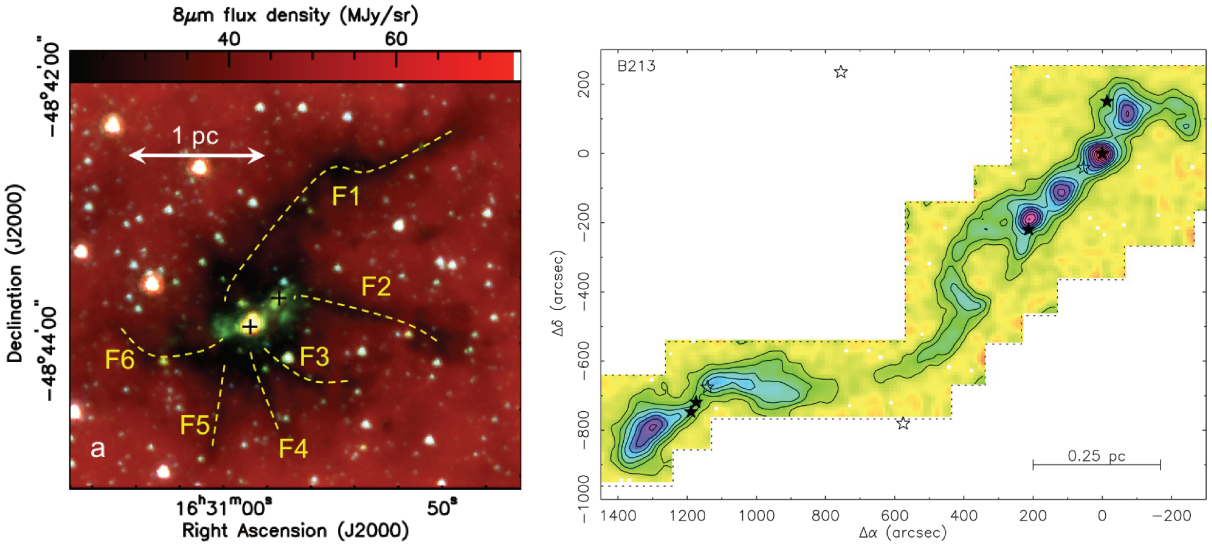


Figure 4.4: *Left:* Mid-infrared *Spitzer* image of the SDC335 dark cloud, which features 6 filaments, marked by dashed yellow lines, meeting in a central hub which contains two massive protostellar cores, marked by black crosses (figure from Peretto et al. 2013). *Right:* Chains of dense cores are visible in the hierarchically fragmenting filament B216 in Taurus. The coordinates are centred on ($04^h19^m42.5s$, $+27°13'36''$), and the star symbols locate the position of YSOs, the class 0/I ones being represented by black symbols (figure from Tafalla & Hacar 2015).

filaments against fragmentation, reducing the gravitational instability growth rate, but the "sausage instability" happens nonetheless, leading to the formation of prolate, rather than spherical, cores (Fiege & Pudritz 2000c; Hartmann 2002).

Since then, the link between filaments and dense cores (and star formation) has been widely confirmed, in particular thanks to *Herschel* observations. The observation of core-less filaments in a rather diffuse environment such as Polaris Flare suggests that core formation happens after filaments have formed, as a later stage of the star formation process. Molinari et al. (2010); Polychroni et al. (2013); André et al. (2014); Könyves et al. (2015) all show the association of prestellar or protostellar cores with dense filaments, with over 70% of the dense cores found in filaments, more specifically in the gravitationally unstable ones with a linear mass $> 16\ M_\odot/pc$. These cores found in filaments are also very likely to be the direct progenitors of stars, as they are mostly gravitationally bound, and their mass probability distribution function (the core mass function, CMF), is very reminiscent of the mass PDF of young stars (the initial mass function, IMF), albeit with core masses a factor ~ 3 higher than those of stars, suggesting mass loss during the star formation process (Fig. 4.5, left, André et al. 2014; Könyves et al. 2015). The picture of filament fragmentation into star-forming cores became more complex when it became visible that fragmentation happens in a hierarchical way, with groups of cores forming along the filaments – despite theoretical works predicting the presence of a single fragmentation scale (Inutsuka & Miyama 1997). Hacar et al. (2013) proposed that the fragmentation of filaments into cores in Taurus happens in successive stages, with the filament first fragmenting into 0.5-pc long, velocity-coherent fibres, which, if massive enough, fragment into cores. In the massive filaments OMC-1 and OMC-3, Takahashi et al. (2013) and Teixeira et al. (2016) also report hierarchical fragmentation, albeit on smaller scales, with tight groups of cores formed out of clumps with sizes and spacings of the order of 0.1–1 pc. Kainulainen et al. (2013b) observe two hierarchical modes of fragmentation in the high-mass ($600\ M_\odot$/pc) filamentary cloud G11.11-0.12. For separations above 0.5 pc, the fragmentation pattern match best a cylindrical instability (Ostriker 1964; Fischera & Martin 2012), but under 0.5 pc, the separations rather match a spherical fragmentation at Jean's length (Jeans 1902), suggesting that at this scales the fragment are less sensitive to the general, filamentary shape of the cloud (Fig. 4.5, right). In the Integral Shaped Filament, Kainulainen et al. (2017) also reveal two main modes of hierarchical fragmentation, which can correspond

to spherical and cylindrical instability modes respectively.

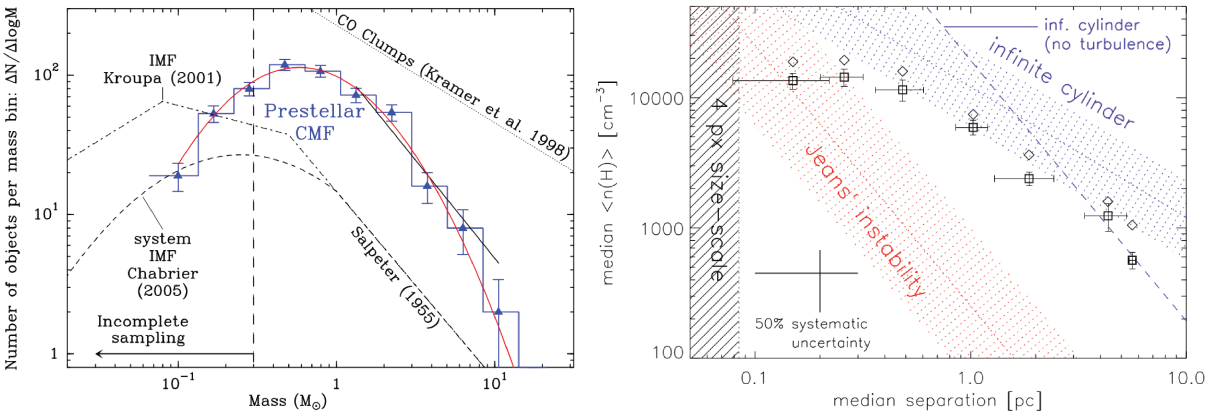


Figure 4.5: *Left:* Prestellar core mass function (CMF, in blue) obtained from *Herschel* observations of Gloud Belt molecular clouds. With its log-normal (red curve) shape extended by a power-law (black line), it is reminiscent of the initial mass function of stars (figure from André et al. 2014). *Right:* Column density and separation of significant structures in the G11.11-0.12 cloud, compared to the predictions of fragmentation length according to spherical (Jeans) or cylindrical gravitational instabilities, showing a transition in fragmentation mode at scales of about 0.5 pc (figure from Kainulainen et al. 2013b).

The growth of specific modes of longitudinal instability is not driven by gravity only. If it were the case, the core mass function would be very narrow (Roy et al. 2015), rather than having a broad peak and a power-law tail (Fig. 4.5) and would correspond to a fragmentation length of roughly 4 times the filament's diameter (Clarke et al. 2016). The instability modes are rather seeded by turbulence. One-dimensional turbulence along the filament (Roy et al. 2015) contributes to creating density fluctuations which can grow due to gravitational instability and create cores of different masses. Turbulence in the accretion flows that feed the filament (Clarke et al. 2016, 2017) can either lead to a two-level hierarchical fragmentation, if gravity dominates turbulence, or produce more arbitrarily spaced cores if turbulence is dominant. Together, these turbulent fluctuations of the core spacing and initial density can lead to a broader dispersion of core masses.

The most massive cores, however, are rather formed through the other star formation mechanism involving filaments, namely hubs (Fig.4.4, left). In that scenario, several filaments merge into a central dense clump, which often harbours one or several very massive protostellar cores. Such hubs, with masses reaching hundreds of solar masses, are potential progenitors of massive stellar clusters such as OB associations. Myers (2009) have noted the strong correlation between young stellar clusters of various masses and so-called hub-filament structures, with such examples as B 59 in the Pipe Nebula, the Serpens South cluster, the Ophiuchus complex of the OMC-1 cluster in Orion A. The general aspect of such structures can be explained by a central clump compressed into a sheet and submitted to self-gravity. Myers (2011) confirmed with more detailed models that the hub-filament configuration, with strings of dense cores in each filament, is the best geometry to reproduce the mass and spacing of stars in young clusters.

Observationally, the strongest result is the presence of velocity gradients along the filaments in such hubs, which suggests large scale flows of gas along the filaments towards the central hub. In the Serpens South cluster, Kirk et al. (2013) identify both accretion flows from the surrounding medium onto filaments, and flows along the filaments that feed the central hub. The accretion rate of the filaments, of the hub and the star formation rate in the hub are of comparable magnitudes. Similarly, Peretto et al. (2013) identified flows along the filaments merging in the SDC335.579-0.272 hub, providing enough material to form O-type star progenitors and possibly turn the dark cloud into an OB association. Peretto et al. (2014) obtain the same kind of results in SDC13, and notice that in addition to gradients, the velocity dispersion also increases along the filaments when nearing the central hub, which suggest that the gravitational

acceleration feeds gravity-driven turbulence (Arzoumanian et al. 2013). A more recent study of SDC13, by Williams et al. (2018), showed that, similarly to the model of Myers (2011), this hub-filament structure features both star formation mechanisms, with gas flowing along the filaments into the hub, but also dense cores forming within each filament. The increase of velocity dispersion around the cores correlates well with their mass (lower in filaments, larger in the hub), which also speak in favour of gravity-driven turbulence enabling the formation of super-Jeans starless cores of increasing masses.

In Orion B, the Flame filament in NGC 2024 (Fig. 4.1) is known to harbour many massive YSOs. Its shape is reminiscent of a hub-filament structure, but it is unclear if it is really one, in the sense of converging flows. The Flame might be a proper hub-filament structure, the short protrusions being the remainders of filaments evaporated or blown apart by the expanding HII region, or it might be a single, massive filament, disrupted and ragged by the influence of the HII region.

4.1.4 Tangled in filaments: a few issues

One of the first questions that arise when studying observations of filaments is checking if the observed objects actually correspond to filamentary objects in the 3D physical space. Two main caveats can be brought up. First, observed filaments could be sheet-like objects seen nearly edge on. Although it might be unlikely to see edge-on a large number of flat sheets that would be mistaken for filaments, it is on the other hand very likely that a undulating, rippled sheet locally appears edge-on. However, an edge-on sheet geometry results in very different volume densities for a given column density than a filament geometry. This in turn has an impact on chemistry and temperature. Observations of cold dense tracers such as N_2H^+ in many filaments have confirmed that the order of magnitude of the volume densities generally match better filamentary geometries than edge-on sheet ones. Second, long filaments could result from the superposition along the line of sight of several less elongated but roughly aligned objects. This question is even more relevant in the context where fibres are detected in observed and simulated filaments (Fig. 4.3). It appears from simulations that the sub-filaments identified in PPP space do not always match the fibres identified in PPV space in observations (Zamora-Avilés et al. 2017). The continuity of the fibres can therefore sometimes be a projection artefact, and the continuity and width of a filament can vary as a result of projection (Fig. 4.3, bottom, and Smith et al. 2016). In general, however, most sub-structures along the line of sight are physically related, and observed filaments are "real" objects (Smith et al. 2014b). However, these projection effects imply that the moderately supersonic character of some filaments might be related to the observational resolution and / or to their actual inner structure, comprising overlapping fibres or not (e.g., the two projections shown lower left and centre of Fig. 4.3 would result in different velocity dispersions due to the different amount of fibres mixed up by the line-of-sight projection).

Following on the subject of sub-structures in filaments, arises the question of the way these sub-structures form. Smith et al. (2014b) and Smith et al. (2016) propose a "fray and gather" scenario, in which fibres are formed first, as a result of turbulence dissipation, and these velocity-coherent, subsonic structures are then swept together either by gravitational forces or by the long-wavelength modes of turbulence. This "fray and gather" scenario is opposed to the "fray and fragment" scenario (Hacar et al. 2013; Tafalla & Hacar 2015; Clarke et al. 2017), in which filaments first form at the sonic scale, and only then fragment radially into fibres under the influence of gravity and vorticity. As a consequence, one scenario proposes that all filaments are necessarily sub-structured, whereas the other one suggests that filaments can be homogeneous and quiet, or fragmented and on the way to star formation. The formation of filaments is thus also intimately linked to their subsequent evolution and their ability to form stars.

Finally, one remaining issue concerns the filament width, or diameter. The ~ 0.1 pc width of filaments brought to light by Arzoumanian et al. (2011) was initially presented as a "typical" value, but it has since often be referred to as a "universal" or a "constant" width of filaments. It might be a bit exaggerated to call this width "constant" given the scatter of the observed filament widths (Sect. 4.1.1), but the distribution of these widths seems to be indeed peaked around this value of 0.1 pc. It is also not fully clear why this width should be universal in the processes that shape star-forming structures in molecular clouds. Part of the difficulties might arise from definition issues, given that there is no universal agreement on what is a "filament" and what is its "width". There is a variety of models used to describe the cross-section of

filaments, and all have different corresponding typical dimensions. For example, Kainulainen et al. (2016) note that depending on the assumed definition, the average width of the Musca filament can vary by a factor 3. Using simulation results, Smith et al. (2014b) also show variations of the filament widths by up to a factor 2, depending on the definition and the profile fitting method. They also point out biases that arise due to resolution limitation and degeneracies between fitting parameters. Panopoulou et al. (2017) also extensively discuss the biases that can arise from the fitting methods as well as from the statistical treatment of the results, which can both lead to an artificially narrow distribution of filament widths. In summary, given the lack of a firm definition and the reported scatter of measurements, which spans a factor ~ 4 above and below 0.1 pc, it is safe to say that molecular filaments have widths *of the order of* 0.1 pc (with variations both within and between filaments) in local molecular clouds. Further research at high angular resolution and in more distant GMC is needed to better constrain the distribution of filament widths and the origin of its shape.

4.2 Characterizing filaments in Orion B

In Orkisz et al. (2018), we have aimed at characterizing the properties of the entire filamentary network present in 1.9 square degrees of the south-western edge of Orion B. As we wanted to avoid detection biases as much as possible, we set quite lax requirement on our definition of filaments, and we used jointly two filament detection schemes, comparable to the ones used by Schisano et al. (2014) and Salji et al. (2015). This approach resulted in a very large sample (over 100 filaments) which we then have studied statistically.

The focus has been set on the gravitational properties of filaments: their density profiles (and thus their width), their linear density (and thus their stability against gravitational collapse), their velocity profiles. We have shown that the filaments in Orion B have widths of the order of $0.12\,\mathrm{pc} \pm 0.04\,\mathrm{pc}$, which is apparently independent from their masses or densities. The distributions of linear densities and volume densities are on average lower than what is generally observed. In conjunction with relatively warm temperatures (of the order of 25 K, compared to the often referenced 10 or 15 K) and moderately supersonic velocity dispersions, it results in filaments that are in general far from being gravitationally collapsing or even bound. The velocity field suggests that the turbulence is being dissipated in the inner parts of the filaments (although not enough to reach a transonic regime), and the filament that shows the clearest signs of longitudinal fragmentation is also gravitationally unbound and does not harbour a single protostar[4]. All these results, combined with the overall low numbers of YSOs in Orion B, speak in favour of a young evolutionary status of the GMC.

[4]One blue star (HD 37845, of type B9.5) happens to be associated with the fragmenting Hummingbird Filament. In the absence of dedicated distance measurements and of clear signs of interaction between the star and the filament, it is difficult to say if the two objects are physically related.

Astronomy & Astrophysics manuscript no. filaments-these
December 17, 2018

A dynamically young, gravitationally stable network of filaments in Orion B⋆

Jan H. Orkisz[1,2,3], Nicolas Peretto[4], Jérôme Pety[2,3], Maryvonne Gerin[3], François Levrier[3], Emeric Bron[5,6], Sébastien Bardeau[2], Javier R. Goicoechea[6], Pierre Gratier[7], Viviana V. Guzmán[8], Annie Hughes[9], David Languignon[5], Franck Le Petit[5], Harvey S. Liszt[10], Karin Öberg[11], Evelyne Roueff[5], Albrecht Sievers[12], and Pascal Tremblin[13]

[1] Univ. Grenoble Alpes, IRAM, F-38000 Grenoble, France
[2] IRAM, 300 rue de la Piscine, F-38406 Saint Martin d'Hères, France
[3] Sorbonne Université, Observatoire de Paris, Université PSL, École normale supérieure, CNRS, LERMA, F-75005, Paris, France
[4] School of Physics and Astronomy, Cardiff University, Queen's buildings, Cardiff CF24 3AA, UK
[5] Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, F-92190, Meudon, France
[6] Grupo de Astrofisica Molecular. IFF-CSIC. Calle Serrano 121-123. 28006, Madrid, Spain.
[7] Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[8] Joint ALMA Observatory (JAO), Alonso de Córdova 3107 Vitacura, Santiago de Chile, Chile
[9] CNRS, IRAP, 9 Av. colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France
[10] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA, 22903, USA
[11] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, 02138, USA
[12] IRAM, Avenida Divina Pastora, 7, Núcleo Central, E-18012 Granada, España
[13] Maison de la Simulation, CEA-CNRS-INRIA-UPS-UVSQ, USR 3441, Centre d'étude de Saclay, F-91191 Gif-Sur-Yvette, France

ABSTRACT

Context. Filaments are a key step on the path that leads from molecular clouds to star formation. However, their characteristics (for instance their width) are heavily debated, and the exact processes that lead to their formation and fragmentation into dense cores still remain to be fully understood.
Aims. We aim at characterizing the mass, kinematics, and stability against gravitational collapse of a statistically significant sample of filaments in the Orion B molecular cloud, which is renown for its very low star formation efficiency.
Methods. We characterize the gas column densities and kinematics over a field of 1.9 deg^2, using $C^{18}O$ ($J = 1 - 0$) data from the IRAM-30m large programme ORION-B, at angular and spectral resolutions of 23.5″ and 49.5 kHz, respectively. Using two different Hessian-based filters, we extract and compare two filamentary networks, each containing over 100 filaments.
Results. Independent of the extraction method, the filament networks have consistent characteristics. The filaments have widths of $\sim 0.12 \pm 0.04$ pc, and show a wide range of linear ($\sim 1 - 100\,M_\odot\,pc^{-1}$) and volume densities ($\sim 2 \times 10^3 - 2 \times 10^5\,cm^{-3}$). Compared to previous studies, the filament population is dominated by low-density, thermally sub-critical structures, suggesting that most of the identified filaments are not collapsing to form stars. In fact, only $\sim$ 1% of the Orion B cloud mass covered by our observations can be found in super-critical, star-forming filaments, explaining the low star formation efficiency of the region. The velocity profiles observed across the filaments show quiescence in the center, and coherency in the plane of the sky, despite being mostly supersonic.
Conclusions. The filaments in Orion B apparently belong to a continuum which contains a few elements comparable to already studied star-forming filaments (e.g. in the IC 5146, Aquila or Taurus regions) as well as many lower-density, gravitationally unbound structures. This comprehensive study of the Orion B filaments shows that the mass fraction in super-critical filaments is a key factor in determining star formation efficiency.

Key words. ISM: clouds - ISM: structure - ISM: kinematics and dynamics - Methods: data analysis - Radio lines: ISM - ISM: individual object: Orion B

1. Introduction

For decades, filaments of interstellar dust and molecular gas have been known to represent an important structural element of star-forming regions in the Galaxy (e.g. Schneider & Elmegreen 1979). The possible mechanisms leading to the formation of interstellar filaments are numerous, and can involve turbulence, gravity, magnetic fields, or any combination of these (e.g. Padoan et al. 2001; Burkert & Hartmann 2004; Hennebelle 2013; Smith et al. 2014; Federrath 2016). The presence of filaments in non-self-gravitating clouds (e.g. Ward-Thompson et al. 2010) suggests that turbulence plays a major role in filament formation. It has been proposed that filaments result from the intersection of sheet-like shock structures in supersonic turbulence (Pety & Falgarone 2000; Padoan et al. 2001). Recently, Hennebelle (2013) showed that in fact, if compression is necessary to accumulate gas in the first place, shear is the main driver behind clump elongation, and magnetic fields help confining filamentary structures and therefore make them more long-lived. Compression from winds of OB associations are also believed to have formed some of the Pipe Nebula filaments (Peretto et al. 2012). External ram pressure is not the only process leading to the formation of filaments, in particular in self-gravitating gas:

⋆ Based on observations carried out at the IRAM-30m single-dish telescope. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain).

self-gravity also has the effect of enhancing density anisotropies, and thus clump elongation (e.g. Hartmann & Burkert 2007; Peretto et al. 2007). Nagai et al. (1998) showed how gas sheets in hydrostatic equilibrium threaded by a magnetic field can fragment into filaments that are parallel or perpendicular to it, depending on their density.

Lately, far-infrared and sub-mm observations of the sky made with the *Herschel* space observatory revealed the tight link between the presence and properties of interstellar filaments and their ability to form stars (e.g. André et al. 2010; Molinari et al. 2010). It has been shown that more than 70% of gravitationally bound cores lie within thermally super-critical filaments, where the linear density $M_{\rm l}$ is larger than a critical value $M_{\rm l}^{\rm crit}$ above which the filaments become gravitationally unstable (Polychroni et al. 2013; Könyves et al. 2015). This suggests that most star-forming cores form as the result of gravitational instabilities occurring within unstable filaments (André et al. 2014). Another key proposition that has emerged from *Herschel* observations of star-forming clouds in the Gould Belt is the potential universality of the width of interstellar filaments at ~ 0.1 pc (Arzoumanian et al. 2011). This width seems to be relatively independent of the central column density of the filaments, which is surprising as the densest filaments would be expected to quickly collapse radially into thin spindles. Arzoumanian et al. (2013) proposed that accretion onto super-critical filaments could maintain a constant filament diameter during the filament's contraction. However, using velocity-resolved maps of the filament gas emission, Hacar et al. (2013, 2018) proposed that the filament width is actually not universal, but depends on environment, with broader filaments in low-mass star forming regions, and narrower filaments in massive star-forming regions. In this picture, all filaments have a linear density that is about critical, close to hydrostatic equilibrium, explaining why they would not collapse into spindles. The "universality" of the filament width would then be an observational bias of dust continuum emission maps that would merge narrower, velocity coherent, filaments (called *fibres* by Hacar et al. 2013) into one, non-velocity coherent elongated structure. Filaments formed in numerical simulations that involve the three key elements (turbulence, gravity, magnetic fields) discussed above seem to agree with *Herschel* results regarding the proposed universality of filament widths (e.g. Kirk et al. 2015; Federrath 2016), possibly linked to the scale at which the gas becomes subsonic, as first proposed by André et al. (2010). However, the predicted decrease of the velocity dispersion towards the inner parts of filaments is not always observed (Williams et al. 2018). This question of the (non-)universality of filament widths is still very much debated (e.g. Panopoulou et al. 2017).

Observationally, the statistical characterisation of filament kinematics is a difficult task, because it requires a large amount of telescope time. Until recently, only studies dedicated to a few individual filaments have been performed (Busquet et al. 2013; Hacar et al. 2013; Kirk et al. 2013; Peretto et al. 2013, 2014; Hacar et al. 2016; Williams et al. 2018). Only in the last few years, large surveys such as the Green Bank Ammonia Survey (GAS, Friesen et al. 2017) and the CARMA-NRO Orion Survey (Kong et al. 2018) have started to observe Gould Belt clouds in molecular lines over several square degrees. Such large datasets are required in order to build statistically significant samples of filaments and characterize their dynamical properties. However, none of these large surveys have yet looked in detail at the filament population. Here, we propose to use observations from the IRAM-30m ORION-B large programme to analyse the properties of the Orion B filaments over an area of 1.9 deg^2. This study will provide, for the first time, a complete picture of the filament population in the region, and shed light on the origin of the low star formation efficiency observed in Orion B.

This paper is organized as follows. In Section 2, we present the IRAM-30m, *Herschel*, and *Planck* data used in this study, and the data processing to which the IRAM-30m position-position-velocity (PPV) cubes were submitted. Section 3 introduces the filamentary network in the molecular cloud, and briefly explains how it is detected – methodological details on the filament identification procedures can be found in Appendix B. A statistical analysis of the physical properties that characterize these filaments is presented in Section 4 , while the implications of these properties for the structure and evolution of the filaments are discussed in Section 5. Section 6 summarizes the results and provides outlooks for this work.

2. Observational data and data processing

We aim at studying the structure and dynamics of the filaments in the south-western part the Orion B cloud. To do this, we need to constrain the mass, the temperature, the kinematics, and the magnetic field in this region. We will use *Planck* data to constrain the latter's geometry, and the map of the C^{18}O ($J = 1 - 0$) line observed as part of the ORION-B large program combined with a dust temperature map deduced from *Herschel* Gould Belt Survey observations, to derive the other quantities. This section presents the datasets and the steps applied to obtain an accurate map of the molecular hydrogen column density.

2.1. Molecular lines from the IRAM-30m ORION-B large programme

The ORION-B project (Outstanding Radio-Imaging of OrioN B, co-PIs: Jérôme Pety and Maryvonne Gerin) aims at mapping about 5 square degrees of the southern part of the Orion B molecular cloud over most of the 3 mm band in about 850 hours of IRAM-30m telescope time. The observing strategy and data reduction are discussed in detail in Pety et al. (2017), and we recall here just a few key numbers. The observations cover a bandwidth of 40 GHz, from 72 to 80 and from 84 to 116 GHz, in 3 tunings of the EMIR receivers. The angular resolution ranges from 36″ to 22.5″ for these frequencies. The median sensitivity of the Fast Fourier Transform Spectrometer (FTS) spectra ranges from 0.1 to 0.5 K in main beam temperature depending on the frequency, at a spectral channel spacing of 195 kHz. The $J = 1-0$ lines of the ^{13}CO and C^{18}O isotopologues are observed with a median noise of 0.12 K, and a spatial resolution of 23″. At the typical distance of ~ 400 pc of the Orion B cloud (Menten et al. 2007; Schlafly et al. 2014; Schaefer et al. 2016), this corresponds to a resolution of about 45 mpc. The higher spectral resolution but narrower bandwidth VESPA auto-correlator was used simultaneously to observe several lines, including ^{13}CO and C^{18}O ($J = 1-0$). The median noise for these data is 0.34 K ($T_{\rm mb}$) for a spectral channel spacing of 40 kHz. The position-position-velocity cubes of all identified molecular lines are all smoothed to a 29″ (60 mpc) resolution by convolution with a Gaussian kernel, and resampled onto identical grids with a resolution of $9'' \times 9'' \times 0.5$ km s^{-1} (or 0.1 km s^{-1} for VESPA). The spatial coordinates of the data cubes are centred onto the photo-dissociation region (PDR) of the Horsehead nebula, and rotated by 14° anticlockwise with respect to RA-Dec (J2000) coordinates, so that the edge of the IC 434 PDR, from which the famous Horsehead nebula emerges, is vertical in the maps.

As the data are still being acquired, the field presented in this paper covers 99′ by 68′, which corresponds to about 11.5 by 7.9 parsecs at the distance of Orion B (Fig. 1). It features objects such as the Horsehead pillar, the NGC 2023 and NGC 2024 (a.k.a. Flame nebula) star-forming regions and the PDRs IC 434 (illuminated by the multiple system σ Ori) and IC 435 (illuminated by HD 38087).

Here we make use of the VESPA data, as it provides a better spectral resolution while still retaining a good sensitivity, similar to the FTS sensitivity when smoothing the data to the same spectral resolution[1]. Most of the analysis focuses on the $C^{18}O$ ($J = 1 - 0$) data-cube, which is our best available molecular tracer for the medium at high visual extinctions typical of filamentary structures: as Pety et al. (2017) have shown, the ^{12}CO but also the ^{13}CO ($J = 1 - 0$) lines are saturated at high visual extinctions, while the $C^{17}O$ ($J = 1 - 0$) line is only detected around dense cores. In situations where we compare the filaments with the more diffuse surrounding medium, we complement the $C^{18}O$ ($J = 1 - 0$) data with ^{13}CO ($J = 1 - 0$).

Figure 1 shows a RGB representation of the complex velocity structure of the denser parts of Orion B, deduced from the $C^{18}O$ ($J = 1 - 0$) position-position-velocity cube. A map of its peak temperature can also be seen in Fig. B.1 (panel 1).

2.2. Modelling the data-cube as a sum of Gaussian profiles

The maximum value of the peak signal-to-noise ratio over the entire field of view in $C^{18}O$ ($J = 1 - 0$) is 63, while its median value is 2.8. It means that many lines of sight over the studied field of view are measured at a relatively low signal-to-noise ratio. Our first analysis step is thus to transform the noisy observational data into a model cube. For that purpose, we perform a multi-Gaussian fit of the spectra for each individual line of sight. This provides a model data-cube at a final spatial resolution of about 60 mpc (see Appendix A). The resulting model cube allows us to have a clean and easily exploitable representation of the noisy signal without the problematic windowing effects that may occur when signal-to-noise ratio masks are applied to the data.

2.3. Independent column density and temperature estimate from the Herschel Gould Belt Survey

By fitting a composite spectral energy distribution built from *Herschel* Gould Belt Survey (HGBS, André et al. 2010; Schneider et al. 2013) as well as *Planck* satellite (Planck Collaboration XIX 2011) continuum observations, Lombardi et al. (2014) have derived a dust temperature map $T_{\rm dust}$ as well as a dust opacity map at 850 µm, τ_{850}. These maps have a resolution of 36" where HGBS coverage is available, and 5' otherwise.

We converted the opacity map to a map of total hydrogen column density $N_{\rm H}$ using the following conversion factors:

$$\begin{aligned} N_{\rm H}/A_{\rm V} &= 1.8 \times 10^{21}\ {\rm cm^{-2}\ mag^{-1}}, \\ A_{\rm V}/\tau_{850} &= 2.7 \times 10^{4} \end{aligned} \qquad (1)$$

These factors are based on Bohlin et al. (1978) and Liszt (2014), and on Rieke & Lebofsky (1985) and Cardelli et al. (1989), respectively (see Pety et al. 2017, for detailed explanations).

The $C^{18}O$ ($J = 1 - 0$) line traces relatively dense ($n_{\rm H} \sim 10^3 - 10^4\ {\rm cm^{-3}}$) material with low FUV illumination (Pety et al. 2017; Gratier et al. 2017; Bron et al. 2018). Although the dust-to-gas coupling is completely established only above $10^5\ {\rm cm^{-3}}$ (Goldsmith 2001), the dust temperature should still be a better proxy for the gas temperature of this relatively cold medium than the excitation temperature of ^{12}CO ($J = 1 - 0$), which traces the warmer, more diffuse envelope of the cloud (Bron et al. 2018). The dust temperature map is shown in Fig. 2 (right).

[1] The data products associated to this paper are available at http://www.iram.fr/~pety/ORION-B

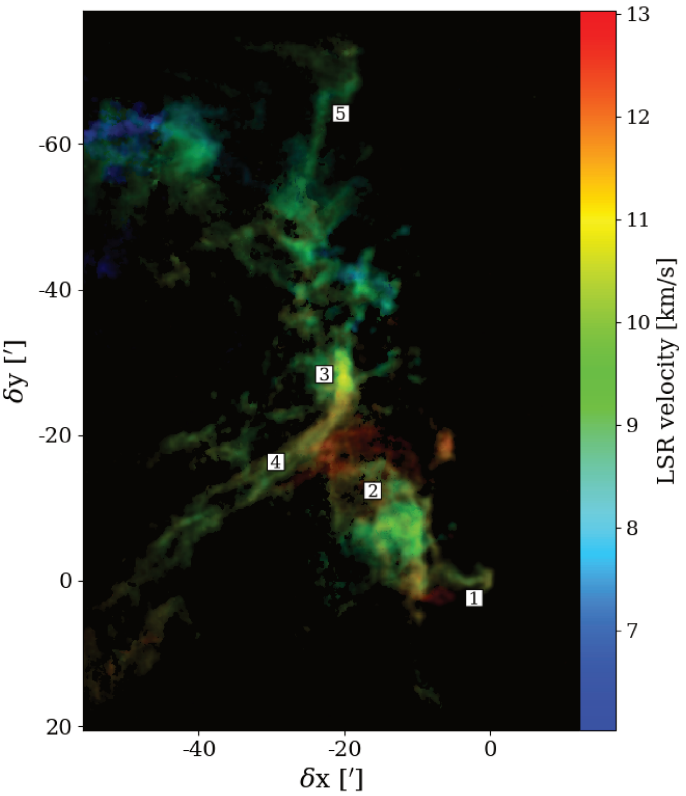


Fig. 1. Composite image displaying the velocity structure present in the $C^{18}O$ ($J = 1 - 0$) cube – each velocity channel is directly encoded as a hue, and the final image is obtained by additive colour synthesis. The brightest points correspond to 4 K in main beam temperature. The labels help locate (1) the Horsehead nebula, (2) NGC 2023, (3) NGC 2024 (the Flame nebula), (4) the Flame filament and (5) the Hummingbird filament. The spatial and kinematic complexity of the cloud appear clearly, in particular in regions such as the north-eastern part of the cloud or around NGC 2023.

2.4. A $C^{18}O$-derived column density

Using the integrated intensities from the modelled $C^{18}O$ ($J = 1 - 0$) cube, an estimate of the column density of $C^{18}O$ was computed assuming local thermodynamic equilibrium (LTE) and an optically thin medium in this line, and using the dust temperature from Lombardi et al. (2014). We followed the standard equations described in Mangum & Shirley (2015), using spectroscopic data from CDMS (Müller et al. 2005). The $C^{18}O$ column density was converted to $N_{\rm H}$ using $N_{\rm H} = 2 \cdot N_{\rm C^{18}O}/5.6 \times 10^{-7}$. This assumes that all the available carbon is locked in gas-phase CO, with a C/H_2 abundance of $\sim 2.8 \times 10^{-4}$ (Sofia et al. 2004; Parvathi et al. 2012; Gerin et al. 2015), and a $^{18}O/^{16}O$ isotopic ratio of $\sim 1/500$ (Wilson & Rood 1994).

The obtained column densities are not expected to trace the totality of the matter present along the line-of-sight, as there

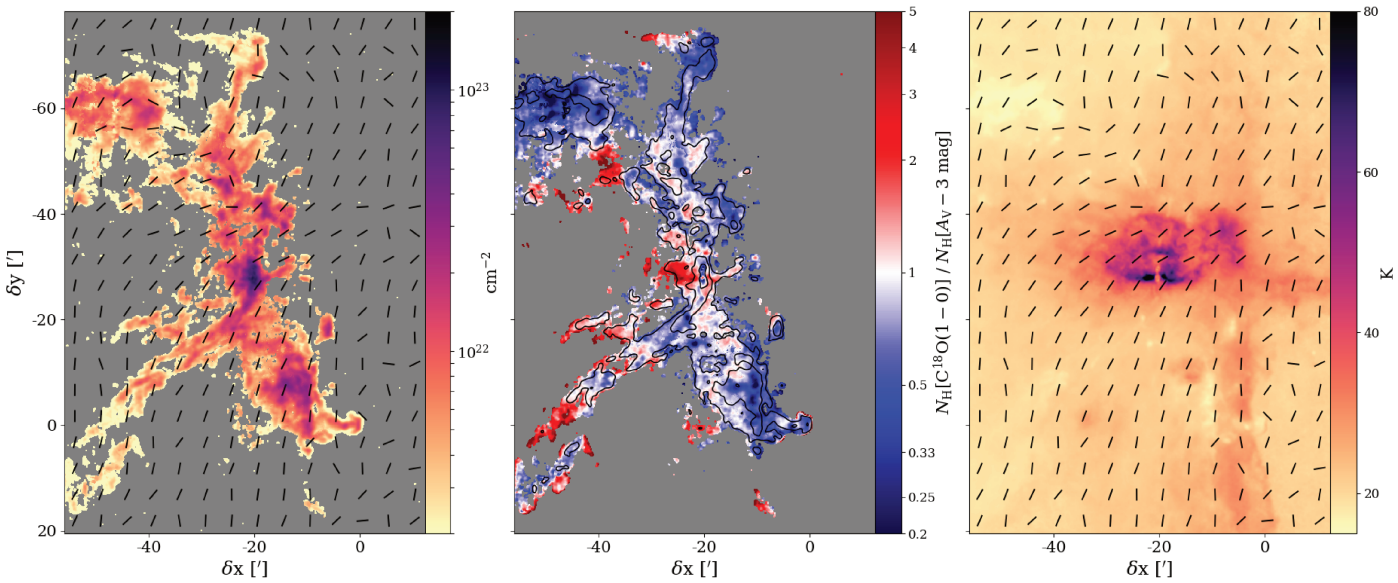


Fig. 2. *Left:* N_H column density map derived from the $C^{18}O$ ($J = 1-0$) integrated intensity (multi-Gaussian model). The orientation of the magnetic field, derived from *Planck* polarimetric data, is shown by the black segments. The spacing of the segments corresponds to the 5′ beam of *Planck*. *Middle:* Map of the ratio between the $C^{18}O$-traced N_H and the dust-traced N_H above 3 magnitudes of A_V in Orion B. Superimposed are contours of the $C^{18}O$ ($J = 1-0$) integrated intensity at 0.75, 2.5 and 7.5 K km s^{-1}. *Right:* Effective dust temperature map computed by Lombardi et al. (2014). Here again, the black segments show the orientation of the magnetic field.

can be both atomic and molecular $C^{18}O$-dark gas in the fore- or background of the molecular emission region. When comparing this molecular emission and the dust extinction A_V derived from *Herschel* data, Pety et al. (2017) showed that the $C^{18}O$ emission starts to be detected at $A_V \sim 3$. We therefore compare the obtained $C^{18}O$-traced column densities to the dust-traced column density above 1 magnitude of A_V, rather than to the total dust-traced column density. The comparison of these tracers in the filamentary regions identified and analysed in this work is shown in Fig. 3. We see a good consistency of the resulting column densities. The ratio of $N_H[C^{18}O]/N_H[A_V - 3\,mag]$ has a mean value of 1.02, and a median of 0.83, with a standard deviation of 1.19.

The middle panel of Fig. 2 shows the spatial distribution of the $N_H[C^{18}O]/N_H\,[A_V - 3\,mag]$ ratio, which is close to one in a large fraction of the map. It is significantly smaller than 1 on the western edge of the cloud, and in particular at the base of the Horsehead pillar. This might be an effect of selective photo-dissociation: at the edge of the IC 434 PDR, the self-shielding of $C^{18}O$ is too weak to prevent its destruction by FUV radiation. The ratio is also much smaller than 1 in a region to the North-East, which is the coldest in the current field of view and is known to harbour dense cores: in that case, $C^{18}O$ depletion is the most probable explanation, because the dust is cold enough for CO to freeze-out on grain surfaces. The $N_H[C^{18}O]/N_H[A_V - 3\,mag]$ ratio is conversely significantly larger than 1 in several low-density regions lying to the East. In these regions, away from the sources of photo-dissociating radiation, A_V ranges from 3.0 to 5.4, with an average of 3.5, close to the chosen extinction threshold of 3 magnitudes. In this region the column density ratio becomes very sensitive to the choice of the extinction threshold, leading to higher uncertainties. The extinction threshold may vary across the field of view and reach somewhat lower values in regions away from the interface with IC 434. The ratio is also larger than 1 around NGC 2024, which is mostly likely due to a layering effect with a significant temperature gradient along the line of sight, poorly rendered by a single value of effective dust temperature. As a consequence, the dust-traced column density is underestimated and the CO-traced column density is overestimated (Pety et al. 2017).

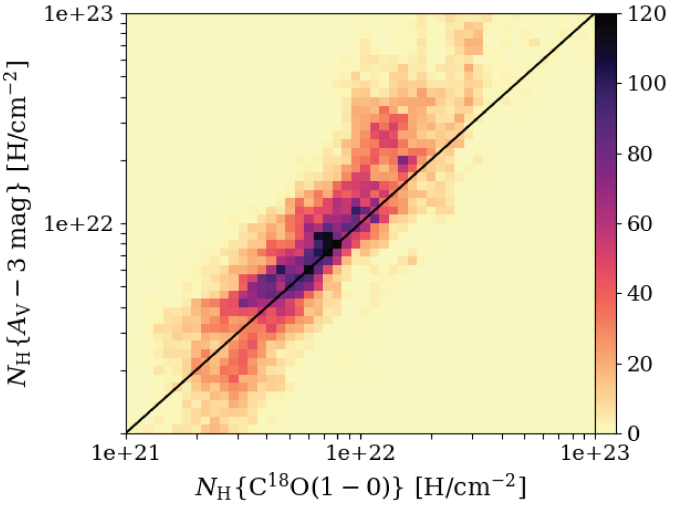


Fig. 3. Joint distributions of the N_H column densities as traced by $C^{18}O$ ($J = 1-0$) against those inferred from A_V with an offsets of 3 magnitudes. This threshold corresponds to the extinctions at which the molecular tracers starts to be detected according to Pety et al. (2017). The distribution is computed in the identified filamentary regions. The 1:1 relation is overplotted as a black line.

The question of depletion by freeze-out is important when studying filaments, because they are expected to be dense and cold, and CO is known to freeze out at temperatures < 20 K. We therefore compare the $N_H[C^{18}O]/N_H[A_V - 3]$ ratio to the temperature (Fig. 4) to search for potential signs of such a systematic depletion effect in the regions identified as filamentary

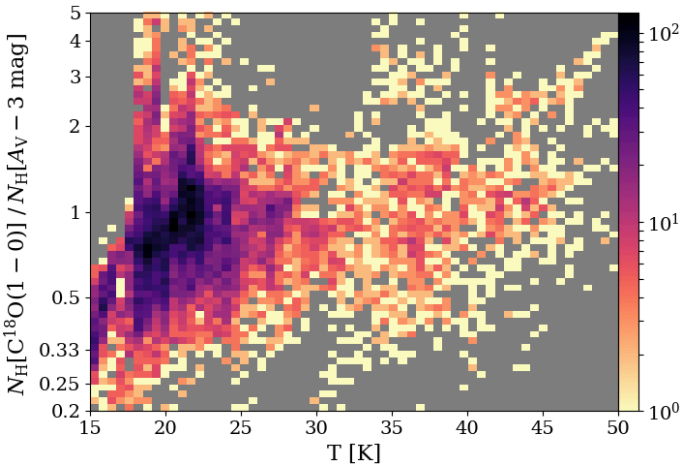


Fig. 4. Joint distribution of the ratio of $C^{18}O$-traced vs. dust-traced N_H column density (Fig. 2 middle) against the dust temperature (Fig. 2 right), taken only in the identified filamentary regions.

(see Sect. 3.2). A "tail", corresponding to both a lower temperature and a lower column density ratio, exists in the distribution. However, the lines of sight where CO is probably depleted (lines of sight with a temperature below 20 K and a column density ratio 50% below its average value) amount to less than 8% of the filamentary regions. Moreover, they almost exclusively lie in the north-eastern cold core region, which does not host major filaments (see Fig. 1 and Fig. 2). For the most part, $C^{18}O\,(J = 1-0)$ is therefore a tracer well suited to recover the gas column densities in the density and temperature regimes of the filamentary regions of the Orion B cloud.

2.5. Magnetic field orientation from the Planck all-sky survey

We also used *Planck* polarization data[2] at 353 GHz to estimate the orientation of the magnetic field in Orion B. The Stokes I, Q, U maps were used at their native resolution of 5' to derive the polarisation angle χ and the magnetic field angle ψ (which is rotated by 90° with respect to χ in the IAU convention). These angles were rotated to match the custom North-South axis of our projection (i.e., 0° points to the top of the presented field, not to the standard North in J2000 equatorial coordinates). The orientation of the magnetic field ψ is presented on the left and right panels of Fig. 2, superimposed on top of the column density map and temperature map respectively.

3. Detection and characterization of the filamentary network

3.1. Qualitative description of the filaments

Figures 1 and 2 suggest that observations of $C^{18}O\,(J = 1-0)$ bring out a complex molecular filamentary network.

At the very centre of the observed field, the star-forming region NGC 2024 shines brightly. This spot has both the warmest temperature and the highest column density in this area. The molecular emission comes from a dense filament, seen in the optical as a dark dusty lane in the foreground of the young HII region, with a characteristic shape that earned the Flame Nebula its name. This large Flame filament can be clearly seen in $C^{18}O\,(J = 1-0)$, as it flows diagonally from NGC 2024 to the South-East. In its more diffuse part, it is clearly sub-structured, made of parallel strands of molecular gas.

At the South-West of our field lies the Horsehead nebula, with its characteristic shape. This nebula is a pillar carved in the Orion molecular cloud by the IC 434 HII region.

Between NGC 2024 and the Horsehead lies the quieter star-forming region NGC 2023. The kinematics of the gas surrounding NGC 2023 is complex, with at least two velocity components (visible in green and orange colours in Fig. 1), and the filaments this medium might host are less obviously distinguishable by eye.

Just North of the Flame Nebula lies a filamentary region exposed to the influence of the NGC 2024 HII region. Further North, the round shape of the HII bubble becomes less visible ($\delta x; \delta y \approx -20'; -40'$). At the northern edge of our field lies another long filament, which we dub the Hummingbird filament. By eye, it is the second longest filament in the cloud after the Flame filament, and it stands out as a quite isolated structure, which makes it a perfect subject to study, e.g., gravitational fragmentation.

Finally, to the North-East lies a blue-shifted turbulent region containing dense cores within a complex velocity structure. Here again, the filaments are not easy to identify by eye.

3.2. Identifying the filaments

Visual inspection is insufficient to precisely and objectively locate and thus study filaments, in particular in the most entangled regions. Therefore, we have implemented several algorithms to identify filamentary structures in a map. These algorithms are presented in detail in Appendix B. Here we describe the concepts we use and we briefly summarize the algorithms that we apply to the N_H map derived from the $C^{18}O\,(J = 1-0)$ integrated intensity.

3.2.1. Morphological definition and extraction algorithms

From an observational point of view, we can qualitatively define filaments as elongated, over-dense structures in the molecular ISM. We thus expect to see them as bright structures with high aspect ratios. If we were looking at the altitude map of a mountainous region (where the altitude would correspond to the brightness of molecular emission), the filaments would correspond to the main mountain ranges of this region. The ridge lines of these mountain ranges would correspond in turn to the filamentary skeleton. These mountain ranges, and these ridge lines, can be simply defined in terms of topography, that is in terms of differential properties of the studied map.

The filaments are characterized by the properties of the Hessian matrix, the second derivative of the map (Schisano et al. 2014): the main directions of variation are identified, and ridges appear as local maxima along one direction, while the perpendicular direction shows negligible, or at least smaller, variation (Table B.1). The first filament extraction procedure was meant to be as straightforward as possible, and simply applied a threshold on the eigenvalues of the Hessian matrix - yielding the skeleton S1. The second procedure aimed at introducing several refinements: the data were rescaled with an arcsinh function prior to the computation of the Hessian matrix, then the eigenvalues were used to compute the local aspect ratio of the structures in

[2] Based on observations obtained with Planck (http://www.esa.int/Planck), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada.

the column density map, and the gradient was also used to refine the ridge detection. The whole analysis was performed in a multi-scale fashion, and yielded the skeleton S2. The details of both approaches are described in Appendix B.

3.2.2. Skeletons, bones, nodes, and filamentary network

To avoid confusion between the different filamentary objects that will be discussed in this paper, we define them as follows. The filament identification process yields a set of 1-pixel wide curves which constitute a graph called the *skeleton* (or *filamentary skeleton*). These lines have endpoints and intersections (these points are *nodes*) and the branches of the graph between two nodes are *bones* (as in Panopoulou et al. 2014). The bones can be "fleshed out" by attributing some of the surrounding gas to each 1-pixel wide curve, thus yielding *individual filaments*. Taken together, these individual filaments form a *filamentary network* (i.e., a fleshed-out skeleton). Some objects visible in the field and spontaneously referred to as filaments can be made of several individual filaments (e.g., the Flame filament) – to avoid confusion these are therefore referred to as *filamentary structures*. Finally, the term *filamentary regions* refers to the regions identified as bright and structured during the filament identification process, regardless of any attribution of the gas to a specific individual filament (binary masks shown in panel 5 of Fig. B.1 and B.2).

3.2.3. Skeleton analysis

The two methods show disparities, some structures can be identified by one method and not the other, or vice-versa. To assess the similarities and differences quantitatively, we superimpose both skeletons (Fig. 5). The different filtering approaches can lead to differences of the order of one or two pixels in the position of the identified structures. But as this does not have much impact on the further analysis of the filaments, such "neighbouring" points are considered as matching points, up to a distance of 2 pixels. While there are some structures that are exclusive to one skeleton or the other, this criterion shows that a large portion (about 68% of S1, 83% of S2) of the skeletons is common to both methods.

Rather than choosing between the two possible skeletons, our approach is to keep both and study them and their statistical properties, in order to assess the variability in the physical properties that can result from a variability in filament identification method. Therefore, four skeletons in total will be compared throughout this work: the skeleton S1 obtained by simple thresholding; the skeleton S2 obtained by the adaptive method with ridge-detection; the "robust skeleton", which is made of the common or neighbouring points of S1 and S2 (and thus corresponds to the intersection S1∩S2); and the "composite skeleton", which is made of all the points of S1 and S2 (and thus corresponds to the union S1∪S2). Since the robust skeleton has, by construction, a thickness of two or three pixels due to the "neighbouring points" being taken into account, it needs to be thinned again (see Appendix B.1) to a single-pixel width. The first three skeletons are used for statistical comparisons, while the last one is useful for displaying simple maps of the physical quantities.

Once the skeletons have been extracted from the observational data, we still need to identify the bones. This sorting of the skeleton into bones and nodes allows us to analyse the local properties of the filamentary skeleton. It also enables us to "clean" (Appendix B.2) the skeleton by removing isolated nodes and short bones (under 0.22 pc long), and those that do not match the definition of what a filament should be: curvature (Appendix B.2.2) and contrast (Appendix B.2.3) can show that some structures are not over-dense, or narrow, or elongated enough to be regarded as filaments. With the exception of the Appendix, the figures and statistics presented in this paper are therefore obtained using the cleaned skeletons.

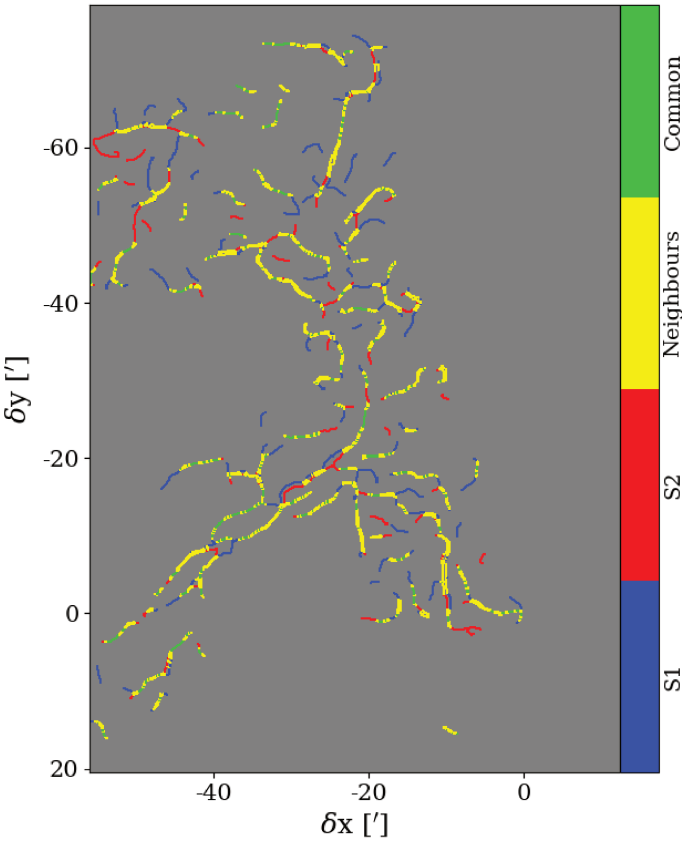


Fig. 5. Comparison of the skeletons obtained with the two different methods, S1 denoting the skeleton obtained by simple thresholding, and S2 the one obtained by the adaptive method with ridge-detection. The structures only identified by one method appear in red or blue, the structures common to both methods appear in green if they perfectly match, and in yellow in the cases where the morphological thinning led to small position offsets (see Appendix B).

3.3. Transverse profile fitting

The Hessian identification of filaments provides us with their position angles. We can thus study the cross-sections (or transverse profiles) of the individual filaments, by plotting the variations of a physical quantity of interest perpendicular to the bone's local major axis – in particular using the hydrogen column density map.

For the sake of simplicity and robustness (see Arzoumanian et al. (2011) and Panopoulou et al. (2014) for a discussion), the column density profiles of the individual filaments were considered to be Gaussian peaks superimposed on top of a linear baseline. Such a profile is therefore constrained by five parameters: the position x_0 of the Gaussian with respect to the reference pixel in the skeleton, the amplitude and the width of the Gaussian, A and w, and the slope and offset of the baseline, α and β:

$$P(x) = \alpha \cdot x + \beta + A \cdot \exp\left(\frac{-(x - x_0)^2}{2w^2}\right) \quad (2)$$

The range over which the profile is fitted, on either side of the skeleton ridge, can have quite a strong influence on the resulting model profile (Panopoulou et al. 2014). Therefore, we tried to use a spatial range as wide as possible, but we were limited by the density of the filamentary network: a very wide range can intersect several individual filaments, which leads to incorrect fitting results. As a compromise between these limitations, we set the fitting range to 6 pixels on each side of the ridge (or about 1.9') to limit this effect. We thus avoid getting a "secondary bump" in the outer parts of the profile, that would be due to a neighbouring individual filament. However, even with this fitting range, the same pixel in the column density map can be attributed to several profiles, either of the same individual filament (e.g., after a turn) or of neighbouring individual filaments.

We started by fitting the transverse profiles individually (i.e. for each pixel of the skeleton). However, this led to strong degeneracies between the various parameters when the baseline deviates from the assumed straight line, which required human supervision to be overcome. As wanted to avoid this in a semi-automated, statistical analysis of the filamentary network, we have set the individual filament width w as a semi-fixed parameter. To determine its value, we fit the mean profile of each individual filament with all five free parameters, since this mean profile has a better signal-to-noise ratio and a more Gaussian shape than the profiles for individual pixels. The obtained value of w is then kept for the transverse profile of each pixel in the individual filament.

We also have taken into account the fact that the skeleton is not always perfectly aligned with the physical ridge lines of the filaments (as shown in Fig. 5, the thinning of the skeletons can lead to positional uncertainties of about one or two pixels). For a given transverse profile, this means that the peak of the profile is off-centred, and that $x_0 \neq 0$. Such offsets can artificially broaden the mean profile. Therefore, a recursive approach is used: after a first fit of the mean individual filament profile which results in a given individual filament width, we fit each individual profile using that width as a fixed parameter. This individual fit yields in turn the position offset of the profile peaks, which allows us to re-align the profiles before recomputing an updated, centred mean profile. This centred mean profile is then fitted (with five free parameters) and yields a better (usually narrower) estimate of the average profile width in the individual filament (which is then used to perform a better fitting – with only α, β, A and x_0 as free parameters – of the individual profiles).

The fit results give us access directly to such quantities as the filament width (Sect. 4.1.1) and contrast (Appendix B.2.3), and to quantities more closely linked to star formation, such as the mass and gravitational stability of the filaments (Sect. 4.1.2).

4. Physical properties of the filaments

4.1. Profile and mass

4.1.1. Filament width

The characteristic width of filaments is a direct output of column density profile fitting, and it is the most commonly measured and discussed quantity for filaments in the ISM (see, e.g., Arzoumanian et al. 2011, 2013; Kainulainen et al. 2016; Panopoulou et al. 2014, 2017). The filament width is measured on the global profile of each individual filament (as described above), using a non-weighted average (i.e. all profiles in the individual filament are normalized to the same amplitude). The obtained result is deconvolved from the synthetic Gaussian beam of the fitted IRAM-30m observations, which corresponds to 29″ or about 60 mpc. The final width is thus given by $w_{\rm deconv} = \sqrt{w_{\rm fit}^2 - 0.06^2}$ pc.

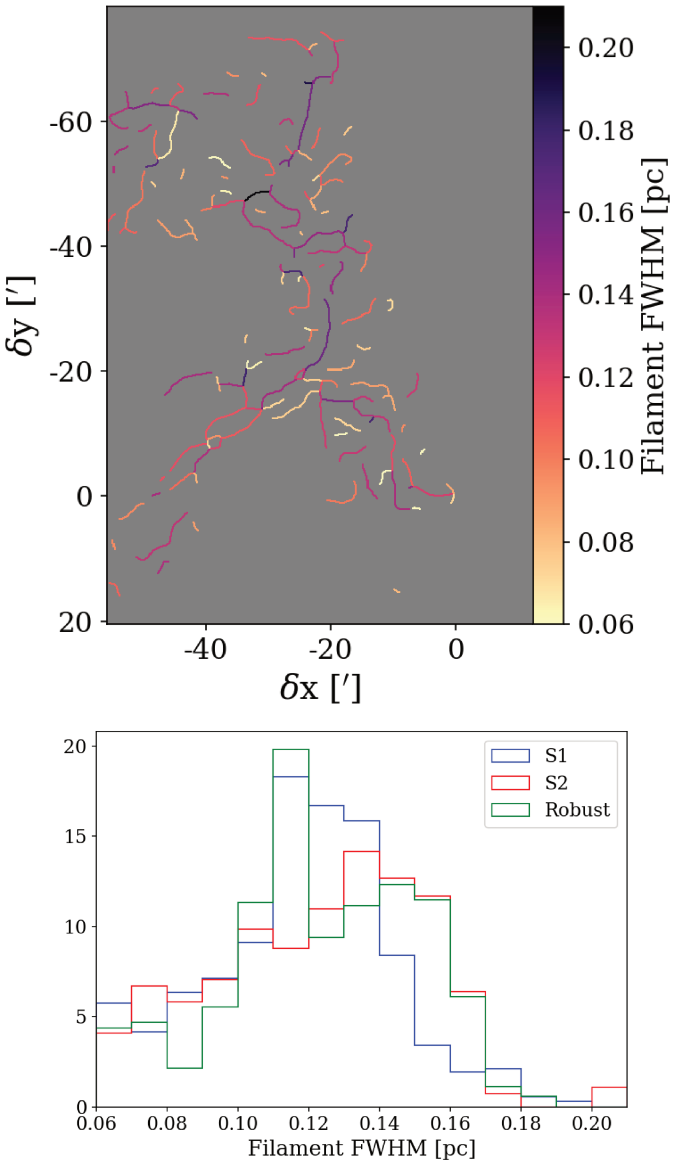


Fig. 6. *Top:* Map of the FWHM of the mean profiles of the individual filaments (combined skeleton). *Bottom:* Histograms of the filament width for the S1, S2 and robust skeletons.

The results are shown in Fig. 6. The typical widths for each skeleton, in terms of mean, median or most probable value are listed in Table 1. Except for a few individual filaments, mostly short ones, for which the fit does not converge or yields oddly high values (which is the sign of filaments with low contrast, see Appendix B.2.3), the spread of widths is rather small.

4.1.2. Linear density and gravitational stability

The next physical quantity derived from the profile fitting is the linear density of the filaments. The linear density $M_{\rm l}$ of a fila-

Table 1. Mean, median and most probable FWHM of filaments for the three studied skeletons, resulting from mean profile fitting of the $C^{18}O$-derived N_H column density.

Skeleton	Mean	Median	Most probable	Std dev
S1	0.11 pc	0.12 pc	0.13 pc	0.04 pc
S2	0.11 pc	0.12 pc	0.13 pc	pc
Robust	0.11 pc	0.12 pc	0.11 pc	0.04 pc

ment should take into account only the matter in the filament itself, not its foreground or background. This is why the linear baseline of Eq. 2 is subtracted from the fitted profile. The linear density for a given line-of-sight is then simply the integral of the corresponding Gaussian transverse profile of surface density.

Knowing the linear density of the individual filaments, and having a proxy for their kinetic temperature T_K thanks to the dust temperature map, we can also estimate the stability of Orion B's filaments against gravitational collapse. The criterion for balance between thermal pressure and gravity is given by $\gamma = M_l/M_l^{crit}$, where M_l^{crit} is determined by Ostriker (1964) as:

$$M_l^{crit} = \frac{2kT_K}{\mu m_H G} = \frac{2c_s^2}{G} \approx 16\,M_\odot\,pc^{-1} \times \left(\frac{T_K}{10\,K}\right). \quad (3)$$

The resulting gravitational instability criterion γ of the filaments is presented in Fig. 7 (left).

However, with this approach, the critical linear density M_l^{crit} is a lower limit, because it only takes into account the thermal (kinetic) pressure. In order to account for support against gravity from both thermal and non-thermal motions of the gas, we need to compute an effective temperature $T_{eff} = T_K + c_{turb}^2 \frac{\mu m_H}{k}$, where c_{turb} is the observed non-thermal velocity dispersion (Arzoumanian et al. 2013; Peretto et al. 2014; Kainulainen et al. 2016). We could have access to the turbulent velocity dispersion $c_{turb} = \Delta v$ thanks to the velocity-weighted moments of the $C^{18}O\,(J = 1-0)$ spectra. However this would not take into account the fact that spectra can contain several velocity components. The spectral signature of the turbulence providing support in the form of an effective pressure is expected to be found in line broadening, rather than in the multiplicity of spectral components, which rather trace the presence of physical sub-structures. Therefore, when computing T_{eff}, the velocity dispersion Δv that we use is rather the typical FWHM of the Gaussian velocity components identified by the multi-Gaussian fit (Sect. 3.3). The corresponding effective critical linear density $M_{l,eff}^{crit}$ is an upper limit this time, because of the implicit assumption that all the non-thermal spectral broadening arises from turbulent motions, which might not necessarily be the case (e.g. accretion or gravitational collapse can also contribute to line broadening). From that, we derive the effective gravitational instability criterion γ_{eff} (Fig. 7, right).

Both the lower limit of the instability criterion, γ_{eff}, and its upper limit, γ, show that the filaments in Orion B are mostly stable against gravitational collapse. This is further discussed in Sect. 5.2.

Since unstable filaments undergoing gravitational collapse are likely to be leading to star formation (Arzoumanian et al. 2011; Hacar et al. 2013), we also compare the spatial distribution of the gravitational instability criteria γ and γ_{eff} in the filamentary network with the positions of the youngest among the young stellar objects (YSOs) identified by Megeath et al. (2016) and by the HOPS survey (Fischer et al. 2013; Furlan et al. 2016). We can see indeed on Fig. 7 that the positions of YSOs, in dense clusters (NGC 2024), in looser groups (between NGC 2023 and the Horsehead) or isolated, tend to correspond to local maxima of the instability criterion γ or γ_{eff}, i.e. regions where the filaments are closest to radially collapsing under the effect of self-gravity.

4.2. Relative alignment of the magnetic field and filaments

As we have access to an estimate of the magnetic field orientation thanks to *Planck* data (Fig. 2), it is straightforward to compare it with the position angle of bones, which is obtained as a by-product of the Hessian filament extraction as mentioned in Sect. 3.3. Given that both the filament position angle and the orientation of the magnetic field are defined only modulo 180°, their relative orientation is between 0° and 90° (Fig. 8). However, the major caveat is that the resolution of *Planck* data does not match that of the IRAM-30m observations (5′ and 31″ respectively).

We thus smoothed the IRAM-30m data to the resolution of the *Planck* magnetic field data, i.e. 5′, and performed the filament extraction on that smoothed data, before comparing it to the magnetic field map. The distribution of this relative orientation is compared to what it would have been if the two orientations were uncorrelated. The uncorrelated distribution and the corresponding error margins are obtained by a Monte-Carlo sampling which randomly associates one value from the observed filament position angle distribution with one value from the observed magnetic field orientation distribution (Fig. 8).

The distribution obtained for uncorrelated quantities is not flat, which is a result of the anisotropy of both the magnetic field (Fig. 2) and the large-scale structures of the gas. The actual distribution of relative orientation shows a modest peak around 20°, and a major peak which lies between 40 and 60°. The latter one corresponds to the large filamentary structures with a position angle of roughly ±45° (the north-eastern extension, part of the the Flame filament, and mostly the NGC 2023-Horsehead complex).

4.3. The velocity field around filaments

Compared to Herschel photometric images, the high spectral resolution of the molecular line data provides access to the motions of the gas in the filaments and their immediate surroundings.

4.3.1. Line-of-sight velocity dispersion

As mentioned in Sect. 4.1.2, we computed the line-of-sight FWHM velocity dispersion Δv using the typical width of a Gaussian component identified by the multi-Gaussian fit. From this velocity dispersion we derive the Mach number, by comparing it with the sound speed $c_s = \sqrt{\gamma k_B T/m}$, where T is the gas temperature (assumed equal to the effective dust temperature) and m is the average molecular mass. The average transverse profiles of the Mach number in the filaments as probed by for $C^{18}O\,(J = 1-0)$ are shown in Fig. 9 (top). For comparison, we also plotted the transverse profiles of the Mach number as probed by $^{13}CO\,(J = 1-0)$ in Fig. 9 (bottom). In that case, the gas temperature used to compute the sound speed was the composite temperature obtained using the effective dust temperature and the $^{12}CO(J = 1-0)$ peak temperature, described in Orkisz et al. (2017).

The $^{13}CO\,(J = 1-0)$ profiles display significantly higher Mach numbers than $C^{18}O\,(J = 1-0)$, but show no significant feature whatsoever. On the other hand, the $C^{18}O\,(J = 1-0)$ profiles show a pronounced decrease in Mach number towards

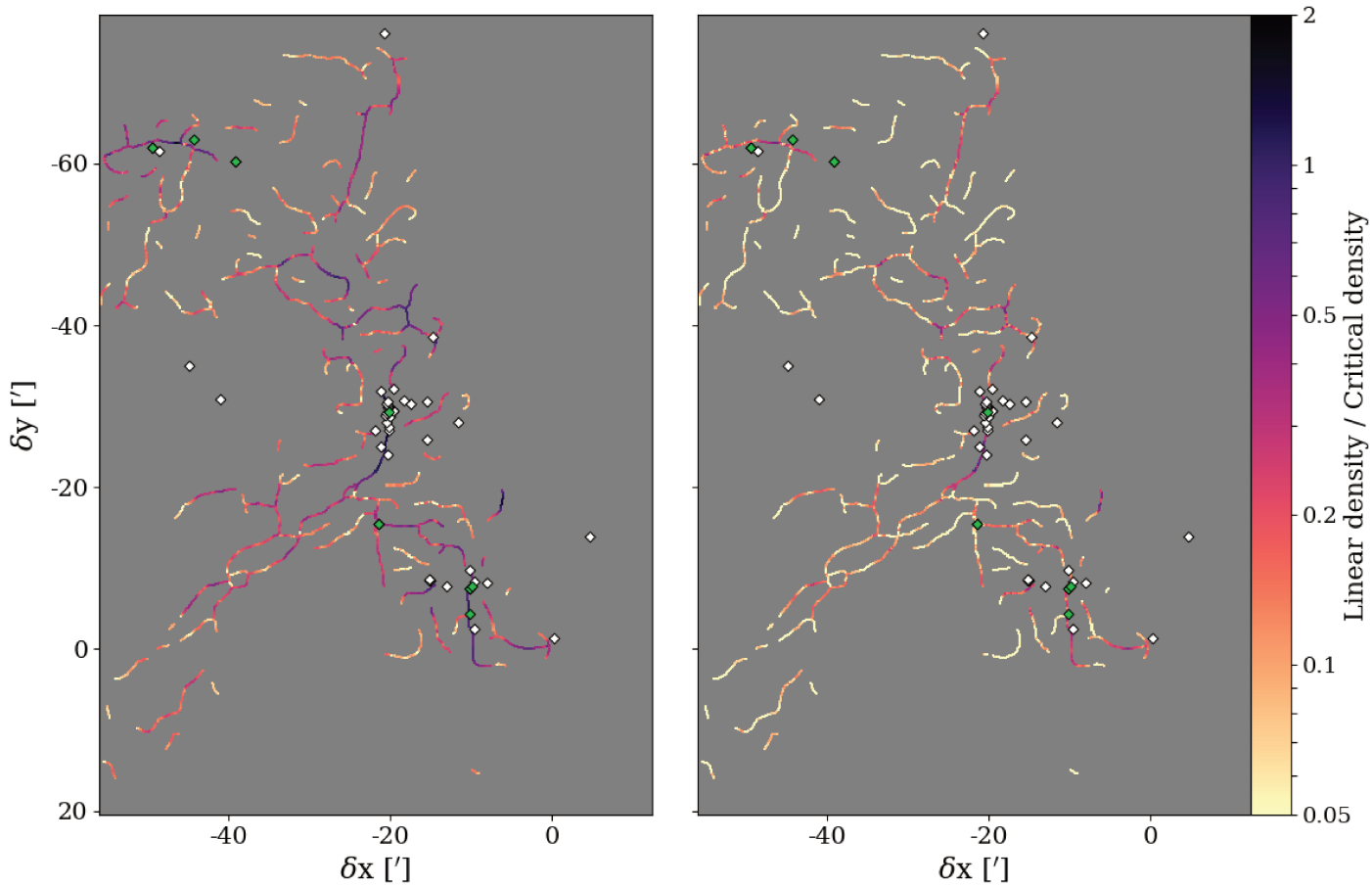


Fig. 7. Gravitational instability criterion of the filaments (combined skeleton), derived from the $C^{18}O$-traced linear density estimation and the thermal pressure (γ, *left*) or the effective pressure that also takes into account the $C^{18}O$ ($J = 1 - 0$) velocity dispersion ($\gamma_{\rm eff}$, right). Overlaid are the positions of protostars from Megeath et al. (2016) in white, and class 0 YSOs from HOPS (Furlan et al. 2016) in green.

the center of the filaments. The width of this feature is similar to the measured filament width.

4.3.2. Centroid velocity gradient

While the velocity dispersion gives access to the kinematics along the line of sight, there is no direct way to observe velocity effects in the plane of the sky. As a proxy for such observations, we have used the gradient of centroid velocity. Its amplitude measures the variability of the velocity field in the plane of the sky, in contrast to the linewidth which probes the velocity dispersion along the line of sight.

Figure 10 shows the average transverse profiles of the amplitude of the centroid velocity gradient, observed in $C^{18}O$ ($J = 1-0$) (top) and in ^{13}CO ($J = 1-0$) (bottom). The gradient amplitude is almost constant across the filaments in $C^{18}O$ ($J = 1 - 0$), with no particular visible trend. The ^{13}CO ($J = 1 - 0$) profiles, on the other hand, exhibit slightly higher amplitudes of the gradient, with a quite pronounced minimum towards the centre of the filaments, which brings the centroid velocity gradient amplitudes of ^{13}CO ($J = 1 - 0$) down to the average value observed for $C^{18}O$ ($J = 1 - 0$).

5. Discussion

5.1. Width of the filaments

5.1.1. On the robustness of our width estimation

Panopoulou et al. (2017) suggest that the narrow distribution of filament widths with a typical scale of 0.1 pc can be an artefact resulting from the analysis method, in particular the measure of the width of the mean profile of a filament rather than the widths of its individual profiles, and the spatial range used for the profile fitting. The use of average filament profiles results in a narrower distribution of widths, but does not modify its peak significantly. In our case, the high spatial density of detected filaments and thus the number of intersections fragments the largest filamentary structures (such as the Flame filament) into shorter individual filaments, reducing the amount of averaging. The obtained distribution of filament widths, and in particular its dispersion, appears as an intermediate between the broad distributions obtained for individual profile widths and the very narrow ones obtained for average filament profiles, as shown in Panopoulou et al. (2017, their Fig. 2). The comparison with measurements on filaments in the Polaris Flare by Panopoulou et al. (2017) suggests that the absence of filaments with widths larger than 0.2 pc in Orion B can be a result of the fitting window which is 1.9′, or 0.21 pc on each side of the filament ridge. However, the median and most likely widths in Fig. 6 and Table 1 are small enough to be confidently measured, but large enough not to be due to

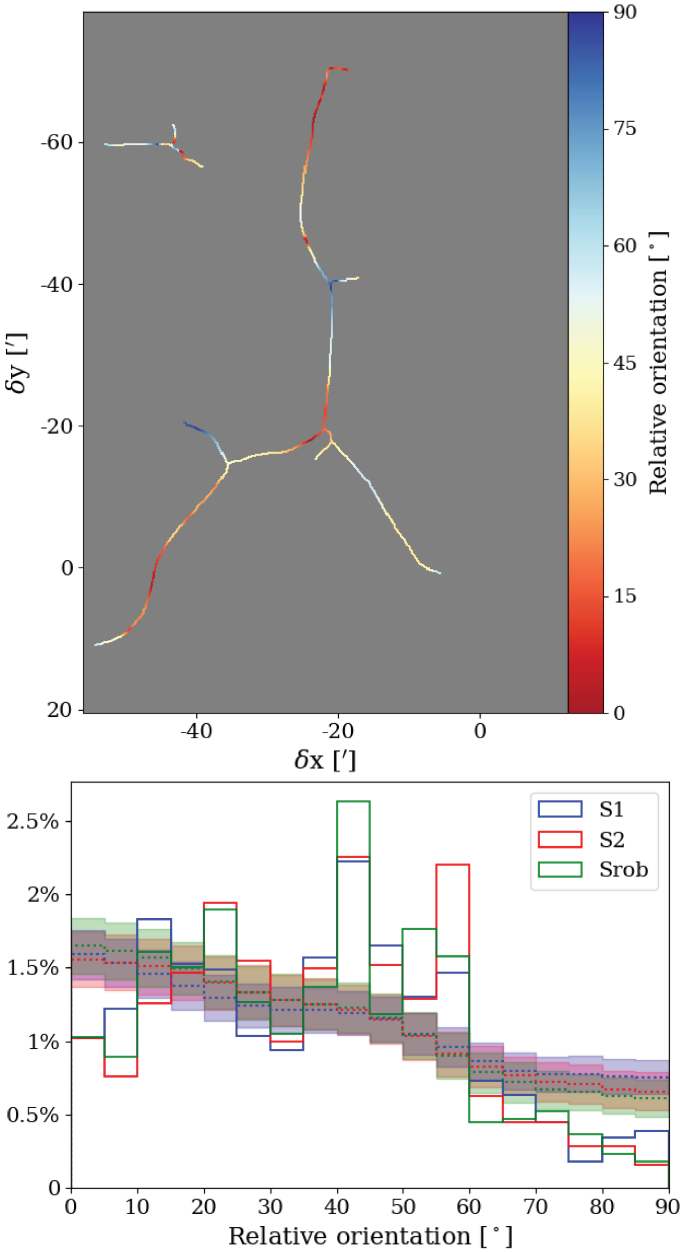


Fig. 8. *Top:* Map of the angle between the magnetic field and the filaments detected in Orion B at a 5′ resolution. The filaments are red when they are parallel to the magnetic field, and blue when perpendicular to it. *Bottom:* Distribution of the relative orientation of the filaments and the magnetic field, for the S1, S2 and robust skeletons. The dotted lines (and the shaded areas) present the distribution (and the corresponding $\pm 1\sigma$ uncertainties) that we would get if the two quantities were uncorrelated. The uncorrelated distribution is obtained by a Monte-Carlo sampling of the magnetic field and filament position angles.

the telescope beam. We can therefore say that, even though the spread of widths of the filaments is probably underestimated in this work, the median and most probable value of the filament width, of the order of 0.12 pc, are reliable.

In addition to this first caveat, one should also stress that a Hessian detection filter behaves to some extent like a wavelet filter, bringing out structures matching the scale of the Gaussian derivative used for the calculation. However, the skeleton S2, thanks to its adaptive nature, can overcome this bias, since the data dictates the scales at which the filter has its strongest response (see Appendix B.1). For the skeleton S1, the use of a single smoothing scale could induce a stronger bias, but the robustness of the results was checked by studying skeletons obtained in the same way as S1, but with a halved or doubled smoothing scale: although the distribution of the obtained filament widths was slightly shifted towards smaller or larger scales respectively (due to the inability of a very narrow filter to pick up very broad structures, and the excessive smoothing by a very wide filter that blurs out very small structures), its peak remained unchanged, thus proving that the main detected structures do not strongly depend on the filter, and that they correspond indeed to filaments with a mean width of the order of 0.12 pc (Table 1).

Finally, the average Mach number profiles derived from the $C^{18}O$ ($J = 1 - 0$) velocity dispersion show a feature of similar width (Fig. 9). The width of feature seen in the average profiles of the ^{13}CO ($J = 1 - 0$) centroid velocity gradient norm is also of the order of 0.1 – 0.2 pc, also it is less clear what this quantity is tracing. These measurements, completely independent from spatial distribution of the column density, argue in favour of the robustness of our estimate for the FWHM of filaments.

5.1.2. A universal filament width?

The distribution of mean filament widths that we obtain in Orion B, no matter which skeleton we consider, is very similar to the one presented by Arzoumanian et al. (2011) for the IC 5146, Polaris and Aquila regions, where a 0.1 pc "typical" width with a 0.03 pc spread was reported, and to simulations results by e.g. Federrath (2016).

In contrast, our statistics of the filament widths are different from what is observed in the Taurus molecular cloud by Panopoulou et al. (2014), based on ^{13}CO ($J = 1 - 0$) observations (Narayanan et al. 2008), with a typical filament width of 0.4 pc. This discrepancy results at least partly from the chosen molecular tracer, which is less adapted for tracing the filamentary material, as discussed in Sect. 2.1. ^{13}CO is susceptible to become optically thick and therefore to better trace the extended, power-law-like envelope of the filament rather than its tubular, central part (Arzoumanian et al. 2011) – which can result in wider FWHM estimations. In the case of Orion B, the measurement of filament widths applied to the ^{13}CO-derived N_H column density yields filament widths of 0.18 ± 0.04 pc, wider than the 0.12 ± 0.04 pc measured with $C^{18}O$.

5.2. Gravitational stability and star formation

5.2.1. Low linear and volume densities

The linear densities of the filaments that we detected range from a few $M_\odot$/ pc to about 100 $M_\odot$/ pc, with a median linear density of $\sim$ 5 $M_\odot$/ pc for all skeletons (Fig. 11, top). This apparently log-normal distribution matches the usual linear densities of interstellar filaments rather well, as observed e.g. in IC 5146 (Arzoumanian et al. 2011), in Taurus (Panopoulou et al. 2014) or in Musca (Kainulainen et al. 2016). It is of course lower than what is observed for high linear-density filaments such as the Integral-Shaped Filament (Stutz & Kainulainen 2015; Kainulainen et al. 2017), for which the linear density is of the order of several 10^2 $M_\odot$/ pc. More precisely, while the upper end of our distribution reaches the typical order of magnitude for linear densities (in the range of a few tens or hundreds of $M_\odot$/ pc), a large fraction of the filamentary network has rather low linear densities.

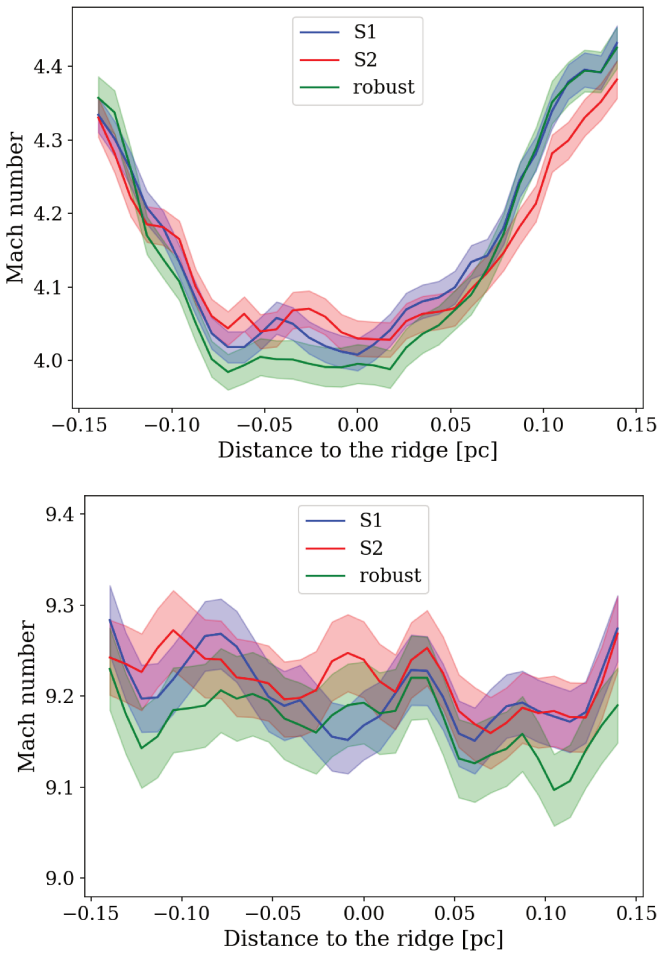


Fig. 9. *Top:* Average transverse profiles of the line-of-sight FWHM velocity dispersion of the filaments, computed for the S1, S2, and robust skeletons, using the modelled $C^{18}O\,(J = 1-0)$ linewidth (the shaded areas show the standard error for each profile). *Bottom:* Same as above, but using the $^{13}CO\,(J = 1-0)$ data-cube for comparison.

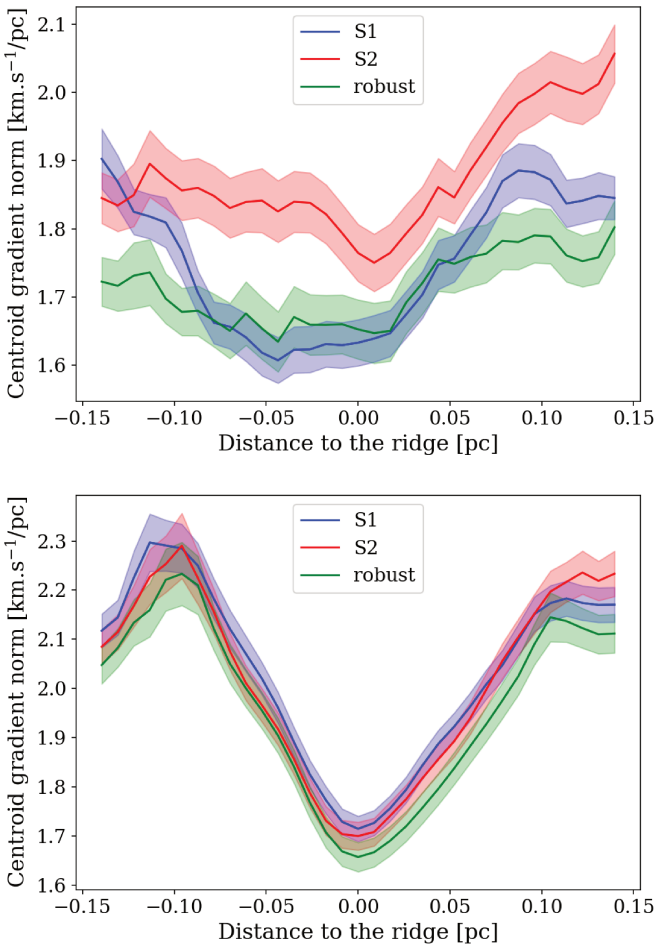


Fig. 10. *Top:* Average transverse profiles of the amplitude of the centroid velocity gradient of the filaments, computed for the S1, S2, and robust skeletons, using the $C^{18}O\,(J = 1-0)$ data-cube (the shaded areas show the standard error for each profile). *Bottom:* Same as above, but using the $^{13}CO\,(J = 1-0)$ data-cube for comparison.

This is also visible when looking at the typical volume densities of the filaments, which are estimated by assigning the linear density to a uniform cylinder, the diameter of which would be the FWHM of the individual filament profile. These volume densities range from 10^4 to 10^5 H cm^{-3}, again with a distribution close to a log-normal one, with a median value of $\sim 2{\times}10^4$ H cm^{-3}(Fig. 11, bottom). This is consistent with the upper end of the volume density distribution in the whole western edge of the Orion B molecular cloud, as presented in Bron et al. (2018). These typical densities are however lower by an order of magnitude than those measured by Teixeira et al. (2016) (and references therein) or Kirk et al. (2015), based on observations of filaments or hydrodynamical simulations, respectively.

We therefore have a set of filaments which contains a few objects matching the usual linear or volume densities found in the literature, but with an excess of low-density elements. There are several possibilities to explain this effect. First, André et al. (2010) and Arzoumanian et al. (2011) note that fitting filament profiles with a Gaussian rather than a Plummer profile can lead to an underestimation of their density by about 20%. However, it will not increase the densities we measured by an order of magnitude, and many studies cited above also used Gaussians to fit the transverse profiles. A more important factor is the detection scheme used in this work. Most studies focus on a small number of well-identified and carefully selected filaments, e.g. by setting high persistence levels when detecting filaments with DisPerSE (Sousbie 2011). In contrast, we use a lax definition of the filaments, resulting in a number of rather faint, but still contrasting and elongated objects to be part of the analysed skeletons. In certain cases, this is desirable (e.g. in the case of the south-eastern extension of the Flame filament, which clearly divides into many sub-structures that we do not want to miss). In other cases, structures that would usually not classify as filaments are retained, such as lower-density striations (i.e. strands of more diffuse gas that is accreting onto a main filament).

5.2.2. No signs of gravitational collapse

The filaments in Orion B are striking by their exceptional stability against gravitational collapse, even when looking at the higher limit of their instability criterion (Fig. 7, left). While the denser filaments in our skeletons match rather well the sample in Arzoumanian et al. (2013), this exceptional stability of the filaments can be explained by several factors.

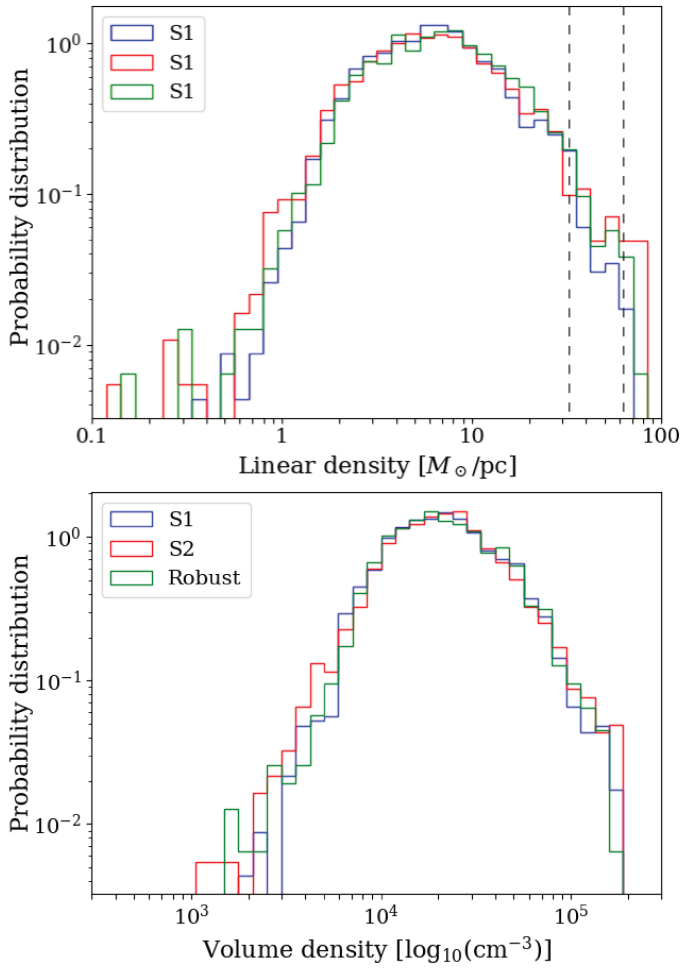


Fig. 11. *Top:* Distribution of the linear density of the filaments. The vertical dashed lines show the most probably critical linear densities corresponding to $M_{\rm l}^{\rm crit}$ and $M_{\rm l,eff}^{\rm crit}$. *Bottom:* Distribution of the volume density of the filaments, derived from the linear density and the FWHM of the transverse profiles by assuming a uniform cylindrical geometry. The typical critical volume densities lie beyond the plot, at $\gtrsim 3 \times 10^5$ cm^{-3}.

First, the western edge of Orion B is a warm environment, heated by the large amount of FUV radiation coming both from outside (σ Ori) and from inside (NGC 2024) the cloud. These high temperatures (Fig. 2, right) lead to high thermal pressures and therefore high critical masses (Eq. 3). Arzoumanian et al. (2013) simply assume a constant temperature of 10 K in the filaments, Teixeira et al. (2016) a constant temperature of 15 K, while the dust temperature in Orion B rarely drops below 20 K. These warm temperatures could either be a layering effect (i.e. the filaments are actually cold but the dust temperature is dominated by the warmer surrounding medium) or an actual specificity of this region of Orion B. In any case, an overestimated or genuinely higher temperature of the gas leads to higher critical linear densities.

Second, turbulence also plays a role in stabilizing the filaments against gravity. Kainulainen et al. (2016) showed that taking turbulence into account brings the Musca filament from a super-critical to trans-critical state. In the case of Orion B, this corresponds to the dramatic difference between the left and right panels of Fig. 7 – although, as mentioned, $\gamma_{\rm eff}$ is a lower limit, as some of the velocity dispersion might come e.g. from infall / collapse (Arzoumanian et al. 2013), and not from rotation or turbulent motions that would support the filaments against gravity. Table 2 summarizes by how much the fraction of the filamentary network prone to gravitational instability would have increased had we assumed no turbulent support or a lower gas temperature, which shows how important velocity-resolved observations of molecular lines are, when trying to determine the stability of filaments.

Table 2. Fraction of super-critical ($\gamma > 1$) or at least trans-critical ($\gamma > 0.5$) filaments depending on the assumptions on their internal pressure, computed for the S1, S2 and robust skeletons.

	$T = 10$ K	$T = T_K$	$T = T_K + T_{\rm turb}$
S1 skeleton			
$\gamma > 1$	9.9%	0.4%	0.1%
$\gamma > 0.5$	33.2%	8.0%	0.4%
S2 skeleton			
$\gamma > 1$	9.3%	1.2%	0.2%
$\gamma > 0.5$	27.2%	7.4%	0.8%
Robust skeleton			
$\gamma > 1$	9.0%	0.6%	0.1%
$\gamma > 0.5$	25.4%	7.6%	0.4%

This overall lack of gravitationally unstable filaments in Orion B correlates well with its known low star-formation efficiency (SFE) (Lada 1992; Carpenter 2000; Federrath & Klessen 2013; Megeath et al. 2016; Orkisz et al. 2017). Moreover, the NGC 2023 and NGC 2024 star-forming regions are among the few regions containing super-critical or trans-critical filaments. This is consistent with the fact that these regions also show the most compressive motions, as measured by Orkisz et al. (2017).

5.2.3. Star formation efficiency

Measurement of the filament masses also enable to check what fraction of the mass of the molecular cloud is contained in the filamentary network, and, in particular, in the gravitationally unstable filaments, as this last fraction directly relates to the SFE of the cloud. Using the ^{12}CO ($J = 1 - 0$) line and following Solomon et al. (1987) and Bolatto et al. (2013) in the same way as in Pety et al. (2017), we obtain a total virial mass of the cloud (for the considered field of view and accounting for the CO-empty IC 434 PDR) comprised between 8400 and 13900 M$_\odot$. For simplicity, we will use the average of these values, at 11100 M$_\odot$. This yields a fraction of mass in the filamentary network of about 4.3 ± 1.1% (474 ± 28 M$_\odot$) in the case of the S1 skeleton, 3.6 ± 0.9% (405 ± 22 M$_\odot$) in the case of S2, and 3.2 ± 0.8% (357 ± 21 M$_\odot$) for the robust skeleton. It is significantly less than the fraction of mass derived for the "environment of filaments" in Pety et al. (2017) (about 40%), which is mostly explained by the sparse character of the filamentary network, compared to an $A_{\rm V}$ extinction mask, and also less than the total mass traced by C^{18}O (about 1200 M$_\odot$, or 11 ± 3% of the cloud's mass). Thus, only about 1/3 of the C^{18}O-traced mass is found in filaments.

The fraction of mass in gravitationally super-critical or trans-critical filaments depends on the definition of the instability criterion. When using γ, we have 266 ± 17 M$_\odot$, or 2.4 ± 0.6% of the cloud's mass in trans-critical filaments, and 111 ± 7 M$_\odot$, or 1.0 ± 0.3% of the mass in super-critical filaments. When using $\gamma_{\rm eff}$, the fraction of the cloud's mass in trans-critical filaments is 0.9 ± 0.2% (95 ± 6 M$_\odot$), with only 0.1% of the mass (21 ± 1 M$_\odot$) in super-critical filaments. To first order, the fraction of the mass

of a molecular cloud contained in gravitationally unstable structures can be directly related to the SFE of that cloud, as we can roughly assume that this mass is going to collapse into cores and into stars. Among the fractions mentioned above, the fraction of the cloud's mass in the trans-critical filaments using the gravitational instability criterion γ can be considered as the upper limit for the SFE. Such a value of $2.4 \pm 0.6\%$ is consistent with previous measurements of the SFE by Lada (1992), Carpenter (2000), Federrath & Klessen (2013) or Megeath et al. (2016), which range from 0.4% to 3%.

Lada et al. (2010) propose that the star formation rate of molecular clouds is proportional to the mass above a threshold of $0.8\,A_K$(which approximately corresponds to 6 magnitudes of A_V, or $1.1 \times 10^{22}\,cm^{-2}$), while André et al. (2014) find prestellar cores only in regions with column densities higher than $1.4 \times 10^{22}\,cm^{-2}$ (i.e. $A_V \approx 8$ mag). These thresholds are higher than the detection limit of $C^{18}O\,(J = 1 - 0)$ (about 3 magnitudes of A_V, as mentioned in Sect. 2.1, which corresponds to ≈ 0.4 magnitudes of A_K), and exclude part of the filamentary regions (Fig. 2 left, Fig. 3). They delimit a "star-forming" mass of the cloud of $\sim 540\,M_\odot$ – less than the total of the $C^{18}O$-traced mass, but much more than what is contained in trans-critical or super-critical filaments. The typical volume density threshold that we obtain for gravitationally unstable filaments, using a typical value of the critical linear density of $30 - 60\,M_\odot/pc$ (Fig. 11, Eq. 3) and assuming a filament diameter of 0.12 pc, is about $5 \times 10^5\,cm^{-3}$. It corresponds to a column density of $\sim 4 \times 10^{22}\,cm^{-2}$, significantly higher than the threshold column densities proposed by Lada et al. (2010) and André et al. (2014). The $\sim 1.1 \times 10^{22}\,cm^{-2}$ threshold can thus be seen as a general absolute minimum under which the gas contained in molecular clouds does not contribute to star formation, but the actual column density threshold over which star formation is likely to occur in a given cloud can depend on intrinsic properties of the cloud, such as its age or kinematics. In the case of Orion B, in which strong solenoidal motions lower the SFE (Orkisz et al. 2017), this actual threshold is particularly high.

5.2.4. One clear example of longitudinal fragmentation

When trying to identify mechanisms of star-formation in the filamentary network, it is difficult to conclude in favour of radial collapse, or longitudinal collapse leading to star formation via accretion onto hubs (Peretto et al. 2014) based on the relative positions of the YSOs and the filamentary skeleton (Fig. 7). However, another widely observed phenomenon leading to the formation of prestellar cores is longitudinal fragmentation (Hacar et al. 2013; Teixeira et al. 2016; Kainulainen et al. 2016; Hacar et al. 2018). Filaments that are above the gravitational instability limit are expected to collapse radially. But when they come close to their gravitational stability limit ($\gamma \lesssim 1$), filaments are mostly susceptible to fragment longitudinally on scales close to their Jeans length (Takahashi et al. 2013; Teixeira et al. 2016) or to four times their FWHM (Inutsuka & Miyama 1992).

Signs of such fragmentation are visible in at least one individual filament in our field – the Hummingbird filament, which, interestingly, does not harbour any known YSO. Oscillations of the linear density are visible along this filament, forming evenly-spaced "beads". Following the recommendations of Schulz-Dubois & Rehberg (1981), we computed the two-point auto-structure function of the linear density along the curvilinear abscissa of the filament's bone. It is defined as follows

$$F(u) = \left\langle (\lambda(x) - \lambda(x+u))^2 \right\rangle_x , \quad (4)$$

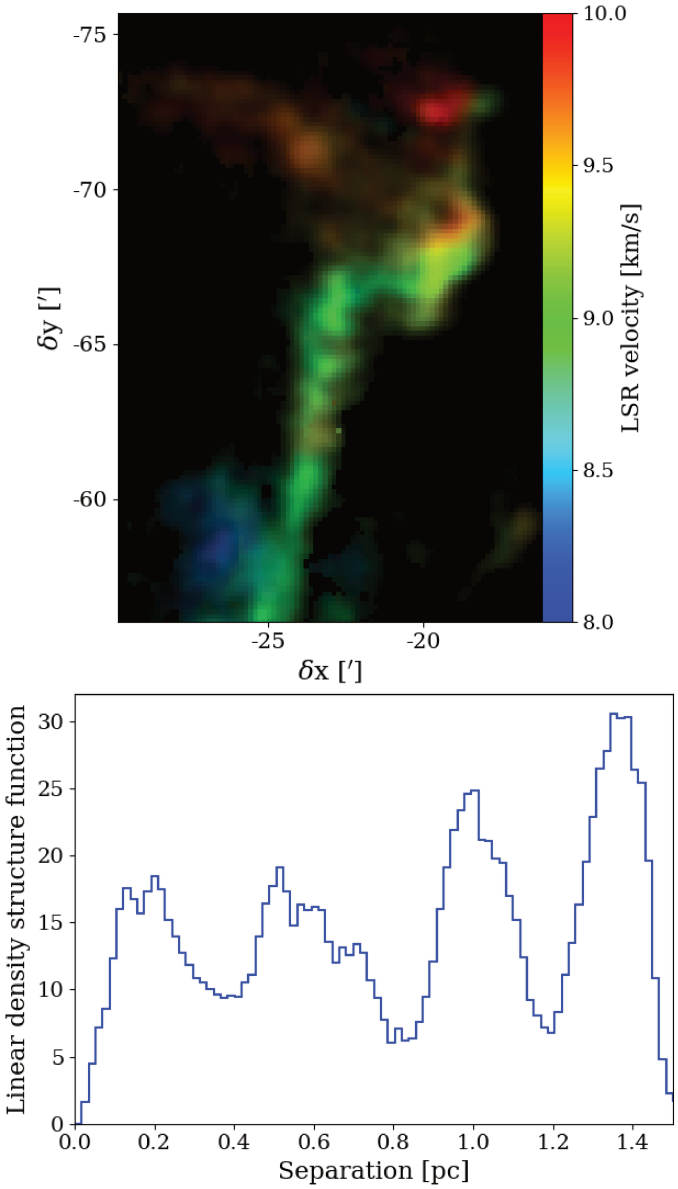


Fig. 12. *Top:* Same as Fig. 1, but with the field and the velocity range zoomed in on the Hummingbird filament, showing "beads" in the rectilinear tail of the Hummingbird. *Bottom:* Two-point structure function of the linear density in the Hummingbird filament (following the robust skeleton), in $(M_\odot/pc)^2$, displaying a periodic pattern which suggests longitudinal fragmentation with a period of 0.4 pc.

where λ is the linear density, x a position along the bone and u the separation between the considered points. The resulting function is plotted in Fig. 12.

The oscillating pattern highlights the presence of evenly-spaced structures with sizes of the order of 0.4 pc. Given that we have a FWHM of 0.11 pc for that individual filament, such a fragmentation length matches well the analytical prediction of Inutsuka & Miyama (1992) for collapsing isothermal filaments, as well as the simulation results of Clarke et al. (2016) for accreting filaments, since both predict a fragmentation with separations of about 4 times the diameter.

Following Spitzer (1998); Takahashi et al. (2013); Teixeira et al. (2016), we compute the Jeans length of the filament,

$l_{\rm Jeans} = \sqrt{(\pi\, c_{\rm eff}^2)/(G\, n_0)}$, where $c_{\rm eff}$ is the effective sound speed, corresponding to the effective temperature , and ρ_0 is the volume density. The mean volume density $n_0 = 1.5 \times 10^4\,{\rm cm}^{-3}$ of the individual filament, derived for the case of a uniform cylinder, combined with an effective temperature $T_{\rm eff} = 43\,{\rm K}$ (while $T_K = 18\,{\rm K}$) yields a Jeans length of 0.54 pc, which is slightly higher than the observed characteristic scale in Fig. 12. If we replace this average density by an estimate of the maximum density (obtained by comparing the average value of a Gaussian over its FWHM to its peak value, and correcting for the transition from two to three dimensions: $\rho_{0,\rm max} = \left\langle \exp(-x^2/2) \right\rangle_{\rm FWHM}^{-3/2} \cdot \rho_0 \approx 1.4 \cdot \rho_0$), we obtain a Jeans length of 0.38 pc, even closer to the observed fragmentation length.

In summary, filaments in Orion B show no signs of radial collapse, no clear evidence (at this stage of the analysis) of longitudinal collapse onto hubs, but at least one good example of longitudinal fragmentation.

5.3. Magnetic field

The features of the distribution of relative orientations between the filaments and the magnetic field (Fig. 8) are difficult to interpret, as they neither match the parallel orientation expected for sub-critical structures, nor the perpendicular orientation expected for super-critical structures (Nagai et al. 1998). Indeed, assuming a random orientation in three-dimensional space of the [filament and magnetic field] pairs of vectors, the observed distribution of relative orientations should feature a narrow peak around 0° if the filaments and the magnetic field were always parallel, or a broad peak around 90° if they were always perpendicular to each other. However, the distribution of relative orientations also suggests that these quantities are not uncorrelated either, as shown by the significant differences between the actually observed and the randomly sampled distributions on Fig. 8.

When analysing these results, we must take into account that we are dealing with magnetic field data at low angular resolution. Therefore, the analysis of Nagai et al. (1998) cannot be directly applied to the results obtained on the stability of the filaments in Orion B, since the filaments are absolutely unresolved at 5′, and the blurred structures detected at this resolution have different mass properties – they are not the direct progenitors of dense cores.

The favoured projected relative orientation of 40–60° can be explained by two phenomena: line-of-sight or beam averaging and projection effects. Averaging would imply that the observed polarization is the mixture of emission from regions with parallel and regions with perpendicular orientations between the filaments and the magnetic field, along the line of sight or in a single beam. This is all the more likely that the polarization fraction is decreasing with increasing density (Planck Collaboration Int. XIX 2015; Planck Collaboration Int. XX 2015), so that the contribution to the polarization orientation from non-filamentary regions can be important. Projection effects can also lead to such an observed relative orientation. This implies a dominating perpendicular orientation in three dimensions, and a favoured projection angle of the [filament and magnetic field] pairs which would result in an observed angle between the two vectors close 50° (or 130°). This projection effect seems possible when looking at the geometry of the magnetic field orientation (Fig. 2), which seems to mostly follow the orientation of the edge of IC 434 PDR, and is disturbed by NGC 2024. If the magnetic field is mostly shaped by the ionized regions on large scales (Planck Collaboration Int. XXXIV 2016), it is possible that most of the observed filaments have rather high inclinations with respect to the plane of the sky and are perpendicular to a magnetic field embedded in planes parallel to the IC 434 PDR. Measurements of the Zeeman splitting of CN lines would enable the determination of the magnetic field intensity along the line of sight, providing further information on the 3D orientation of the magnetic field and thus constraining the geometry of the magnetic field.

5.4. Kinematics

The ^{13}CO ($J = 1 - 0$) line traces the moderately dense gas that forms the bulk of molecular clouds and surrounds the filamentary network. Towards filament ridges, a substantial fraction of the ^{13}CO ($J = 1 - 0$) emission can start to originate from the filaments, traced by the C^{18}O ($J = 1 - 0$) emission. The kinematics of ^{13}CO could thus trace the transition between the turbulent environment and the quiescent inner part of filaments (e.g., Hatchell et al. 2005; Federrath 2016). Towards the centre of the filaments, the norm of the centroid velocity gradient decreases, reaching at its minimum the same value as for C^{18}O ($J = 1 - 0$). The "wings" of these profiles have an uncertain origin: they are not analysis artefacts, as no other profiles show a similar behaviour, and might instead be a result of the crowded field, showing the influence of neighbouring filaments. We would expect the Mach number to decrease as well, but it shows flat profiles instead. This could be explained by the fact that the optical depth of ^{13}CO ($J = 1 - 0$) increases towards the center of the filaments, resulting in a non-negligible opacity broadening which compensates for the decrease in linewidth due to a lower velocity dispersion. Indeed, Orkisz et al. (2017) have shown that the ^{13}CO ($J = 1 - 0$) opacity broadening is negligible except in the dense regions that we are interested in here. For the column densities of ^{13}CO typical of the densest filaments ($N_{^{13}\rm CO} \sim 1 \times 10^{17}\,{\rm cm}^{-2}$), the opacity can reach $\tau_{^{13}\rm CO} \gtrsim 3$ and lead to a line broadening by a factor ~ 1.5, which is consistent with the absence of variation of the FWHM across the filaments, the decrease in velocity dispersion being compensated by an increase in opacity broadening. The overall high Mach number traced by ^{13}CO ($J = 1 - 0$), compared to the average of ~ 6 given in Orkisz et al. (2017), is mostly due to presence of cold but strongly agitated regions in the East which were not part of the field studied by Orkisz et al. (2017).

The C^{18}O ($J = 1 - 0$) results are contrasting with ^{13}CO ($J = 1 - 0$). The velocity dispersion *along the line of sight* is supersonic (Mach number $\mathcal{M} \sim 4$), much higher that the transonic regime which is usually observed or predicted (e.g., Arzoumanian et al. 2013; Federrath 2016). However, the Mach number decreases significantly inside the filaments, which can be the sign of the dissipation of turbulence (Hennebelle 2013). The norm of the centroid velocity gradient possibly shows a similar, but far less pronounced decrease towards the centre of the filaments. One can also notice that, with filament widths of the order of 0.1 pc and centroid velocity gradients of $< 2\,{\rm km\,s^{-1}\,/\,pc}$, the centroid velocity variations *in the plane of the sky* are subsonic, given typical sound speeds of the order of 0.3–0.7 km s^{-1}. This is reminiscent of the results of (Smith et al. 2016), who show that filaments are structures which move coherently on large scales (and thus have near-constant centroid velocities), regardless of the small-scale turbulence.

In the context of the velocity field of filaments, one element often discussed in the presence of fibres (Hacar et al. 2013; Panopoulou et al. 2014; Hacar et al. 2018). One would expect from the possible presence of unresolved fibres that there would

be multiple spectral components detected in many lines of sight in the filaments. However, it appears to be rarely the case. Only the part of the Flame filament closest to the NGC 2024 star-forming region displays a long portion that consistently has two or more identified spectral components. This questions the presence of fibres, since it would mean that they are not only unresolved spatially, but also spectrally, since in most cases, their presence would merely broaden the $C^{18}O$ ($J = 1-0$) line instead of showing more spectral peaks (which could then possibly explain partly the high Mach numbers even at the center of the filaments).

A further understanding of the dynamics of these filaments using the ORION-B data-set would imply to perform a three-dimensional identification of the structures, which would improve the understanding of the spectral multiplicity of the filaments and a better view of their crossings and hubs, and a study of the velocity not only across, but also along the filaments, studying their position-velocity diagrams, in a fashion similar to, e.g., Peretto et al. (2014). For example, the observed longitudinal fragmentation in the Hummingbird filament appears to be associated with a specific velocity pattern that will be explored in detail in a future paper.

5.5. Statistical influence of the filament extraction method

The study presented here shows that the difference between the statistical results of the different skeletons are modest, if not negligible. This could be expected from the significant overlap of the skeletons, which is in turn related to the consistent manner in which the tuning of the detection parameters was done, be it for the initial extraction (Appendix B.1) or during the cleaning process (Appendix B.2), for both the S1 and S2 skeletons.

When comparing Fig. 5 and 7, one can see that most of the individual filaments that are exclusive to either S1 or S2 are particularly stable, i.e. they are among the least dense structures in the field of view. The only significant structures that are not in common are the Horsehead PDR and a small portion of the Flame filament in NGC 2024. This last point can explain the difference in percentage of mass in super-critical or trans-critical filaments observed between S1, S2 and the robust skeleton, as the missing fragment of NGC 2024 in S1 contains a significant amount of mass above the gravitational instability threshold. The similarity of the results obtained with the robust skeleton on the one hand compared to those with the S1 and S2 skeletons on the other hand is also important, because it shows that the structures not detected by either S1 or S2 do not play a major role in the statistical properties of the filamentary network.

Provided that the definition of a filament is consistent with what was used in this paper (in particular in terms of density or length requirement) it is safe to assume that filament detection schemes do not play a significant role in the difference between different studies of interstellar filaments.

6. Conclusion

In this paper, we have used velocity-resolved line maps obtained through IRAM-30m observations in the context of the ORION-B project to trace the dense, filamentary matter in the Orion B molecular cloud with the $C^{18}O$ ($J = 1-0$) line. Using two different extraction methods, we have identified the network of filaments in the cloud, and have used both extracted skeletons as well as their intersection to compare the statistical properties of the filaments they contain.

The main results regarding these filaments are the following.

1. Given the very coherent filament detection criteria between the two extraction methods (and despite the technical disparities) the statistical properties of the detected structures in any skeleton can be considered as quasi-identical.
2. The filaments display a typical width of $\sim$ 0.12 pc, with a narrow spread of $\pm$0.04 pc (Fig. 6). This value seems to be free from detection bias, and is supported by the width of the variations in the velocity field (Figs. 4 and 10);
3. The upper ends of the distributions of linear densities and volume densities of the filaments are consistent with observations and simulations of interstellar filaments. However, many extracted filaments have lower densities. This suggests that the criteria generally used to identify filaments are too restrictive.
4. The filaments in Orion B are stable against gravitational collapse, due to their relatively lukewarm temperatures and their supersonic velocity dispersions. This is consistent with the very low star formation efficiency of Orion B, at about 1%;
5. At least one filament (the Hummingbird filament, lying north of NGC 2024) shows visible signs of periodic longitudinal fragmentation, despite being clearly gravitationally subcritical.
6. In the vicinity of the filaments, the velocity dispersion is larger in ^{13}CO ($J = 1-0$) than in $C^{18}O$ ($J = 1-0$). The surroundings of the filaments are thus more turbulent than their inner part. However, the $C^{18}O$ ($J = 1-0$) velocity dispersions, both along the line of sight and in the plane of the sky, seem to increase near the centre of the filaments. This feature can be explained either by accretion and/or shocks, or by the presence of unresolved fibres which make up the structure of the filament at smaller scales.
7. It is difficult to clearly check whether the relative orientation of the filaments and the magnetic field differs for subcritical and trans-critical filaments, because the resolution of the *Planck* data is one order of magnitude coarser than the measured width of the Orion B filaments. We here tentatively favour a perpendicular orientation combined with a significant inclination of the filaments.

The results 3 to 6 speak in favour of the filamentary structures in Orion B being in a young evolutionary stage, meaning that the cloud might eventually evolve into a more active environment like its neighbour Orion A. This is consistent with the relatively young age of the HII regions within the cloud, estimated at $\sim$ 100 000 and 200 000 years for NGC 2023 and NGC 2024 respectively (Tremblin 2014, priv. comm.), and also with the fact that the fraction of very young protostars among all YSOs is significantly larger in the south-western part of Orion B than in Orion A (24% vs. 1.5%, Stutz et al. 2013).

This paper mostly focused on the transverse properties of the filaments. The longitudinal properties were preliminarily tackled in the specific case of the Hummingbird filament, but they deserve an in-depth study as well, because signs of longitudinal fragmentation are suspected at other locations in the cloud, and are apparently accompanied by longitudinal velocity patterns. The Hummingbird filament itself will be the object of a follow-up study at higher angular resolution, which will shed light onto the entire filamentary network of Orion B.

Polarimetry measurements at higher angular resolution are also required to better understand how the magnetic field interacts with gas structures on the $\sim$ 0.1 pc scale. Measurements of the dust continuum polarization using the NIKA-2 (Catalano

et al. 2018; Adam et al. 2018) camera at the IRAM-30m telescope could provide a major improvement in that regard.

From a methodological point of view, the Hessian approach (Appendix B.1) can be easily extended to perform the filament detection in three dimensions. Such a three-dimensional detection making full use of the velocity information would facilitate the analysis of the longitudinal velocity structure of the filaments, and possibly enable the detection of fibres (Hacar et al. 2013).

Thanks to the multi-tracer nature of the ORION-B project, the filamentary skeletons could be used to stack the observed spectra in order to reveal faint molecular lines and characterize the molecular signature of filamentary regions, in a manner similar to Gratier et al. (2017) and Bron et al. (2018). Conversely, classification tools such as the clustering used by Bron et al. (2018) could possibly reveal the presence of different families of filaments, based on their properties derived in this paper, which could point at different evolutionary stages or scenarios for the filaments in Orion B.

Acknowledgements. We thank Pr. Maciej Orkisz for pointing us to the work of Frangi et al. (1998) and helping us with the understanding and implementation in an astrophysical context of this analysis method coming from the field of medical imaging. We also thank Pr. Jouni Kainulainen for stimulating discussions and helpful remarks. This research has made use of data from the Herschel Gould Belt survey (HGBS) project (http://gouldbelt-herschel.cea.fr). The HGBS is a Herschel Key Programme jointly carried out by SPIRE Specialist Astronomy Group 3 (SAG 3), scientists of several institutes in the PACS Consortium (CEA Saclay, INAF-IFSI Rome and INAF-Arcetri, KU Leuven, MPIA Heidelberg), and scientists of the Herschel Science Center (HSC). This work was supported by the CNRS/CNES program "Physique et Chimie du Milieu Interstellaire" (PCMI). We thank the CIAS for its hospitality during the three workshops devoted to this project. PG thanks ERC starting grant (3DICE, grant agreement 336474) for funding during this work. JRG thanks the Spanish MINECO for funding support under grant AYA2017-85111-P.

References

Adam, R., Adane, A., Ade, P. A. R., et al. 2018, A&A, 609, A115
André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, Protostars and Planets VI, 27
André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
Arzoumanian, D., André, P., Didelon, P., et al. 2011, A&A, 529, L6
Arzoumanian, D., André, P., Peretto, N., & Könyves, V. 2013, A&A, 553, A119
Aylward, S. R. & Bullitt, E. 2002, IEEE transactions on medical imaging, 21, 61
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Bron, E., Daudon, C., Pety, J., et al. 2018, A&A, 610, A12
Burkert, A. & Hartmann, L. 2004, ApJ, 616, 288
Busquet, G., Zhang, Q., Palau, A., et al. 2013, ApJ, 764, L26
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Carpenter, J. M. 2000, AJ, 120, 3139
Catalano, A., Adam, R., Ade, P. A. R., et al. 2018, Journal of Low Temperature Physics
Clarke, S. D., Whitworth, A. P., & Hubber, D. A. 2016, MNRAS, 458, 319
Federrath, C. 2016, MNRAS, 457, 375
Federrath, C. & Klessen, R. S. 2013, ApJ, 763, 51
Fischer, W. J., Megeath, S. T., Stutz, A. M., et al. 2013, Astronomische Nachrichten, 334, 53
Florack, L. M., ter Haar Romeny, B. M., Koenderink, J. J., & Viergever, M. A. 1992, Image and Vision Computing, 10, 376 , information Processing in Medical Imaging
Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. 1998, in Medical Image Computing and Computer-Assisted Intervention (Springer Berlin Heidelberg), 130–137
Friesen, R. K., Pineda, J. E., co-PIs, et al. 2017, ApJ, 843, 63
Furlan, E., Fischer, W. J., Ali, B., et al. 2016, ApJS, 224, 5
Gerin, M., Ruaud, M., Goicoechea, J. R., et al. 2015, A&A, 573, A30
Goldsmith, P. F. 2001, ApJ, 557, 736
Gratier, P., Bron, E., Gerin, M., et al. 2017, A&A, 599, A100
Hacar, A., Kainulainen, J., Tafalla, M., Beuther, H., & Alves, J. 2016, A&A, 587, A97
Hacar, A., Tafalla, M., Forbrich, J., et al. 2018, A&A, 610, A77
Hacar, A., Tafalla, M., Kauffmann, J., & Kovács, A. 2013, A&A, 554, A55
Hartmann, L. & Burkert, A. 2007, ApJ, 654, 988
Hatchell, J., Richer, J. S., Fuller, G. A., et al. 2005, A&A, 440, 151
Hennebelle, P. 2013, A&A, 556, A153
Inutsuka, S.-I. & Miyama, S. M. 1992, ApJ, 388, 392
Kainulainen, J., Hacar, A., Alves, J., et al. 2016, A&A, 586, A27
Kainulainen, J., Stutz, A. M., Stanke, T., et al. 2017, A&A, 600, A141
Kirk, H., Klassen, M., Pudritz, R., & Pillsworth, S. 2015, ApJ, 802, 75
Kirk, H., Myers, P. C., Bourke, T. L., et al. 2013, ApJ, 766, 115
Koch, E. W. & Rosolowsky, E. W. 2015, MNRAS, 452, 3435
Koenderink, J. J. 1984, Biological cybernetics, 50, 363
Kong, S., Arce, H. G., Feddersen, J. R., et al. 2018, ArXiv e-prints
Könyves, V., André, P., Men'shchikov, A., et al. 2015, A&A, 584, A91
Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687
Lada, E. A. 1992, ApJ, 393, L25
Liszt, H. 2014, ApJ, 783, 17
Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, A&A, 566, A45
Mangum, J. G. & Shirley, Y. L. 2015, PASP, 127, 266
Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, AJ, 151, 5
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Molinari, S., Swinyard, B., Bally, J., et al. 2010, A&A, 518, L100
Müller, H. S., Schlöder, F., Stutzki, J., & Winnewisser, G. 2005, Journal of Molecular Structure, 742, 215
Nagai, T., Inutsuka, S.-i., & Miyama, S. M. 1998, ApJ, 506, 306
Narayanan, G., Heyer, M. H., Brunt, C., et al. 2008, ApJS, 177, 341
Orkisz, J. H., Pety, J., Gerin, M., et al. 2017, A&A, 599, A99
Ostriker, J. 1964, ApJ, 140, 1056
Padoan, P., Juvela, M., Goodman, A. A., & Nordlund, Å. 2001, ApJ, 553, 227
Panopoulou, G. V., Psaradaki, I., Skalidis, R., Tassis, K., & Andrews, J. J. 2017, MNRAS, 466, 2529
Panopoulou, G. V., Tassis, K., Goldsmith, P. F., & Heyer, M. H. 2014, MNRAS, 444, 2507
Parvathi, V. S., Sofia, U. J., Murthy, J., & Babu, B. R. S. 2012, ApJ, 760, 36
Peretto, N., André, P., Könyves, V., et al. 2012, A&A, 541, A63
Peretto, N., Fuller, G. A., André, P., et al. 2014, A&A, 561, A83
Peretto, N., Fuller, G. A., Duarte-Cabral, A., et al. 2013, A&A, 555, A112
Peretto, N., Hennebelle, P., & André, P. 2007, A&A, 464, 983
Pety, J. & Falgarone, É. 2000, A&A, 356, 279
Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98
Planck Collaboration Int. XIX. 2015, A&A, 576, A104
Planck Collaboration Int. XX. 2015, A&A, 576, A105
Planck Collaboration Int. XXXII. 2016, A&A, 586, A135
Planck Collaboration Int. XXXIV. 2016, A&A, 586, A137
Planck Collaboration XIX. 2011, A&A, 536, A19
Polychroni, D., Schisano, E., Elia, D., et al. 2013, ApJ, 777, L33
Rieke, G. H. & Lebofsky, M. J. 1985, ApJ, 288, 618
Salji, C. J., Richer, J. S., Buckle, J. V., et al. 2015, MNRAS, 449, 1782
Schaefer, G. H., Hummel, C. A., Gies, D. R., et al. 2016, AJ, 152, 213
Schisano, E., Rygl, K. L. J., Molinari, S., et al. 2014, ApJ, 791, 27
Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, ApJ, 786, 29
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Schneider, S. & Elmegreen, B. G. 1979, ApJS, 41, 87
Schulz-Dubois, E. O. & Rehberg, I. 1981, Applied Physics, 24, 323
Smith, R. J., Glover, S. C. O., & Klessen, R. S. 2014, MNRAS, 445, 2900
Smith, R. J., Glover, S. C. O., Klessen, R. S., & Fuller, G. A. 2016, MNRAS, 455, 3640
Sofia, U. J., Lauroesch, J. T., Meyer, D. M., & Cartledge, S. I. B. 2004, ApJ, 605, 272
Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730
Sousbie, T. 2011, MNRAS, 414, 350
Spitzer, L. 1998, Physical Processes in the Interstellar Medium, 335
Stutz, A. M. & Kainulainen, J. 2015, A&A, 577, L6
Stutz, A. M., Tobin, J. J., Stanke, T., et al. 2013, ApJ, 767, 36
Takahashi, S., Ho, P. T. P., Teixeira, P. S., Zapata, L. A., & Su, Y.-N. 2013, ApJ, 763, 57
Teixeira, P. S., Takahashi, S., Zapata, L. A., & Ho, P. T. P. 2016, A&A, 587, A47
Ward-Thompson, D., Kirk, J. M., André, P., et al. 2010, A&A, 518, L92
Williams, G. M., Peretto, N., Avison, A., Duarte-Cabral, A., & Fuller, G. A. 2018, ArXiv e-prints
Wilson, T. L. & Rood, R. 1994, ARA&A, 32, 191
Zhang, T. Y. & Suen, C. Y. 1984, Communications of the ACM, 27, 236

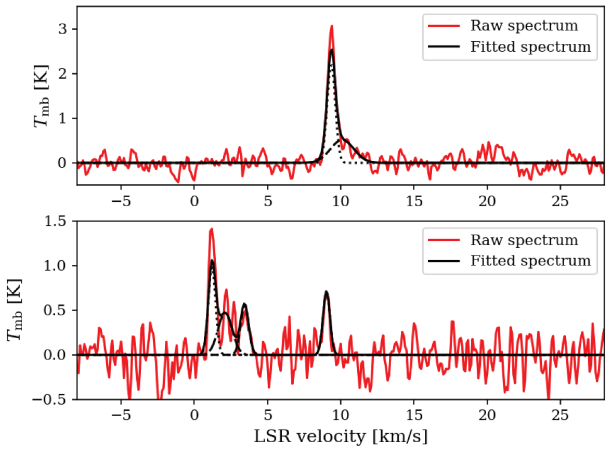


Fig. A.1. Examples of spectra fitted with multiple Gaussians, at signal-to-noise levels of about 15 (*top*) and 5 (*bottom*). The individual Gaussian components are marked in dashed, dotted, or dash-dotted lines.

Appendix A: Multi-Gaussian fitting of the molecular spectra

In order to transform the noisy observational data into a model cube, we first perform a Gaussian smoothing the $C^{18}O$ ($J = 1–0$) VESPA spectra with a width of 3 velocity channels to improve the peak detection. Significant peaks in the spectrum, above a threshold of 4σ, are detected, then a multi-Gaussian fit is performed on the spectrum, with one Gaussian component per identified peak. After that, the reduced χ^2 is computed over 6 channels on either side of the peak (13 channels in total), and compared to the noise level of the spectrum.

If the χ^2 is lower than 0.5σ, an unnecessary Gaussian component is removed. When it is larger than a threshold of 2σ, an extra Gaussian component is added in the corresponding velocity range. Our initial determination of the number of significant peaks tends to be overestimated. For that reason, during the iterative process, only one extra component can be added. Conversely, the number of components can decrease to a minimum of one. This allows the fitting process to take into account, e.g., components separated by less than their average velocity dispersion, which do not present one separate peak per component. But it also prevents over-fitting, whereby a single spectral component would be reproduced by many Gaussian peaks. Finally, we visually inspect the residual cube to check that no obvious signal feature remains unfitted. Figure A.1 shows two examples of the fit of multicomponent spectra.

Spatial correlations between neighbouring pixels are not taken into account during the fit. However, the retrieved moment maps, and in particular the integrated intensity, do not show any significant incoherence (Fig. 2, left).

Appendix B: Structural detection of the filaments: methodological details

B.1. Step by step extraction of the filaments

After obtaining from the raw observational data a modelled cube and a column density estimate (Fig. B.1, panels 1, 2 and 3), we use this column density map to identify the filamentary structures in Orion B. This section details the extraction of the skeletons used throughout this paper. We restrict ourselves to study filamentary structures in a 2D map but the concepts and implementations can be generalized to a position-position-velocity cube.

Table B.1. Topography of a two-dimensional dataset depending on the eigenvalues of its Hessian matrix ε_1 and ε_2. L indicates a low value, H- and H+ a highly negative (respectively positive) value.

ε_1	ε_2	Structure
L	L	No preferred direction
L	H-	Ridge
L	H+	Valley
H-	H-	Peak (local maximum)
H-	H+	Pass (saddle point)
H+	H+	Hole (local minimum)

B.1.1. The Hessian approach

We have qualitatively and observationally defined filaments as elongated, over-dense structures in the molecular ISM. Their identification can be compared with that of mountain ranges in an altitude map. These mountain ranges, and their ridge lines, can be simply defined in terms of topography, that is in terms of differential properties of the studied map.

The Hessian matrix (i.e., the matrix of the second order partial derivatives) provides the most useful insight into the topography: it measures the variations of the slopes, and therefore enables us to locate the valleys, summits and ridges. The eigenvectors of the Hessian indicate the main directions of variations of the slope, and the eigenvalues (ε_1 and ε_2 in our two-dimensional case) whether the terrain is passing through a minimum or a maximum in these two directions. In the case of filaments or ridges, we expect the "altitude" to vary very little in one direction, and to reach a local maximum in the other direction – we therefore expect a second derivative with a value close to zero along the ridge, and a strongly negative one perpendicular to the ridge (if we were working with absorption data, the reasoning would remain the same, albeit with reversed signs). We can summarize the behaviours of ε_1 and ε_2 in terms of characteristic structures in Table B.1.

The computation of the Hessian itself makes use of the concepts of linear scale space theory (Florack et al. 1992; Koenderink 1984), as advised by Frangi et al. (1998). Differentiation is thus performed by convolving the field with derivatives of n-dimensional Gaussians, which allows to simultaneously smooth out the noise and adapt to the typical spatial scale of searched structures (Frangi et al. 1998; Salji et al. 2015) – in our case, as no initial assumption is made on the length of filaments, the only relevant scale is their width. Once the smooth Hessian has been obtained and diagonalized, we can use the eigenvalues as a filter to extract filamentary regions from the map. In this work two implementations of such filters have been used.

B.1.2. Filament extraction by thresholding

The first approach is meant to be as simple as possible, and is similar to the one presented in Planck Collaboration Int. XXXII (2016). The Hessian is directly computed for the map of $C^{18}O$-derived column density (Fig. B.1, panel 4), and the analysis focused on a single spatial scale, set to 0.13 pc (corresponding to

0.11 pc after deconvolution). This choice of scale is the result of an iterative approach. The first guess for the size of the filaments was assumed to be 0.1 pc, following Arzoumanian et al. (2011). The distribution of filament widths yielded a peak consistent with this initial guess – which could have been a detection bias. The analysis was therefore repeated with a Hessian smoothing scale of 0.05 pc and 0.2 pc, and in both cases the peak of the filament width distribution was of 0.11±0.01 pc. Therefore the final choice of detection scale is set to match as well as possible the scale properties of the dataset.

We then threshold the eigenvalues. This is a quantitative way to transcribe the characteristics of a filament in terms of eigenvalues, as seen in Table B.1. If the eigenvalues are first sorted so that $|\varepsilon_1| < |\varepsilon_2|$, then selecting the pixels where $\varepsilon_2 < \tau < 0$ (τ being a global threshold) ensures that we are in the vicinity of a local maximum in the direction perpendicular to the filament. No condition is set on the smaller eigenvalue, so that we allow any kind of peak or saddle point, as long as the relief is not steeper along the ridge than perpendicular to the ridge (Fig. B.1, panel 5). We find the best value of the threshold τ to be 4% of the lowest (negative) eigenvalue in the map, based on visual inspection (Fig. B.1 left).

B.1.3. Adaptive filament extraction with ridge-detection

For the second approach, several additional features refine the filament detection method. Koch & Rosolowsky (2015) apply an arctan transform to the intensities in order to suppress bright point sources (e.g., dense cores) that can dominate the fainter filamentary structures which are searched for. In our case, we find it better to use the asinh transform used in Gratier et al. (2017) (Fig. B.2, panel 1), which is linear at small intensities, and logarithmic at high intensities. The asymptotic behaviour of the arctan transform would flatten the bright regions and thus render the Hessian approach ineffective. We therefore use

$$\widetilde{N_{\rm H}} = a \cdot \mathrm{asinh}(N_{\rm H}/a) \tag{B.1}$$

as an input for filament detection, with the parameter a set to $a = 5.18 \times 10^{21}$ cm^{-2}, i.e. the median of non-zero values of the C^{18}O-derived column density.

This second filament extraction method takes into account the *local aspect ratio* of the studied field. The concepts behind this method have been first described in the field of medical imaging, for the purpose of identifying blood vessels in magnetic resonance imaging (MRI) or computer tomography (CT) images (Frangi et al. 1998), and have been adapted to astrophysical data and extensively illustrated by Salji et al. (2015). The map of Hessian eigenvalues computed from $\widetilde{N_{\rm H}}$ is filtered with the following function:

$$V_0 = \begin{cases} 0 & \text{if } \varepsilon_2 > 0, \\ \exp\left(-\dfrac{R^2}{2b^2}\right)\left(1 - \exp\left(-\dfrac{S^2}{2c^2}\right)\right) & \text{otherwise.} \end{cases} \tag{B.2}$$

where $R = \varepsilon_1/\varepsilon_2$ is the local aspect ratio (here again $|\varepsilon_1| < |\varepsilon_2|$), and $S = \sqrt{{\varepsilon_1}^2 + {\varepsilon_2}^2}$ is the local amplitude of the second derivative. In other terms, the filter function will emphasize pixels where the signal varies significantly ($S > c$), and the local aspect ratio is large ($R < b$).

Following the recommendations of Frangi et al. (1998), the parameter b was set to $b = 0.5$, and c was set to half the maximum value of the Hessian norm S over the field of view. The resulting filtered map can be seen on Fig. B.2 (panel 2).

An extra step is added by implementing a *ridge-detection* function, which is, as far as we know, a novel addition, but can be compared to the centreline extraction method of Aylward & Bullitt (2002). The goal here is to narrow down the response of the filter, so that the obtained structures are as close as possible to the ridge lines (i.e. the skeleton) of the filaments, rather than being broad filamentary regions. This function makes use of the fact that the component of the gradient perpendicular to the filament should go through zero at the ridge line, hence the norm of the gradient should reach a local minimum also close to zero if the variations along the ridge line are negligible. We therefore compute the norm of the gradient of our map, $|\boldsymbol{g}| = |\boldsymbol{\nabla}\widetilde{N_{\rm H}}|$, here again through convolution with Gaussian derivatives. The final filter is:

$$V = V_0 \cdot \exp\left(\frac{-|\boldsymbol{g}|^2}{2d^2}\right) \tag{B.3}$$

where the parameter d ensures that the slope $|\boldsymbol{g}|$ is close enough to zero. This parameter was set to 10% of the maximum value of $|\boldsymbol{g}|$ over the field of view (Fig. B.2, panel 3).

To avoid favouring a single scale, we perform this structure extraction in a multi-scale fashion, as advised by Frangi et al. (1998). We run the detection algorithm to obtain the filtered map $V(s)$ with smoothing scales s from 0.06 pc to 0.32 pc in steps of 0.02 pc, and for each pixel we pick up the maximal response among these filters to achieve a scale-adapted filament detection. Thus, the wider structures are detected strongly and stand out well, while the narrower objects in the field are still picked up (Fig. B.2, panel 4). Here again the choice of scales is the result of an iterative approach. A first computation was performed with a 0.1 pc scale only, yielding filaments widths ranging from 0.06 pc (our resolution limit) to about 0.3 pc. Although the widest filaments were rather poorly identified, we kept a 0.06 – 0.32 pc detection range to avoid as much as possible a detection bias, while keeping the scale range reasonably limited.

After that, a global threshold $\tau_{\rm f}$ is applied on the resulting scale-adapted filtered map (which now has values ranging from 0 to 1) to select regions which are close enough to the ridge lines of filaments. By visual inspection, $\tau_{\rm f}$ was set to 0.03 (Fig. B.2, panel 5).

B.1.4. Skeletonization

Once binary masks that identify the filamentary regions have been obtained, they are thinned down until we are left with single-pixel wide skeletons. One of the most common methods of skeletonization is known as morphological thinning, which we perform following the algorithm of Zhang & Suen (1984), as implemented in the Python module `skimage`. The resulting skeletons are shown on the last panels of Fig. B.1 and B.2. As this step of the analysis only makes use of the geometry of the binary masks, and does not take into account the underlying physical map, the extracted centrelines do not necessarily match the ridge lines of the filaments, especially if they do not have a cylindrical geometry. It is therefore all the more useful to try and narrow down the mask as accurately as possible before the skeletonization, as it is done with the gradient filter in Sect. B.1.3.

B.2. Cleaning the skeleton

Separating a filamentary skeleton into individual filaments enables cleaning it by checking if each individual filament matches the assumed definition. For that purpose, a geometrical analysis

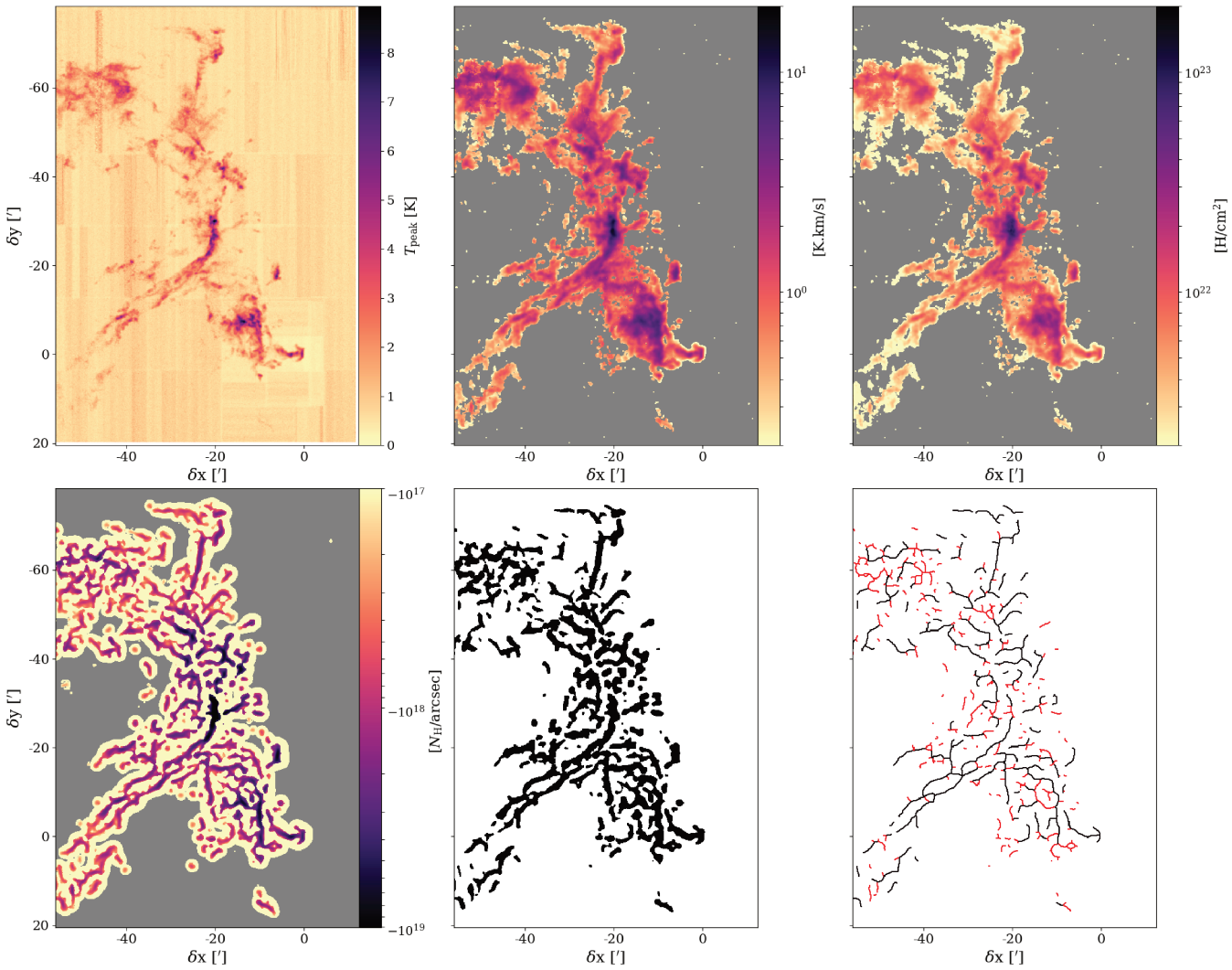


Fig. B.1. *From left to right and top to bottom:* (1) Peak temperature of the raw $C^{18}O$ ($J = 1 - 0$) data; (2) Integrated intensity of the multi-Gaussian fit result; (3) Column density estimate derived from this integrated intensity and the dust temperature; (4) Largest (in absolute value) eigenvalue ε_2 of the Hessian matrix computed at each point of the column density map; (5) Regions identified as filamentary, by applying a threshold onto the ε_2 map; (6) Skeleton obtained by morphological thinning of these regions. The filaments that end up being eliminated at some point in the cleaning process are marked in red.

allows us to distinguish between regular points and nodes (or vertices). The regular points have exactly two neighbours, while the nodes have fewer (if they are endpoints) or more (if they are intersections) than two. The individual filaments are therefore strings of points belonging to the skeleton and linking two nodes. The cleaning process relies on the following criteria: skeleton geometry, curvature radius, relative contrast and width of the individual filaments.

B.2.1. Geometrical cleaning

The first stage of geometrical cleaning is applied immediately after the first extraction of the filamentary skeleton. It consists in removing isolated nodes (i.e. single pixels) and very short filaments: to be able to distinguish between clumps and filaments, we require an aspect ratio of at least 2 and therefore we set a lower length limit at 0.22 pc, i.e. twice the typical filament width. As a dense skeleton can contain many short individual filaments due to the frequency of the intersections, the length selection criterion is only applied to individual filaments that are either isolated or that are so-called spurious branches, i.e. filaments that stick out off the side of a structure, with only one of their extremity being an intersection, the other one being an endpoint.

This stage of cleaning is also reapplied after each of the other stages of cleaning, usually twice in a row to take into account the structure differences that can result from the elimination of spurious branches. Experience shows that two repetitions of this geometrical cleaning are enough to converge to a stable cleaned skeleton – the second repetition is even often superfluous.

B.2.2. Curvature radius

One of the first possible characterizations of the individual filaments is a quantitative analysis of their shape, namely whether the individual filaments are rather straight or very curvy. Several methods are possible to measure how straight or curvy a filament is. A simple one is the measurement of the ratio of the distance between the extremities of the filament and its curvilinear length – but this does not discriminate between a large smooth loop and a jagged structure that is more likely to be an artefact

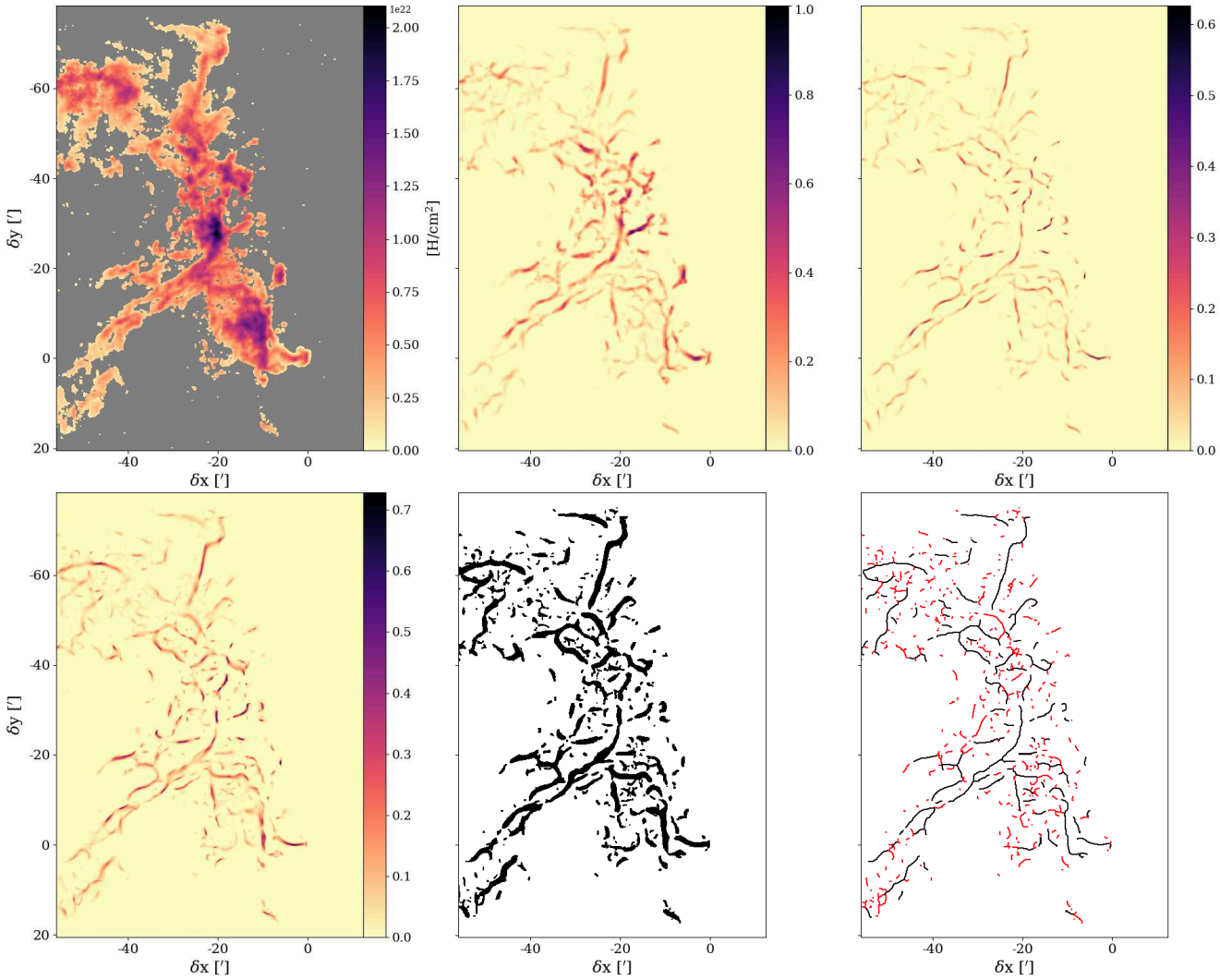


Fig. B.2. *From left to right and top to bottom:* (1) Column density (Fig. B.1, panel 3) after an asinh transform – note that the colour scale is linear, not logarithmic; (2) Structures resulting from the aspect-ratio filter of Frangi et al. (1998) applied to the Hessian eigenvalues of the transformed column density, with a single detection scale of 0.14 pc; (3) The same structures after applying a gradient-based ridge-detection filter, with the same detection scale; (4) The result of the multi-scale (0.06 to 0.3 pc) filtering; (5) Regions identified as filamentary, by applying a threshold onto the multi-scale filtered map; (6) Skeleton obtained by morphological thinning of these regions. The filaments that end up being eliminated at some point in the cleaning process are marked in red.

of skeletonization. Another approach is to compute the circular variance of the filament position angles, but again a large smooth loop might be discarded as its position angles are spread around 360°, while a poorly-identified filament with a position angle varying wildly from −70 to +70° might have a lower circular variance and could be retained. The chosen approach was thus to use the mean curvature radius of the filament – which would discard structures with rapid position angle variation, while keeping large-scale loops.

To that purpose, the first step is estimating the position angle of the filaments at each position. Two possible approaches are possible: either the position angle is directly deduced from the eigenvectors output by the Hessian matrix diagonalization, or it can be geometrically computed on a pixel-by-pixel basis, by comparing the position of each point in the filaments with its nearest neighbours. The first option has the advantage of yielding continuous angles (whereas the second one yields angles in steps of 22.5°) and is reasonably reliable despite some misalignments of the skeletons and the position angle map, so that it is the Hessian angle that is used in the entire paper, and in particular for curvature measurements.

The curvature radius is then simply defined as the inverse of the derivative of the position angle computed along the filament. This curvature radius R_c is then averaged over the individual filament, and is compared to the width w of the individual filament (Sect. 4.1.1). The individual filaments are rejected if they have a R_c/w ratio lower than 1.5, thus ensuring that the transverse profiles of the filament will not be contaminated by a further part of the filament after a sharp turn. This selection affects a large number of individual filaments, mostly short ones, as can be see in Table B.2.

B.2.3. Contrast

The value of the baseline at the peak of the filament profile, obtained from the Gaussian fit (Eq. 2), gives us an estimate of the surface density of the underlying background. An instinctive way of estimating the *contrast* of the filaments, i.e. how

much they stand out from the background, is to take the ratio of the linear density of the filament to the surface density of the background for each line-of-sight. The resulting quantity has the dimension of a length, and can be interpreted as the distance from the filament one needs to go to accumulate as much mass from the surrounding medium as there is in the filament. We can make this contrast dimensionless by defining the *relative contrast*, which is the ratio of the contrast to the filament width. To summarize:

$$\begin{aligned} \text{Contrast} &= \frac{\text{Filament linear density}}{\text{Background surface density}} \\ \text{Relative contrast} &= \frac{\text{Contrast}}{\text{Filament width}} \end{aligned} \tag{B.4}$$

The relative contrast measures how many times denser the filament is relative to its surrounding medium: for instance, a relative contrast of 10 indicates that one needs a region about 10 times broader than the filament to accumulate as much mass as contained in the corresponding portion of filament. This accumulation of mass might be interesting when studying the power-law-like extension of the profile of an isolated filament, but in a skeleton as crowded as the ones studied in this paper, we are mostly concerned by the presence of a clearly visible inner part of the filaments – hence a lower limit imposed on the relative contrast at ~ 1. The distribution of relative contrasts in the non-cleaned filamentary skeletons featured a gap at a relative ratio of 0.6, which we thus chose as a limit under which filaments were rejected. As it can be seen in Table B.2, this criterion only had a moderate impact on the cleaning of the skeleton.

B.2.4. Final selection and summary

In addition to the selection criteria described above, the widths of the filaments are also taken into account to reject filaments that we deem not reliably detected. Two limits are set: one lower limit at 0.06 pc, which corresponds to the resolution of the data (structures of this width or less are unresolved and thus cannot be characterized reliably), and an upper limit of 0.22 pc, which corresponds to a gap in the filament width distribution – the few wider individual filaments are considered as outliers. The lower limit mostly affects the smallest-scale structures detected in the S2 skeleton. These narrow rejected objects are quite numerous. Indeed, the Hessian filter is a derivative, which by essence enhances small-scale, noisy structures, and when the Gaussian smoothing scale comes down to the 0.06 pc spatial resolution of the data, it stops preventing this noise enhancement, and thus many noisy structures at very low column densities are picked up by the filter. The upper limit, on the other hand, mostly affects "bridges" between two neighbouring filaments, which are artefacts of the skeletonization process and are perpendicular to the orientation of well-detected filaments running close to each other.

Table B.2 summarizes all stages of the cleaning process. After the first detection step, the geometrical cleaning is run twice, and the resulting skeletons are used as a starting point (initial state) for the first analysis, which starts by measuring the width and curvature of the filaments, leading to the rejection of a large amount of structures (some of them can be rejected by more than one criterion). After this, the geometrical cleaning is reapplied twice, yielding intermediate skeletons which then undergo the contrast cleaning. This stage of cleaning is the last one, because of the relative contrast computation is quite sensitive to the quality of the transverse profile fitting, and therefore can lead to results too extreme to be analysed for some of the poorly defined structures that get eliminated in the previous stages of cleaning. After this, two last passes of geometrical cleaning yield the final skeletons, which are then used for the rest of the analysis in this paper. Note that the numbers of pixels and filaments in Table B.2 do not always add up properly. This is due to the fact that some structures can be eliminated based on several criteria, and to the geometrical cleaning which removes some isolated nodes left out by other cleaning stage, and which can reduce the number of filaments either by removing the smallest ones or by merging them into larger ones when a spurious branch (and thus an intersection) is eliminated.

Table B.2. Evolution of the skeletons S1 and S2 after successive stages of cleaning, in terms of number of pixels (Pix) and filaments (Fil). The initial skeleton is first geometrically cleaned of short, isolated filaments, then we exclude filaments with too strong curvatures, too low relative contrasts, too broad or too narrow column density profiles. See text for details on why the numbers do not add up properly.

Skeleton	S1		S2	
	Pix	Fil	Pix	Fil
Raw state	5256	457	4837	447
Geometry cleaning	-646	N/A	-909	N/A
Initial state	4610	291	3928	246
Curvature cleaning	-814	-88	-924	-96
Width cleaning	-472	-43	-675	-67
Intermediate state	3664	157	2833	119
Contrast cleaning	-325	-20	-112	-9
Final state	3315	130	2708	106

B.3. Qualities and limitations of the filament detection methods

Having used in parallel skeletons obtained with two different methods, we need to compare the relative merits of each, with respect to the implementation, detection quality and physical implications.

The first method (which yields the skeleton S1) presents the major advantage of being very straightforward in its implementation, providing a simple yet efficient way to identify filamentary regions in a molecular cloud. However, it is not very specific in its geometric requirements when identifying filaments. This means that regions with strong spatial features that are not filamentary are detected (only to be later rejected during the cleaning process, see the filaments marked in red in Fig. B.1, panel 6), and that the identified filamentary regions are quite broad. This broadness is an advantage in terms of completeness, e.g. if one wants to use the mask of Fig. B.1 (panel 5) to obtain molecular line ratios inside and outside the filaments, but it is a disadvantage during the skeletonization, as the position of the skeleton is less precise. This uncertainty on the position of the skeleton can be a problem when measuring the properties of the filament profiles, in particular their widths (Sect. 4.1.1). The binary aspect of the detection method achieved with a single threshold makes also the method unsuitable for a multi-scale approach.

The second method (which yields the skeleton S2) is much more demanding as far as the properties of the identified structures are concerned, which also comes with its lot of pros and cons. The rescaling of the data (which could easily be transposed to the first method) makes the whole detection process less sensi-

tive to the brightest/densest regions, allowing for a better tuning of the filtering parameters. The first stage of filtering achieves a better distinction between filament-like and blob-like structures using the local aspect ratio. Together with the second stage of filtering, it yields much thinner filamentary regions (Fig. B.2, panel 5), enabling a more accurate skeletonization. The use of a continuous filter (with values ranging from 0 to 1) before thresholding makes the method suitable for a multi-scale approach, which reduces possible detection biases. However, as can be seen from the filaments marked in red in Fig. B.2 (panel 6), this approach also yields a large amount of small-scale structures that end up being rejected during the cleaning (19% of rejection in the first geometrical cleaning stage, compared to only 12% in the case of the S1 skeleton). The multi-stage filtering process also requires more parameters to be adjusted by hand (at least two, d and τ_f) than the first method. Another issue arises from the smoothing that occurs during the computation of the gradient: the gradient can be distorted in the vicinity of intersections of filaments, which can result in a distorted (misaligned) skeleton, but more often the filter simply misses the intersections, truncating the filaments.

In summary, one can consider to first order that S1 is an upper limit for filament detection, and S2 is a lower limit. This would be completely true if S2 were nested in S1 (i.e. if we had S2 = robust $\subset$ S1). However, one can see in Fig. 5 that this is not the case, but that nonetheless filaments exclusive to S2 are far less numerous than those exclusive to S1.

It is important to note that the final quality of the obtained skeletons does not only depend on the initial detection method, but also on the cleaning process. In that respect, one can see that a larger proportion of the S2 skeleton (44%) is rejected than in the S1 skeleton (37%). In both cases it is a significant fraction, which shows that the skeletons should not be used without cleaning, lest the statistical results be contaminated by many unwanted structures. One can also add a qualitative remark on the rejected fraction of the skeletons: although the fraction rejected in S1 is smaller, the filaments rejected in S2 seem to be mostly very small structures, with the longer ones being retained. It can be a sign that the multi-scale detection has a drawback tightly linked to its main quality: by being unbiased, it picks up many small features that, on closer inspection, end up not matching the requirements that we set on *bona fide* filaments. It is thus not easy to rule in favour of one or the other detection scheme – at least not in their current state.

One can also question the nature of the detected structures, regarding their geometry and their origin. Filaments are most often described as elongated, almost unidimensional condensations in the three-dimensional turbulent medium – but this mechanism is not the only one that can form elongated structures in the molecular ISM. Geometry and projection effects are a first issue: for example, it is possible that some detected filaments are actually two-dimensional structures seen edge-on. This is most probably the case for at least one filamentary structure detected in Orion B, namely the vertical filament at the base of the Horsehead nebula: rather than a unidimensional structure, it is the edge of the IC 434 ionization front, a wall seen edge-on. Thus, its high observed density is the result of a larger dimension along the line of sight than for other filaments. Another structure can raise the question of its evolutionary scenario: the Horsehead nebula itself. Rather than a condensation in a 3D turbulent medium, it is rather a pillar carved by the IC 434 ionization front.

The reasons to keep all structures in the filamentary network, even when the knowledge of the region points to the fact that they might not match the usual definition of filaments, are twofold. First, this a priori knowledge of the structure can be absent when studying another filamentary region, and it is important to treat the skeleton statistically in an unbiased way, to see whether and how the physical properties of the filaments can distinguish different populations. The results of this paper suggest that all the above mentioned structures should be treated in the same way. Second, mostly in the case of the Horsehead nebula, the fact that an elongated structure had one formation scenario or another does not necessarily determine its further evolution, hence the necessity of keeping the entire variety of structures in the sample.

4.3 Comments and prospectives

While the geometry of the filaments in Orion B, and in particular their width, is in line with the results commonly found in the literature, their specificity is their high stability against gravitational collapse, which results in part from their particularly low linear and volume densities for filaments found in a GMC. In particular, while the distributions of densities in filaments display a log-normal shape (Orkisz et al. 2018, Fig. 11), reminiscent of the distribution of column densities in the turbulent ISM (Sect. 3.2.1, see also Roy et al. 2015), one can note the absence of a visible power-law tail at high densities, which would be characteristic of self-gravitating regions (Tremblin et al. 2014a). This also points at the lack of gravitationally super-critical filaments in Orion B.

The rarity of dense, super-critical, star-forming filaments is coherent with the low star-formation activity in Orion B, but the origin of the large number of detected low-density filaments is unclear. It might be that the filaments found in Orion B are really particularly not dense, which would correlate well with observations of non-star-forming clouds such as the Polaris Flare. Whether part of the low-density filaments could gradually increase their mass or if most of them would disperse quickly is unknown. It might also be that the detection of so many low-density filaments is a consequence of the filament definition and the adopted detection scheme. As mentioned in Sect. 4.1.1 and 4.1.4, there is no universal agreement on the physical nature of filaments, and therefore on how should be defined and identified in observations. In some cases, faint elongated structures (striations) can be shown to be physically different from the larger filament they are flowing into (Palmeirim et al. 2013), which sets a lower limit for the density of filaments. But this is not a universally definition, and, for example, unlike the filaments vs. striations observed in Taurus, the population of filaments in Orion B seems to follow a continuous distribution. In practice, the definition of the studied filaments in various works is set by the parameters of the filament extraction method used. This approach is understandable for observers who try to make the best of the data they have with a given quality and a given field of view, but progresses in the understanding of filaments will be made more easily when confusion in the studied objects will disappear, and this requires to be able to find a firm, physical definitions of the objects we are looking for. Besides, the many existing filament identification methods – piecewise linear fitting (Hacar et al. 2013), *getfilaments* (Men'shchikov 2013), *FilFinder* (Koch & Rosolowsky 2015), DisPerSE (Sousbie 2011), or various Hessian-based approaches (Schisano et al. 2014; Salji et al. 2015; Orkisz et al. 2018), etc. – despite sometimes sharing some similarities, have not really been benchmarked against each other, which also raises the question of methodological biases.

From the methodological point of view, few studies make use of the full three-dimensional information when dealing with observational PPV data. In particular, Hessian based methods can be transposed to three dimensions, but the asymmetry between the two spatial dimensions and the spectral one makes the analysis a bit more complex than in 2D, because of possible variations of the aspect ratio of the "tubes" that need to be identified in PPV space, if a given filament width corresponds to different velocity dispersions. However, such a 3D analysis could disentangle the filamentary skeleton identified in Orion B and distinguish between filaments that actually cross or merge, and those which merely happen to be superimposed along the line of sight. In particular, such a diagnosis could determine whether the filaments are mostly embedded in a sheet (seen almost face-on) or if there are many "layers" of filaments along the line of sight.

The 3D detection of filaments would also be a ideal preliminary to study the longitudinal velocity field of the filaments, in addition to the transverse one studied in Orkisz et al. (2018). This could put constraints on the way filaments merge and possibly help identify hubs. In particular, as mentioned in Sect. 4.1.3, the Flame nebula might be such a massive star-forming hub. Besides, preliminary kinematic analysis has shown that the Flame filament show several interesting longitudinal velocity features: a very low velocity divergence, kinks in the centroid velocity at merging points with secondary filaments, a red-shifted acceleration near its most massive part that might point at gravitational infall, and signs of interaction with the expanding NGC 2024 HII region.

Longitudinal velocity analysis could also reveal the dynamics associated with the fragmentation of

the Hummingbird filament. This object will benefit from a high-resolution follow-up, since telescope time has been granted for NOEMA interferometric observations of this object, which will provide in the near future a mosaic of the entire fragmenting region at a 4″ resolution.

At the scale of the ORION-B project, the detection of filaments mostly provides masks of the filamentary regions, which will allow to search for a chemical and/or kinematic signature of these structures by making use of the entire ORION-B dataset rather than simply using the ^{13}CO ($J = 1 - 0$) and C^{18}O ($J = 1 - 0$) lines. Further statistical analysis of the filamentary network is also foreseen, and in particular clustering should reveal whether all filaments belong to the same family of objects, or if there are distinct sub-populations which can be identified based on certain observational criteria. Such results would provide answers to the question of the definition of molecular filaments.

Chapter 5

Conclusions

5.1 The Orion B Giant Molecular Cloud as it had never been seen before

5.1.1 A unique dataset

Giant Molecular Clouds are the cradles of stars, and as such they are the direct link between star formation and the evolution of the galaxies. They are thus a cornerstone for studies of the interstellar medium. Giant Molecular Clouds are particularly challenging to study observationally, because of their remarkable physical and chemical complexity, and because of the tight link that exists between all the physical scales of a GMC, from the galactic influences down to star formation and stellar feedback.

This is why the dataset acquired over the course of the ORION-B project is of such importance for this field of research. It will provide a wide-field view of the Giant Molecular Cloud (12×18 pc) which covers about half of the cloud, at a spatial resolution (~ 60 mpc) better than the sonic scale, thus delivering a dynamic range of ~ 200. While this kind of coverage is routinely achieved by continuum observations (e.g., *Herschel* Gould Belt observations in the infrared), the IRAM-30m observations also provide a rich view of the chemical diversity in the cloud. The 40 GHz spectral coverage with a sensitivity of ~ 0.1 K allows us to detect tens of molecular lines, with a maximal signal-to-noise ratio of ~ 400 (at full spectral and spatial resolution) for the brightest lines. Besides, the 195 kHz spectral resolution, which corresponds to about $0.6\,\mathrm{km\,s^{-1}}$ in the 3 mm band, gives access to the detailed line-of-sight velocity structure of each molecular line. The position-position-velocity-chemistry structure of the ORION-B dataset is thus far more multi-dimensional, than, say, the combination of a column density and an effective dust temperature map.

The various combinations of the available "dimensions of information" thus enables a wide range of studies, from purely chemical ones (which thus benefit from large statistics offered by the spectral coverage) to purely spatial ones (which use the chemical tracers as ways to observe specifically different structures in the cloud)... The multiplicity of possible approaches, tackled by a large team of specialists from various domains, gradually reveals an unprecedented view of the Orion B Giant Molecular Cloud.

5.1.2 Statistical characterization of the molecular environment

The number of wide-field maps of different spectral lines enables a statistical approach of the molecular tracers in the GMC. In particular, it is possible with such a dataset to quantitatively assess the relation between the emission of a given molecular line and environments with specific physical conditions.

In a preliminary analysis (Pety et al. 2017), we have tested the ^{12}CO ($J = 1 - 0$) line as a tracer of the total mass of the molecular cloud. Contrarily to the expectations, it was found that when using the standard $X_{\mathrm{CO}} = 2.0 \times 10^{20}\,\mathrm{cm^{-2}/(K\,km\,s^{-1})}$ conversion factor, the ^{12}CO-traced mass was about three times higher than the dust-traced mass derived from *Herschel* FIR observations. It can be explained by the high FUV illumination of the south-western edge of Orion B, which results in warmer dust and brighter ^{12}CO. We also quantified the fraction of emission of each molecular line that came from regions of increasing extinction or increasing illumination, to test which tracers are the most sensitive to gas density

and gas temperature. This analysis was also performed for line ratios, to provide observables free from calibration uncertainty and independent of the distance of the target. N_2H^+ appears as the best tracer of dense and cold gas, C_2H as the best tracer of illuminated gas. The results also show that molecules which are commonly used as tracers of dense gas because of their high critical densities, such as HCO^+, HCN or HNC, have actually most of their flux coming from regions of moderate or even low extinction, showing that a detailed understanding of the abundance and excitation conditions of a molecule is required to use some of its transitions as diagnosis of specific ISM conditions.

Further statistical analysis was applied by Gratier et al. (2017) to the integrated intensities of the 12 brightest detected lines. Principal Component Analysis revealed the main correlations between the brightnesses of the different molecular tracers, and which linear combinations of lines (Principal Components) carry most of the variance present in the dataset. The obtained Principal Components were then compared with physical parameters derived from independent data (column density, volume density and UV illumination) to test whether it is possible to reconstruct such physical parameters from molecular emission alone. It appears that, to first order, all molecular tracers are positively correlated with each other, and with column density: more matter along the line of sight implies more emission, for any line. The next components involved tracers of dense gas or illuminated gas identified by Pety et al. (2017), and correlated indeed with measurements of volume density and UV illumination. Thus, using only the three main components derived from twelve molecular lines, we can obtain a synthetic view of the molecular cloud which shows the amount of gas along the line of sight, its density and the radiation field it is subjected to.

Yet another statistical approach, implemented by Bron et al. (2018), aimed at identifying characteristic regions in the cloud using the molecular emission. Applying the MeanShift clustering algorithm to integrated intensity maps of CO isotopologues brings out a nested pattern of regions of increasing gas density, while also providing a distinction between cold and warm dense regions. When two extra molecular tracers, CN and HCO^+, are added, more details emerge in the high density regions: HCO^+reveals regions with typical densities an order of magnitude higher than CO isotopologues do, and CN provides an excellent view of the strongly illuminated photo-dissociation regions. This identification of typical regions in the molecular cloud enables more detailed studies of the corresponding environments, as stacking of spectra in the selected regions can bring out fainter lines and help characterize the chemistry at an effective signal-to-noise ratio better than that of the original observational data.

These studies can directly have applications for the study of the extragalactic ISM, where usually only the mean spectra of entire molecular clouds are available. By comparing the molecular signature, and in particular the line ratios, with spatially resolved quantitative measurement in Orion B, it will be possible to relate the extragalactic emission to certain physical conditions of the molecular gas, such as, in particular, its density and incident radiation field.

5.1.3 Turbulence and filaments: the dynamics of star formation

In addition to "purely chemical" statistical approaches, the spatial dynamic range combined with velocity-resolved line observations allowed for a detailed analysis of the structures and motions of the gas, in particular regarding how they relate to the star formation activity.

The bright and widely detected ^{13}CO ($J = 1 - 0$) line is a good proxy to study the kinematics of the bulk of the gas, given its high signal-to-noise ratio and its limited saturation, compared to the optically thick ^{12}CO. In first approximation, the ^{13}CO ($J = 1 - 0$) data-cube can thus represent the plane-of-the-sky projection of the line-of-sight component of the momentum density $\rho\vec{v}$ of the molecular cloud. Assuming this momentum density field is isolated and statistically isotropic, the fraction of solenoidal or compressive motions it contains can be determined. In a hypersonic case (Mach number > 5), a full mixing of modes is expected, yielding at equipartition 2/3 of solenoidal modes and 1/3 of compressive ones. In Orion B, on the largest scales of the cloud, the velocity field is hypersonic, and yet there is an excess of solenoidal motions, which indicates that turbulence is dominated by solenoidal ordered motions. These motions might be driven by a large scale rotation of the cloud, and they hinder gravitational collapse, which is in line with the particularly low star formation efficiency of the GMC. On the other hand, zooming into

the regions of brightest ^{13}CO emission, i.e., the star forming regions NGC 2023 and NGC 2024, reveals highly compressive motions, which are less efficiently mixed with solenoidal modes due to lower Mach numbers. The compressive motions can be due both to contraction that leads to star formation, and to expansion motions generated by young stars, in particular the expansion of the young HII regions. The Horsehead nebula, submitted to the pressure of the IC 434 PDR, and the northern edge of the NGC 2024 nebula, show signs of being particularly compressed. This suggests that *triggered* star formation might occur under the influence of these pressure fronts, notably in the dense northern edge of NGC 2024, which, so far, is starless.

The study of filaments using C^{18}O ($J = 1 - 0$) as a tracer of these dense structures also draws a picture consistent with the low star formation activity of Orion B, which is possibly related to a young evolutionary state of the cloud (e.g., compared to Orion A, where the star formation activity, possibly gradually triggered by the feedback from the earliest generations of stars, is much more efficient). The detection of a large filamentary network (over 100 filaments) in Orion B provides a statistical insight into the properties of the filaments in the molecular cloud. Most of the filaments are gravitationally stable, which can be both explained by the particularly low linear and volume densities of the filaments (compared to previous studies of filaments in other clouds), and by their relatively high temperature and velocity dispersions. The only gravitationally unstable filaments are, again, associated with the star-forming regions NGC 2023 and NGC 2024. The analysis of the velocity field around the filaments shows that they are quieter than their environment, but not subsonic, and their inner regions either contain unresolved fibres, or are stirred by accretion shocks. In addition to that, longitudinal fragmentation non-associated with prestellar objects is observed in at least one filament. The presence of accretion shocks and/or fibres and/or longitudinal fragmentation in starless filaments is particularly interesting, as it suggests that the filamentary network is young and further studies might help set constraints on the formation mechanisms of molecular filaments.

5.1.4 What next in the ORION-B project?

The studies carried out so far in the ORION-B project, rather than providing final answers and closing the topics, have rather opened new avenues and raised new questions, which need further investigation. This is all the more true that novel methodological approaches were applied to a fraction of total ORION-B field of view only. The next obvious step is therefore to expand the applied methods to the entire, 5-square-degree field of view, and refine them whenever possible (non-linear or piecewise linear PCA, different clustering methods, better rendition of the momentum density field from PPV data-cubes...). In particular, the study of the mean spectra of typical environments, and investigations of the variations of the $X_{\rm CO}$ factor as a function of the field of view, are of prime importance for applications to extragalactic studies. Having access to the motions of the largest scales will also help correlate the nature of the turbulence in Orion B with its galactic environment.

In addition to the development of the previous works, new statistical approaches are also being developed to try and predict accurately physical parameters such as the total column density or the ionization fraction using the molecular emission. Unlike the previous statistical studies, these new approaches do not work solely with the ORION-B dataset, but are trained on results from astro-chemical simulations. More generally, the project is gearing towards more interaction with numerical and theoretical astro-chemistry. For example, the exchange reaction between HCN and HNC in collision with H, and the usability of the HCN/HNC abundance ratio as a thermometer for the molecular gas is being investigated.

The statistical studies of the entire cloud will also benefit more and more from detailed studies of a limited number of lines of sights, or of small regions at high angular resolution. For example, detailed studies of the ionization fraction in a few cores are currently under way, and NOEMA interferometric observations of a specific filament are scheduled for the coming months. Although no detailed structure analysis was carried out so far except for filaments, other studies at high angular resolution can a priori be envisioned. More generally, a range of archival data should progressively be added to the ORION-B studies, as they offer data at other frequencies that the 3 mm band, and therefore provide other transitions of the molecules detected by ORION-B observations – albeit over far smaller fields of view, which

include the Horsehead nebula, NGC 2023, NGC 2024 and the Orion B9 cores. Having access to multiple transitions per molecule should enable should enable more accurate kinetic temperatures and molecular column density determinations. Such a detailed benchmarking in a few smaller regions should be very beneficial for the wide field statistical analysis. On the large scales, the project would also benefit from a thorough analysis of its galactic environment, such as a more extensive review of the possible ionizing sources and embedded young stars, using *Gaia* distance measurements. A detailed study of possible foreground and background contamination is also needed, as recent results (Lallement et al. 2018) suggest that that a small contribution to the extinction might be located 200 pc in front of Orion B, and a larger one 200 pc behind it, but this needs confirmation with a more extended dataset.

Longer term perspectives for the ORION-B project probably imply a gradual convergence of the different studies that will have tackled the dataset from many different viewpoints. For example, statistical (machine-learning) studies have so far never made use of the velocity information contained in the data-cubes. A complete PPV statistical analysis of the dataset, based on a spatially continuous and multi-line spectral decomposition, could reveal the variations of chemical abundances between different velocity components, and get access to (part of) the three-dimensional physical and chemical structure of the Orion B Giant Molecular Cloud.

5.2 Outlooks

Outside of the ORION-B project, the future of radio-astronomy of the interstellar medium is that of larger and larger datasets (Fig. 2.7), which will provide not only a more and more detailed view of the studied objects, but also will allow to sample a statistically significant portion of the interstellar medium.

Let us take the example of the 25-pixel receiver that is being developed and should be installed at the IRAM-30m telescope in the coming years, and consider rule-of-thumb estimates of the possibilities such a receiver would offer. If, for the sake of the argument, this receivers actually delivers a 25 times higher data-rate at the same sensitivity as the current EMIR receiver, then a project such as ORION-B would only take between 30 and 40 hours, the duration of an average observing programme. Conversely, using the same amount of time as the ORION-B observations, one could observe the same 5 deg^2 at a 5 times better sensitivity and thus probably map 5 times as many molecular lines (Guzmán 2013) – however, the results of Gratier et al. (2017), Bron et al. (2018) and the ongoing machine-learning work show that more lines are not necessarily needed to obtain a good understanding of the physics of the cloud. One could also observe the entire 3 mm band *and* 1 mm band, to have access to several transitions per species, and do that over 3 times the field of view of the ORION-B project (given that 1 mm observations are roughly 7 times slower than at 3 mm due to sensitivity and resolution) – but developing tools to obtain good temperature and density modelling from lines in the 3 mm band only is on good tracks. Or, finally, one could simply observe 25 fields equivalent to the ORION-B one, and thus obtain a statistically significant sample of GMC environments. This would offer the possibility to compare clouds of different ages, to disentangle the effects of galactic environment, metallicity, radiation field, etc... and thus it would help building a robust classification of clouds and a coherent picture of their evolution, maybe to the point of bringing out a “main sequence” of Giant Molecular Clouds in the way it has been done for stars (Fig. 2.8) and for galaxies.

However, such progress in the understanding of the ISM and molecular clouds in particular will only be possible with appropriate interactions between the observers, theoreticians and numericists. To improve comparisons between observations and simulations, efforts are being made in the development of techniques that would constrain the “missing dimensions” in observational data (discussed in Sect. 3.3.1) and would approximately reconstruct the 3D structure of the molecular clouds. On the largest scales, *GAIA* gives an unprecedented access to the 3D extinction structure of the Galaxy, and thus to the spatial extent along the line of sight of the largest absorbing structures (Capitanio et al. 2017; Lallement et al. 2018). Progress is also made in the analysis of continuum data, regarding the contribution of structures at different temperatures and densities along the line of sight (Marsh et al. 2017). The comparison of structures detected in position-position-velocity space with their position-position-position counterparts is

also a hot topic (Burkhart et al. 2013; Koch et al. 2017).

On the numerical side of things, simulations of the ISM are making steady progress, leading to the emergence of more and more realistic models of galaxies, molecular clouds, star-forming regions... Despite the constant increase in numerical resolution, simulations accurately representing all scales from the entire galaxy to the protostar remain however out of reach. Nevertheless, the quality and accuracy of the simulations are improved by incorporating better and better sub-grid models (notably for the chemistry of the gas) and using re-simulations, which are used in particular to perform higher resolution studies of the star formation processes with realistic initial conditions, and thus can also provide realistic stellar feedback.

The combination of realistic simulations, able to identify key observables that can distinguish between different scenario of star formation and molecular cloud evolution, with large sets of diverse observations, able to provide robust statistics on the molecular clouds of the Milky Way (and, to some extent, on extragalactic clouds as well), will reveal how these complex stellar factories are formed in galaxies, how they evolve under many influences and how they are eventually dispersed by their stellar offspring.

Appendix A

Co-authored papers

The appendix contains the full text of the co-authored peer-reviewed papers of the ORION-B team published during this PhD work, i.e., in the September 2015 – September 2018 period.

Two of these (Pety et al. 2017; Gratier et al. 2017) were published in a single issue together with the Orkisz et al. (2017) paper. This triple publication, which publicly launched the ORION-B project, was accompanied by a press release. Bron et al. (2018), while published later, was still limited to data from the first phase of observations, and is thus the last of the "early results" papers of the ORION-B project.

A&A 599, A98 (2017)
DOI: 10.1051/0004-6361/201629862

Astronomy & Astrophysics

The anatomy of the Orion B giant molecular cloud: A local template for studies of nearby galaxies

Jérôme Pety[1,2], Viviana V. Guzmán[3,4], Jan H. Orkisz[5,1,2], Harvey S. Liszt[6], Maryvonne Gerin[2], Emeric Bron[7,9], Sébastien Bardeau[1], Javier R. Goicoechea[9], Pierre Gratier[8], Franck Le Petit[7], François Levrier[2], Karin I. Öberg[3], Evelyne Roueff[7], and Albrecht Sievers[10]

[1] IRAM, 300 rue de la Piscine, 38406 Saint-Martin-d'Hères, France
e-mail: pety@iram.fr
[2] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, École normale supérieure, 75005 Paris, France
[3] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, 02138, USA
[4] Joint ALMA Observatory (JAO), Alonso de Cordova 3107 Vitacura, Santiago de Chile, Chile
[5] Univ. Grenoble Alpes, IRAM, 38000 Grenoble, France
[6] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA, 22903, USA
[7] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, 92190 Meudon, France
[8] Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[9] ICMM, Consejo Superior de Investigaciones Cientificas (CSIC), 28049 Madrid, Spain
[10] IRAM, Avenida Divina Pastora, 7, Núcleo Central, 18012 Granada, España

Received 7 October 2016 / Accepted 8 November 2016

ABSTRACT

Context. Molecular lines and line ratios are commonly used to infer properties of extra-galactic star forming regions. The new generation of millimeter receivers almost turns every observation into a line survey. Full exploitation of this technical advancement in extra-galactic study requires detailed bench-marking of available line diagnostics.
Aims. We aim to develop the Orion B giant molecular cloud (GMC) as a local template for interpreting extra-galactic molecular line observations.
Methods. We use the wide-band receiver at the IRAM-30 m to spatially and spectrally resolve the Orion B GMC. The observations cover almost 1 square degree at 26″ resolution with a bandwidth of 32 GHz from 84 to 116 GHz in only two tunings. Among the mapped spectral lines are the ^{12}CO, ^{13}CO, C^{18}O, C^{17}O, HCN, HNC, ^{12}CN, C_2H, HCO^+, N_2H^+ (1−0), and ^{12}CS, ^{32}SO, SiO, c-C_3H_2, CH_3OH (2−1) transitions.
Results. We introduce the molecular anatomy of the Orion B GMC, including relationships between line intensities and gas column density or far-UV radiation fields, and correlations between selected line and line ratios. We also obtain a dust-traced gas mass that is less than approximately one third the CO-traced mass, using the standard X_{CO} conversion factor. The presence of over-luminous CO can be traced back to the dependence of the CO intensity on UV illumination. As a matter of fact, while most lines show some dependence on the UV radiation field, CN and C_2H are the most sensitive. Moreover, dense cloud cores are almost exclusively traced by N_2H^+. Other traditional high-density tracers, such as HCN(1−0), are also easily detected in extended translucent regions at a typical density of ∼500 H_2 cm^{-3}. In general, we find no straightforward relationship between line critical density and the fraction of the line luminosity coming from dense gas regions.
Conclusions. Our initial findings demonstrate that the relationships between line (ratio) intensities and environment in GMCs are more complicated than often assumed. Sensitivity (i.e., the molecular column density), excitation, and, above all, chemistry contribute to the observed line intensity distributions, and they must be considered together when developing the next generation of extra-galactic molecular line diagnostics of mass, density, temperature, and radiation field.

Key words. galaxies: ISM – ISM: clouds – HII regions – radio lines: galaxies – astrochemistry

1. Introduction

The star formation process from interstellar gas raises many outstanding questions. For instance, the relative roles of micro-physics and galactic environment on star formation efficiency are debated; more precisely, the respective roles of magnetic fields, self-gravity, and turbulence (see e.g., Hennebelle & Chabrier 2011; Hennebelle 2013) compared with the external pressure, position in galactic spiral arm/interarm regions and spiral-arm streaming motions (e.g., Meidt et al. 2013; Hughes et al. 2013). Other points hotly discussed include the impact of feedback from H II region expansion and supernovae (Kim et al. 2013), the identification of the key dynamical parameters (Mach number, virial parameter, amount of energy in solenoidal/compressive modes of the turbulence, see Federrath & Klessen 2012, 2013, for example), the amount of CO-dark molecular gas and whether this biases global estimation of the mass of the molecular reservoir at cloud scales (Wolfire et al. 2010; Liszt & Pety 2012), the relative fractions of diffuse (∼100−500 cm^{-3}, 80 K) and dense (∼10^4 cm^{-3}, 10 K) gas in a giant molecular cloud (GMC), and

the determination of the true fraction of star-forming dense gas (Lada et al. 2010, 2012, 2013).

All these questions also arise in extra-galactic studies with the additional difficulty that GMCs are unresolved at the typically achieved angular resolution (1″ corresponds to 15 pc for a 3 Mpc distant galaxy). It is therefore crucial to first understand how the average spectra of molecular lines relate to actual physical properties when the line emission is spatially resolved. By mapping a significant fraction of a GMC at a spatial resolution of ∼50 mpc and a spectral resolution of 0.6 km s^{-1}, we address some of the following issues: what linear resolution must be achieved on a GMC to correctly derive its global properties including star formation rate and efficiency (Leroy et al. 2016)? For instance, are usual extra-galactic line tracers of the various molecular cloud density regimes reliable (Bigiel et al. 2016)? Do we get a more accurate estimate of the mass by resolving the emission? More generally, can we derive empirical laws that link tracer properties averaged over a GMC to its internal star-forming activity?

With the advent of wide-bandwidth receivers associated to high-resolution spectrometers, any observation now simultaneously delivers emission from many different tracers. Moreover, the increased sensitivity makes it possible to cover large fields of view (FoVs). The possibility to map many different lines in many different environments allows us to start answering the questions presented above. The essence of the ORION-B (Outstanding Radio Imaging of OrioN B, PI: J. Pety) project is to recast the science questions of star formation in a statistical way. Wide-field hyper-spectral mapping of Orion B is used to obtain an accurate 3D description of the molecular structure in a GMC, a key for defining chemical probes of the star-formation activity in more distant Galactic and extragalactic sources.

Approximately thirty 3 mm lines are detected in only two frequency tunings with the same sensitive radio single-dish telescope at a typical resolution of 26″ over almost 1 square degree. The field of view (5.6 × 7.5 pc) would fall in a single resolution element of a map of the Orion B molecular cloud observed at 3 mm with a telescope of similar diameter as the IRAM-30 m from the Small or Large Magellanic Clouds. The spectra, averaged over the field of view, would then represent the spectra of Orion B as seen by an alien from the Magellanic Clouds. Conversely, our imaging experiment allows us to reveal the detailed anatomy of a molecular emission that is usually hidden behind these mean spectra in nearby galaxy studies. The south-western edge of the Orion B molecular cloud (a.k.a. Barnard 33 or Lynds 1630) represents an ideal laboratory for this kind of study. It forms both low-mass and massive stars, and contains regions of triggered or spontaneous star formation, photon-dominated regions and UV-shielded cold gas, all in a single source.

In companion papers, Gratier et al. (2017) study a Principal Component Analysis (PCA) of the same dataset to understand the main correlations that exist between the different lines. Orkisz et al. (2017) quantify the fractions of turbulent energy that are associated with the solenoidal/compressive modes, and they relate these values to the star formation efficiency in Orion B. In this paper, we present the observational results of the ORION-B project, focusing on the mean properties of this GMC and evaluate the diagnostic power of commonly used line tracers and ratios.

We present the targeted field of view, as well as the observations and data reduction process in Sect. 2. Typical properties such as UV-illumination, mean line profiles, CO-traced, dust-traced and virial masses are computed in Sect. 3. In Sect. 4, we investigate the fraction of flux arising in different gas regimes for each line. In Sect. 5, we compare the visual extinction map with the line integrated intensities and compute the luminosity per proton of the different line tracers. The properties of various line ratios are discussed in Sect. 6. A discussion is presented in Sect. 7, focusing on possible biases introduced by the characteristics of the observed field of view, and whether the HCO^+, HCN, and HNC (1−0) lines are good tracers of dense gas. We end the discussion by comparing the observed line ratios in Orion B with extra-galactic observation results. Section 8 summarizes the results and concludes the paper.

2. The Orion B GMC

2.1. Targeted field of view

Figure 1 displays a composite image of the ^{12}CO (blue), ^{13}CO (green), and C^{18}O (red) (1−0) peak-intensity main-beam temperatures. It shows the south-western edge of the Orion B molecular cloud. This region samples the interaction of the molecular cloud with at least four H II regions. First, σOri is an O9.5V star that illuminates the western edge of the Orion B cloud. It creates the IC 434 nebula from which the Horsehead pillar emerges. Second, NGC 2023 and NGC 2024 are two younger H II regions embedded in the Orion B molecular cloud, powered by B1.5V (HD 37903) or late O, early B (IRS2b) stars, respectively (Bik et al. 2003). NGC 2024 covers 20 arcmin at the northern edge of the mapped FoV, while NGC 2023 is situated approximately halfway between IRS2b and the Horsehead. The B5V HD 38087 star creates the IC 435 nebula at the south-eastern edge of the FoV. Finally, one of the three Orion Belt stars, the O9.71b star Alnitak (a.k.a. ζOri), falls in the observed FoV. Table 1 lists the characteristics of these exciting stars. To guide the eye on the right panel of Fig. 1, we overlaid crosses at the position of the main exciting stars, and circles at the approximate boundaries of the different H II regions. These visual markers are used throughout the paper.

2.2. IRAM-30 m observations

The observations were taken with the IRAM-30 m telescope in four observing runs: August 2013, December 2013, August 2014, and November 2014 for 133 h in total (telescope time) under average summer weather (6 mm median water vapor) and good winter weather (3 mm median water vapor). During all these runs, we observed with a combination of the 3 mm sideband separated EMIR receivers and the Fourier transform spectrometers, which yields a total bandwidth of ∼32 GHz per tuning (i.e., ∼8 GHz per sideband and per polarization) at a channel spacing of 195 kHz or 0.5−0.7 km s^{-1}. The two tuned frequencies were 102.519 GHz and 110.000 GHz at the 6.25 MHz intermediate frequency of the upper sideband, resulting in local oscillator frequencies of 96.269 and 103.750 GHz, respectively. This allowed us to observe nearly the entire 3 mm band from 84.5 to 116.5 GHz.

We used the on-the-fly scanning strategy with a dump time of 0.25 s and a scanning speed of 17″/s to ensure a sampling of five dumps per beam along the scanning direction at the 21.2″ resolution reached at the highest observed frequency, that is, 116 GHz. We covered the full FoV (∼0.9 square degrees) with 103 tiles of approximately 110″ × 1000″ size. The rectangular tiles had a position angle of 14° in the Equatorial J2000 frame to adapt the mapping strategy to the global morphology of the western edge of the Orion B molecular cloud. These tiles were covered with rasters along their long axis (almost the Dec direction). The

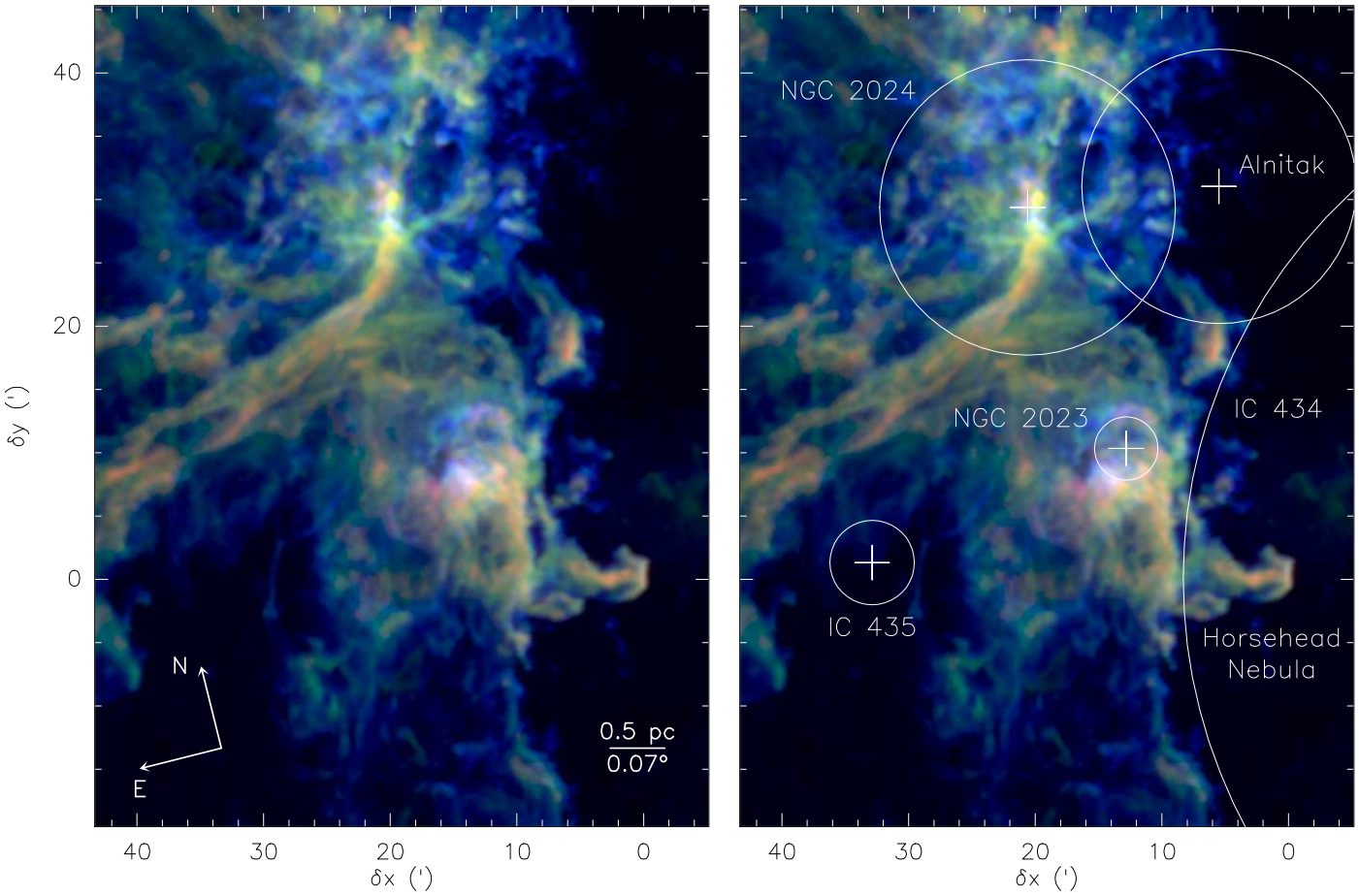


Fig. 1. Composite image of the ^{12}CO (blue), ^{13}CO (green), C^{18}O (red) (1–0) peak-intensity main-beam temperature. The circles show the typical extensions of the H II regions and the crosses show the position of the associated exciting stars (see Table 1). The σ-Ori star that excites the IC 434 H II region is located 0.5° east of the Horsehead nebula.

Table 1. Properties of the stars exciting the H II regions in the observed field of view.

H II region	Star	Type	α, δ (J2000)	$(\delta x, \delta y)$ ($''$, $''$)	Parallax mas	Distance pc	$V_{\rm LSR}$ km s^{-1}
IC 434	σOri	O9.5V B	$05^{\rm h}38^{\rm m}44.779^{\rm s}$, −02°36′00.12″	(−33.35, 00.07)	2.5806 ± 0.0088$^{(1)}$	387.5 ± 1.3$^{(1)}$	15.0 ± 1.6
IC 435	HD 38087	B5V D	$05^{\rm h}43^{\rm m}00.573^{\rm s}$, −02°18′45.38″	(+32.87, 01.33)	5.90 ± 1.29$^{(2)}$	169 ± 37$^{(2)}$	18.1 ± 4.5
NGC 2023	HD 37903	B1.5V C	$05^{\rm h}41^{\rm m}38.388^{\rm s}$, −02°15′32.48″	(+13.72, 09.42)	2.776 ± 0.271$^{(3)}$	362 ± 35$^{(3)}$	−7.7 ± 2
NGC 2024	IRS2b	O8V-B2V	$05^{\rm h}41^{\rm m}45.50^{\rm s}$, −01°54′28.7″	(+20.54, 29.43)	–	415$^{(4)}$	–
	Alnitak	O9.7Ib+B0III C	$05^{\rm h}40^{\rm m}45.527^{\rm s}$, −01°56′33.26″	(+05.49, 31.04)	3.4 ± 0.2$^{(5)}$	294 ± 21$^{(5)}$	3.7 ± 1.3

Notes. (1) Schaefer et al. (2016); (2) van Leeuwen (2007); (3) *Gaia* DR1: Gaia Collaboration (2016b), Lindegren et al. (2016), Fabricius et al. (2016), Gaia Collaboration (2016a); (4) Anthony-Twarog (1982); (5) Hummel et al. (2013).

separation between two successive rasters was $\lambda/2D = 8.46''$ to ensure Nyquist sampling perpendicular to the scanning direction. The scanning direction was reversed at the end of each line (zigzag mode). This implied a tongue and groove shape at the bottom and top part of each tile. We thus overlapped the top and bottom edges of the tiles by 30″ to ensure correct sampling. Furthermore, the left and right edges of the tiles were adjusted to avoid any overlap, that is, to maximize the overall scanning speed. The FoV was covered only once by the telescope, except for the tiles observed in the worst conditions (low elevation and/or bad weather) that were repeated once.

The calibration parameters (including the system temperature) were measured every 15 min. The pointing was checked every two hours and the focus every four hours. Following Mangum et al. (2007) and Pety et al. (2009), we used the optimum position switching strategy. A common off reference position was observed for 11 s every 59 s with the following repeated sequence OFF-OTF-OTF-OFF, where OFF is the reference position and OTF stands for On-The-Fly. No reference position completely devoid of ^{12}CO (1–0) emission could be localized in the close neighborhood of the Orion B western edge. As this reference position is subtracted from every OTF spectrum in order to remove the common atmospheric contribution, the presence of signal in the reference position results in a spurious negative contribution to the signal everywhere in the final cube. Searching for a reference position farther away in the hope that it is devoid of signal would degrade the quality of the baseline because the atmospheric contribution would vary from the OTF spectra to the reference position. We thus tested several nearby potential reference positions using the

Table 2. Typical properties of the south-western edge of Orion B.

Parameter	Value	Notes
Distance	400 pc	$1'' = 2$ mpc
Systemic velocity	10.5 km s^{-1}	LSR, radio convention
Projection center	$05^h40^m54.270^s$, $-02°28'00.00''$	α, δ(J2000), mane of the Horsehead
Offset range and field of view	$[-5.2', +43.3'] \times [-19.5', 45.3']$	$49' \times 65'$ or 5.6×7.5 pc
$W_{CO}^{min}-W_{CO}^{mean}-W_{CO}^{max}$	0–61–288 K km s^{-1}	in $[-2, +18]$ km s^{-1}
$A_V^{min}-A_V^{mean}-A_V^{max}$	0.7–4.7–222 mag	$A_K/A_V = 0.13$
$T_d^{min}-T_d^{mean}-T_d^{max}$	16–26–99 K	
$G_0^{min}-G_0^{mean}-G_0^{max}$	4–45–36 000	Inter-Stellar Radiation Field (Habing 1968)
CO-traced mass	11 000 $M_\odot$	Standard X_{CO} and helium dealt with
Dust-traced mass	3900 $M_\odot$	Standard N_H/A_V and H I gas negligible
Virial traced mass	Between 6200 and 9500 $M_\odot$	Depending on the assumed density radial profile
Imaged surface	43 pc^2	$=S$
Typical volume	280 pc^3	$=S^{3/2}$
CO-traced mean column density	260 $M_\odot$ pc^{-2} | 12×10^{21} H$_2$ cm^{-2}	Standard X_{CO} and helium dealt with
Dust-traced mean column density	92 $M_\odot$ pc^{-2} | 4×10^{21} H$_2$ cm^{-2}	Standard N_H/A_V and H I gas negligible
CO-traced mean volume density	40 $M_\odot$ pc^{-3} | 590 H$_2$ cm^{-3}	Standard X_{CO} and helium dealt with
Dust-traced mean volume density	14 $M_\odot$ pc^{-3} | 210 H$_2$ cm^{-3}	Standard N_H/A_V and H I gas negligible

frequency-switched observing mode that does not require a reference position. This is possible because the observed lines have narrow linewidth. We then selected the nearest position that has the minimum line integrated emission in ^{12}CO (1–0). Offsets of this position are $(-500'', -500'')$ with respect to the projection center given in Table 2. The ^{12}CO, and ^{13}CO (1–0) peak intensities at this position are ∼1 and 0.05 K, respectively. The correction of the negative contribution from the reference position to the final cube requires a good observation of the reference position. We therefore observed this reference position using the frequency-switched observing mode in both tunings, a few minutes per observing session.

2.3. IRAM-30 m data reduction

Data reduction was carried out using the GILDAS[1]/CLASS software. The data were first calibrated to the T_A^* scale using the chopper-wheel method (Penzias & Burrus 1973) before being converted to main-beam temperatures (T_{mb}) using the forward and main-beam efficiencies (F_{eff} and B_{eff}) listed in Table A.1. The B_{eff} values are derived from the Ruze's formula

$$B_{eff}(\lambda) = B_{eff}^0 \exp\left\{-\left(\frac{4\pi\sigma}{\lambda}\right)^2\right\}, \quad (1)$$

$$\text{with} \quad B_{eff}^0 = 0.863, \quad \text{and} \quad \sigma = 65.6\,\mu\text{m}, \quad (2)$$

where λ presents the wavelength dependence[2]. The resulting amplitude accuracy is ∼10%. A 12 to 20 MHz-wide subset of the spectra was first extracted around each line rest frequency. We computed the observed noise level after subtracting a first order baseline from every spectrum, excluding the velocity range from 0 to 18 km s^{-1} LSR, where the gas emits for all observed lines, except ^{12}CO and HCO$^+$ (1–0) for which the excluded velocity range was increased from −5 to 20 km s^{-1} LSR. A systematic comparison of this noise value with the theoretical noise computed from the system temperature, the integration time, and the channel width allowed us to filter out outlying spectra (typically 3% of the data).

To correct for the negative contribution from the reference position to the final cube, we first averaged all the observations of the reference position 1) to increase the signal-to-noise ratio of the measured profiles, and 2) to decrease the influence of potential calibration errors. Signal in the reference position was only detected for the ^{12}CO and ^{13}CO (1–0) lines. The correction was thus applied only for these two lines. The averaged spectra at these frequencies were fitted by a combination of Gaussians after baseline subtraction, in order to avoid adding supplementary noise in the final cube. This fit was then added to every on-the-fly spectrum.

The spectra were then gridded into a data cube through a convolution with a Gaussian kernel of $FWHM \sim 1/3$ of the IRAM-30 m telescope beamwidth at the rest line frequency. To facilitate comparison of the different line cubes, we used the same spatial (pixels of 9″ size) and spectral (80 channels spaced by 0.5 km s^{-1}) grid. The position-position-velocity cubes were finally smoothed at the common angular resolution of 31″ to avoid resolution effects.

2.4. Map of visual extinction and dust temperature from Herschel and Planck data

In this paper, we observationally check the potential of line intensities and of ratios of line intensities to characterize physical properties of the emitting gas. Ancillary data are thus needed to deliver independent estimates of these physical properties. We use recent dust continuum observations to provide estimates of the column density of material and of the far UV illumination.

After combining the *Herschel* Gould Belt Survey (André et al. 2010; Schneider et al. 2013) and *Planck* observations (Planck Collaboration I 2011) in the direction of Orion B, Lombardi et al. (2014) fitted the spectral energy distribution to yield a map of dust temperature and a map of dust opacity at 850 μm (τ_{850}). Hollenbach et al. (1991) indicates that the equilibrium dust temperature at the slab surface of a 1D

[1] See http://www.iram.fr/IRAMFR/GILDAS for more information about the GILDAS softwares (Pety 2005).

[2] The values of B_{eff}^0 and σ can be found at http://www.iram.es/IRAMES/mainWiki/Iram30mEfficiencies

Table 3. Line intensities and luminosities in the [−2, +18 km s^{-1}] velocity range and including all pixels.

Species	Simplified[a] quantum numbers	Complete[b] quantum numbers	A_{ij} s^{-1}	E_u/k K	Intensity mK km s^{-1}	Relative to ^{12}CO	Luminosity $L_\odot$
^{12}CO	1–0	$J = 1–0$	7.2×10^{-8}	5.5	60 430	100.00	1.0×10^{-2}
^{13}CO	1–0	$J = 1–0$	3.2×10^{-8}	5.3	9 198	15.22	1.4×10^{-3}
HCO$^+$	1–0	$J = 1–0$	4.2×10^{-5}	4.3	1 630	2.70	1.3×10^{-4}
HCN	1–0	$J = 1–0, F = 2–1$	2.4×10^{-5}	4.3	1 540	2.55	1.2×10^{-4}
CN	1–0	$N = 1–0, J = 3/2–1/2, F = 5/2 - 3/2$	1.2×10^{-5}	5.4	776	1.28	1.3×10^{-4}
C^{18}O	1–0	$J = 1–0$	6.3×10^{-8}	5.3	556	0.92	8.0×10^{-5}
^{12}CS	2–1	$J = 2–1$	1.7×10^{-5}	7.0	513	0.85	5.3×10^{-5}
C_2H	1–0	$N = 1–0, J = 3/2–1/2, F = 2–1$	1.5×10^{-6}	4.2	457	0.76	3.2×10^{-5}
HNC	1–0	$J = 1–0, F = 2–1$	2.7×10^{-5}	4.4	445	0.74	3.6×10^{-5}
^{32}SO	2–1	$J = 3–2, K = 2–1$	1.1×10^{-5}	9.2	283	0.47	3.0×10^{-5}
C^{17}O	1–0	$J = 1–0, F = 7/2 - 5/2$	6.7×10^{-8}	5.4	215	0.36	3.3×10^{-5}
c-C_3H_2	2–1	$J = 2–1, K_+ = 1–0, K_- = 2–1$	2.3×10^{-5}	6.4	149	0.25	1.1×10^{-5}
N_2H^+	1–0	$J = 1–0, F1 = 2–1, F = 3–2$	3.6×10^{-5}	4.5	67	0.11	6.0×10^{-6}
CH_3OH	2–1	$J = 2–1, K = 0 - 0$	3.4×10^{-6}	7.0	65	0.11	6.4×10^{-6}
H^{13}CN	1–0	$J = 1–0, F = 2–1$	2.2×10^{-5}	4.1	48	0.08	3.3×10^{-6}
H^{13}CO$^+$	1–0	$J = 1–0$	3.9×10^{-5}	4.2	25	0.04	1.8×10^{-6}
HN^{13}C	1–0	$J = 1–0, F = 2–1$	1.9×10^{-5}	4.2	–	–	–
SiO	2–1	$J = 2–1$	2.9×10^{-5}	6.3	–	–	–

Notes. The lines are sorted by decreasing value of their intensity. [(a)] Simplified transition used everywhere else in the paper. [(b)] Complete list of quantum numbers associated to the transition whose frequency is listed in Table A.1. This frequency is the one used to fix the velocity scale.

Photo-Dissociation Region (PDR) is linked to the incident far UV field, G_0 at $A_V = 0$, through

$$T_d = 12.2\, G_0^{0.2}\ \mathrm{K}, \quad (3)$$

where the G_0 value is given in units of the local interstellar radiation field (ISRF, Habing 1968). We invert this equation to give an approximate value of the far-UV illumination. This value is likely a lower limit to the actual G_0 in most of the mapped region. Indeed, it is the far-UV field at the surface of the PDR, while there are embedded H II regions in the FoV. However, Abergel et al. (2002) estimates a typical $G_0 \sim 100$ for the western edge of L 1630, which is a large scale edge-on PDR. Using Eq. (3), this value is compatible with the typical dust temperature fitted towards this edge, that is, approximately 30 K.

Lombardi et al. (2014) compared the obtained 850 μm opacity map to an extinction map in the K band (A_K) of the region. A linear fit of the scatter diagram of A_K and τ_{850} gives $A_K = 3460\,\tau_{850}$ for Orion B (Lombardi et al. 2014, name this factor γ). They used a value of $A_K/A_V = 0.112$ from Rieke & Lebofsky (1985). However, this value, including their estimated $R_V \simeq 3.1$, is not based on observations towards Orion stars. Cardelli et al. (1989) measured the properties of dust absorption E_{B-V} and R_V towards two stars of our FoV. Using their parametrization, we yield $A_K/A_V = 0.1254$ for $R_V = 4.11$ towards HD 37903, and $A_K/A_V = 0.1335$ for $R_V = 5.30$ towards HD 38087. We here take an average of both values, $A_K/A_V = 0.13$, that is, a 20% larger value than Rieke & Lebofsky (1985). We therefore have

$$A_V = 2.7 \times 10^4\, \tau_{850}\ \mathrm{mag}. \quad (4)$$

The dust properties (both the temperature and visual extinction) are measured at an angular resolution of 36″.

2.5. Noise properties, data size, percentage of signal channels, and line integrated intensities

The median noise levels (computed on the cubes that have 0.5 km s^{-1} channel spacing and 31″ angular resolution) range from 100 to 180 mK (T_{mb}) depending on the observed frequency. Details can be found in Appendix A. The reduced data cube amounts to approximately 160 000 images of 325×435 pixels or 84 GB of uncompressed data. It would make a movie of 1h50m at 24 images per second. However, approximately 99.5% of the channels show mostly noise because of the limited sensitivity of our observation. The 0.5% of the bandwidth where clear signal is detected includes the emission from low J lines of CO, HCO$^+$, HCN, HNC, CN, CS, SO, C_2H, c-C_3H_2, N_2H^+, CH_3OH, SiO, and some of their isotopologues, in particular, CO isotopologues (see Table 3 and Fig. 2).

Most of this paper studies the properties of the line integrated intensity[3] defined as $W = \int T(v)\,\mathrm{d}v$. To produce reliable spectral line maps, we used all the pixels matching two conditions: 1) its own signal-to-noise ratio is larger than four, and 2) the signal-to-noise ratio of at least 25% of its neighbors are larger than four. Residual striping may be seen along the vertical scanning direction, in particular at low intensity on the images of line integrated intensities. Indeed, baselining corrects for the striping to first order. Hence residual striping is more visible for faint lines and/or lines for which the overlapping of the velocity and the hyperfine structures (HCN ground state transition, for example) requires the definition of wider baselining windows. Finally, we used two different flavors of the integrated intensity. On one hand, we use the line intensity integrated over the full line profiles when we aim to study the gas properties along the full line of sight. This happens, for instance, when we compute the CO-traced mass in Sect. 3.4 and the correlations between

[3] The data products associated with this paper are available at http://www.iram.fr/~pety/ORION-B

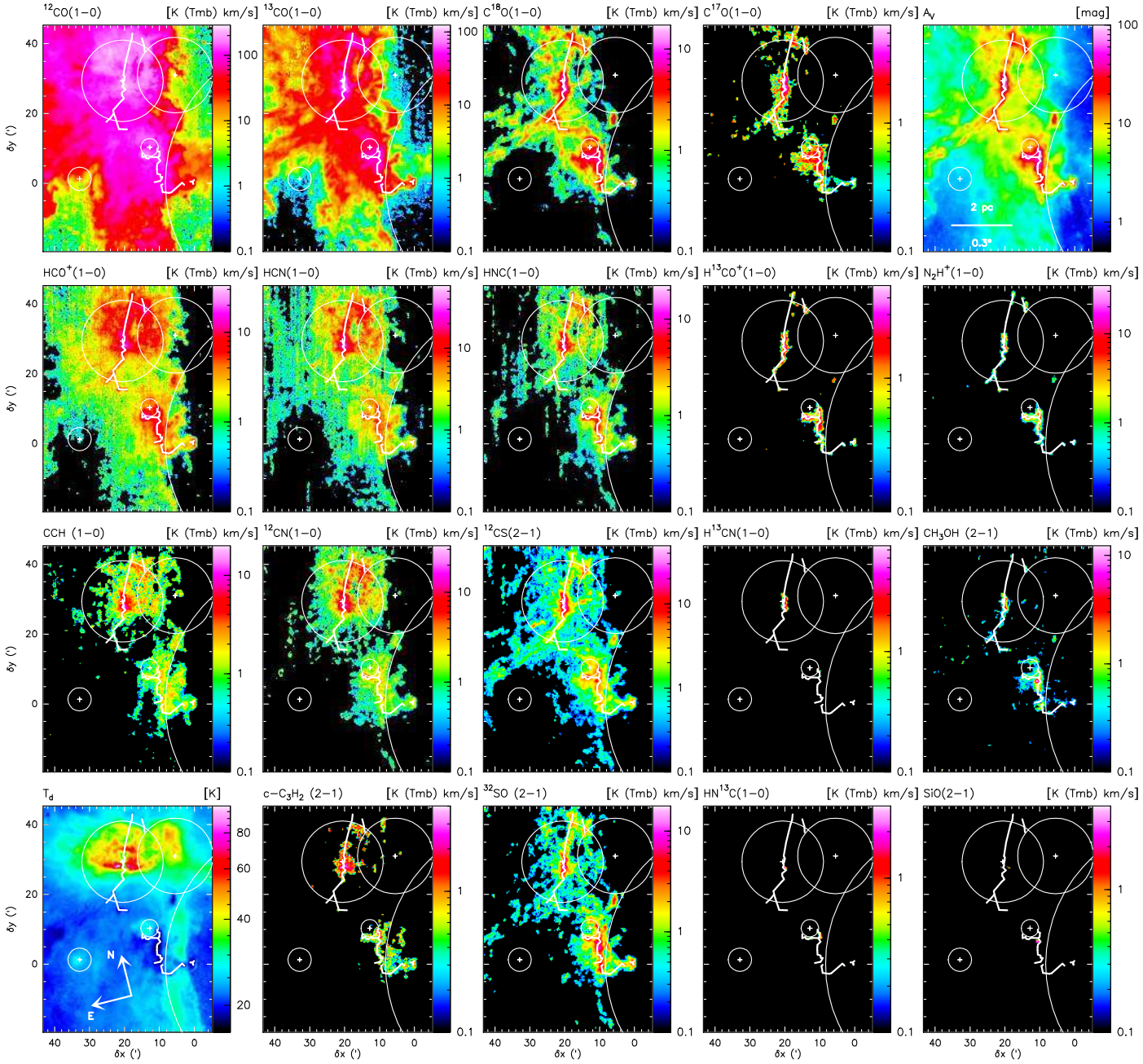


Fig. 2. Spatial distribution of the line integrated intensity for some of the detected lines in the 3 mm band, plus the dust temperature (*bottom left panel*) and the visual extinction (*top right* corner). Continuum data comes from the publicly available SED fit done by Lombardi et al. (2014) on the *Herschel* Gould Belt Survey data (PI: P. André). The color-scales are logarithmic to reveal the distribution of faint signal. Pixels with a signal-to-noise lower than four were blanked out. In addition to the circles and crosses that show the approximate boundaries of the H II regions and the associated exciting stars, we overlaid broken lines that were somewhat arbitrarily drawn by connecting the N_2H^+ (1−0) emission.

the column density of material along the full line of sight and the line integrated intensity in Sect. 5. On the other hand, lines are detected over different velocity ranges. Using the same velocity range for all lines, $[-2, +18\,\mathrm{km\,s^{-1}}]$ for example, results in noisy integrated intensities for tracers that have the narrowest lines. In contrast, adapting the velocity range to each line could bias the results. We thus adopted a compromise for sections where we can restrict our investigations to the bulk of the gas: we computed the line integrated intensity over the velocity range where the core of the line can be found for each species and transition over the measured FoV. This velocity range is $[9, 12\,\mathrm{km\,s^{-1}}]$.

3. Mean properties

From this section on, we only study the properties of the (1−0) line for the CO isotopologues (^{12}CO, ^{13}CO, C^{18}O, and C^{17}O), HCO^+, HCN, HNC, and their ^{13}C isotopologues, ^{12}CN, C_2H, and N_2H^+, as well as the (2−1) transition for ^{12}CS, ^{32}SO, CH_3OH, and SiO.

3.1. Geometry, spatial dynamic, typical visual extinction, dust temperature, and far UV illumination

Table 2 lists the typical properties of the observed FoV. At a typical distance of ∼400 pc (Menten et al. 2007; Schlafly et al. 2014),

the mapped FoV corresponds to 5.6 × 7.5 pc. This corresponds to a surface of 43 pc^2. Assuming that the depth along the line of sight is similar to the dimension projected on the plane of sky, we get a volume equal to the surface at the power 3/2, or 280 pc^3.

The angular resolution ranges from 22.5 to 30.4″ at 3 mm while the typical 30 m position accuracy is ∼2″. All the cubes were smoothed to 31″ angular resolution, that is, 60 mpc or ∼10^4 AU. We thus explore a maximum spatial dynamic range of 125 for all the lines.

The visual extinction ranges from 0.7 to 222 mag with a mean value of 4.7 mag. This is associated to a range of ^{12}CO (1−0) integrated intensity from 0 to 288 K km s^{-1} with a mean value of 61 K km s^{-1}. In other words, the FoV contains all kind of gas from diffuse, without CO emission, to highly visually extinct with bright CO emission, but most of the gas is in the higher end of the translucent regime ($2 \lesssim A_V \lesssim 6$).

The SED-fitted dust temperature along the line of sight ranges from 16 to 99 K with a mean value of 26 K. This translates into a typical far UV illumination, G_0, ranging from 4 to 3.6 × 10^4 using the Inter-Stellar Radiation Field (ISRF) definition by Habing (1968). The G_0 mean value is 45. This confirms that the observed FoV is on average strongly far-UV illuminated by the different massive exciting stars listed in Table 1 (see Sect. 2.1).

3.2. Distribution of line integrated intensities

Figure 2 presents the spatial distribution of the line integrated intensities. We also included the spatial distribution of the dust temperature in the bottom left panel and the visual extinction in the top right panel to give reference points on the underlying nature of the gas that emits each line tracer (see Sect. 2.4).

The spatial distributions of the molecular lines presented here are different. The (1−0) line of the CO isotopologues themselves show a very different behavior. The line of the rarer isotopologue, C^{17}O, has a spatial distribution that is similar to that of the N_2H^+ (1−0) line, which is a known tracer of the cold and dense regions in molecular clouds (Bergin & Tafalla 2007). Indeed, N_2H^+ and C^{17}O are seen only towards lines of sight of high extinction (A_V > 20−30). The (1−0) line emission of the slightly more abundant isotopologue C^{18}O is more extended and clearly traces the dense and cold filaments of the cloud. Moreover, the C^{18}O (1−0) emission is similar to the extinction map shown in Fig. 2, a property consistent with the known linear correlation of C^{18}O (1−0) integrated intensity with the visual extinction (Frerking et al. 1982). The (1−0) emission of the second most abundant CO isotopologue, ^{13}CO, traces gas in the extended envelope surrounding the filaments traced by the C^{18}O (1−0) emission. The emission of the main CO isotolopogue no longer traces the dense gas and is largely dominated by the extended and more diffuse or translucent gas because it is then strongly saturated.

The HCO^+, HCN, and HNC (1−0) lines are usually considered to be good tracers of dense molecular gas because of their high spontaneous emission rates and large critical densities. Among these three species, the HNC (1−0) map bears the closest resemblance to the C^{18}O (1−0) map. Emission in the ground state lines of HCO^+ and HCN exhibits a more extended component and more closely resembles the ^{13}CO (1−0) map. All three lines as well as CN, present bright emission towards high-extinction lines of sight. Their emission also seems to trace the edges of the H II regions. A clear difference between CN and the other N-bearing species and HCO^+ is the larger contrast between the warmer northern region, near NGC 2024, and the cooler southern region near the Horsehead. In contrast to their main isotopologue, the ^{13}C isotopologues of HCO^+, HCN, and HNC are only clearly detected towards the dense cores. The methanol emission is slightly more extended than the N_2H^+ (1−0) emission but it is clearly seeded by the dense cores as traced by N_2H^+.

The emission of the sulfur-bearing species, in particular the ^{12}CS (2−1) line, has similar spatial distributions to that of the C^{18}O (1−0) line. Finally, the SiO (2−1) line is only detected at the position of two previously known outflows. The first is located at the south-west of NGC 2023 around the class-0 NGC 2023 mm1 protostars located at 05^{h}41^{m}24.9^s, −02°18′09″ (J2000, Sandell et al. 1999), while the second is located on both sides of the FIR5 young stellar object located at 05^{h}41^{m}44.6^s, −01°55′38″ (J2000, Richer 1990; Chernin 1996), near the center of NGC 2024. This confirms that SiO is, above all, a shock tracer.

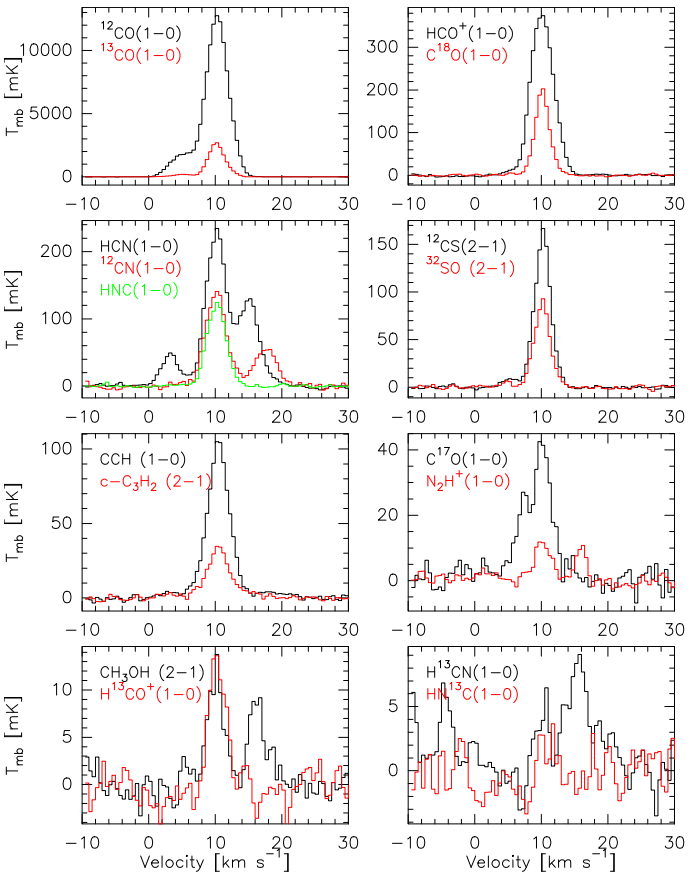


Fig. 3. Spectra averaged over the mapped field of view.

3.3. Mean line profiles over the observed FoV

Figure 3 shows the spectra of the main detected lines averaged over the mapped FoV. Several spectra show multiple components for different reasons. First, the multi-peak nature of the CN, HCN, C^{17}O and N_2H^+ ground state lines is a consequence of the resolved hyperfine structure of these transitions. Second, the faintest spectra (e.g., HN^{13}C) are detected at low signal-to-noise ratio, implying a noisy profile. Finally, the western side of the Orion B cloud displays two velocity components: the main one at approximately 10.3 km s^{-1} and a satellite one, ten times fainter, at approximately 4.9 km s^{-1}. The two components have

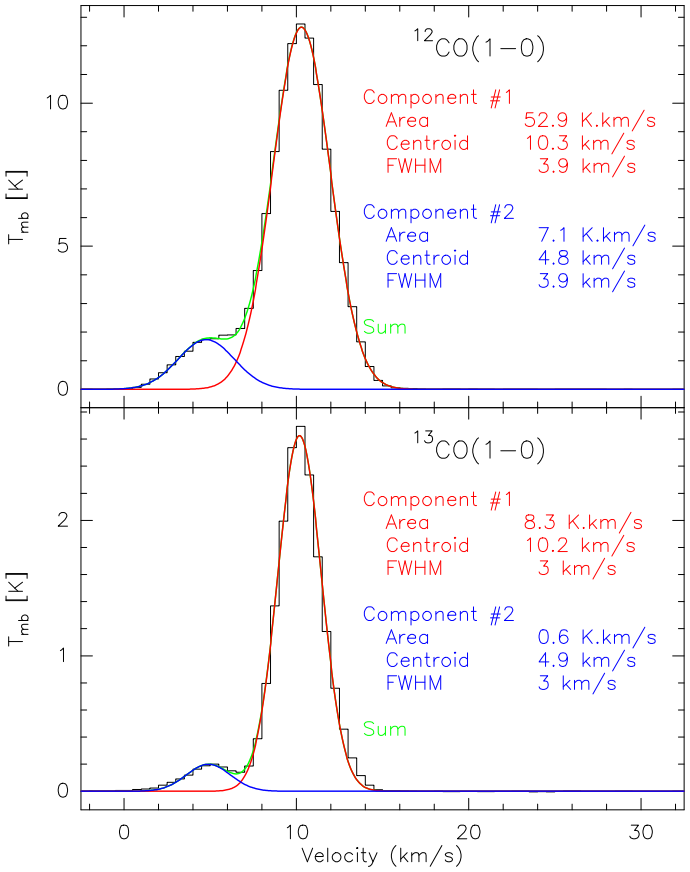


Fig. 4. Gaussian fits of the two main velocity components that appear in the western edge of Orion B.

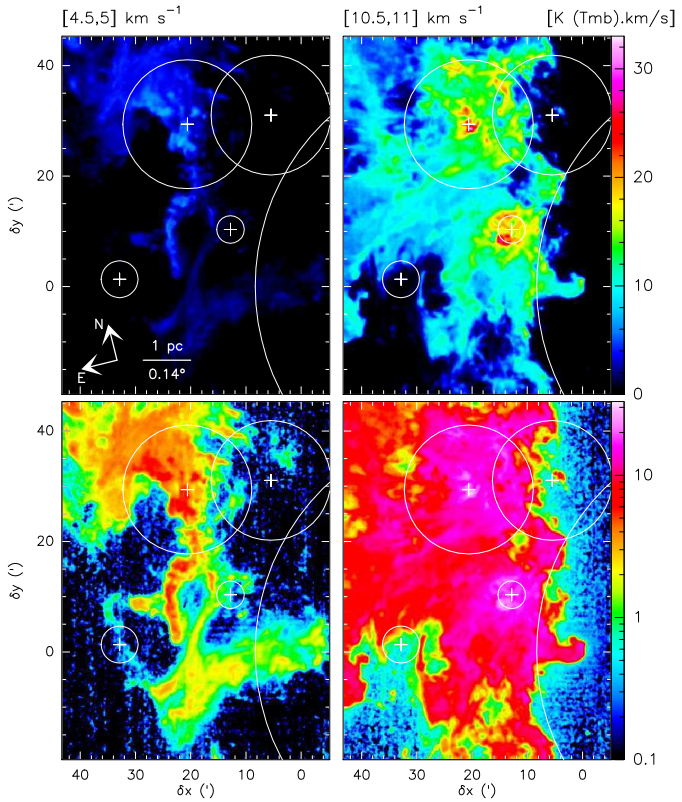


Fig. 5. Spatial distribution of the ^{12}CO (1−0) emission integrated over two different 0.5 km s^{-1} velocity ranges on linear (*top row*) and logarithmic (*bottom row*) color scales.

similar line width (3−4 km s^{-1}) and they overlap between 5 and 9.5 km s^{-1}. Figure 4 displays the fit for these two components on the ^{12}CO and ^{13}CO (1−0) spectra averaged over the FoV. The values of the ^{12}CO/^{13}CO integrated line intensity ratios are 6.4 and 11.8 for the main and the satellite components, respectively. The difference in line ratios suggests that the satellite velocity component corresponds to lower column density material (see Sect. 6.1). The satellite component is barely detected in C^{18}O, HCO$^+$, and ^{12}CS, and remains undetected for the other lines. Figure 5 shows the ^{12}CO (1−0) emission in two channel maps belonging to the two velocity components, both in linear and logarithmic color scales. The use of a logarithmic transfer function shows that bright emission is surrounded by a halo of faint emission. This shows that the fainter velocity component still covers a large fraction of the observed surface.

We now argue that both velocity components along the line of sight are associated with the Orion B GMC. Figure 5 shows that the spatial distribution of both components overlap on most of the observed field, and that they are in close interaction with the massive stars of known distance listed in Table 1. Furthermore, the 3D structure of interstellar extinction has been studied by Lallement et al. (2014) and Green et al. (2015) using differential reddening of stars at known distances. Towards Orion B, the reddening steeply increases between 300 and 500 pc, and most importantly, there is no significant reddening detected at closer or larger distances in this direction of the sky (Lallement, priv. comm.). These results are in excellent agreement with the distance determination of Orion through maser parallax (Menten et al. 2007). Overall the 3D structure of the Orion clouds is complex and could extend over several tens of parsec along the line of sight, a dimension comparable to the projected size on the plane of the sky.

Table 3 lists the integrated line intensities, W, and luminosities, L, computed as

$$\frac{L}{L_\odot} = 3.4\times10^{-8}\left(\frac{3\,\mathrm{mm}}{\lambda}\right)^3\left(\frac{W}{1\,\mathrm{K\,km\,s^{-1}}}\right)\left(\frac{D}{400\,\mathrm{pc}}\right)\left(\frac{\Omega}{1'\times1'}\right),$$

where $k_{\rm bolt}$ is Boltzmann constant, λ is the line rest wavelength, D is the source distance, and Ω is the FoV angle. The dynamic range of reliable integrated intensity is approximately 2400. Moreover, the typical intensity ratios of the (1−0) lines would be ^{12}CO/^{13}CO = 6.7, ^{13}CO/C^{18}O = 16.7, C^{18}O/^{12}CO = 0.9%, and HCO$^+$/^{12}CO ∼ HCN/^{12}CO ∼ 3%. ^{12}CN, C^{18}O, ^{12}CS, C$_2$H, and HNC emit approximately 1% of the ^{12}CO intensity. The low-J lines of ^{32}SO, C^{17}O, c-C$_3$H$_2$, N$_2$H$^+$, H^{13}CN, H^{13}CO$^+$ are up to 25 times fainter than the previous family, exemplified by C^{18}O. The HN^{13}C and SiO integrated intensity can not be reliably measured.

3.4. CO-traced, dust-traced, and virial-traced mass and densities

In this section, we compute the typical gas mass and densities using three common approaches: 1) the ^{12}CO (1−0) luminosity, 2) the dust continuum luminosity, and 3) the virial theorem. Table 2 lists the found values.

The direct sum of the pixel intensity over the mapped FoV and between the $[-2,+18]$ km s^{-1} velocity range indicates that the data cube contains a total CO luminosity of ∼2500 K km s^{-1} pc^2. Using the standard CO-to-H$_2$ conversion factor, $X_{\rm CO}$ = 2.0 × 10^{20} cm^{-2}/(K km s^{-1}) or

$4.35\,M_\odot\,\mathrm{pc}^{-2}/(\,\mathrm{K\,km\,s}^{-1})$ (this includes the factor 1.36 to account for the presence of helium, Bolatto et al. 2013), this corresponds to a gas mass of $1.1 \times 10^4\,M_\odot$. The total surface covered was 0.86 square degrees, that is, $43\,\mathrm{pc}^2$. The mean intensity and mean surface density are $61\,\mathrm{K\,km\,s}^{-1}$ and $260\,M_\odot\,\mathrm{pc}^{-2}$, respectively. The associated column density of gas is approximately $10^{22}\,\mathrm{H_2\,cm}^{-2}$. This in turn gives a typical volume density of $40\,M_\odot\,\mathrm{pc}^{-3}$ or $590\,\mathrm{H_2\,cm}^{-3}$.

Using the Gould Belt Survey data and its SED fits, we can derive values for the same quantities from dust far infrared emission. To do this, in Sect. 2.4 we first used a value different from the standard one for the conversion factor from A_K to A_V because 1) this is an observational quantity that can be measured relatively easily, and 2) we mainly deal with molecular gas, while the standard value is derived in diffuse gas. While this value depends on the optical properties (grain composition, grain shapes, and size distribution, which leads to the extinction curve) of the dust in Orion B, it is independent of any assumption about the gas properties. On the other hand, to derive the dust traced mass, we also need to use a value for the N_H/A_V ratio. Assuming that the dependency of this ratio on the dust optical properties is only a second order effect, this ratio mainly depends on the gas-to-dust ratio, that is, on how many grains there are per unit mass of gas. We thus use the standard value, $N_\mathrm{H}/A_V = 1.8 \times 10^{21}\,\mathrm{H\,cm}^{-2}\,\mathrm{mag}^{-1}$, for this ratio. This directly leads to a dust-traced mass of the mapped FoV of $\sim 3900\,M_\odot$, a mean surface density of $92\,M_\odot\,\mathrm{pc}^{-2}$ or $4 \times 10^{21}\,\mathrm{H_2\,cm}^{-2}$, and a mean volume density of $14\,M_\odot\,\mathrm{pc}^{-3}$ or $210\,\mathrm{H_2\,cm}^{-3}$.

The column density of neutral atomic hydrogen measured by integrating across profiles of the 21 cm H I line taken by the LAB all-sky H I survey (Kalberla et al. 2005) is

$$N(\mathrm{H\,I}) = 1.823 \times 10^{18}\,\mathrm{cm}^{-2} \int T_B \mathrm{d}v \approx (1.8 \pm 0.2) \times 10^{21}\,\mathrm{cm}^{-2}, \quad (5)$$

in the optically thin limit. This corresponds to approximately 1.0 mag of visual extinction using the usual conversion $N_\mathrm{H}/E_{B-V} = 5.8 \times 10^{21}\,\mathrm{cm}^{-2}\,\mathrm{mag}^{-1}$ derived by Bohlin et al. (1978) and $R_V = E_{B-V}/A_V = 3.1$. The total expected foreground gas contribution for a source at a distance of 400 pc is $N_\mathrm{H} = 1.2 \times 10^{21}\,\mathrm{cm}^{-2}$ for a local mean gas density $\langle n(\mathrm{H})\rangle = 1.15\,\mathrm{cm}^{-3}$ (Spitzer 1978) corresponding to $A_V = 0.75$ mag using the same conversion from column density to extinction. The minimum value of the visual extinction across the observed FoV, (i.e., 0.7) is therefore in good agreement with the expected contribution of diffuse material along the line of sight. As the mean visual extinction is 4.7 mag, correcting for this diffuse component would result in decreasing the molecular part of the dust-traced mass and densities by less than 20%. We choose to consider this difference negligible, that is, to consider that all the dust-traced mass refers to gas where hydrogen is molecular.

Following Solomon et al. (1987) and Bolatto et al. (2013), we can also compute a mass assuming that turbulent pressure and gravity are in virial equilibrium. Bolatto et al. (2013) indicate that the virial mass, M_vir, is given by

$$M_\mathrm{vir} = f\,R\,\sigma^2, \quad (6)$$

where R is the projected radius of the measured FoV, σ is the 1D velocity dispersion (full width at half maximum of a Gaussian divided by 2.35), and f is a factor that takes into account projection effects. This factor depends on the assumed density profile of the GMC. For a spherical volume density distribution with a power-law index k, that is,

$$\rho(r) \propto r^{-k}, \quad (7)$$

f is 1 160, 1 040, and 700 $M_\odot\,\mathrm{pc}^{-1}\,(\mathrm{km\,s}^{-1})^{-2}$, for $k = 0$, 1, and 2, respectively. In our case, $R \simeq 0.5\,\sqrt{5.6 \times 7.5} = 3.3$, and $\sigma \simeq 3.9/2.35 = 1.7\,\mathrm{km\,s}^{-1}$, when we only take into account the Gaussian fit of the main velocity component at approximately $10.5\,\mathrm{km\,s}^{-1}$. We thus obtain a virial mass between 6200 and $9500\,M_\odot$.

We find that, contrary to expectations, the CO-traced mass is typically three times the dust-traced mass, and that the virial mass is lower than the CO-traced mass but is much higher than the dust-traced mass. Throughout the paper, we propose that this discrepancy is related to the strong far-UV illumination of the mapped FoV (see Sect. 3.1). In the meantime, we will take an average between the CO-traced and dust-traced mass and densities when we require an order of magnitude estimate for these quantities.

4. Fraction of line fluxes from different gas regimes

In this section, we explore which fraction of the line fluxes comes from more or less dense gas, and from more or less far-UV illuminated gas.

4.1. Flux profiles over different A_V ranges

We chose four ranges of A_V, representing diffuse ($1 \le A_V < 2$), and translucent ($2 \le A_V < 6$) gas, the environment of filaments ($6 \le A_V < 15$), and dense gas ($15 \le A_V$). Table 4 lists the physical properties of the different regions based on their ^{12}CO (1–0) and far infrared emission. While the different regions have, by construction, increasing values of their mean visual extinction (1.4, 4, 9, and 29, respectively), they present similar mean dust temperature and far UV illumination. As expected the minimum dust temperature decreases when the range of visual extinction increases. In contrast, the maximum dust temperatures, and thus far UV illuminations, are also found in the masks of highest visual extinctions. This is related to the presence of very dense (probably cold) molecular gas in front of young massive stars that excite H II regions (see, e.g., the dark filament in front of IRS2 that excites the NGC 2024 nebula). This could also be due to the presence of embedded heating sources.

Contrary to standard expectations, the dust and CO-traced mass are similar for diffuse and dense regions, while they differ by a factor three, mostly in the translucent gas and filament environment. Moreover, both the dust and CO-traced matter indicate that approximately 50% of the gas lies in diffuse and translucent gas. Dense cores ($15 \le A_V$) represent between 10 and 20% of the mass but only 3% of the surface and 0.6% of the volume. The sum of the volume fractions of the four regions only amounts to 55% because of the simplified way the volumes are computed ($V = S^{3/2}$). This implies that volume densities can only be interpreted as typical values. Finally, the volume densities increase from ~ 100 to $7300\,\mathrm{H_2\,cm}^{-3}$ for diffuse and dense gas, respectively. Translucent gas and the environment of filaments have typical density values of ~ 500 and $1500\,\mathrm{H_2\,cm}^{-3}$, respectively. The volume density increases by a factor of between three and five from one gas regime to the next. We will use this fact to statistically identify high/low A_V lines of sight with high/low density gas, respectively.

Figure 6 presents the A_V masks and displays flux profiles integrated over regions of different extinction ranges. To better quantify the different behavior of the fluxes integrated over these different regions, Table 5 and Fig. 7 present, for each line, the percentage of the total line flux that comes from the different A_V

Table 4. Properties of the A_V masks sorted by increasing range of visual extinction.

Parameter	Unit	$1 \leq A_V < 2$	$2 \leq A_V < 6$	$6 \leq A_V < 15$	$15 \leq A_V < 222$
W_{CO}^{min}–W_{CO}^{mean}–W_{CO}^{max}	K km s^{-1}	0–6.8–46.9	1.3–62.6–211	24.8–122–261	41.5–137–288
A_V^{min}–A_V^{mean}–A_V^{max}	mag	1–1.4–2	2–3.6–6	6–8.5–15	15–28.9–222
T_d^{min}–T_d^{mean}–T_d^{max}	K	19–24–45	19–26–67	18–29–99	16–26–95
G_0^{min}–G_0^{mean}–G_0^{max}	ISRF (Habing 1968)	8.9–30–680	8.4–46–5100	6–72–36 000	4–47–28 000
CO-traced mass	$M_\odot$	320 (3%)	5600 (51%)	4400 (40%)	830 (8%)
Dust-traced mass	$M_\odot$	300 (8%)	1500 (38%)	1400 (36%)	790 (20%)
Emitting surface	pc^2	11 (25%)	21 (48%)	8.4 (20%)	1.4 (3.3%)
Typical volume	pc^3	36 (13%)	93 (33%)	24 (8.6%)	1.7 (0.6%)
CO-traced mean column density	$M_\odot$ pc^{-2}|10^{21} H$_2$ cm^{-2}	30 | 1.4	270 | 12	530 | 24	590 | 27
Dust-traced mean column density	$M_\odot$ pc^{-2}|10^{21} H$_2$ cm^{-2}	28 | 1.3	71 | 3.3	160 | 7.6	560 | 26
CO-traced mean volume density	$M_\odot$ pc^{-3}|H$_2$ cm^{-3}	8.9 | 130	60 | 890	180 | 2700	500 | 7500
Dust-traced mean volume density	$M_\odot$ pc^{-3}|H$_2$ cm^{-3}	8.4 | 130	16 | 230	57 | 850	470 | 7100

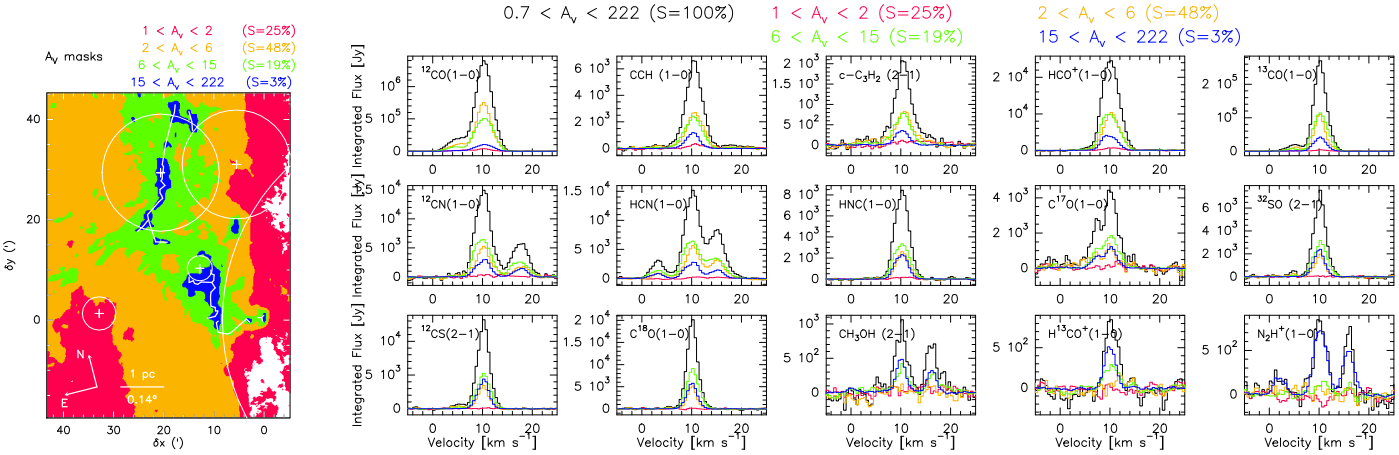


Fig. 6. *Left*: spatial distribution of the four following masks: $1 \leq A_V < 2$ in red, $2 \leq A_V < 6$ in orange, $6 \leq A_V < 15$ in green, and $15 \leq A_V < 222$ in blue. The percentages in the legend list the fraction of the surface contained in the different masks. *Right*: flux is integrated over the masks as a function of velocity. The spectrum of different colors shows the evolution of the flux in each line as a function of the mask used: all pixels observed in black, all pixels with $1 \leq A_V < 2$ in red, $2 \leq A_V < 6$ in orange, $6 \leq A_V < 15$ in green, and $15 \leq A_V < 222$ in blue.

masked regions (F_{A_V}/F_{tot}). In all cases, the fluxes are integrated in the [9, 12 km s^{-1}] velocity range. The lines were sorted by decreasing value of the sum of $F(1 \leq A_V < 2)$ and $F(2 \leq A_V < 6)$. This value represents the flux coming from both diffuse and translucent gas for each line. The layout of the panels in Fig. 6 also follows this order. We can group the lines into four categories depending on how the line flux is divided between regions of very low ($1 \leq A_V < 2$), low ($2 \leq A_V < 6$), intermediate ($6 \leq A_V < 15$), or high ($15 \leq A_V$) visual extinction.

In the first category of lines, the regions of low and intermediate visual extinctions contribute more than ~45% of the total flux, and regions of high visual extinction contribute less than ~20% of the flux. In this category, the total flux is predominantly coming from translucent lines of sight ($2 \leq A_V < 6$). This is the case of the (1–0) lines of ^{12}CO, HCO$^+$, C$_2$H, and the c-C$_3$H$_2$ (2–1) line. From these species, ^{12}CO is the one with the largest contribution (55%) from diffuse and translucent gas ($A_V \leq 6$).

In the 2nd category, the total flux is now predominantly coming from regions of intermediate visual extinction (45% coming from $6 \leq A_V < 15$), but the diffuse and translucent gas still contributes a similar fraction (35–40%) of the total flux, and dense gas does not contribute more than 20% of the total flux. This is the case of the (1–0) lines of ^{13}CO, HCN, and CN.

In the third category, the flux comes predominantly from regions of intermediate visual extinction as in the 2nd category, but the regions of low and high visual extinctions both contribute similar fractions of the total flux (approximately 30%). The (1–0) lines of HNC, C^{18}O, C^{17}O, and the lines of the sulfur species, namely the (2–1) line of ^{12}CS, and ^{32}SO, belong to this category.

The (1–0) lines of N$_2$H$^+$ and H^{13}CO$^+$, as well as the (2–1) lines of CH$_3$OH form the last category. In this one, the flux is predominantly coming from the regions of high visual extinction ($15 \leq A_V$). These lines all present a small surface filling factor (≲5%) and negligible contribution from the translucent and diffuse gas. In this category, N$_2$H$^+$ plays a special role. This is the only easily mapped line, where the flux is completely dominated (at 88%) by regions of high visual extinction; probably dense cores.

4.2. Flux profiles over different T_d ranges

We chose four ranges of T_d, representing cold dust ($16 \leq T_d < 19.5$ K) that corresponds to gas that is shielded from the UV field (e.g., the dense cores), lukewarm dust ($19.5 \leq T_d < 23.5$ K), warm dust ($23.5 \leq T_d < 32$ K), and hot dust ($32 \leq T_d < 100$ K). It is clear that this fitted dust temperature is biased toward the

Table 5. Percentage of the total line fluxes inside the four A_V mask regions, integrated over $[9, 12\,\mathrm{km\,s^{-1}}]$.

Species	Transition	$0 \le A_V < 222$	$1 \le A_V < 2$	$2 \le A_V < 6$	$6 \le A_V < 15$	$15 \le A_V < 222$
^{12}CO	1–0	100%	2.5%	52%	38%	7.6%
C_2H	1–0	100%	4.4%	41%	37%	17%
c-C_3H_2	2–1	100%	4.5%	40%	38%	17%
HCO^+	1–0	100%	2.8%	41%	40%	16%
^{13}CO	1–0	100%	1.4%	38%	45%	15%
HCN	1–0	100%	1.7%	36%	44%	18%
^{12}CN	1–0	100%	2.6%	33%	45%	19%
HNC	1–0	100%	2.1%	29%	41%	27%
C^{17}O	1–0	100%	5.6%	25%	43%	26%
^{12}CS	2–1	100%	0.68%	25%	42%	32%
^{32}SO	2–1	100%	0.86%	24%	44%	31%
C^{18}O	1–0	100%	0.49%	23%	48%	29%
CH_3OH	2–1	99%	4.4%	5.8%	41%	48%
H^{13}CO$^+$	1–0	98%	0.67%	7.1%	34%	56%
N_2H^+	1–0	100%	−11%	8.2%	17%	88%

Notes. The lines are sorted by decreasing value of the flux coming from the diffuse and translucent lines of sight.

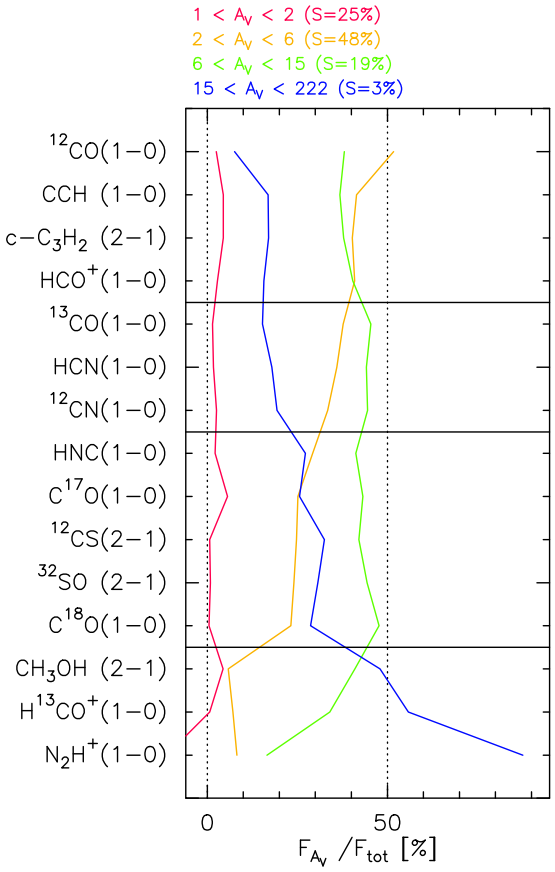


Fig. 7. For each line, flux integrated over each of the four A_V masks divided by the flux computed over the observed field of view. All fluxes are computed between 9 and 12 km s^{-1}. The black horizontal lines define the groups of lines described in Sect. 4.1.

presence of warm dust because the dust emissivity increases rapidly with temperature in the far infrared. Hence, cool dense gas is probably present along the line of sight of highest extinction, even though the fitted dust temperature is relatively high.

Here we use the dust temperature as a proxy for the typical far-UV illumination along the line of sight (see Sect. 2.4). In fact, using Eq. (3), we obtain a mean far-UV illumination of 9, 18, 50, and 400 for the cold, lukewarm, warm, and hot dust masks, respectively. This implies very different kinds of PDRs present along the line of sight. Moreover, the four masks of dust temperature display a morphology very different from that of the masks of visual extinction. Only the dense cores in Horsehead and near NGC 2023 are clearly delineated in both families of masks, while the intermediate density filamentary structure and diffuse/translucent gas are present in all four masks of dust temperature. Instead, the morphology of these temperature masks coincides well with the boundaries of the different H II regions. We thus interpret the cold, lukewarm, warm, and hot dust masks as very low, low, medium, and high far-UV illumination masks.

The FoV is dominated by intermediate far-UV illumination PDRs (83% of the surface have a G_0 between 10 and 120). Less than 2% of the lines of sight have $G_0 \le 10$ and approximately 15% have $G_0 > 120$. Moreover, the CO-traced mass is 0.86, 2.6, 3.1, and 3.5 times the dust-traced mass in the cold, lukewarm, warm, and hot dust regions, respectively. Similar ratios are found for the volume densities. This confirms that the discrepancy between CO and dust traced mass is linked to the enhanced far-UV illumination of the south-western edge of Orion B.

Figure 8 presents the T_d masks and displays the flux profiles integrated over regions of different far-UV illumination. To better quantify the different behavior of the fluxes integrated over these different regions, Table 7 and Fig. 9 presents, for each line, the percentage of the total line flux that comes from the different T_d masked regions ($F_{T_\mathrm{d}}/F_\mathrm{tot}$). In all cases, the fluxes are integrated in the $[9, 12\,\mathrm{km\,s^{-1}}]$ velocity range. The lines were sorted by decreasing distance between the sum of the flux coming from the highest far-UV illumination regions (red and orange masks) and the sum of the flux coming from the lowest far-UV illumination regions (green and blue masks). The layout of the panels in Fig. 8 also follows this order. While oscillations on the four individual curves of Fig. 9 are present, the general tendency is that the percentage of flux coming from the highest far-UV illuminated regions decreases from top to bottom. We can thus group the lines into four categories depending on whether the line flux

Table 6. Properties of the T_d masks sorted by increasing range of dust temperature.

Parameter	Unit	$16 \leq T_d < 19.5$	$19.5 \leq T_d < 23.5$	$23.5 \leq T_d < 32$	$32 \leq T_d < 100$
$W_{CO}^{min}-W_{CO}^{mean}-W_{CO}^{max}$	K km s^{-1}	0.2–63.8–149	0–46.5–197	0–57.4–273	0.2–114–288
$A_V^{min}-A_V^{mean}-A_V^{max}$	mag	1.1–17.2–127	0.7–4–69	0.8–4.1–222	0.8–7.2–186
$T_d^{min}-T_d^{mean}-T_d^{max}$	K	16–19–20	20–22–24	24–27–32	32–40–99
$G_0^{min}-G_0^{mean}-G_0^{max}$	ISRF (Habing 1968)	4–8.6–10	10–18–26	26–50–120	120–400–36 000
CO-traced mass	$M_\odot$	220 (2%)	3900 (35%)	4300 (39%)	2700 (25%)
Dust-traced mass	$M_\odot$	260 (7%)	1500 (38%)	1400 (36%)	770 (20%)
Emitting surface	pc^2	0.8 (1.8%)	19 (45%)	18 (41%)	5.5 (13%)
Typical volume	pc^3	0.7 (0.3%)	84 (30%)	73 (26%)	13 (4.6%)
CO-traced mean column density	$M_\odot$ pc^{-2}|10^{21} H$_2$ cm^{-2}	280 | 13	200 | 9.3	250 | 12	490 | 23
Dust-traced mean column density	$M_\odot$ pc^{-2}|10^{21} H$_2$ cm^{-2}	330 | 16	78 | 3.6	81 | 3.7	140 | 6.4
CO-traced mean volume density	$M_\odot$ pc^{-3}| H$_2$ cm^{-3}	310 | 4600	46 | 690	59 | 890	210 | 3100
Dust-traced mean volume density	$M_\odot$ pc^{-3}|H$_2$ cm^{-3}	380 | 5600	18 | 270	19 | 290	59 | 890

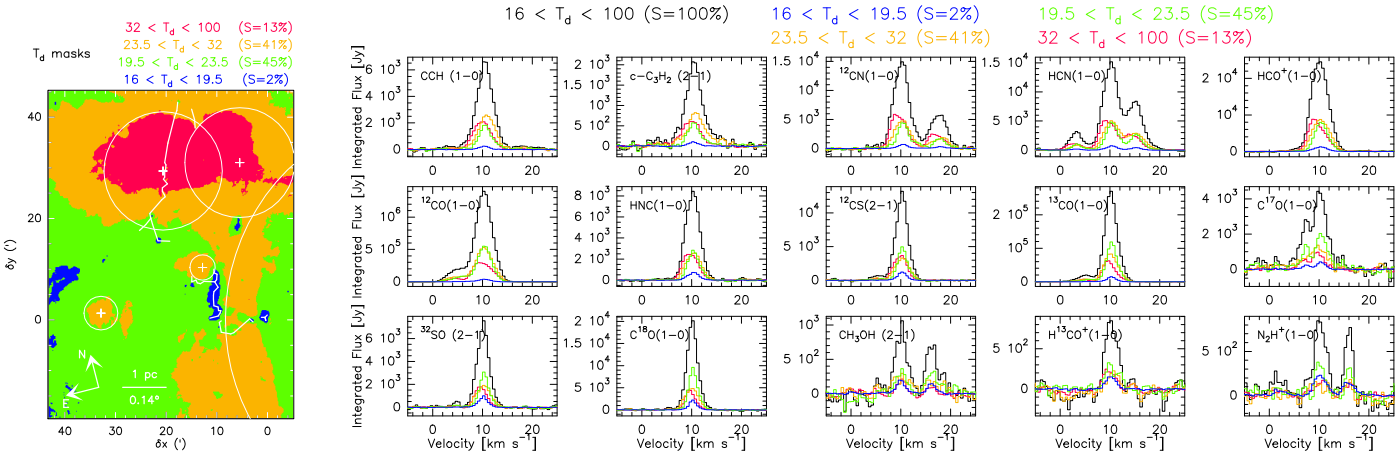


Fig. 8. *Left*: spatial distribution of the four following masks: $16 \leq T_d < 19.5$ K in blue, $19.5 \leq T_d < 23.5$ K in green, $23.5 \leq T_d < 32$ K in orange, and $32 \leq T_d < 100$ K in red. The percentages in the legend list the fraction of the surface contained in the different masks. *Right*: flux integrated over the masks as a function of velocity. The spectrum of different colors illustrates the evolution of the flux for the same line as a function of the mask used: all pixels observed in black, all pixels with $16 \leq T_d < 19.5$ K in blue, all pixels with $19.5 \leq T_d < 23.5$ K in green, all pixels with $23.5 \leq T_d < 32$ K in orange, and all pixels with $32 \leq T_d < 100$ K in red.

comes predominantly from the very low, low, medium, or high far-UV illumination regions.

In the first category, the regions of high far-UV illumination ($G_0 > 27$) contribute approximately 70% of the total line flux and the region of very low illumination ($G_0 < 10$) contributes less than 5%. High ($G_0 > 120$) and medium ($27 < G_0 < 120$) illumination regions contribute approximately equally to the total flux. The fundamental lines of the C_2H, c-C_3H_2, ^{12}CN, and HCN belong to this category.

In the second category, containing the HCO^+, HNC and ^{12}CO (1–0) lines, the line flux comes predominantly (65–75%) from intermediate far UV illumination regions ($10 < G_0 < 120$). The highest illumination region still contributes 20–30% of the total flux, while the lowest illumination region contributes less than 10%.

In the third category, the flux mostly comes from the region where $10 < G_0 < 27$. Quantitatively, this is the category where the flux coming from $G_0 < 27$ starts to dominate compared to medium and intermediate illumination regions. The (1–0) line of ^{13}CO and $H^{13}CO^+$, as well as the (2–1) line of ^{12}CS, and ^{32}SO belong to this category.

In the last category, the flux coming from regions where $G_0 < 27$ contributes between 52 and 66% of the total flux. This contains the CH_3OH (2–1), and the (1–0) line of the rarest CO isotopologues and N_2H^+.

5. Molecular low-*J* lines as a probe of the column density

5.1. Visual extinction vs. line integrated intensities

Figure 10 presents the joint distributions of the visual extinction and line integrated intensities for the studied molecular tracers. As the visual extinction is proportional to the amount of matter along the line of sight, it is desirable to make a comparison of all the matter traced by the molecules along this line of sight. Hence, the line profiles are integrated over the full velocity range where the line is measured, that is, not just integrated anymore on the [9, 12 km s^{-1}] velocity range. The visual extinctions are defined over the full FoV. In contrast, the line integrated intensities are well defined for only a fraction of the FoV. The joint distributions were thus only computed where the line integrated intensities are well defined (the criteria can be found in Sect. 2.5). For each distribution, the additional statistics (in particular for the visual extinction) are computed on this fraction of points. Table 8 lists these statistics.

The first obvious trend in Fig. 10 is the global correlation between the visual extinction, A_V, and the line integrated intensities, W. This correlation is clearly visualized through the comparison of the black curves, which show the typical behavior of the variations of A_V as a function of W, with the white lines

Table 7. Percentage of the total line fluxes inside the four T_d mask regions, integrated over [9, 12 km s^{-1}].

Species	Transition	$16 \leq T_d < 100$	$16 \leq T_d < 19.5$	$19.5 \leq T_d < 23.5$	$23.5 \leq T_d < 32$	$32 \leq T_d < 100$
C_2H	1–0	100%	3.2%	25%	39%	32%
c-C_3H_2	2–1	100%	4.2%	25%	39%	31%
^{12}CN	1–0	100%	4%	27%	33%	36%
HCN	1–0	100%	4.2%	29%	34%	32%
HCO^+	1–0	100%	4.6%	30%	36%	29%
HNC	1–0	100%	7.9%	31%	32%	29%
^{12}CO	1–0	100%	2.4%	38%	38%	22%
^{12}CS	2–1	100%	8.4%	35%	28%	29%
^{13}CO	1–0	100%	4.6%	41%	33%	21%
^{32}SO	2–1	100%	11%	38%	28%	23%
$H^{13}CO^+$	1–0	100%	16%	33%	24%	26%
$C^{17}O$	1–0	100%	7.9%	44%	29%	19%
CH_3OH	2–1	100%	18%	35%	29%	18%
$C^{18}O$	1–0	100%	9.1%	44%	27%	20%
N_2H^+	1–0	100%	27%	39%	15%	20%

Notes. The lines are sorted by decreasing value of the flux coming from the warm and hot dust lines of sight.

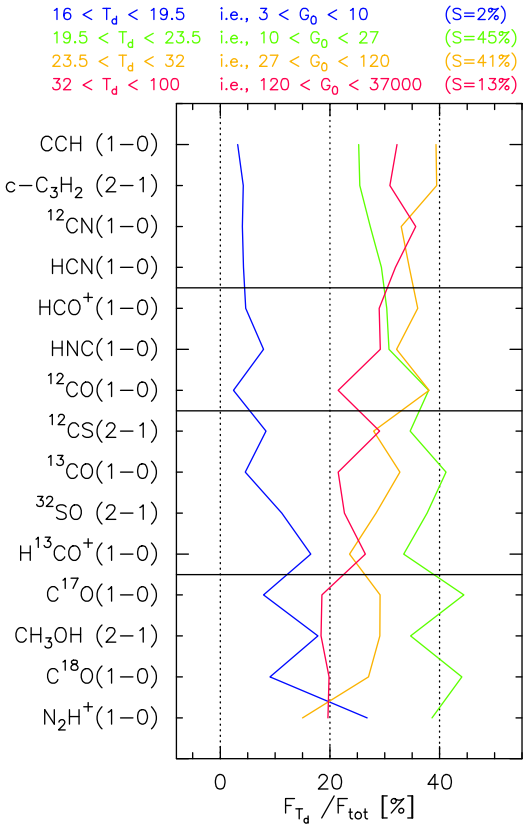


Fig. 9. For each line, flux integrated over each of the four T_d masks divided by the flux integrated over the observed field of view. All fluxes are computed between 9 and 12 km s^{-1}. The black horizontal lines define the groups of lines described in Sect. 4.2.

that represent a linear relationship between these two quantities. While the lines are often overly bright with respect to the white line at low extinction, and their integrated intensity sometimes saturates at high visual extinction, a correlation is clearly present for a large fraction of the measured lines of sight between these two regimes. More precisely, the $C^{18}O$, and HNC (1–0) lines are the best tracers of the visual extinctions when the integrated line intensity is above 1 K km s^{-1}. Indeed, there is an excellent agreement between the black line and the white line when W is above the intensity median value, and the scatter is low around these curves for both transitions. The $C^{18}O$, and HNC (1–0) lines are followed by ^{32}SO and ^{12}CS (2–1) lines. But these start to show a second twofold behavior at high visual extinction, a fraction of the pixels showing a saturation of the line integrated intensity at high visual extinction. This saturation branch is amplified for the HCN, ^{12}CN, and C_2H (1–0) lines. The ^{13}CO (1–0) line is also a good tracer of the visual extinction as it has a clearly monotonic (though non-linear) relationship with low scatter from $A_V \sim 2$ to $A_V \sim 20$.

The second trend concerns the visual extinction thresholds at which the lines become clearly detected. Lines that are detected over a smaller fraction of the mapped FoV show up at a higher A_V than lines with a more extended spatial distribution. Moreover, this threshold behavior is amplified when the position-position-velocity cubes are not smoothed at a common angular resolution in the first place. We emphasize two particular examples. First, the N_2H^+ (1–0) line has a surface filling factor of 2.4% and is detected at a median visual extinction of 26, while the filling factor of the HCO^+ (1–0) is 68% and this line is detected at a median visual extinction of 4.4, close to the median visual extinction at which ^{12}CO (1–0) is emitted. Second, this A_V-threshold behavior is also clear for the suite of CO isotopologues, where ^{12}CO and ^{13}CO (1–0) are detected at visual extinctions even lower than one, while $C^{18}O$ and $C^{17}O$ (1–0) are mostly detected for visual extinctions above three and six, respectively. The obvious explanation is related to detection limits. Rarer isotopologues produce weaker lines per unit column density, hence require a larger total gas column density to produce a signal above the detection threshold.

However, the A_V-flat asymptotes at low values of the integrated intensities are also evidence that it is not just a detection problem. Indeed, a linear relation is expected between the visual extinction and the integrated intensity at low values, that is, in the optically thin regime. The linear trend should thus only

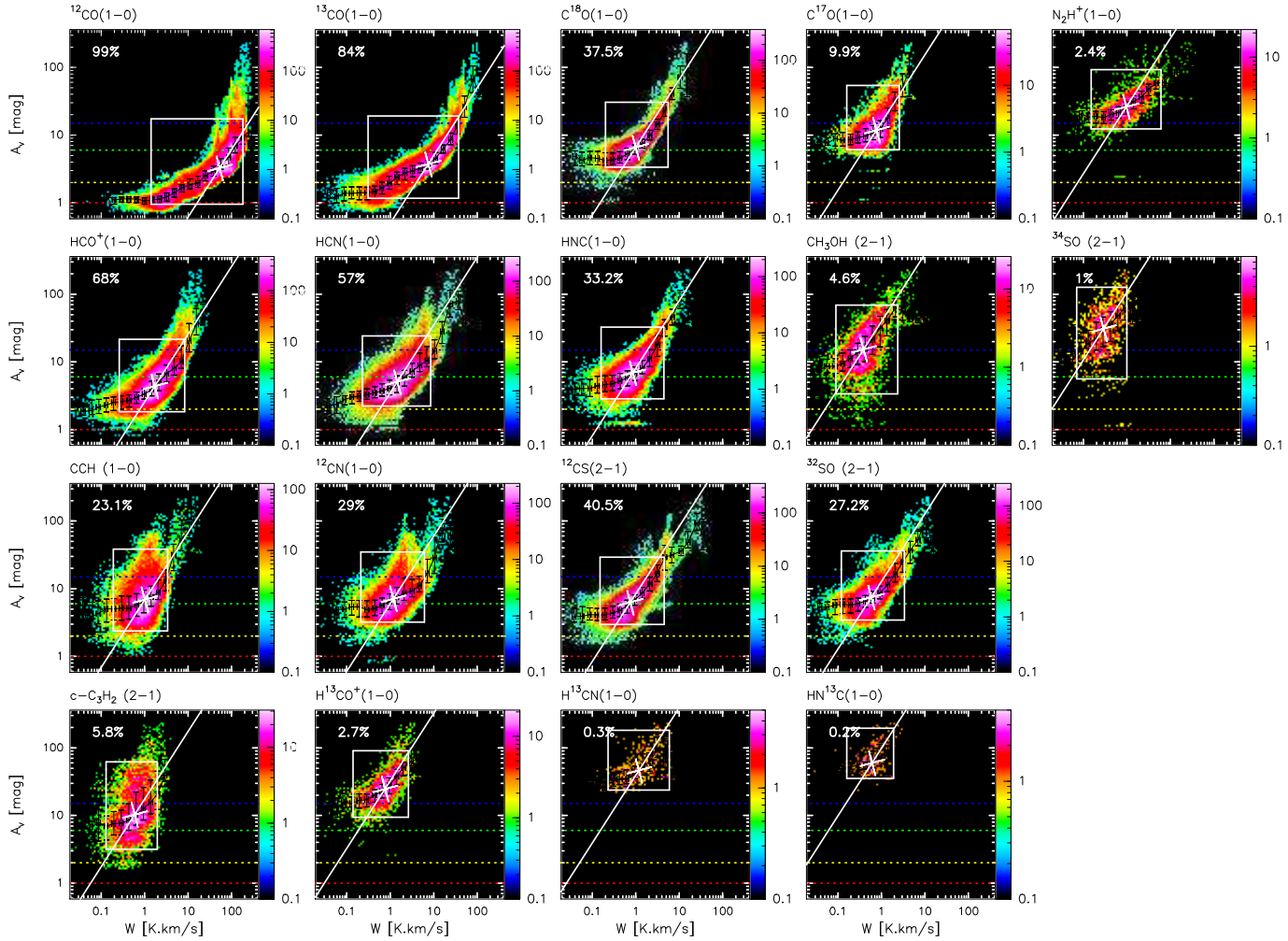


Fig. 10. Joint distributions of the visual extinction as a function of the line integrated intensity for a selection of the detected lines. The percentage in the top left corner indicates the surface over which the joint distributions can be reliably computed. These distributions contain both the global trend for the bulk of the gas and extreme behavior at low and high visual extinctions. The number of sightlines falling in a given 2D bin of the distribution is color-coded using a logarithmic scale to emphasize the extreme behavior (in particular the dense cores) that occupies a small fraction of the observed field of view. In contrast, the white rectangle displays the region of the distribution where 90% of the points are located: 2.5% of the points are outside this rectangle on each side. This allows us to define more robust global trends for the bulk of the gas. The white point shows the median of the two marginalized distributions. A line of unit slope, that is, a linear relationship between visual extinction and the line integrated intensities, is overlaid as the white plain line going through the white cross. The black points show the median values of all data points falling in a regularly sampled interval of the logarithm of the line integrated intensity. The black error bars show the range of values where 50% of the points in the current bin are located. This allows us to ask whether molecular lines are good tracers of the visual extinction. All these parameters are listed in Table 8. The red, orange, green, and blue horizontal dashed lines show the visual extinction limits (1, 2, 6, and 15 mag, respectively) used in the masks of Fig. 6. This enables us to visualize the amount of well detected pixels that falls in each of the masks for each line.

be interrupted at the detection threshold. In contrast, there is a A_V-threshold above which the species starts to emit. This is corroborated by intensity ratios that do not match the values expected from the known carbon isotope ratios, even at low visual extinction where optical depth effects are negligible. The A_V-thresholds could either be explained in terms of chemical or dynamical reasoning. Turbulent mixing between the phases of the ISM or the existence of dense but diffuse globulettes at the edge of H II regions belong to the latter category. In the former category, we have selective chemistry.

In summary, the (1−0) or (2−1) lines of molecular tracers are, to first order, sensitive to different ranges of visual extinction when detected at a similar noise level. In addition, they are, overall, well correlated with the amount of matter along the line of sight. This behavior will be quantified in a following paper concerning the PCA of the dataset (Gratier et al. 2017).

5.2. Tracer luminosities per proton

Figure 11 shows the spatial distribution of the ratio of the line integrated intensity to the visual extinction. The panels show these ratios for the molecular tracers ordered in the same way as the figure displaying the line integrated intensities (Fig. 2). We also added the spatial distribution of the visual extinction and dust temperature as the top right and bottom left panels, respectively, for reference. The intensity ratios are normalized by their median values and the intensities are displayed using a logarithmic scale symmetrically stretched around 1. This eases the visualization of departure of the ratio by a multiplicative factor, such as 1/2 and 2, for example. The luminosity per proton is easily computed by dividing the W/A_V ratios by the standard value of $N_{\rm H}/A_V = 1.8 \times 10^{21}\,{\rm H\,cm^{-2}/\,mag}$.

Table 8. Properties of the joint distributions of visual extinction and line integrated intensities.

Species	Transition	Filling factor %	$W^{2.5\%}-W^{\rm med}-W^{97.5\%}$ K km s^{-1}	$A_V^{2.5\%}-A_V^{\rm med}-A_V^{97.5\%}$ mag
^{12}CO	1–0	99	1.4–53–180	0.95–3.3–17
^{13}CO	1–0	84	0.31–7.5–38	1.2–3.7–19
HCO$^+$	1–0	68	0.26–1.7–8.3	1.8–4.4–22
HCN	1–0	57	0.23–1.5–8.5	2.2–5.1–24
^{12}CS	2–1	41	0.15–0.75–4.5	2.9–6.3–29
C^{18}O	1–0	38	0.2–1–5.5	3.4–6.6–31
^{12}CN	1–0	29	0.21–1.2–6.1	3.2–7.1–35
C_2H	1–0	23	0.19–1–3.4	2.4–6.9–38
^{32}SO	2–1	27	0.12–0.59–3.4	3.4–7.5–36
HNC	1–0	33	0.16–0.86–4.4	2.9–6.9–33
C^{17}O	1–0	9.9	0.16–0.77–2.6	6.1–12–54
c-C_3H_2	2–1	5.8	0.13–0.59–2	3.2–10–62
CH_3OH	2–1	4.6	0.089–0.39–2.4	3.4–15–69
H^{13}CO$^+$	1–0	2.7	0.14–0.76–2.6	9.4–25–91
N_2H^+	1–0	2.4	0.15–0.97–6.1	12–26–93
^{34}SO	2–1	1	0.07–0.29–1	5.6–31–130
H^{13}CN	1–0	0.32	0.23–1.2–5.9	24–46–180
HN^{13}C	1–0	0.18	0.16–0.61–1.9	35–60–190

Notes. The lines are sorted by decreasing value of their surface filling factor.

The $W(^{12}\mathrm{CO})/A_V$ and $W(^{13}\mathrm{CO})/A_V$ present a similar pattern, that is, a luminosity per proton higher than the median value in translucent gas and lower in dense gas. The luminosity per proton decreases again at the very edge of the molecular cloud. The $W(\mathrm{C}^{18}\mathrm{O})/A_V$ ratio shows less variation by a factor of two to three. The $W(\mathrm{HCO}^+)/A_V$, $W(\mathrm{HCN})/A_V$, $W(\mathrm{HNC})/A_V$, and $W(^{12}\mathrm{CN})/A_V$ ratios show maxima associated with the Orion B Eastern edge and with the NGC 2024 H II bubble. The dark filament in front of NGC 2024 delineates the frontier between ratios higher/lower than the median. This can be interpreted as an excitation effect due to higher electron density or an abundance effect. The west/east asymmetry of the ratio is more marked for the $W(\mathrm{C_2H})/A_V$ ratio, in particular around the NGC 2023 region. The $W(^{12}\mathrm{CS})/A_V$ shows a specific pattern with a maximum at the center of NGC 2024 and a minimum between NGC 2024 and NGC 2023.

5.3. Typical abundances

As the low-J molecular lines are, overall, well correlated to the column density of molecular gas, the luminosities per proton could, in principle, be used to estimate the abundance of the different species. To do this, we computed the column density of each species, $N_{\rm species}$, that is required to produce an integrated intensity of 1 K km s^{-1} assuming that the gas is at local thermal equilibrium. The values of $N_{\rm species}$ vary by less than 20% when the temperature increases from 20 to 30 K. Typical abundances with respect to the proton number can then be computed with

$$[\mathrm{species}] = \frac{W}{A_V}\,\frac{A_V}{N_{\rm H}}\,\frac{N_{\rm species}}{1\,\mathrm{K\,km\,s^{-1}}}. \quad (8)$$

Table 9 lists the minimum, maximum, and median values of the so-called abundances for each line. The deduced abundances are reasonable for all the studied lines except ^{12}CO (1–0), which delivers abundances that are too low by one order of magnitude because this line is highly optically thick.

Table 9. Minimum, median, and maximum values of the abundances derived for each species.

Species	Transition	$(W/N_{\rm H})\times(N_{\rm species}/1\,\mathrm{K\,km\,s^{-1}})$ [Pseudo-Abundance] min–med–max
^{12}CO	1–0	1×10^{-7}–1×10^{-5}–3×10^{-5}
^{13}CO	1–0	1×10^{-7}–1×10^{-6}–4×10^{-6}
C^{18}O	1–0	2×10^{-8}–1×10^{-7}–1×10^{-6}
C_2H	1–0	1×10^{-9}–1×10^{-8}–1×10^{-7}
C^{17}O	1–0	1×10^{-8}–5×10^{-8}–4×10^{-7}
CH_3OH	2–1	1×10^{-10}–1×10^{-9}–3×10^{-8}
^{32}SO	2–1	1×10^{-10}–1×10^{-9}–5×10^{-9}
^{12}CN	1–0	2×10^{-10}–2×10^{-9}–2×10^{-8}
HCN	1–0	1×10^{-10}–1×10^{-9}–4×10^{-9}
^{12}CS	2–1	1×10^{-10}–4×10^{-10}–5×10^{-9}
c-C_3H_2	2–1	2×10^{-11}–3×10^{-10}–3×10^{-9}
HCO$^+$	1–0	4×10^{-11}–3×10^{-10}–2×10^{-9}
N_2H^+	1–0	2×10^{-11}–2×10^{-10}–3×10^{-9}
HNC	1–0	3×10^{-11}–2×10^{-10}–2×10^{-9}
H^{13}CO$^+$	1–0	1×10^{-11}–3×10^{-11}–4×10^{-10}

Notes. The lines are sorted by decreasing value of the median pseudo-abundance.

6. Line ratios as tracers of different physico-chemical regimes

Line intensity ratios are commonly used to study the physical and chemical properties of the gas in different environments. The advantage of using line ratios instead of absolute line intensities is that this method allows removal of calibration uncertainties (when lines are observed simultaneously). It is then easier to compare one source to another. Line ratios may also remove

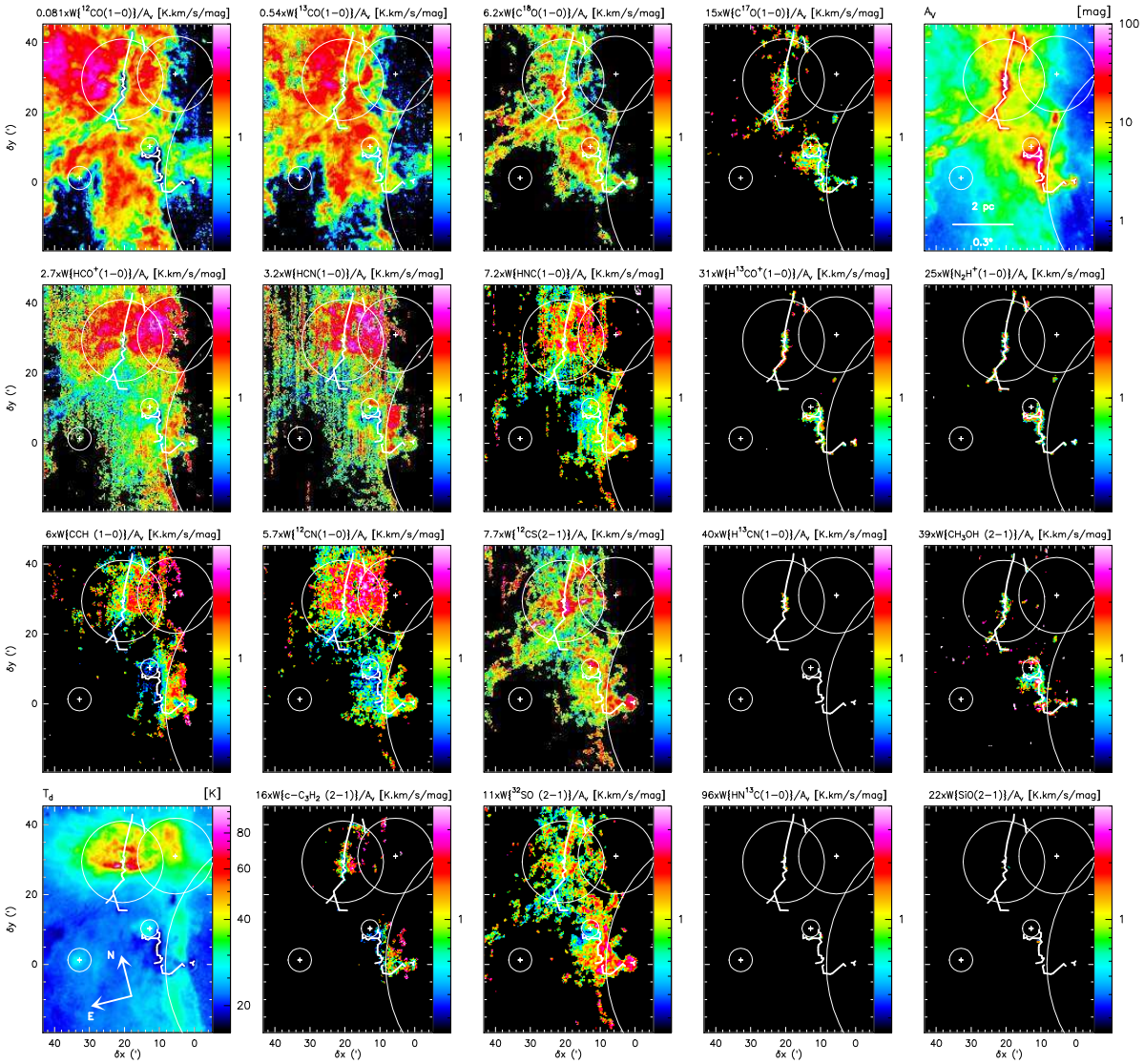


Fig. 11. Spatial distribution of the ratio of the line integrated intensity to the visual extinction for some of the detected lines in the 3mm band, plus the dust temperature (*bottom left panel*) and the visual extinction (*top right corner*). The ratios are normalized by their median value that hence appears as 1 on the color look-up table. The color scale shows ratio values between 0.25 and 4 times the median value for all the ratio panels.

excitation effects and bring forward actual chemical variations. Our knowledge of the chemistry of the gas then allows us to use line ratios to constrain the physical properties of the gas.

An important basic property we wish to determine easily from observations is the density of the gas. Forming dense gas is a required step of star formation, and the availability to form dense gas may regulate star formation efficiency (Lada et al. 2013). Line ratios of HCN and HCO^+ with respect to ^{12}CO and ^{13}CO are commonly used to trace the fraction of dense gas in galactic and extragalactic GMCs (e.g., Lada et al. 2012). This is because ^{12}CO and ^{13}CO can be excited at low densities ($\sim 10^2$ cm^{-3}) compared to HCN and HCO^+, which are expected to be excited only at high densities ($\sim 10^5$ cm^{-3}). Indeed, the HCN/^{12}CO ratio is observed to be well correlated with the star-formation efficiency, traced by IR/HCN in M 51 (e.g., Bigiel et al. 2016).

In this section we first show line ratios involving the brightest detected lines, that is, ^{12}CO and ^{13}CO (1−0), and we conclude with a few other interesting ratios. For this, we discuss 2D-histograms of the ratio denominator vs. the ratio numerator, the spatial distribution of the ratios, and the 2D-histogram of the ratio vs. the visual extinction. This will allow us to study the correlations present before computing the ratio, in order to visually assess the correlations of the line ratios with different kinds of regions, and to quantitatively study potentially remaining correlations with the visual extinction.

6.1. Ratios with respect to ^{12}CO and ^{13}CO (1–0)

The 2D-histograms shown in Fig. 12 display the relation of the integrated intensity of different lines as a function of the integrated intensity of the ^{12}CO (1−0) line. While the eye is mainly caught by the saturation of the ^{12}CO line, that is, other tracer intensities increase by a large factor when $W(^{12}CO) \sim 100-200$ K km s^{-1}, most of the data follows a different trend. The running median and running interval containing 50% of the data, materialized as black points and error bars, indicate that most of the tracers have first a relatively constant integrated intensity as $W(^{12}CO)$ increases, and then their integrated intensity is well correlated to $W(^{12}CO)$. As shown by the white rectangles that display the part of the 2D-histogram populated by 90% of the points, most of the tracers only emit when the ^{12}CO (1−0)

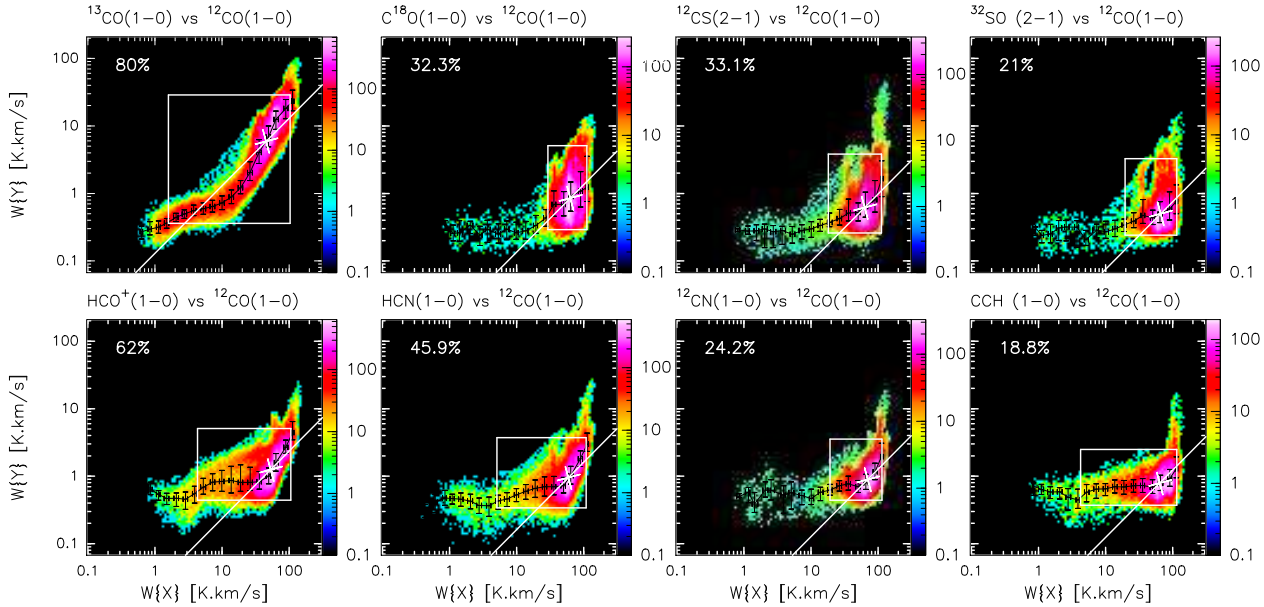


Fig. 12. Joint distributions of various line intensities integrated over [9, 12 km s^{-1}]. Markers have the same signification as in Fig. 10. The x-axis is always $W\left\{^{12}\mathrm{CO}(1\text{–}0)\right\}$.

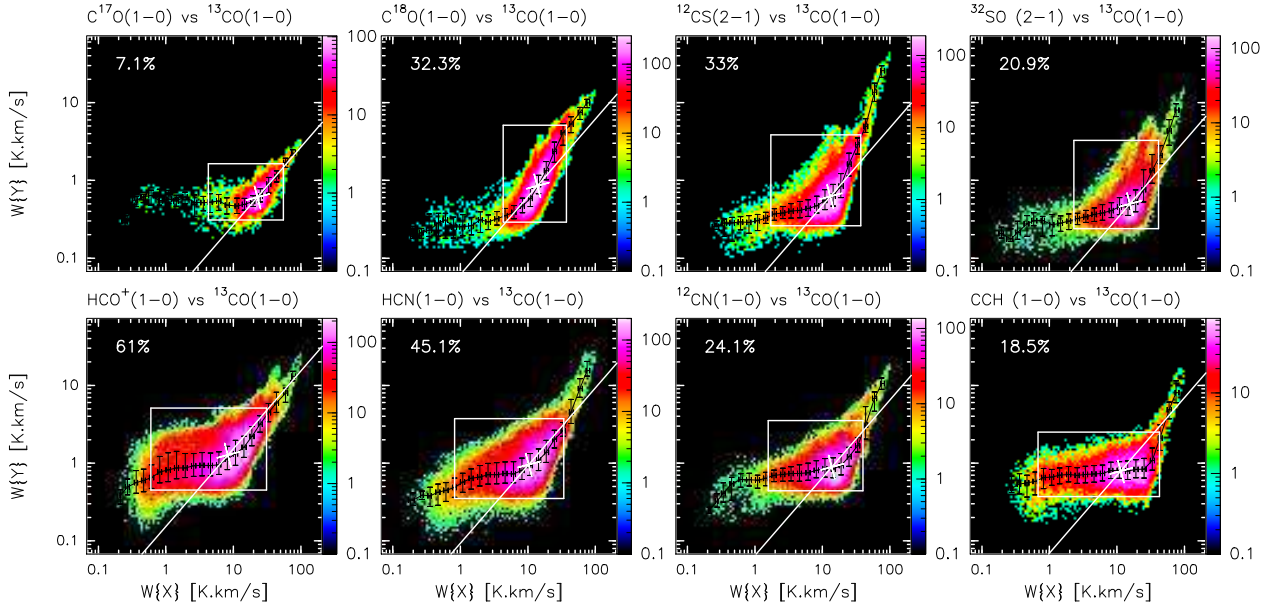


Fig. 13. Same as Fig. 12, except that the x-axis is now $W\left\{^{13}\mathrm{CO}(1\text{–}0)\right\}$.

line is already relatively bright at ∼10–20 K km s^{-1}. On the other hand, the CCH, HCN, HCO$^+$, and ^{13}CO (1–0) lines show a significant fraction of the data at ^{12}CO (1–0) integrated intensity between 1 and 10 K km s^{-1}. The ^{13}CO (1–0) line has a specific behavior as it is under-luminous with respect to a linear correlation going through the median behavior at intermediate ^{12}CO (1–0) intensities ($5 \lesssim W(^{12}\mathrm{CO}) \lesssim 30$ K km s^{-1}).

Figure 13 shows the same 2D-histograms as before but with respect to $W(^{13}\mathrm{CO})$. In general, the same trends seen for ^{12}CO are seen for ^{13}CO. The species integrated intensities have a relatively constant or slightly increasing integrated intensity as $W(^{13}\mathrm{CO})$ increases up to ∼10 K km s^{-1}. Their intensity is then well correlated to $W(^{13}\mathrm{CO})$. The effects of the ^{13}CO (1–0) saturation are visible but less pronounced than with respect to the ^{12}CO (1–0) line.

Figure 14 presents the spatial distribution of the line ratios involving ^{12}CO. The ratios are normalized by their median value to emphasize the symmetric departure of the ratios compared to the general trend. The ratios all show minimum values in the dense regions associated with NGC 2024, NGC 2023, and Horsehead. This probably reflects the saturation of the ^{12}CO emission in regions with the highest column density. These regions are also the densest regions, which implies that the molecular tracers are easily produced and excited. As the ^{12}CO line becomes saturated, the available energy gets carried away by other ^{12}CO transitions or other molecular species. A west-east gradient is superimposed on the first pattern for the ratios involving HCO$^+$, HCN, C_2H, ^{12}CN, and HNC (not shown), for both ^{12}CO and ^{13}CO. The minimum values are obtained on the western edge; the maximum values in the eastern diffuse

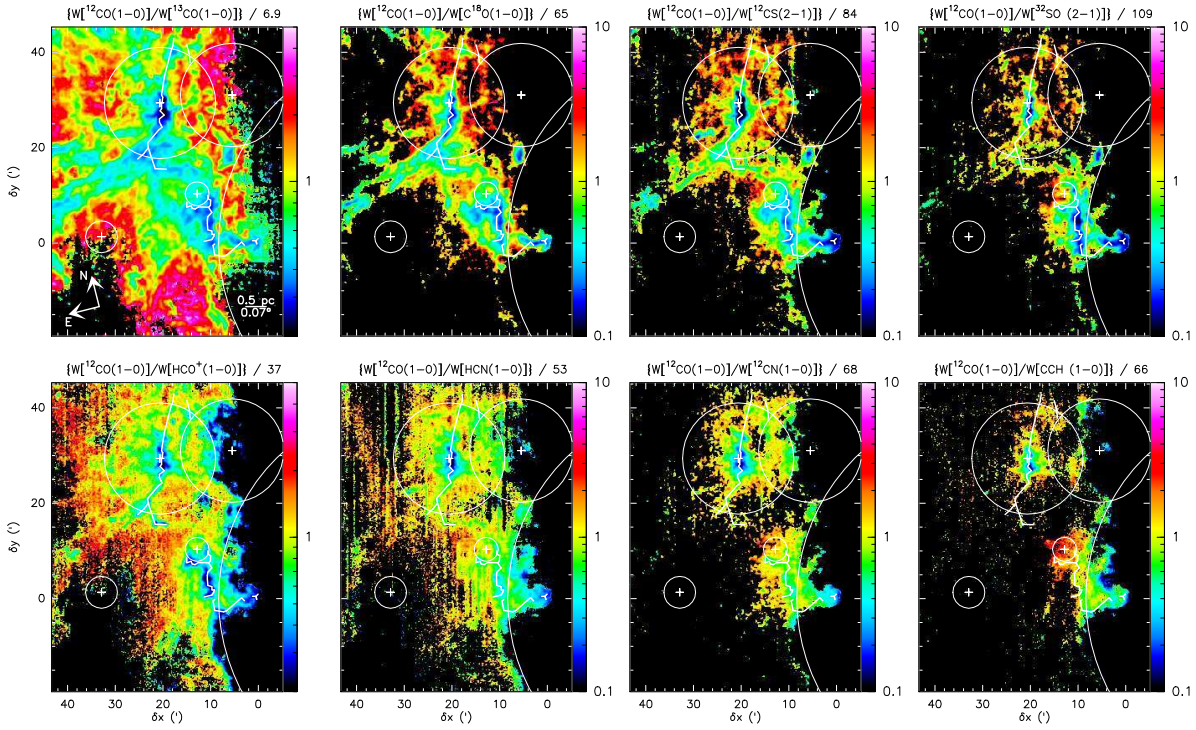


Fig. 14. Spatial distribution of the ratios of line intensity integrated over [9, 12 km s^{-1}]. The ratios are normalized by the median value of the ratio. The numerator is always $W\left\{^{12}\mathrm{CO}(1-0)\right\}$. The color scale shows ratio values between 0.1 and 10 for all the ratio panels, except the ^{12}CO/^{13}CO and ^{12}CO/HCO$^+$ panels where the color scale goes from 0.2 to 5.

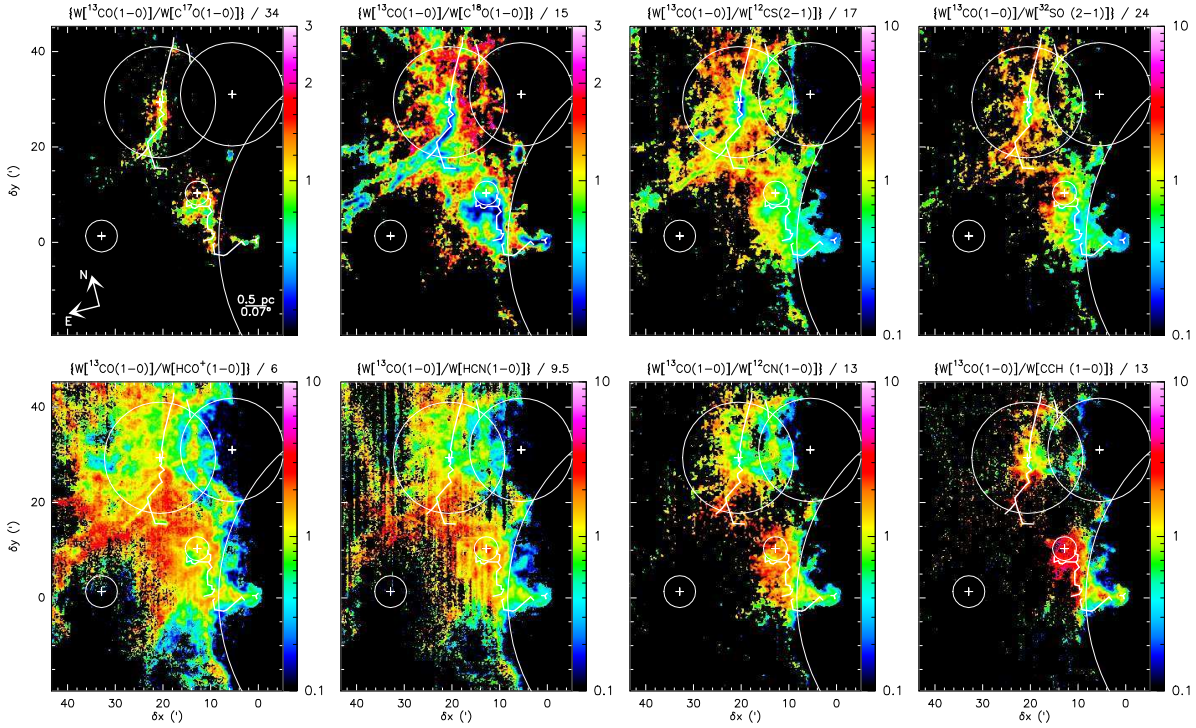


Fig. 15. Same as Fig. 14, except that the numerator is now $W\left\{^{13}\mathrm{CO}(1-0)\right\}$. The color scale shows ratio values between 0.1 and 10 for all the ratio panels, except the ^{13}CO/C^{17}O and ^{13}CO/C^{17}O panels where the color scale goes from 0.33 to 3.

region. This pattern probably indicates a gradient in excitation and abundance in UV-illuminated regions for molecules sensitive to the far-UV radiation. Finally an approximately circular structure around NGC 2024 with a luminosity deficit in C^{18}O, ^{12}CS, and marginally ^{32}SO probably traces UV-illuminated material near NGC 2024.

Similar spatial behavior is also seen for the ratios with respect to ^{13}CO (1−0) (see Fig. 15), but slightly attenuated because the saturation of the ^{13}CO (1−0) line is less pronounced. The first pattern (minimum ratio values in regions of highest density) for the ratios including C^{18}O, ^{12}CS, and ^{32}SO supports the interpretation in terms of opacity for the densest/brightest regions.

Table 10. Minimum, median, and maximum values of the ratios of diverse line integrated intensities.

Species #1	Trans.	Species #2	Trans.	W(#1)/W(#2) min−med−max
^{12}CO	1−0	^{13}CO	1−0	1.2−6.9−46
^{12}CO	1−0	HCO$^+$	1−0	1.6−37−140
^{12}CO	1−0	C^{18}O	1−0	2.9−65−410
^{12}CO	1−0	^{12}CN	1−0	1.9−68−210
^{12}CO	1−0	C_2H	1−0	1.8−66−320
^{12}CO	1−0	HNC	1−0	1.8−66−320
^{12}CO	1−0	^{12}CS	2−1	2.4−84−350
^{12}CO	1−0	^{32}SO	2−1	2.7−110−400
^{13}CO	1−0	HCO$^+$	1−0	0.16−6−30
^{13}CO	1−0	C_2H	1−0	0.4−13−91
^{13}CO	1−0	^{12}CN	1−0	0.47−13−57
^{13}CO	1−0	C^{18}O	1−0	0.57−15−50
^{13}CO	1−0	HNC	1−0	0.4−13−91
^{13}CO	1−0	^{12}CS	2−1	0.71−17−70
^{13}CO	1−0	^{32}SO	2−1	0.67−24−98
^{13}CO	1−0	C^{17}O	1−0	0.65−34−85
HNC	1−0	^{12}CN	1−0	0.33−1.7−6.4
^{12}CS	2−1	HNC	1−0	0.22−1.1−7.3
^{12}CN	1−0	HNC	1−0	0.22−1.1−3.9
C^{18}O	1−0	^{12}CS	2−1	0.14−1.2−6.2
C^{18}O	1−0	HNC	1−0	0.16−1.3−8.5
HCN	1−0	^{12}CN	1−0	0.33−1.7−6.4
HCO$^+$	1−0	HCN	1−0	0.33−1.7−6.4
HCN	1−0	HNC	1−0	0.33−1.7−6.4
HCO$^+$	1−0	HNC	1−0	0.33−1.7−6.4
C^{18}O	1−0	C^{17}O	1−0	0.43−4.3−8.6
HCO$^+$	1−0	N_2H^+	1−0	0.33−1.7−6.4

Notes. In each group, the lines are sorted by increasing value of the W(#1)/W(#2) median.

The east-west pattern is even more pronounced for the HCO$^+$ species with an excess emission of the molecular tracers in the UV-illuminated regions and a deficit in the diffuse/translucent gas. This may be a combined effect of lower heating, moderate density, and an increase of ^{13}CO due to isotopic fractionation (^{12}CO + ^{13}C$^+$ $\longrightarrow$ ^{13}CO + ^{12}C$^+$). Finally, the east-west pattern does not reach the translucent regions on the eastern side for C_2H, and CN. In these cases, we mostly see the increase of the line ratio in the high extinction gas, including the compressed western edge (Schneider et al. 2013).

Finally, Fig. 16 shows the 2D-histograms of A_V as a function of the line ratios involving ^{12}CO. The line ratios have a bimodal behavior relative to A_V, with values lower than the median (marked by the white cross) found both for high- and low-A_V regions. Values of the ratios higher than the median are associated with a small range of visual extinctions, either the translucent ($2 < A_V < 6$) or filamentary ($6 < A_V < 15$) gas. The bimodal trend is present in all ratios, but more pronounced for those involving lines presenting an extended emission (^{13}CO/^{12}CO and HCO$^+$/^{12}CO).

The increasing branch (in A_V) is the dominating one for the ratios involving ^{13}CO, C^{18}O, ^{12}CS, ^{32}SO, and ^{12}CN. This means that the other decreasing branch, while existing, represents a small number of points in our FoV. Low values of these ratios thus mostly point to high-density regions. The increasing branch is most likely a consequence of the saturation of the ^{12}CO (1−0) line compared to the other lines at large gas column densities. In addition, many molecular species will become more abundant at large column densities, as they become shielded from UV-radiation. Both effects will produce lower ratios at large A_V. Because all tracers are correlated to first order to the column density, one would expect to remove a correlation with A_V by taking the ratio of two lines. However, a (anti-)correlation may remain. This is due to the fact that the correlation between the integrated intensity of weaker lines and W(^{12}CO) have a non-linear behavior, probably because these lines have a lower opacity than the ^{12}CO (1−0) line for large gas column densities.

In contrast, the decreasing branch dominates for the ratios involving the HCO$^+$, HCN, and C_2H (1−0) lines. Low values of these ratios point to the lower visual extinction range. This is clear for ^{12}CO/C_2H, whose running median almost monotonically increases from an A_V of ∼2 to 10 mag. When the ^{12}CO(1−0)/HCO$^+$(1−0) and ^{12}CO(1−0)/HCN(1−0) ratios increase, the running median of the visual extinction first increases from values lower than 2 mag up to ∼8 mag, and it then starts to decrease again. One interesting result is that all decreasing branches sample values of the visual extinction as low as 1−2 mag. This is consistent with the fact that all the associated species are detected in diffuse clouds through absorption against extra-galactic continuum sources (Lucas & Liszt 1996, 2000; Liszt & Lucas 2001). This probably means that the strongly polar species, C_2H, and HCO$^+$, reach a radiative regime where they emit more efficiently than ^{12}CO, the weak excitation regime described by Liszt & Pety (2016).

Figure 17 shows the 2D-histograms of the visual extinction as a function of the ratio involving ^{13}CO. We see similar global results as for ^{12}CO, that is, a globally constant visual extinction at values of the ratio higher than the median, and an increasing and a decreasing A_V branch when the ratio decreases below the median value. In contrast with the ratios involving ^{12}CO (1−0), the decreasing A_V branch dominates for most of the ratios. Indeed, the running median for all ratios except C^{18}O/^{13}CO decrease for low visual extinctions. The decrease is even almost monotonic for ratios involving the HCO$^+$, C_2H, ^{12}CN, and HCN (1−0) lines. This behavior is consistent with the fact that the correlation between the integrated intensity of the different species and W(^{13}CO) is more linear. Taking the ratios thus better removes the correlations with the gas column density.

6.2. Various other line ratios

Figures 18 to 20 show the same plots as the previous section but for other line ratios including tracers of the column density (C^{18}O, ^{12}CS, and HNC), and other highly studied ratios in extra-galactic observations (HCO$^+$, HCN, HNC, and CN).

The 2D-histograms that display the integrated intensity of C^{17}O, HNC, and ^{12}CS with respect to W(C^{18}O) show two main behaviors (see Fig. 19). The C^{17}O, HNC, and ^{12}CS (2−1) lines are over-luminous compared to the C^{18}O (1−0) line at low intensities. Their intensities then become linearly correlated. In addition, the ^{12}CS (2−1) line becomes much brighter than both the C^{18}O and HNC (1−0) lines at high-intensity values. The integrated brigthnesses of the HCN-HCO$^+$, HNC-HCO$^+$, HNC-CN, and HNC-HCN (1−0) line pairs are all linearly correlated. The best correlation is found for the HCN-CN pair of lines.

Describing the spatial patterns of the ratios shown in Fig. 19 is less straightforward because the fraction over the FoV where the two lines are detected at enough signal-to-noise ratio is

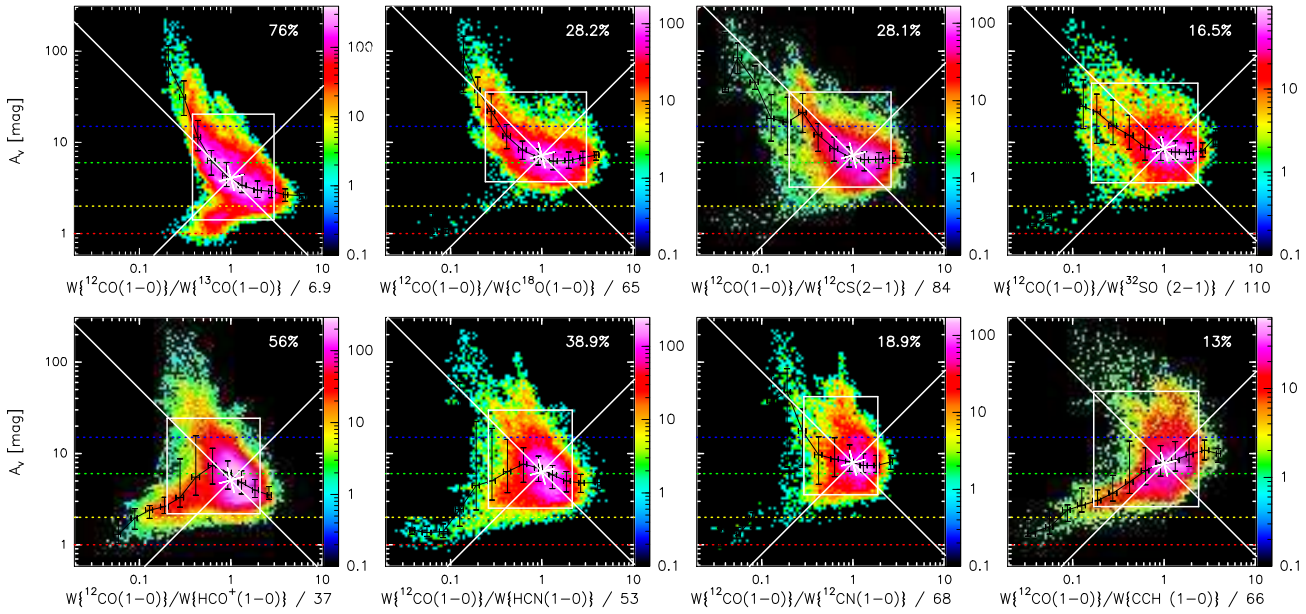


Fig. 16. Joint distributions of the visual extinction as a function of the ratios of line integrated intensities. The ratios are normalized by their median value. The ratio numerator is always here $W\left\{^{12}\mathrm{CO}(1-0)\right\}$. Markers have the same signification as in Fig. 10.

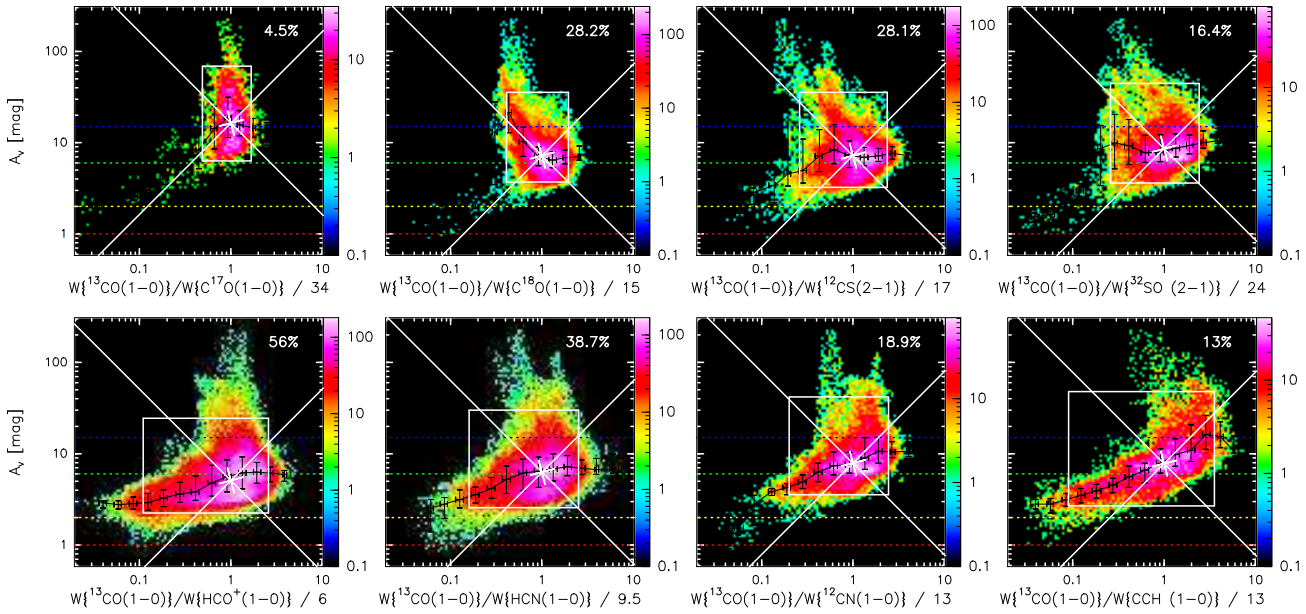


Fig. 17. Same as Fig. 16, except that the ratio numerator is now $W\left\{^{13}\mathrm{CO}(1-0)\right\}$.

lower. The $C^{18}O/C^{17}O$ ratio, shown in the upper left panel, is fairly flat with no clear spatial pattern. The east-west pattern is particularly marked on the $C^{18}O/^{12}CS$ and $HNC/C^{18}O$ ratios. The $^{12}CS/HNC$, $^{12}CN/HNC$, and HCN/HNC ratios all show an approximately circular structure around NGC 2024 with a deficit of HNC integrated intensity. We relate the latter behavior to an isomerisation of HNC into HCN when the gas temperature increases.

Computing the ratio for these lines removed almost any linear (anti-)correlation with the visual extinction, except for the HCN/HNC and CN/HNC ratios (see Fig. 20). For the filamentary gas (A_V between 6 and 15 mag) the ratio spans a large range of values, up to one order of magnitude for $C^{18}O/^{12}CS$ and $C^{18}O/HNC$.

7. Discussion

7.1. Typical line intensities in a strongly UV-illuminated part of a GMC

The FoV sampled here is not a random one square degree part of any GMC. The left panels of Fig. 21 show the spatial distribution of the visual extinction and dust temperature over a much larger fraction of the Orion B molecular cloud than the one presented in this paper, which is shown as the black rectangle. The right panels compare the PDFs of dust properties over our FoV with the PDFs of two other regions with the same surface area. Table 11 lists the minimum, median, and maximum values of the associated distributions, as well as their 5 and 95% quantiles. This clearly shows that three different kinds of environment

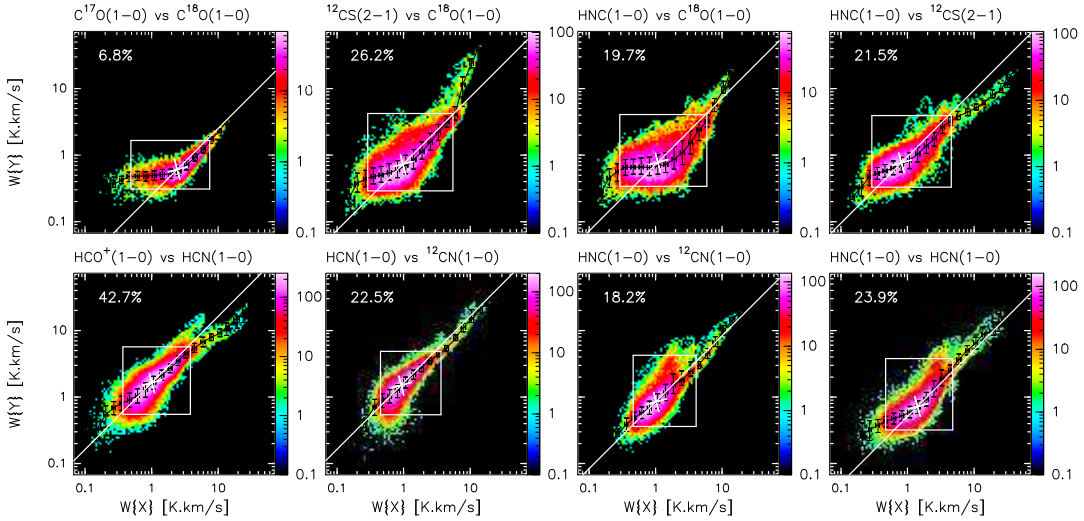


Fig. 18. Same as Fig. 12, except that there is no common *x*-axis in this figure.

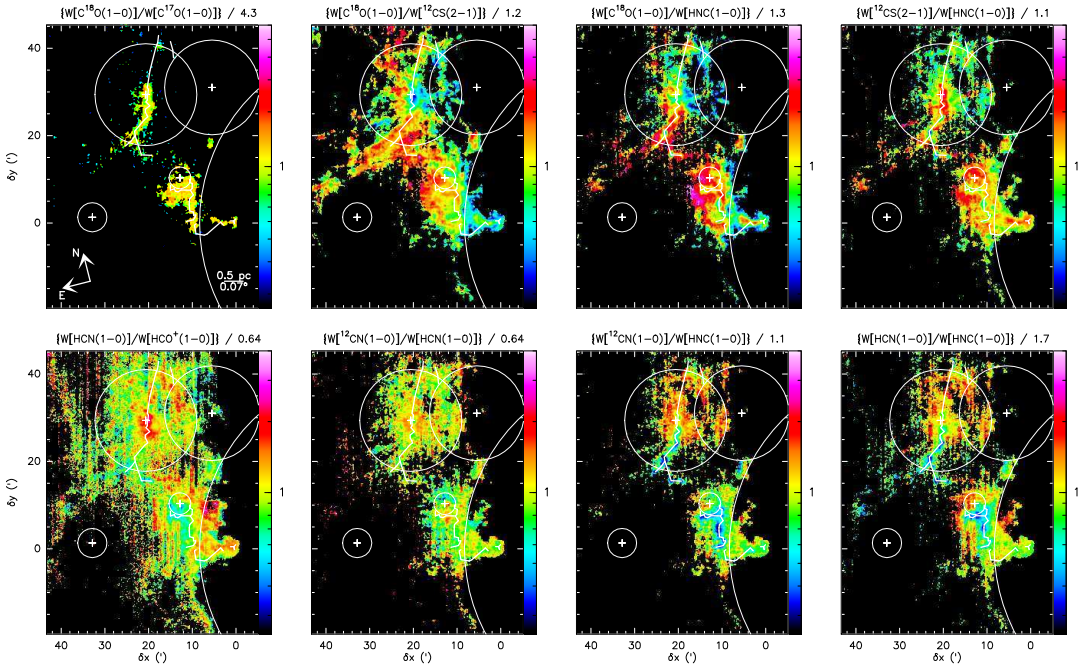


Fig. 19. Same as Fig. 14, except that there is no common numerator in this figure.

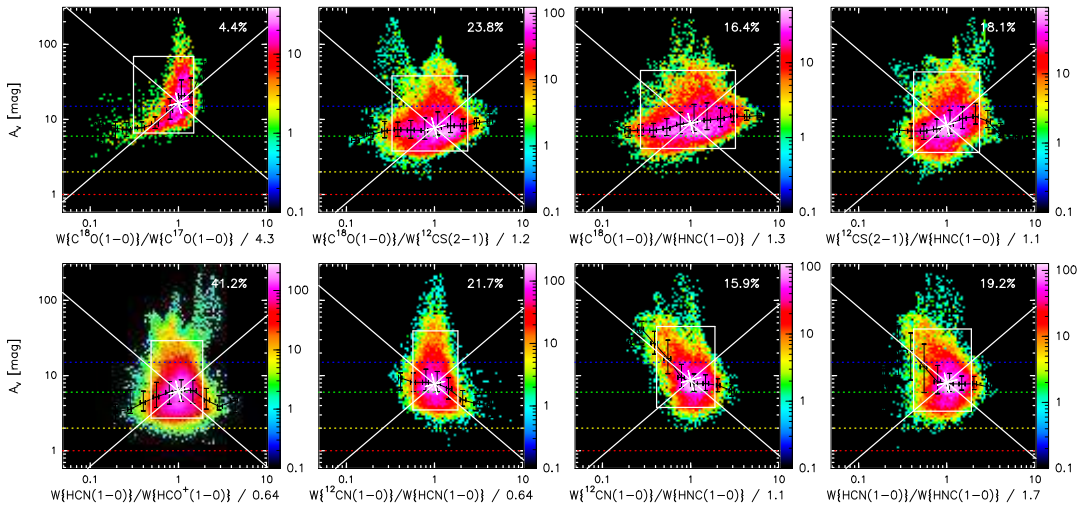


Fig. 20. Same as Fig. 14, except that there is no common ratio numerator in this figure.

Table 11. Visual extinction, dust temperature, and far UV illumination for three different regions of 1 square-degree area in Orion B.

Environment	$A_V^{\rm min}$–$A_V^{5\%}$–$A_V^{\rm med}$–$A_V^{95\%}$–$A_V^{\rm max}$ mag	$T_{\rm d}^{\rm min}$–$T_{\rm d}^{5\%}$–$T_{\rm d}^{\rm med}$–$T_{\rm d}^{95\%}$–$T_{\rm d}^{\rm max}$ K	$G_0^{\rm min}$–$G_0^{5\%}$–$G_0^{\rm med}$–$G_0^{95\%}$–$G_0^{\rm max}$ ISRF (Habing 1968)
UV illuminated	0.76–1–3.3–12–230	16.4–20.1–24.1–40.5–101	4.4–12–30–400–38 000
Translucent	1.1–1.5–2.4–3.5–7.7	16–17.2–17.9–18.6–20.3	3.9–5.6–6.9–8.1–13
UV shieded	1.5–2.2–3.4–7.9–34	12.9–15.2–17.1–18.5–20.3	1.3–3–5.3–8–13

exist in the Orion B molecular cloud. First, the lowest dust temperatures are associated with relatively high visual extinctions (red rectangle). Second, the blue rectangle identifies a translucent region (all pixels have $A_V \lesssim 8$), associated with a typical dust temperature of approximately 18 K. In both cases, the distribution of visual extinction and dust temperature are single peaked with a narrow full width at half maximum. In contrast, our FoV displays wide A_V and $T_{\rm d}$ distributions, and is associated with the highest dust temperatures with a median value of ~24 K ($G_0 \sim 30$). The presence of high gas temperatures is confirmed by the large ^{12}CO (1–0) peak temperatures that are lower limits of the kinetic temperature (Orkisz et al. 2017). These properties are associated with the presence of at least four H II regions (see Sect. 2.1) that imply a large UV-illumination (see the fourth column of Table 11). In particular, the minimum dust temperature in our field (16.4 K) is rather high in Orion B compared to the Taurus molecular cloud (Marsh et al. 2014).

Table 3 indicates that, under these sampling conditions and at the typical sensitivity achieved in studies of nearby galaxies ($3\sigma = 1\,{\rm K\,km\,s^{-1}}$), only the (1–0) line of ^{12}CO, ^{13}CO, HCO$^+$, HCN would easily be detected by a single-dish radio-telescope of 30 m-diameter with a single-beam receiver. A ten times better sensitivity (100 longer integration) is required to detect the (1–0) or (2–1) lines of HCN, CN, C^{18}O, ^{12}CS, C$_2$H, HNC, ^{32}SO, C^{17}O, and c-C$_3$H$_2$. Finally, another order-of-magnitude increase of the sensitivity ($3\sigma = 0.01\,{\rm K\,km\,s^{-1}}$) would be needed to detect N$_2$H$^+$, CH$_3$OH, H^{13}CO$^+$, and H^{13}CN. This means that detecting rare isotopologues of HCO$^+$, HCN, or HNC in nearby galaxies is difficult to achieve with a single-dish radio-telescope because of the dilution of the signal in the beam.

7.2. The influence of the UV field on the determination of molecular mass

The average visual extinction and CO integrated intensity for the observed FoV are 4.7 mag and $61\,{\rm K\,km\,s^{-1}}$. This turns into a $W_{\rm CO}/A_V = 13.0\,{\rm K\,km\,s^{-1}}$/ mag, while the standard $X_{\rm CO}$ factor, that is, $2 \times 10^{20}\,{\rm H_2\,cm^{-2}/(K\,km\,s^{-1})}$, corresponds to $4.7\,{\rm K\,km\,s^{-1}}$/ mag when we assume a standard $N_{\rm H}/A_V$ factor and fully molecular gas. The H I emission indicates that diffuse gas accounts for approximately 1 mag of extinction towards Orion B (see Sect. 3.4). Assuming that contribution from atomic hydrogen to the mass is negligible towards the mapped FoV overestimates the dust-traced molecular mass by 27%, increasing the discrepancy between the CO and dust-traced mass. Therefore, we neglect this subtlety and directly compare the CO and dust-traced mass. We find that the CO-traced mass (and thus the associated surface and volume density) is approximately three times higher than the dust-traced mass.

The origin of this discrepancy lies in the intense UV illumination of the gas by massive stars. The bottom left panel of Fig. 22 compares the spatial distribution of the CO integrated intensity per visual extinction. The standard value of $4.7\,{\rm K\,km\,s^{-1}}$/mag corresponds to the transition between yellow

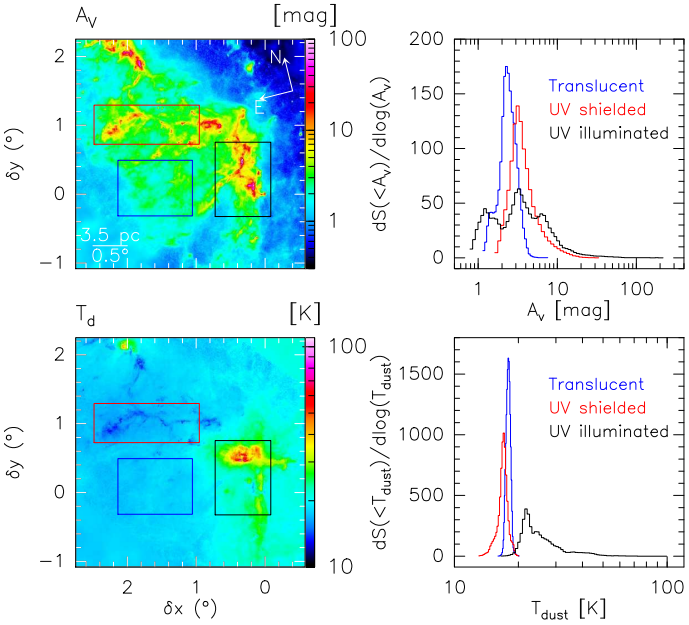


Fig. 21. *Left panels*: spatial distribution of the visual extinction and of the dust temperature. *Right panels*: probability distribution functions (PDFs) of the visual extinction and the dust temperature for the regions inside the black, blue, and red rectangles, respectively.

and green. Only diffuse gas or the UV-shielded dense gas have $X_{\rm CO}$ values close to or lower than standard. This is confirmed by the joint distribution of the $W_{\rm CO}$ and A_V where most of the points (lines of sight with $2 \lesssim A_V \lesssim 15$) lie above the white line of slope $4.7\,{\rm K\,km\,s^{-1}/mag}$. When the visual extinction increases, the CO intensity saturates. At the lowest visual extinction ($A_V \lesssim 2$), CO is destroyed, giving C$^+$. The spatial distribution of G_0 clearly shows that most of the gas lies in regions with $G_0 > 10$, the mean value of G_0 being 45. Under such conditions, dust and gas are heated to higher temperatures than in the standard interstellar radiation field. In the physical conditions of Orion B, the CO emission per H$_2$ molecule is increasing with the kinetic temperature, leading to a possible bias in the mass determination. The bottom right panel of Fig. 22 shows that the CO intensity per A_V increases with G_0.

While it is tempting to conclude that only the CO-traced mass is widely overestimated in such conditions, the dust-traced mass is in fact also underestimated, as indicated by the range of virial mass that we estimated from the FoV size and CO linewidth (see Table 2). It is known that using a single dust temperature to fit the spectral energy distribution on a line of sight that contains dust at different temperatures hides the presence of cold dust along the line of sight, because the luminosity of dust increases extremely quickly with its temperature (Shetty et al. 2009).

All in all, the typical volume density we infer for the regions between 200 and $600\,{\rm H_2\,cm^{-3}}$, is typical of galactic

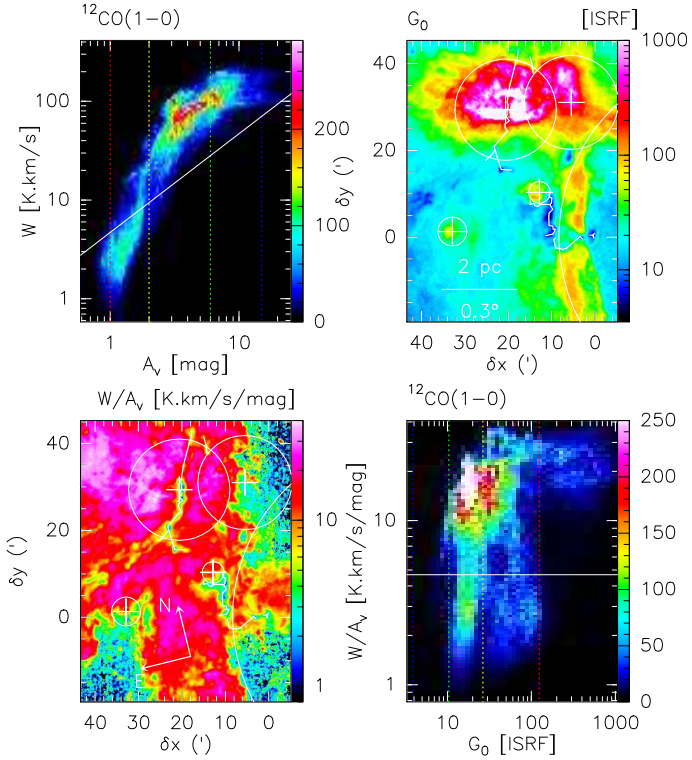


Fig. 22. *Top, left panel*: joint distribution of the ^{12}CO (1−0) line integrated intensity as a function of the visual extinction. The number of sightlines falling in a given 2D bin of the distribution is color-coded using a linear scale to emphasize the bi-modal nature of the distribution. The white line shows the location of points that follows $N(H_2) = X_{CO}\,W$. The red, orange, green, and blue vertical dashed lines show the visual extinction limits (1, 2, 6, and 15 mag, respectively) used in the masks of Fig. 6. *Top, right panel*: spatial distribution of the far UV illumination in units of the ISRF (Habing 1968). *Bottom, left panel*: ratio of the ^{12}CO (1−0) line integrated intensity to the visual extinction. *Bottom, right panel*: joint distribution of this ratio as a function of the far UV illumination. The horizontal white line corresponds to the standard value of the X_{CO} factor. The blue, green, orange, and red vertical dashed lines show the far UV illumination limits (4, 10, 26, and 120, respectively), which corresponds to the dust temperature limits used in the masks of Fig. 8. The color scales of the two images show the same ranges as the axes of the bottom right joint distribution.

GMCs (Heyer & Dame 2015). While the local values of N_H/A_V and X_{CO} are uncertain, and they could well be different from their standard values in Orion B, here we wish to study Orion B as if it was observed from nearby galaxies. In these studies, standard values are used when the metallicity is similar to that of the Milky Way (e.g., see the PAWS project Schinnerer et al. 2013; Pety et al. 2013). From a practical viewpoint, we thus proceeded with standard values, knowing that the correct result is probably in between the CO-traced and dust-traced masses, surface densities, and volume densities.

Enhanced far-UV fields heat large masses of CO gas that turns over-luminous with respect to the standard X_{CO} factor, that is, the average behavior of the CO gas in our Milky way (Planck Collaboration XIX 2011). This effect could compensate for the presence of CO-dark gas (Wolfire et al. 2010; Planck and Fermi Collaborations Int. XXVIII 2015), as proposed by Liszt & Pety (2012) with different observations. Therefore the standard value of X_{CO} may well be applicable to galaxies with a higher than average, yet moderate, massive star-formation rate. A straightforward test would be to increase the size of the mapped FoV to the full Orion B cloud to test at which scale the CO and dust-traced masses derived with standard values of X_{CO} and the N_H/A_V ratio get reconciled to better than a factor of three. Complementary CII observations would also help to settle this point (Goicoechea et al. 2015).

Table 12. Critical density, and percentage of total flux originating from gas in filaments and dense cores for each line.

Species	Transition	n_{crit} cm^{-3}	$F_{6\le A_V<15}$ % of F_{tot}	$F_{15\le A_V}$ % of F_{tot}
^{12}CO	1−0	2×10^3	38	7.6
^{13}CO	1−0	2×10^3	45	15
HCO^+	1−0	2×10^5	40	16
C_2H	1−0	1×10^5	37	17
c-C_3H_2	2−1	1×10^6	38	17
HCN	1−0	3×10^6	44	18
^{12}CN	1−0	3×10^5	45	19
C^{17}O	1−0	2×10^3	43	26
HNC	1−0	4×10^5	41	27
C^{18}O	1−0	2×10^3	48	29
^{32}SO	2−1	2×10^5	44	31
^{12}CS	2−1	2×10^5	42	32
CH_3OH	2−1	3×10^4	41	48
H^{13}CO$^+$	1−0	2×10^5	34	56
N_2H^+	1−0	2×10^5	17	88

Notes. The lines are sorted by increasing value of $F_{15\le A_V}/F_{tot}$ (last column). The typical volume density of the regions with $6 \le A_V < 15$, and $15 \le A_V$ are 1500, and 7300 H_2 cm^{-3}, respectively.

7.3. Dense gas tracers

The brightness of a molecular line depends on the column density of the species, which is affected by the chemistry, and the excitation properties of the line. These in turn depend on the physical properties of the gas (density, temperature, ionization fraction, etc.). Indeed, two conditions must be satisfied for a line to be detected. First, the molecule has to be abundant enough, and second, the excitation conditions must be favorable for the line to be excited and produce bright line emission. It is therefore often assumed that lines with high critical densities, such as the (1−0) lines of HCN and HCO^+, are good tracers of dense gas because these species are abundant and their emission is only expected to be seen in regions where the density is high enough to excite the line. Table 12 lists the critical density of each molecular line as well as the percentage of total flux that is emitted from regions of intermediate ($6 \le A_V < 15$) and high ($15 \le A_V$) visual extinction, as measured in Sect. 4.1. These two regimes are representative of the gas arising in filaments and dense cores, respectively. We do not find any clear correlation between the critical density of the lines and the percentage of flux emitted.

For instance, the lines of the CO isotopologues have nearly equal critical densities (that is, similar excitation conditions) but the percentage of flux coming from the densest regions ($n_{H_2} \sim 7300\,cm^{-3}$) varies from 8% for ^{12}CO (1−0) to 29% for C^{18}O (1−0). The higher fraction of flux coming from high-density regions for the rarer isotopologues is the result of both sensitivity and chemistry. The intrinsic lower abundance of ^{13}CO, C^{18}O and C^{17}O compared to the main isotope will result in weaker emission for the rarer isotopologues everywhere in the cloud. Also, in the UV-illuminated layers, ^{12}CO will survive longer than the

Table 13. Flux ratios in nearby galaxies and in Orion B.

Source	HCN/HCO$^+$	^{12}CO/HCN	^{12}CO/HCO$^+$	^{12}CO/N$_2$H$^+$	^{12}CN/HCN	C$_2$H/^{13}CO	HCN/HNC	Observ.	Ref.
ULIRGs	1.5–2.4	6.3–8.6	9.7–21	40	0.4–0.5	0.8–1.4	1.4–3.0	MOPRA-22 m	1
M 51 P2	1.2	32	39	225	0.5	0.04	2.7	IRAM-30 m	2
AGNs	0.8–2.0	3.4–25	5.7–20	41–68	0.9–1.6	0.2–0.7	2.2–3.0	IRAM-30 m	3
Starbursts	0.7–1.2	16–23	16–24	100–325	1.2–1.4	0.2–0.7	1.9–2.5	IRAM-30 m	3
M 82	0.5	–	–	–	–	–	–	CARMA	4
NGC 253 P7	1.2	6.0	7.0	–	0.6	–	–	ALMA	5
Maffei2B	0.6	–	–	–	–	–	4.3	BIMA,OVRO	6
LMC	0.4–0.7	11–120	6.0–48	59–167	0.2–0.3	0.03–0.27	1.1–3.4	IRAM-30 m	7
Orion B	0.9	39	37	900	0.5	0.05	3.5	IRAM-30 m	8

Notes. [(1)] Nishimura et al. (2016); [(2)] Watanabe et al. (2014); [(3)] Aladro et al. (2015); [(4)] Salas et al. (2014); [(5)] Meier et al. (2015); [(6)] Meier & Turner (2012); [(7)] Nishimura et al. (2016). [(8)] This work.

CO isotopologues due its capacity to self-shield (process known as selective photo dissociation), although this effect is probably not resolved by the observations.

Lines with much higher critical densities (>10^5 cm^{-3}) than CO (1–0) (~10^3 cm^{-3}), such as HCO$^+$, HCN, and HNC (1–0), which are expected to trace dense gas, emit only 16, 18, and 27% of their total flux in mapped Orion B regions of high density. Most of the emission then arises in lower density gas. Shirley (2015) extensively discusses the relevance of the notion of critical density. He highlights that the critical density is the density at which collisional de-excitation equals the net radiative emission. He emphasizes that it is computed in the optically thin limit, implying that it is only an upper limit in the presence of photon-trapping. The fact that these lines can be excited and thus detected in diffuse gas was, in addition, recently discussed by Liszt & Pety (2016). While their results cannot be quantitatively applied to the observed FoV because the line peak intensities are slightly outside the range of applicability of the weak excitation regime, the underlying physical process is still present. At the limit of detectability, Liszt & Pety (2016) showed that the intensity of low-energy rotational lines of strongly polar molecules, such as HCO$^+$ and HNC, is proportional to the product of the total gas density and the molecule column density, independent of the critical density, as long as the line intensity does not increase above a given value. This implies that for any given gas density there is a column density that will produce an observable line intensity. In the observed field of Orion B, we are in an intermediate radiative transfer regime where the fundamental lines of HCO$^+$, HCN, and HNC can be excited in regions of density much smaller than their critical densities.

The lines with the largest fraction of their emission arising in the densest gas are CH$_3$OH and N$_2$H$^+$. In particular, N$_2$H$^+$ (1–0) that presents a high critical density similar to that of HCO$^+$, HCN, and HNC (1–0), has the highest proportion of its flux coming from the densest regions (88%). Moreover, in contrast to all the other lines, only 17% of the total N$_2$H$^+$ (1–0) flux is associated with regions of intermediate A_V ($6 \leq A_V < 15$) which have typical densities of ~1500 cm^{-3}. This shows that the N$_2$H$^+$ (1–0) line is the best molecular tracer of dense regions among the lines studied in this paper.

Despite the similar critical densities between N$_2$H$^+$ and HCO$^+$, their behavior is completely different. First, N$_2$H$^+$ is detected over only 2.4% of the observed FoV, while HCO$^+$ emission covers 68% of the field. Second, the percentages of the flux coming from dense/translucent regions are 88/8% for N$_2$H$^+$, and 15/41% for HCO$^+$. These differences can only be understood by their different chemistry. HCO$^+$ can be formed from ion-molecule reactions involving C$^+$ and other cations, notably CH_2^+ and CH_3^+, in addition to the protonation of CO. On the contrary, the sole reaction producing N$_2$H$^+$ is the protonation of N$_2$. Furthermore, the destruction of HCO$^+$ by dissociative recombination with electrons produces CO while N$_2$H$^+$ can react with CO to produce HCO$^+$. Hence, N$_2$H$^+$ only survives in regions where the electron abundance is low (to prevent dissociative recombination) and where CO is frozen on dust grains (to prevent proton transfer to CO), that is, in cold and dense cores.

In summary, species that have a chemical reason to only be present in dense gas are the only really reliable high density tracers. More generally, the knowledge of the chemical behavior is fundamental in understanding how molecular species can be used to trace the different physical environments.

7.4. Typical line ratios in Orion B and other galaxies

Molecular line ratios have the potential to be powerful probes of physical properties related to star formation activity. Thanks to current observing capabilities, many unbiased line surveys have recently been carried out toward nearby galaxies, and more than 27 species have been detected in the 3 mm band (Meier & Turner 2012; Salas et al. 2014; Watanabe et al. 2014; Aladro et al. 2015; Meier et al. 2015; Nishimura et al. 2016). Common line ratios include the HCN(1–0)/HCO$^+$(1–0), ^{12}CO(1–0)/HCN(1–0), HCN(1–0)/HNC(1–0) and the CN(1–0)/HCN(1–0), which are proposed tracers of density, temperature and radiation field, respectively.

Figure 23 shows observed line ratios in nearby galaxies and in the Orion B molecular cloud. The comparison includes line ratios obtained with both single-dish telescopes and interferometers, and galaxies with distances between 3.3 and 170 Mpc, in addition to the LMC, the nearest external galaxy (50 kpc). The sample therefore spans two orders of magnitude in spatial resolution (written at the bottom of the y-axis in units of pc). To ease the comparison with Orion B, the ratios are normalized by the Orion B ratios of the lines integrated over the full observed FoV (that is, at a resolution of 10 pc).

Line ratios observed in Orion B are, in general, comparable to what is observed in nearby galaxies. Noticeably, line ratios (except ^{12}CO/N$_2$H$^+$) in Orion B at a resolution of 10 pc are very similar to those observed at a resolution of 1000 pc in the spiral arm of the famous whirlpool galaxy (M 51), a prototype of grand design spiral galaxy.

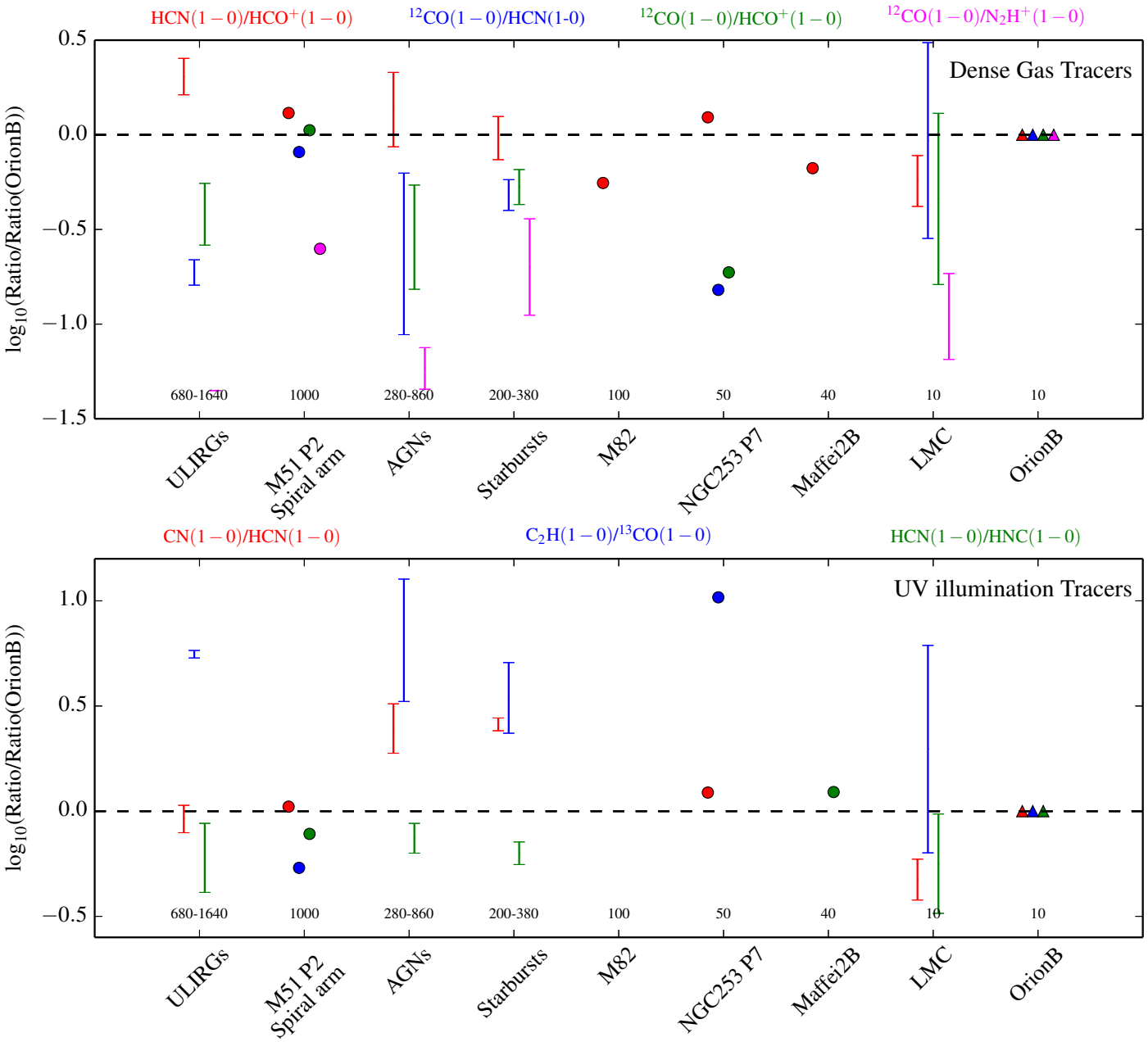


Fig. 23. Line ratios observed towards nearby galaxies and the Orion B molecular cloud. The *upper and lower panels* include line ratios which are proposed dense gas and UV-illumination tracers, respectively. A point is plotted for individual measures, while a range of values is given for the Ultra-Luminous Infra-Red Galaxies (ULIRG), Active Galactic Nuclei (AGN), and starburst samples, which contain measures for several galaxies. The ULIRGs include Arp 220 and Mrk 231, while AGNs include M 51, NGC 1068 and NGC 7469, and the starburst galaxies include M 83, NGC 253 and M 82. The sources are sorted by decreasing spatial resolution, which is given at the bottom of the figure for each source. The Orion B values correspond to the ratios of the lines integrated over the full region to simulate a 10 pc spatial resolution allowing a better comparison with the other galaxies.

The HCN/HCO^+ ratio (shown in red in the upper panel) is assumed to trace dense gas that will eventually form stars, and thus is often used to trace the star formation activity in other galaxies. Both lines are among the brightest lines observed and are thus easily detected. ULIRGs and AGNs present the largest HCN/HCO^+ ratios, but there are no major differences between the sources. The resolved spatial distribution in Orion B (see Fig. 19) shows that in the case of UV-illuminated gas, the HCN/HCO^+ ratio does not trace the high-density gas. Another proposed tracer of dense gas is the inverse of the $^{12}CO/HCN$ ratio (shown in blue in the upper panel). This ratio is higher in Orion B, the LMC and in the spiral arm of M 51 than in the other galaxies. The $^{12}CO/HCO^+$ ratio behaves in a similar way. While the resolved spatial distribution in Orion B (see Fig. 14) shows that these ratios actually separate diffuse from dense gas, the quantitative analysis shows that HCO^+ and HCN (1−0) fluxes mostly trace densities around 500−1500 H_2 cm^{-3} instead of $\sim 10^4$ cm^{-3}.

Better tracers of high-density gas are ratios involving N_2H^+, such as $CO(1-0)/N_2H^+(1-0)$ (shown in magenta), because N_2H^+ resides solely in dense gas ($>10^4$ cm^{-3}), contrary to HCN that can be present at lower densities (see Sect. 4.1). Starbursts (including M 51) and ULIRGs form many stars, requiring the presence of many dense cores, and thus a high N_2H^+ (1−0)

brightness relative to ^{12}CO (1−0) that traces the total reservoir of molecular gas. In contrast, Orion B has a low star-formation efficiency (Lada et al. 2010; Megeath et al. 2016), probably implying a low number of dense cores and thus a lower relative N_2H^+ brightness. The low surface filling factor of N_2H^+ makes it difficult to detect at high signal-to-noise ratio with single-dish telescopes in external galaxies. Fortunately, the much better resolving power of NOEMA and ALMA relieves this difficulty and N_2H^+ starts to be detected in nearby galaxies.

The HCN/HNC is another popular line ratio measured in nearby galaxies (shown in green in the bottom panel). Both species are abundant in cold clouds, but at temperatures higher than approximately 30 K, HNC starts to be converted to HCN through reactions with H (see, e.g., Schilke et al. 1992; Graninger et al. 2014). The HCN/HNC ratio is thus increasing at higher temperatures. The sources included in this comparison present similar HCN/HNC ratios within a factor of two. The spatial distribution shown in Fig. 19 is consistent with the proposed temperature dependence of this ratio, whose lowest values are found in the cold filaments, and which correlates well with the map of the dust temperature seen in Fig. 2.

Finally, the CN/HCN intensity ratio (shown in red in the bottom panel) is expected to trace UV-illuminated gas because CN is a major product of HCN photodissociation. Indeed, higher abundance ratios are found in PDRs and XDRs compared to cold dark clouds (Fuente et al. 2005; Baan et al. 2008). However, in spatially resolved observations, there is no clear correlation of the CN/HCN flux ratio with radiation field (see Fig. 19). Furthermore, in our comparison, AGNs and starbursts present the largest CN/HCN ratios, and ULIRGs have similar ratios to the Orion B cloud and the spiral arm of M 51. This is contradictory with CN/HCN being a radiation field tracer. In fact, the ratios that involve the HCN (1−0) line must be taken with care because this line is known to be pumped by IR photons (Aalto et al. 2007). This will produce brighter HCN lines, and thus could explain the low CN/HCN ratio observed in ULIRGs. A better tracer of radiation field is the $C_2H(1-0)/^{13}CO(1-0)$ ratio. Carbon chains, such as C_2H and C_3H_2, have been observed to form efficiently in UV-illuminated regions (Pety et al. 2005; Guzmán et al. 2015). ULIRGs, AGNs and starburst galaxies present the largest $C_2H/^{13}CO$ ratios, consistent with their high star-formation activities. Moreover, the spatial distribution of the $C_2H/^{13}CO$ ratio (see Fig. 15) shows a clear gradient between the illuminated edge (right) and the UV-shielded (left) side of the Orion B molecular cloud. Therefore, the $C_2H(1-0)/^{13}CO(1-0)$ ratio is a potential tracer of the presence of massive stars.

8. Conclusion

The ORION-B project aims at imaging a statistically significant fraction of the Orion B GMC over the 3 mm atmospheric window, starting with approximately one square degree towards NGC 2024, NGC 2023, Horsehead, and IC 434. The mean dust temperature in the mapped region is 26 K, and the probability distribution function of the visual extinction shows a wide distribution of magnitudes, from less than 1 to a few 100, with most of the surface and volume lying below a magnitude of 15. These dust properties suggests that the south-western edge of Orion B is permeated by far-UV fields from massive stars at relatively large spatial scales. Most of the cloud mass is contained in regions of relatively low A_V, implying that photon-dominated regions are everywhere in the FoV. This is the reason why CO is over-luminous, resulting in a CO-traced mass that is approximately three time as much as the dust traced mass in this region.

Over the 84 to 116 GHz frequency range, we easily detected the fundamental lines of CO isotopologues (from ^{12}CO to C^{17}O), HCO^+, HCN, HNC, and their ^{13}C isotopologues, ^{12}CN, C_2H, and N_2H^+, as well as higher J lines of ^{12}CS, ^{32}SO, SiO, c-C_3H_2, and methanol. The faintest averaged spectra of the species presented in this article, the spectra of HN^{13}C (1−0), is ∼2400 fainter than ^{12}CO (1−0) line. Still fainter lines, tentatively detected in the averaged spectra, are present. They will be discussed in a future paper.

The main CO isotopologues and a number of other species clearly display two velocity components: the main component centered at ∼10 km s^{-1}, where most of the gas lies, and a more diffuse component centered at ∼5 km s^{-1} that contains approximately 10% of the molecular gas along the line of sight. Tomography studies of the interstellar medium using visible absorption against background stars indicates that both components belong to Orion B; their relative distance is compatible with the projected extension of Orion B.

A significant fraction of the HCO^+, HCN, and HNC (1−0) flux integrated over the observed FoV is coming from translucent (41, 36, and 29%, respectively) and from gas forming the filamentary structure (typically 40−45%). Only 16, 18, and 29% (for HCO^+, HCN, and HNC, respectively) of the flux is coming from dense cores ($A_V > 15$). The common assumption that lines of large critical densities (∼10^5 cm^{-3}) can only be excited by gas of similar density is clearly incorrect. Another unrelated result is that these lines are also sensitive to the amount of far-UV illumination, even though less sensitive than the fundamental line of small hydrocarbon chains and ^{12}CN.

While the HCO^+ and N_2H^+ (1−0) lines have similar critical densities and similar peak temperatures over the mapped FoV, the repartition of their flux has a completely different behavior: The N_2H^+ (1−0) line is only emitted from regions with $A_V > 15$, while the HCO^+ (1−0) line is also emitted from regions of visual extinction as low as 1−2 mag. This explains why the surface filling factor of detected emission highly differs between these two species. This is in part due to the fact that there exists another regime of radiative transfer (the weak excitation limit) that produces detectable lines of strongly polar molecules (Liszt & Pety 2016, and reference therein). The difference of behavior between HCO^+ and N_2H^+ can easily be explained by the way these molecules are produced and destroyed in diffuse and dense gas.

We observe a strong correlation of the line integrated intensities with A_V, that is, the line strength increases with the quantity of material along the line of sight. This correlation will be quantified in Gratier et al. (2017) through a PCA that allows us to find significant correlations beyond this one with other physical parameters, such as UV illumination. However, the best tracers of the column density are the C^{18}O and HNC (1−0) lines, followed by the ^{12}CS (2−1) line. This validates the use of these lines to normalize the intensities in extra-galactic studies. The different species clearly show A_V threshold values above which they start to emit. Dividing the line integrated intensities by the visual extinction enables us to remove this correlation and to bring forward chemical differences between the species.

When the visual extinction is not available, making the ratio of the line integrated intensities also emphasizes different chemistry behavior. However, proposed dense-gas tracers such as the $HCN(1-0)/HCO^+(1-0)$ or $HCN(1-0)/^{12}CO(1-0)$ line ratios should be superseded by $N_2H^+(1-0)/^{12}CO(1-0)$ that more accurately traces the fraction of dense gas (≳10^4 cm^{-3}). The fact

that the CN(1−0)/HCN(1−0) line ratio has a pretty flat spatial distribution in our FoV, may be because the gas is permeated by far-UV fields. However, we find that the $C_2H(1-0)/^{13}CO(1-0)$ line ratio is an excellent tracer of the variations of the far-UV illumination in our FoV.

The increasing capabilities of millimeter receivers make it possible to observe multiple molecular lines in large FoVs, and thus to use the spatial distribution of low-J line ratios to classify each line of sight depending on its molecular content.

Acknowledgements. We thank the referee, J. G. Mangum, for his careful reading of the manuscript and useful comments that improved the article. We thank R. Lallement for useful discussions about the distance of the gas in Orion B. We thank P. Andre and N. Schneider for giving us access to the *Herschel* Gould Belt Survey data, and M. Lombardi for delivering his fit of the spectral energy distribution of the *Herschel* data. We thank CIAS for their hospitality during the two workshops devoted to this project. This work was supported by the CNRS program "Physique et Chimie du Milieu Interstellaire" (PCMI). J.R.G. thanks MINECO, Spain, for funding support under grant AYA2012-32032. V.V.G. thanks for support from the Chilean Government through the Becas Chile program. P.G.'s postdoctoral position was funded by the INSU/CNRS. P.G. thanks ERC starting grant (3DICE, grant agreement 336474) for funding during this work. NRAO is operated by Associated Universities Inc. under contract with the National Science Foundation. This paper is based on observations carried out at the IRAM-30 m single-dish telescope. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain). This research also used data from the *Herschel* Gould Belt survey (HGBS) project (`http://gouldbelt-herschel.cea.fr`). The HGBS is a *Herschel* Key Programme jointly carried out by SPIRE Specialist Astronomy Group 3 (SAG 3), scientists of several institutes in the PACS Consortium (CEA Saclay, INAF-IFSI Rome and INAF-Arcetri, KU Leuven, MPIA Heidelberg), and scientists of the *Herschel* Science Center (HSC).

References

Aalto, S., Spaans, M., Wiedner, M. C., & Hüttemeister, S. 2007, A&A, 464, 193
Abergel, A., Bernard, J. P., Boulanger, F., et al. 2002, A&A, 389, 239
Aladro, R., Martín, S., Riquelme, D., et al. 2015, A&A, 579, A101
André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
Anthony-Twarog, B. J. 1982, AJ, 87, 1213
Baan, W. A., Henkel, C., Loenen, A. F., Baudry, A., & Wiklind, T. 2008, A&A, 477, 747
Bergin, E. A., & Tafalla, M. 2007, ARA&A, 45, 339
Bigiel, F., Leroy, A. K., Jiménez-Donaire, M. J., et al. 2016, ApJ, 822, L26
Bik, A., Lenorzer, A., Kaper, L., et al. 2003, A&A, 404, 249
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Chernin, L. M. 1996, ApJ, 460, 711
Fabricius, C., Bastian, U., Portell, J., et al. 2016, A&A, 595, A3
Federrath, C., & Klessen, R. S. 2012, ApJ, 761, 156
Federrath, C., & Klessen, R. S. 2013, ApJ, 763, 51
Frerking, M. A., Langer, W. D., & Wilson, R. W. 1982, ApJ, 262, 590
Fuente, A., García-Burillo, S., Gerin, M., et al. 2005, ApJ, 619, L155
Gaia Collaboration (Brown, A. G. A., et al.) 2016a, A&A, 595, A2
Gaia Collaboration (Prusti, T., et al.) 2016b, A&A, 595, A1
Goicoechea, J. R., Teyssier, D., Etxaluze, M., et al. 2015, ApJ, 812, 75
Graninger, D. M., Herbst, E., Öberg, K. I., & Vasyunin, A. I. 2014, ApJ, 787, 74
Gratier, P., Bron, E., Gerin, M., et al. 2017, A&A, 599, A100
Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2015, ApJ, 810, 25
Guzmán, V. V., Pety, J., Goicoechea, J. R., et al. 2015, ApJ, 800, L33
Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421
Hennebelle, P. 2013, A&A, 556, A153
Hennebelle, P., & Chabrier, G. 2011, ApJ, 743, L29
Heyer, M., & Dame, T. M. 2015, ARA&A, 53, 583
Hollenbach, D. J., Takahashi, T., & Tielens, A. G. G. M. 1991, ApJ, 377, 192
Hughes, A., Meidt, S. E., Colombo, D., et al. 2013, ApJ, 779, 46
Hummel, C. A., Rivinius, T., Nieva, M.-F., et al. 2013, A&A, 554, A52
Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775
Kim, C.-G., Ostriker, E. C., & Kim, W.-T. 2013, ApJ, 776, 1
Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687
Lada, C. J., Forbrich, J., Lombardi, M., & Alves, J. F. 2012, ApJ, 745, 190
Lada, C. J., Lombardi, M., Roman-Zuniga, C., Forbrich, J., & Alves, J. F. 2013, ApJ, 778, 133
Lallement, R., Vergely, J.-L., Valette, B., et al. 2014, A&A, 561, A91
Leroy, A. K., Hughes, A., Schruba, A., et al. 2016, ApJ, 831, 16
Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, 595, A4
Liszt, H., & Lucas, R. 2001, A&A, 370, 576
Liszt, H. S., & Pety, J. 2012, A&A, 541, A58
Liszt, H. S., & Pety, J. 2016, ApJ, 823, 124
Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, A&A, 566, A45
Lucas, R., & Liszt, H. 1996, A&A, 307, 237
Lucas, R., & Liszt, H. S. 2000, A&A, 358, 1069
Mangum, J. G., Emerson, D. T., & Greisen, E. W. 2007, A&A, 474, 679
Marsh, K. A., Griffin, M. J., Palmeirim, P., et al. 2014, MNRAS, 439, 3683
Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, AJ, 151, 5
Meidt, S. E., Schinnerer, E., García-Burillo, S., et al. 2013, ApJ, 779, 45
Meier, D. S., & Turner, J. L. 2012, ApJ, 755, 104
Meier, D. S., Walter, F., Bolatto, A. D., et al. 2015, ApJ, 801, 63
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Nishimura, Y., Shimonishi, T., Watanabe, Y., et al. 2016, ApJ, 818, 161
Orkisz, J. H., Pety, J., Gerin, M., et al. 2017, A&A, 599, A99
Penzias, A. A., & Burrus, C. A. 1973, ARA&A, 11, 51
Pety, J. 2005, in SF2A-2005: in Semaine de l'Astrophysique Française, eds. F. Casoli, T. Contini, J. M. Hameury, & L. Pagani, 721
Pety, J., Teyssier, D., Fossé, D., et al. 2005, A&A, 435, 885
Pety, J., Bardeau, S., & Reynier, E. 2009, IRAM-30 m EMIR time/sensitivity estimator, Tech. Rep., iRAM Memo 2009-1
Pety, J., Schinnerer, E., Leroy, A. K., et al. 2013, ApJ, 779, 43
Planck Collaboration I. 2011, A&A, 536, A1
Planck Collaboration XIX. 2011, A&A, 536, A19
Planck and Fermi Collaborations Int. XXVIII. 2015, A&A, 582, A31
Richer, J. S. 1990, MNRAS, 245, 24P
Rieke, G. H., & Lebofsky, M. J. 1985, ApJ, 288, 618
Salas, P., Galaz, G., Salter, D., et al. 2014, ApJ, 797, 134
Sandell, G., Avery, L. W., Baas, F., et al. 1999, ApJ, 519, 236
Schaefer, G. H., Hummel, C. A., Gies, D. R., et al. 2016, AJ, 152, 213
Schilke, P., Walmsley, C. M., Pineau Des Forets, G., et al. 1992, A&A, 256, 595
Schinnerer, E., Meidt, S. E., Pety, J., et al. 2013, ApJ, 779, 42
Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, ApJ, 786, 29
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Shetty, R., Kauffmann, J., Schnee, S., Goodman, A. A., & Ercolano, B. 2009, ApJ, 696, 2234
Shirley, Y. L. 2015, PASP, 127, 299
Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730
Spitzer, L. 1978, Physical processes in the interstellar medium (New York: Wiley-Interscience)
van Leeuwen, F. 2007, A&A, 474, 653
Watanabe, Y., Sakai, N., Sorai, K., & Yamamoto, S. 2014, ApJ, 788, 4
Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191

Appendix A: Noise properties

Table A.1 summarizes the single-dish observations described in Sect. 2.2. Figure A.1 shows the average system temperature as a function of the rest frequency over the observed bandwidth. The mean (solid line) was computed over all the calibration measurements made during the observation and the shaded backgrounds display the $\pm 3\sigma$ interval that reflects both variations in the quality of the tuning and of the elevation of the source. The variations with the frequency come from a combination of hardware and atmosphere chromaticity effects. The optic and mixer performances vary with the observed radio-frequency and the cryogenic low noise amplifier performances vary with the instantaneous intermediate frequency. The atmospheric transparency mostly varies because of ozone lines present in the bandpass and because of the 118.75 GHz ($N = 1-1$, $J = 1-0$) oxygen line whose wings strongly increase the $T_{\rm sys}$ at the upper end of the 3 mm atmospheric window.

Figure A.2 shows the spatial distribution of the noise at an angular resolution of 31″ for two lines, each one representing one of the two tunings. The regions around the Horsehead and NGC 2023 were observed at the start of the project, before our observing strategy converged to the one described in Sect. 2.2. In particular, this area was covered at least once along the δx axis, and once along the δy axis. This explains why the noise properties of this region are different from other parts of the map. Except for this part of the map, the variations in the noise level reflect the variations of weather or elevation during the observations.

Appendix B: Mean line profiles from different gas regimes

In Sect. 4, we derived the fraction of fluxes that comes from different gas regimes in the lines averaged over the observed FoV. Here we intend to deliver the tools to estimate the typical line fluxes when varying the number of lines of sight that sample the different gas regimes. We thus compute the mean line profiles over the same gas regimes as defined in Sects. 4.1 and 4.2. Mixing these line profiles in different proportions would deliver a proxy for the line profile observed in GMCs that have different ratios of diffuse/dense gas or cold/warm gas.

Figures B.1 and B.3 show the mean line profiles for the different A_V and $T_{\rm d}$ regimes. Tables B.1 and B.2 present, for each transition, line integrated intensity averaged over the different A_V and $T_{\rm d}$ masked regions. Figures B.2 and B.4 show these line integrated intensities as percentages of the integrated intensity averaged over the full FoV. These later plots indicate the potential intensity gain that can be expected when resolving the regions that are compact.

As expected, the intensity increases with the visual extinction, that is, with the column density of matter. However, the increase is only a factor two for the ^{12}CO (1–0) line while it reaches a factor of almost 30 for the N_2H^+ (1–0) line. All lines except N_2H^+ (1–0) have a mean integrated intensity that increases by a factor of two in the filamentary gas ($6 \leq A_V < 15$). The mean spectra at 10 pc resolution are closer to the mean spectra over the translucent gas for all the lines, except for the $H^{13}CO^+$, CH_3OH, and N_2H^+ low-J lines.

The C_2H, c-C_3H_2, ^{12}CN, ^{12}CO, HCN, and HCO^+ low-J lines have similar mean integrated intensities in regions of cold and hot dust, probably because some of the high visual extinction lines of sight are also associated with high dust temperature. The mean spectra of hot dust regions have significantly wider linewidths than the mean spectra of the cold dust regions, confirming that the gas temperature is significantly higher in at least the PDR parts along these lines of sight. The HNC, ^{13}CO, ^{12}CS, $C^{17}O$, ^{32}SO, $C^{18}O$, $H^{13}CO^+$, CH_3OH, and N_2H^+ low-J line mean integrated intensities are between 2 and 20 times brighter in the cold than in the hot dust regions. These lines are thus more characteristic of cold dense gas than the previous category. Finally, the mean line integrated intensity over the full FoV is closer to the line intensity integrated over the regions of warm and lukewarm dust regions than over the cold or hot dust regions.

Table A.1. Observation parameters.

Species	Transition	Frequency GHz	Setup	$F_{\rm eff}$	$B_{\rm eff}$	$T_{\rm sys}$ K	Beam[a] $''$	Vel. res.[b] km s^{-1}	Int. time[c] hr	Noise[d] K
^{12}CO	1–0	115.271202	110/USB	0.95	0.78	287	22.5/31	0.51/0.5	40.4/62.4	0.49/0.18
CN	1–0	113.490970	110/USB	0.95	0.79	188	22.8/31	0.52/0.5	40.4/62.4	0.31/0.11
C^{17}O	1–0	112.358982	110/USB	0.95	0.79	167	23.1/31	0.52/0.5	40.4/62.4	0.27/0.10
^{13}CO	1–0	110.201354	110/USB	0.95	0.79	118	23.5/31	0.53/0.5	40.4/62.4	0.17/0.07
C^{18}O	1–0	109.782173	110/USB	0.95	0.79	114	23.6/31	0.53/0.5	40.4/62.4	0.17/0.07
^{32}SO	2–1	99.299870	110/LSB	0.95	0.80	95	26.1/31	0.59/0.5	40.4/62.4	0.13/0.06
^{12}CS	2–1	97.980953	110/LSB	0.95	0.80	90	26.5/31	0.60/0.5	40.4/62.4	0.12/0.06
CH$_3$OH-A	2–1	96.741375	110/LSB	0.95	0.81	93	26.8/31	0.60/0.5	40.4/62.4	0.11/0.06
CH$_3$OH-E	2–1	96.739362	110/LSB	0.95	0.81	93	26.8/31	0.60/0.5	40.4/62.4	0.11/0.06
N_2H^+	1–0	93.173764	110/LSB	0.95	0.81	100	27.8/31	0.63/0.5	40.4/62.4	0.13/0.07
HNC	1–0	90.663568	102/LSB	0.95	0.81	115	28.6/31	0.64/0.5	44.9/70.5	0.12/0.08
HCO^+	1–0	89.188525	102/LSB	0.95	0.81	130	29.1/31	0.66/0.5	44.9/70.5	0.13/0.09
HCN	1–0	88.631848	102/LSB	0.95	0.81	124	29.3/31	0.66/0.5	44.9/70.5	0.12/0.09
C_2H	1–0	87.316898	102/LSB	0.95	0.82	132	29.7/31	0.67/0.5	44.9/70.5	0.15/0.11
HN^{13}C	1–0	87.090825	102/LSB	0.95	0.81	137	28.6/31	0.64/0.5	44.9/70.5	0.14/0.11
SiO	2–1	86.846960	102/LSB	0.95	0.82	142	29.9/31	0.67/0.5	44.9/70.5	0.14/0.11
H^{13}CO$^+$	1–0	86.754288	102/LSB	0.95	0.81	136	28.6/31	0.64/0.5	44.9/70.5	0.14/0.10
H^{13}CN	1–0	86.340184	102/LSB	0.95	0.81	136	28.6/31	0.64/0.5	44.9/70.5	0.13/0.10
c-C_3H_2	2–1	85.338893	102/LSB	0.95	0.82	123	30.4/31	0.69/0.5	44.9/70.5	0.12/0.10

Notes. [(a)] Listed as natural/smoothed resolution. [(b)] Listed as natural/oversampled channel spacing. [(c)] Listed as on-source time/telescope time. [(d)] Listed as measured on the natural/resampled resolution cubes.

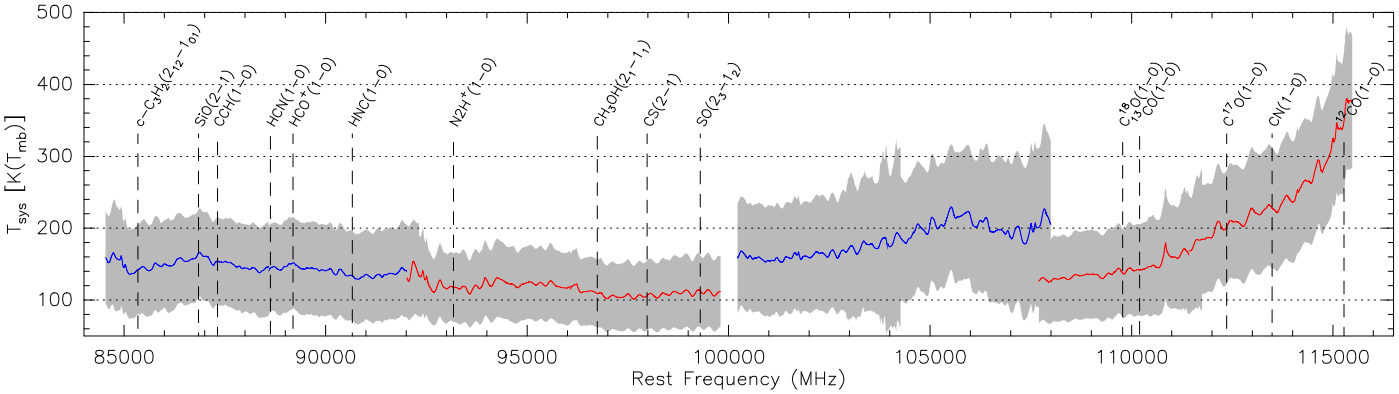


Fig. A.1. $T_{\rm sys}$ as a function of the frequency. The solid lines display the average $T_{\rm sys}$ and the gray shaded backgrounds show the $\pm 3\sigma$ interval at each frequency. The blue and red colors present the instantaneous 8 GHz-bandwidth of the lower and upper sidebands covered by the two tunings. The vertical dashed lines show the frequencies of the brightest lines studied here.

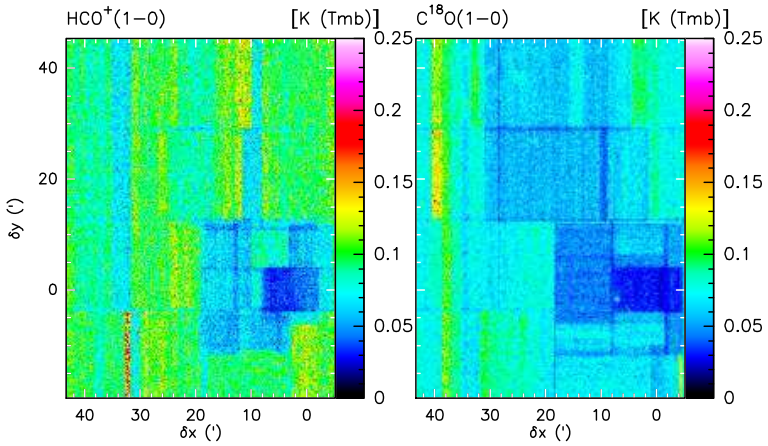


Fig. A.2. Typical spatial distribution of the RMS noise for two lines, each one belonging to one of the two tuning setups.

Table B.1. Line averaged intensities in mK km s^{-1} and in percentage of the total flux inside the four A_V mask regions.

Species	Transition	$0 \leq A_V < 222$	$1 \leq A_V < 2$	$2 \leq A_V < 6$	$6 \leq A_V < 15$	$15 \leq A_V < 222$
^{12}CO	1–0	33 000 (100%)	3300 (9.9%)	36 000 (110%)	65 000 (200%)	77 000 (230%)
^{13}CO	1–0	6200 (100%)	350 (5.7%)	4900 (79%)	15000 (230%)	29 000 (470%)
HCO$^+$	1–0	980 (100%)	110 (11%)	840 (85%)	2000 (210%)	4700 (480%)
HCN	1–0	590 (100%)	40 (6.9%)	440 (75%)	1300 (230%)	3200 (550%)
C^{18}O	1–0	430 (100%)	8.4 (1.9%)	210 (49%)	1100 (240%)	3800 (880%)
^{12}CS	2–1	370 (100%)	9.9 (2.7%)	190 (52%)	800 (220%)	3700 (990%)
^{12}CN	1–0	350 (100%)	36 (10%)	250 (70%)	810 (230%)	2100 (590%)
HNC	1–0	310 (100%)	26 (8.4%)	190 (61%)	650 (210%)	2600 (830%)
C_2H	1–0	270 (100%)	47 (18%)	230 (87%)	510 (190%)	1400 (520%)
^{32}SO	2–1	200 (100%)	6.8 (3.4%)	100 (50%)	460 (230%)	1900 (940%)
C^{17}O	1–0	100 (100%)	23 (22%)	53 (53%)	230 (220%)	800 (780%)
c-C_3H_2	2–1	87 (100%)	15 (18%)	73 (84%)	170 (190%)	450 (520%)
H^{13}CO$^+$	1–0	33 (100%)	0.87 (2.6%)	4.9 (15%)	57 (170%)	560 (1700%)
N_2H^+	1–0	29 (100%)	−13 (−45%)	5 (17%)	25 (85%)	780 (2700%)
CH_3OH	2–1	27 (100%)	4.7 (17%)	3.3 (12%)	57 (210%)	390 (1500%)

Notes. The lines are sorted by decreasing value of their intensity.

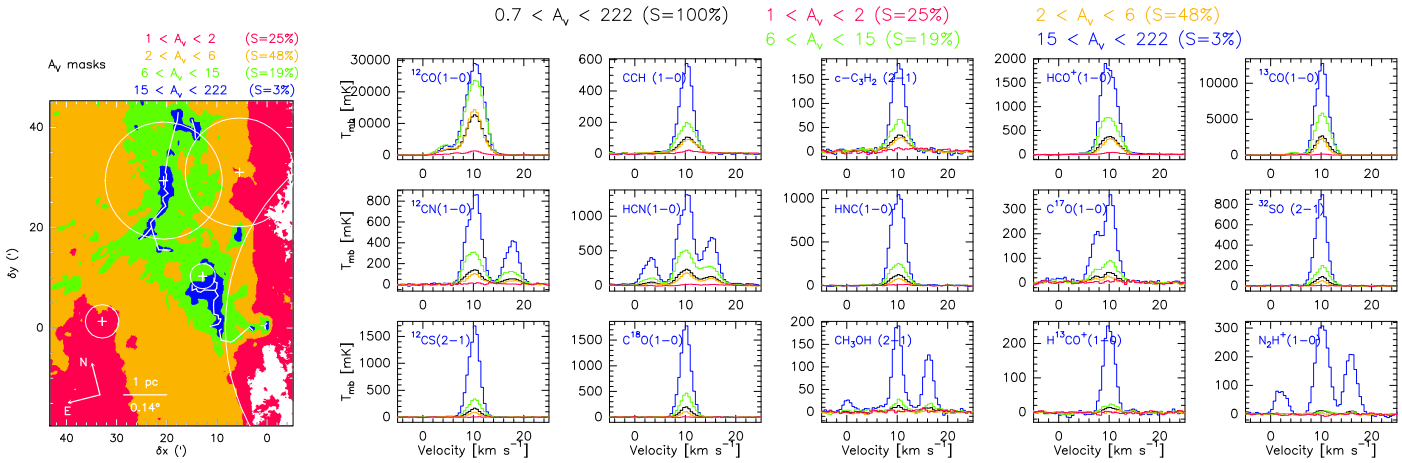


Fig. B.1. Same as Fig. 6, except that the spectra show the temperature intensity averaged over the different masks.

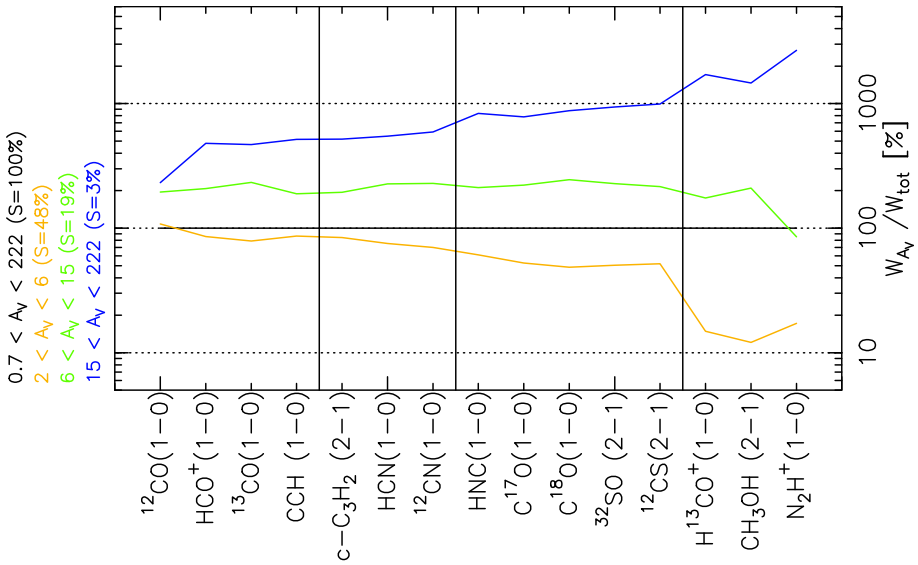


Fig. B.2. For each transition, line integrated intensity computed over the same velocity range divided by the line integrated intensity computed over the surface where $A_V \geq 15$. All line integrated intensities are computed between 9 and 12 km s^{-1}. The black vertical lines define the groups of lines described in Sect. 4.1.

Table B.2. Line intensities averaged over [9, 12 km s^{-1}] inside the four T_d mask regions in mK km s^{-1} and in percentage of the intensity of the cold dust region.

Species	Transition	$16 \leq T_d < 100$	$16 \leq T_d < 19.5$	$19.5 \leq T_d < 23.5$	$23.5 \leq T_d < 32$	$32 \leq T_d < 100$
^{12}CO	1–0	33 000 (100%)	44 000 (130%)	28 000 (85%)	31 000 (94%)	56 000 (170%)
^{13}CO	1–0	6200 (100%)	15000 (250%)	5800 (92%)	5000 (81%)	10000 (170%)
HCO$^+$	1–0	980 (100%)	2500 (250%)	670 (68%)	870 (89%)	2200 (220%)
HCN	1–0	590 (100%)	1400 (230%)	390 (66%)	500 (85%)	1500 (250%)
C^{18}O	1–0	430 (100%)	2100 (490%)	420 (99%)	290 (67%)	660 (150%)
^{12}CS	2–1	370 (100%)	1700 (450%)	290 (78%)	250 (69%)	830 (230%)
^{12}CN	1–0	350 (100%)	770 (220%)	220 (61%)	290 (81%)	970 (280%)
HNC	1–0	310 (100%)	1300 (430%)	210 (69%)	240 (79%)	690 (230%)
C_2H	1–0	270 (100%)	460 (170%)	150 (57%)	260 (97%)	670 (250%)
^{32}SO	2–1	200 (100%)	1200 (610%)	170 (84%)	140 (70%)	350 (180%)
C^{17}O	1–0	100 (100%)	440 (430%)	100 (99%)	73 (72%)	150 (140%)
c-C_3H_2	2–1	87 (100%)	200 (230%)	50 (57%)	85 (97%)	210 (240%)
H^{13}CO$^+$	1–0	33 (100%)	290 (890%)	25 (75%)	19 (58%)	67 (210%)
N_2H^+	1–0	29 (100%)	420 (1500%)	25 (86%)	11 (37%)	44 (150%)
CH_3OH	2–1	27 (100%)	260 (970%)	21 (78%)	19 (72%)	38 (140%)

Notes. The lines are sorted by decreasing value of their intensity.

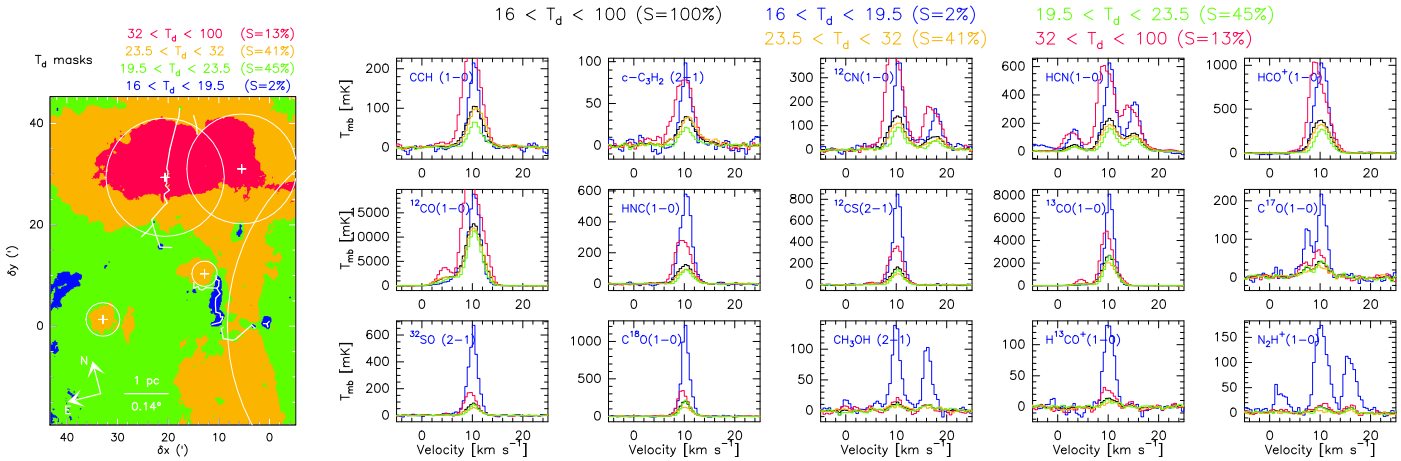


Fig. B.3. Same as Fig. 8, except that the spectra show the temperature intensity averaged over the different masks.

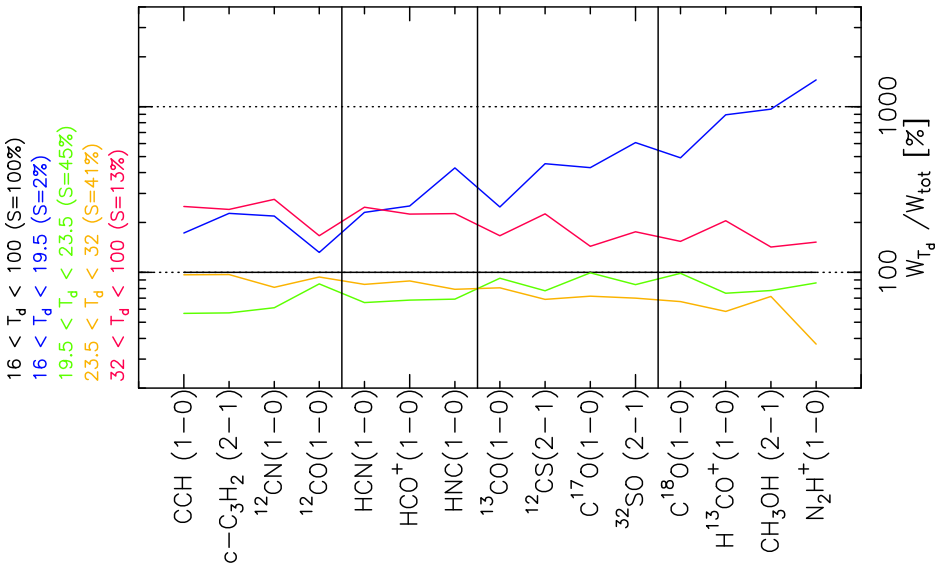


Fig. B.4. For each transition, line integrated intensity divided by the line integrated intensity computed over the surface where $T_d < 20$ K. All line integrated intensities are computed between 9 and 12 km s^{-1}. The black vertical lines define the groups of lines described in Sect. 4.2.

A&A 599, A100 (2017)
DOI: 10.1051/0004-6361/201629847

Astronomy & Astrophysics

Dissecting the molecular structure of the Orion B cloud: insight from principal component analysis⋆

Pierre Gratier[1], Emeric Bron[2,3], Maryvonne Gerin[4], Jérôme Pety[5,4], Viviana V. Guzman[6], Jan Orkisz[5,4], Sébastien Bardeau[5], Javier R. Goicoechea[2], Franck Le Petit[3], Harvey Liszt[7], Karin Öberg[6], Nicolas Peretto[8], Evelyne Roueff[3], Albrech Sievers[5], and Pascal Tremblin[9]

[1] Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, Allée Geoffroy Saint-Hilaire, 33615 Pessac, France
e-mail: pierre.gratier@u-bordeaux.fr
[2] ICMM, Consejo Superior de Investigaciones Cientificas (CSIC), 28049 Madrid, Spain
[3] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, 92190 Meudon, France
[4] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, École normale supérieure, 75005, Paris, France
[5] IRAM, 300 rue de la Piscine, 38406 Saint-Martin d'Hères, France
[6] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[7] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[8] School of Physics and Astronomy, Cardiff University, Queen's buildings, Cardiff CF24 3AA, UK
[9] Maison de la Simulation, CEA-CNRS-INRIA-UPS-UVSQ, USR 3441, Centre d'étude de Saclay, 91191 Gif-Sur-Yvette, France

Received 5 October 2016 / Accepted 13 January 2017

ABSTRACT

Context. The combination of wideband receivers and spectrometers currently available in (sub-)millimeter observatories deliver wide-field hyperspectral imaging of the interstellar medium. Tens of spectral lines can be observed over degree wide fields in about 50 h. This wealth of data calls for restating the physical questions about the interstellar medium in statistical terms.
Aims. We aim to gain information on the physical structure of the interstellar medium from a statistical analysis of many lines from different species over a large field of view, without requiring detailed radiative transfer or astrochemical modeling.
Methods. We coupled a non-linear rescaling of the data with one of the simplest multivariate analysis methods, namely the principal component analysis, to decompose the observed signal into components that we interpret first qualitatively and then quantitatively based on our deep knowledge of the observed region and of the astrochemistry at play.
Results. We identify three principal components, linear compositions of line brightness temperatures, that are correlated at various levels with the column density, the volume density and the UV radiation field.
Conclusions. When sampling a sufficiently diverse mixture of physical parameters, it is possible to decompose the molecular emission in order to gain physical insight on the observed interstellar medium. This opens a new avenue for future studies of the interstellar medium.

Key words. ISM: molecules – ISM: clouds – photon-dominated region (PDR) – ISM: individual objects: Orion B – methods: statistical

1. Introduction

Molecular clouds have a complex structure, with filaments hosting dense cores and immersed in a low density diffuse envelope. Large-scale dust continuum maps obtained with *Herschel* have provided a breakthrough, by showing the tight relationship between the filaments and the dense cores. These maps however do not provide information on the gas dynamics or its chemical composition. Furthermore the relationship between the submillimeter dust emission and the gas column density is affected by the dust temperature and possible variations of the dust emissivity. Molecular line emission maps provide alternative means to study molecular cloud structure and relate it to the flow kinematics. Molecular line emission is linked to the underlying physical properties of the interstellar medium (ISM), such as density, gas and dust temperatures, UV radiation field, and cosmic ray ionization rate. However these relationships are complex and their detailed study is a field in itself, namely astrochemical modeling (Le Bourlot et al. 2012; Agúndez & Wakelam 2013). Further complexity arises when considering radiative transfer to derive line intensities from the local chemical composition and physical structure.

The last few years have seen the installation of new wideband receivers and spectrometers at millimeter and sub-millimeter radiotelescopes. With these instruments, line surveys of several GHz bandwidth and several tens of thousands of spectral channels are the new default mode of observations. Combined with wide field imaging capabilities both for single dish and

⋆ Based on observations carried out at the IRAM-30 m single-dish telescope. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain).

Table 1. Properties of the observed spectral lines. The last six columns show the statistics of the data before and after asinh reparametrization.

Molecule	Transitions	Frequency (MHz)	Noise (K)	Original data Min. (K)	Median (K)	Max. (K)	Std. (K)	After asinh reparametrization Min. (K)	Median (K)	Max. (K)	Std. (K)
^{12}CO	$J = 1 \rightarrow 0$	115 271.202	0.09	−0.39	13.40	57.11	10.18	−0.37	2.39	3.32	0.91
^{13}CO	$J = 1 \rightarrow 0$	110v201.354	0.04	−0.19	1.38	36.43	3.27	−0.19	0.97	3.03	0.74
CS	$J = 2 \rightarrow 1$	97 980 .953	0.06	−0.36	0.06	15.53	0.48	−0.35	0.06	2.48	0.23
HCN	$J = 1 \rightarrow 0$, $F = 2 \rightarrow 1$	88 631 .848	0.10	−0.58	0.15	10.32	0.39	−0.52	0.15	2.22	0.25
HCO^+	$J = 1 \rightarrow 0$	89 188 .525	0.09	−0.45	0.26	8.07	0.47	−0.42	0.25	2.07	0.30
SO	$N = 3 \rightarrow 2$, $J = 2 \rightarrow 1$	99 299 .870	0.06	−0.43	0.04	6.46	0.24	−0.40	0.04	1.92	0.17
CN	$N = 1 \rightarrow 0$, $J = 3/2 \rightarrow 1/2$, $F = 5/2 \rightarrow 3/2$	113 490.970	0.09	−0.58	0.10	6.33	0.27	−0.52	0.09	1.91	0.20
HNC	$J = 1 \rightarrow 0$	90 663 .568	0.08	−0.49	0.07	6.01	0.27	−0.45	0.07	1.88	0.19
CCH	$N = 1 \rightarrow 0$, $J = 3/2 \rightarrow 1/2$, $F = 2 \rightarrow 1$	87 316 .898	0.12	−0.62	0.08	5.72	0.22	−0.55	0.08	1.85	0.18
C^{18}O	$J = 1 \rightarrow 0$	109 782 .173	0.06	−0.30	0.06	5.55	0.42	−0.29	0.06	1.83	0.26
N_2H^+	$J = 1 \rightarrow 0$, $F1 = 2 \rightarrow 1$, $F = 3 \rightarrow 2$	93 173 .764	0.08	−0.44	0.00	4.53	0.13	−0.41	0.00	1.70	0.10
CH_3OH	$J = 2 \rightarrow 1$, $K = 0 \rightarrow 0$, $(A+)$	96 741 .375	0.06	−0.34	0.01	2.24	0.08	−0.32	0.01	1.26	0.08

interferometers, hyperspectral imaging is now routinely carried out with these instruments.

The analysis and interpretation of these large datasets, consisting of thousands of spatial positions and tens of thousands of spectral channels, will benefit from the use of statistical tools. Principal component analysis (PCA) is one of the most widely used multivariate analysis method. It has been used to study the ISM using molecular emission maps (Ungerechts et al. 1997; Neufeld et al. 2007; Lo et al. 2009; Melnick et al. 2011; Jones et al. 2012).

In this paper, we address the following question: can PCA provide a method to study the underlying physics of the ISM when applied to a large dataset of molecular emission, without performing either radiative transfer or astrochemical modeling? The article is divided as follows. In Sect. 2 we present the data used in this study. In Sect. 3 we describe the statistical method used in this paper and its implementation. Results are presented in Sect. 4 first by analyzing the output of the PCA, and further by comparing theses outputs with independent maps of physical conditions in Orion B in Sect. 5. The last section discusses these results.

2. Data

The data used in this paper was selected from the ORION-B project (PI: J. Pety), which aims at mapping with the IRAM-30 m telescope a large fraction of the south-western edge of the Orion B molecular cloud over a field of view of 1.5 square degrees in the full 3 mm atmospheric window at 200 kHz spectral resolution. Pety et al. (2016) describe in detail the data acquisition and reduction strategies. Table 1 lists the 12 transitions selected in this paper from the already observed frequency range (from 84 to 116 GHz), based on the inspection of the full data cube.

For each line, we focused on emission coming from a limited 1.5 km s^{-1}-velocity range centered on the peak velocity (i.e., 10.5 km s^{-1}) of the main velocity component along the line of sight. Averaging the three 0.5 km s^{-1} velocity channels allowed us to get a consistent dataset from the radiative transfer and kinematics viewpoints. In particular, we avoided the need to disentangle 1) the effects of hyperfine structures of some lines, and 2) the complex velocity structure of the source (Orkisz et al. 2017).

The observed field of view covers 0.81 deg × 1.10 deg and contains the Horsehead nebula, and the H II regions NGC 2023, NGC 2024, IC 434, and IC 435. The angular resolution ranges from 22.5 to 30.5″. The 12 resulting maps have a common pixel size of 9″ that corresponds to a Nyquist sampling for the highest frequency line observed (^{12}CO(1–0) at 115.27 GHz). The maps thus contain 315×420 pixels. At a distance of ~400 pc (Menten et al. 2007), the maps give us access to physical scales between ~50 mpc and 10 pc.

Figure 1 shows the 12 maps of the resulting brightness temperature multiplied by an ad hoc factor in order that they can share the same color look-up table, even though the intrinsic brightness temperatures of the different lines differ by more than one order of magnitude. The relative calibration of the different lines is excellent because they were observed with the same telescope at almost the same time, since the observed bandwidth was covered in only two frequency tunings. The noise for each map, along with the minimum, median, maximum, and variance values are listed in Table 1. The noise was computed by fitting a gaussian function to the negative part of the histogram of pixel brightnesses. This enabled us to compute the noise without needing to mask out the emission.

3. Principal component analysis

3.1. Principle

We use the following standard statistical terms: the dataset is composed of "samples" each described by individual "features". In our case, each spatial pixel is a sample and each line intensity is a feature (the full dataset thus corresponds to a data matrix of 132 300 samples times 12 features).

PCA is a widely-used multi-dimensional analysis technique (Jolliffe 2002), that can be defined in several mathematically equivalent ways. It aims at finding a new orthogonal basis of the feature space (whose axes are called principal components, or PCs), so that for each k, the projection onto the hyperplane defined by the first k axes is optimal in the sense that it preserves most of the variance of the dataset (or equivalently that the error caused by this projection is minimal). PCA thus defines successive approximations of the dataset by hyperplanes of increasing dimension.

This is equivalent to the diagonalization of the covariance matrix, so that the principal components are naturally uncorrelated. It can be thought of as finding the principal axes of inertia

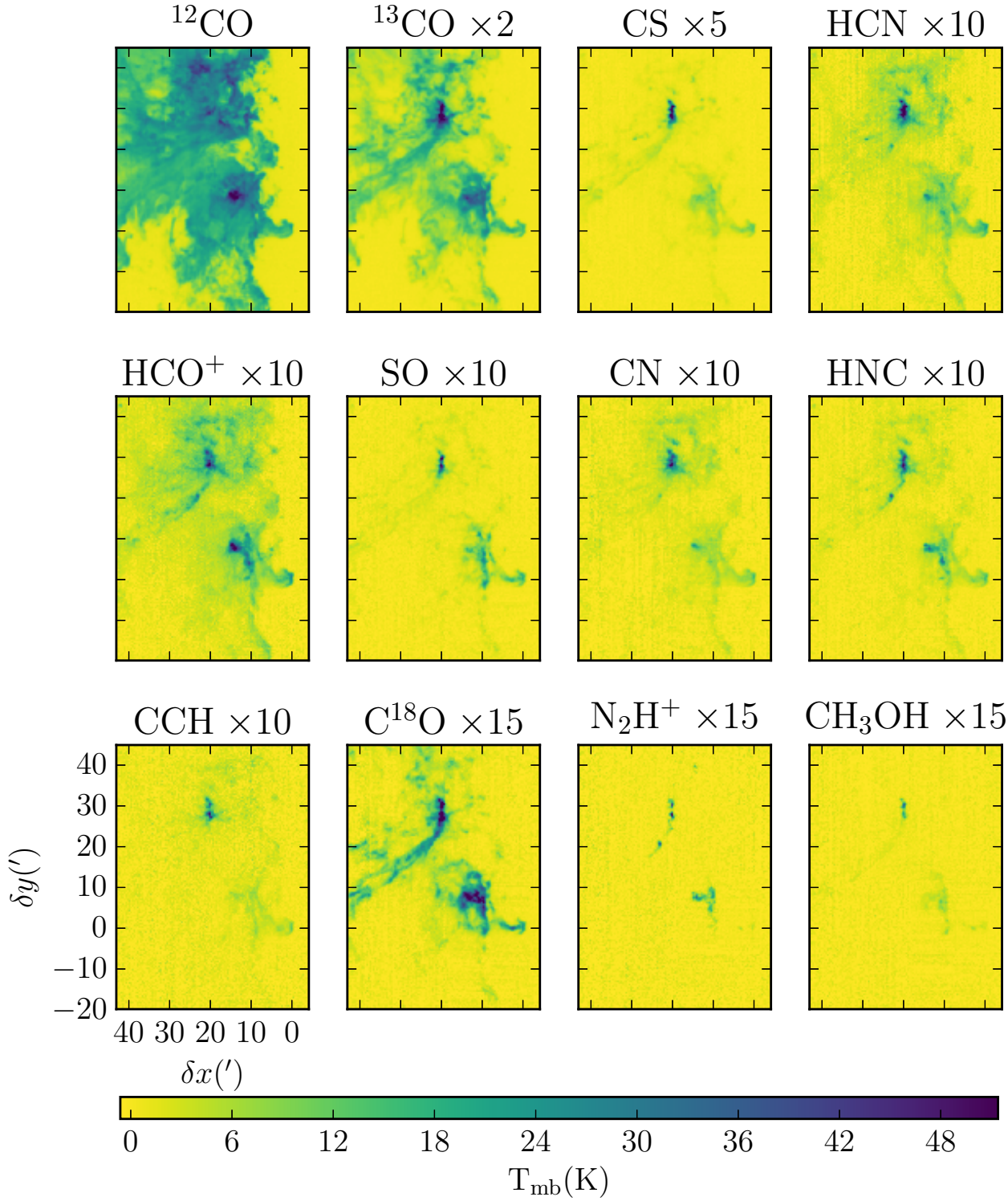


Fig. 1. Maps of molecular emission in Kelvin main beam temperatures.

of the cloud of samples about their mean in feature space, and is thus a way to analyze the covariance structure of the data. The principal components are ordered by decreasing projected variance. As a result PC1 is the axis of largest variance in the data. PC2 is then the axis of largest variance at constant PC1 (orthogonal to PC1) and so forth. Neglecting the axes of lowest variance then allows the definition of a low-dimensional hyperplane in which the dataset is approximately embedded. An important property to keep in mind is the linearity of PCA, namely that it defines low-dimensional hyperplanes, and not general low-dimensional hypersurfaces.

A common variant[1] in the application of PCA is to normalize the variations of the dataset around the mean by the standard deviation of each features, before applying the PCA. This amounts to diagonalizing the correlation matrix instead of the

[1] This variant in the application of PCA goes back to one of the two earliest descriptions of PCA: Hotelling (1933).

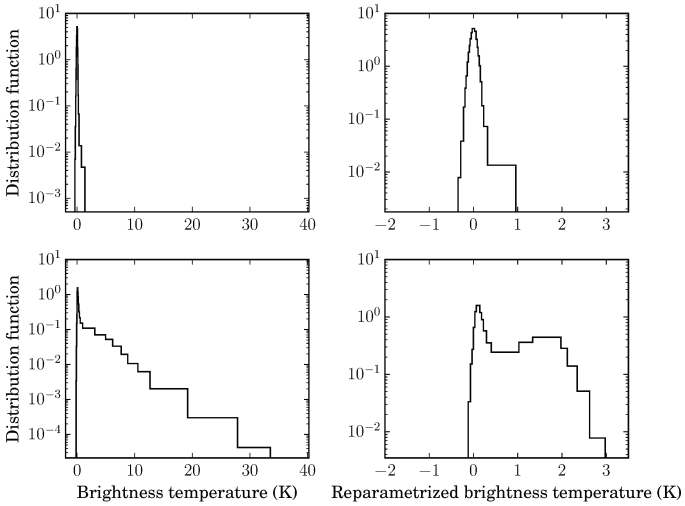


Fig. 2. Effect of the asinh renormalization on the intensity distributions. *Left column*: before renormalization, *right column*: after renormalization, *top row*: N_2H^+, *bottom row*: ^{13}CO.

covariance matrix. The correlation-based variant allows to avoid having one feature dominate the variance, and is appropriate if the relative scales of the features are not relevant for the purpose of the analysis. As the relative intensity scales of the different molecules used here are largely affected by properties (dipole moment, elemental abundances, ...) that are not relevant for our analysis of the chemical variations across the map, we use here the correlation-based version of PCA.

In this work, we used the PCA implementation available in the `Python` package `scikit-learn` (Pedregosa et al. 2011), which uses a singular-value decomposition to compute the principal component axes.

3.2. Reparametrization of input data

3.2.1. The need of a reparametrization

As seen in Table 1, some of the tracers have large dynamical ranges (two orders of magnitude for $^{13}CO(1–0)$ and $^{12}CO(1–0)$).

Figure 2 shows the histogram of the brightnesses temperatures of two lines with contrasting behavior in our dataset, namely $^{13}CO(1–0)$, and $N_2H^+(1–0)$. As the dynamic range is large both in intensity and number of pixels per bin, these histograms use the Bayesian blocks algorithm (Scargle et al. 2013, using here the `Python` implementation from AstroML; see Vanderplas et al. 2012; Ivezić et al. 2014), which adapts the bin width to the underlying distribution. Although the histogram of the $N_2H^+(1–0)$ is Gaussian to first order, the histogram of $^{13}CO(1–0)$ exhibits heavy tails similar to power laws. As a result, extreme intensity values might dominate the covariance structure of the data, and hide the variations at the more common lower intensity values.

From the physical viewpoint, taking the logarithm of the brightness temperature is also desirable. PCA is a linear technique which decomposes the data as a sum of uncorrelated components. Applying it to the logarithm of the data allows a decomposition as a product of factors, and thus describes the data structure in terms of ratios, products, and power laws, which is more adapted to the underlying radiative transfer and chemical effects. Taking the logarithms of the data is the equivalent in astrochemistry to performing color−color magnitude diagram analysis in optical or UV studies. Pety et al. (2016) show that the line-integrated brightness temperatures of our dataset are, to first order, correlated to the column density of matter along the line of sight. We expect this aspect to appear in our PCA analysis, whereas second order chemical variations around this trend, which would be revealed by line ratios, are thus better described as multiplicative (rather than additive) factors.

3.2.2. Impact of noise

The presence of noise causes the possibility of negative values in those pixels where some lines are undetected. The logarithm transform cannot be applied to these negative noise values. In addition, because the logarithm stretches the lowest values compared to the largest ones, it will also tend to stretch the positive noise values of the undetected pixels, and this gives them more weight in the covariance of the data.

There may be two different reasons for a non-detection: either the measurement is not sensitive enough to detect the line or the region just does not contain the species that emits the line. The latter case happens particularly in H II regions (e.g., IC 434), in which $^{12}CO(1–0)$ is photo-dissociated by the far UV photons. In this particular case, we could remove from our dataset all the samples (pixels) where no $^{12}CO(1–0)$ is detected. This just assumes that no $^{12}CO(1–0)$ detections at high sensitivity imply the absence of molecular gas. However, this method is not generic. For instance, $N_2H^+(1–0)$ is only detected in dense cores (Pety et al. 2016), and restricting ourselves to these regions would drastically limit the scope of our study. If we wish to use this important tracer of the molecular gas while still covering the range of different chemical regimes present in our full map, we thus need to find an alternative. Moreover, in the 3 mm band, radio recombination lines, which emit in H II regions, could in principle be added to our dataset to study the formation of molecular gas.

Adding a thresholding step before taking the logarithm will only worsen the scope of undefined values. Although there are PCA methods (e.g. Ilin & Raiko 2010) that can take missing datapoints into account, they rely on the fact that these missing points have the same statistics than the measured points (i.e., when a value is missing, it is independent of the actual value). This is clearly not the case here because the missing values (undetected lines) are missing due to them being below the sensitivity threshold. These are called “censored values” in statistics. We thus searched for a function that is linear around zero (in the noise-dominated domain), and that is asymptotically equal to a logarithm for large values (compared to the noise level). The inverse hyperbolic sinus function, $\mathrm{asinh}(x)$, fulfills these conditions. We thus used the following function to reparametrize the data before applying PCA

$$T(x) = a\,\mathrm{asinh}(x/a), \tag{1}$$

where the parameter a is the typical value for which the function’s behavior changes from a linear to a logarithmic regime (Fig. 3).

The only free parameter in the method is the threshold a. Appendix A discusses our choice, namely $a = 8 \times 0.08 = 0.64$ K = eight times the median noise of the dataset. In short, we select the value of a that maximizes the correlations of the first three principal components with independent known measurements of the column density, volume density, and UV illumination (Sect. 5). This appendix also demonstrates that our results are quite insensitive to the exact value of a.

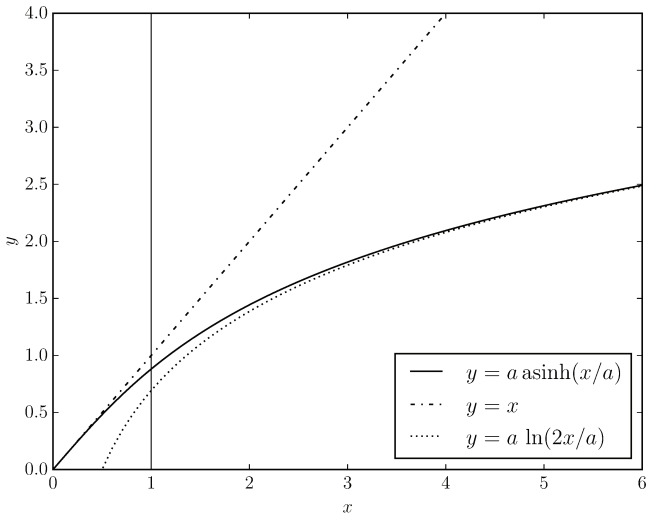


Fig. 3. Plot of the asinh function (solid line) showing the asymptotes when $x \to 0$ (dash dotted line) and $x \to +\infty$ (dotted line). The parameter a (here $a = 1$) is traced with a thin vertical line.

The right column of Fig. 2 shows the result of the asinh transformation on the intensity distributions of two transitions representative of bright (^{13}CO(1–0)) and weak (N_2H^+(1–0)) averaged lines. In the case of the bright lines, the dynamic range is drastically reduced, with the heavy tail being transformed into a second peak in the distribution but with no values above three. In the case of N_2H^+ the distributions before and after reparametrization are very similar. Figure 4 shows the 12 maps of the molecular emission after reparameterization by the asinh function, but before the normalization step of the PCA analysis. The brightness temperatures of all the maps have been compressed between about −0.5 and 3.5, with low signal-to-noise brightness temperatures between −0.5 and 0.5 being mostly untransformed.

4. Results

The PCA method exposes the correlations between the line brightness temperatures. The derived PCs give the main axes of correlated variations in the data set. As such, PCA does not directly yield physical information underlying the dataset. In this section, we describe the results of the application of this statistical method to our dataset and we start to discuss their possible physic interpretation based on our a priori astrochemical knowledge. The possible relations between the PCs and physical variables are investigated in a later section.

4.1. Correlation fraction explained by the different principal components

Figure 5 shows the percentage of the correlation explained by each PC (as a function of the principal component number) along with the cumulative explained correlation as a function of the number of principal components kept in the decomposition.

The first principal component explains the majority (60%) of the total correlation present in the original dataset. Thus a large part of the variations in the dataset occur along a single axis (i.e. all lines are strongly correlated to each other). The second principal component accounts for about 10% of the correlation. It is significantly less than the first component, but more than any other components. PCs 3, 4 and 5 correspond to similar amounts of correlations (around 5% each) and PC6 slightly less (3.3%). PC1 to 6 collectively explain more than 90% of the correlation in our dataset. The remaining PCs have similar low amounts of explained correlation (from 2% for PC7 to 0.9% for PC11).

4.2. Discussion of the principal components

The PCs defined by our analysis represent new axes in the feature space (the full 12 PCs are simply a rotation of the initial basis of the feature space), deduced from the data itself. They can thus be expressed in terms of the original axes, as a linear combination of the (transformed) line intensities. Figure 6 displays the quantitative contribution of each initial feature (line) to each PC. An alternative view of the relationship between the PCs and the line intensities, namely the correlation wheels, is presented in Fig. 7.

Each sample (pixel brightness) can then be projected on the new axes, providing new coordinates commonly called "component scores". The PCA method considers the pixels as independent samples, and thus ignores the spatial structure of the molecular emission. It is nevertheless possible to reconstruct the maps of the component scores. Figure 8 shows these projected maps. The chosen color look-up table emphasizes that positive and negative values of the projected maps, which correspond to variations above and below the average along the considered axis respectively, clearly extract a different spatial pattern per principal component.

The first principal component is a linear combination of all tracers, with similar positive weights for all lines (Fig. 6). It thus describes correlated variations of all molecules, and these account for most of the variations in the dataset, which is a natural consequence of all lines being well correlated (positively) to each other. Pety et al. (2016) show that the emission of all lines is correlated to first order with the column density of matter along the line of sight. The first component is thus probably related to the total column density, whose increase causes, to first order, an increase in all lines. This is because in the linear approximation of PCA, non linear effects such as saturation of the ^{12}CO line are not captured. The corresponding component map (Fig. 8) indeed resembles a map of column density. This relation between PC1 and the total column density of matter will be investigated more quantitatively in Sect. 5. We note that as PC1 has only positive coefficients for all lines, orthogonality ensures that all other PCs will represent contrasts between different lines.

The second principal component represents the axis of largest variation at constant PC1 (orthogonal to PC1). This axis of variations is dominated by positive contributions of N_2H^+ and CH_3OH, and negative contribution of ^{12}CO and ^{13}CO. The first two tracers are chemically associated with dense and cold regions of the ISM. For instance, because N_2H^+ is easily destroyed by CO, it can thus only be abundant in the gas phase when CO has been depleted on the grain surfaces (Pety et al. 2016). The component map shows strong positive values highlighting known dense cores, including the clumps in the head and in the neck of the Horsehead (Ward-Thompson et al. 2006).

The third principal component shows positive contributions of CCH and CN that are known to be sensitive to UV illumination, and negative contributions of N_2H^+, CH_3OH and of the CO isotopologues, that trace gas shielded from the UV field. This component thus probably traces the chemical specificities of UV illuminated gas. The component map clearly shows positive values at the eastern edge of the cloud, illuminated by σ Ori, and in the star-forming region NGC 2024.

The fourth principal component is particular in the fact that its map almost only shows large (positive or negative) values in

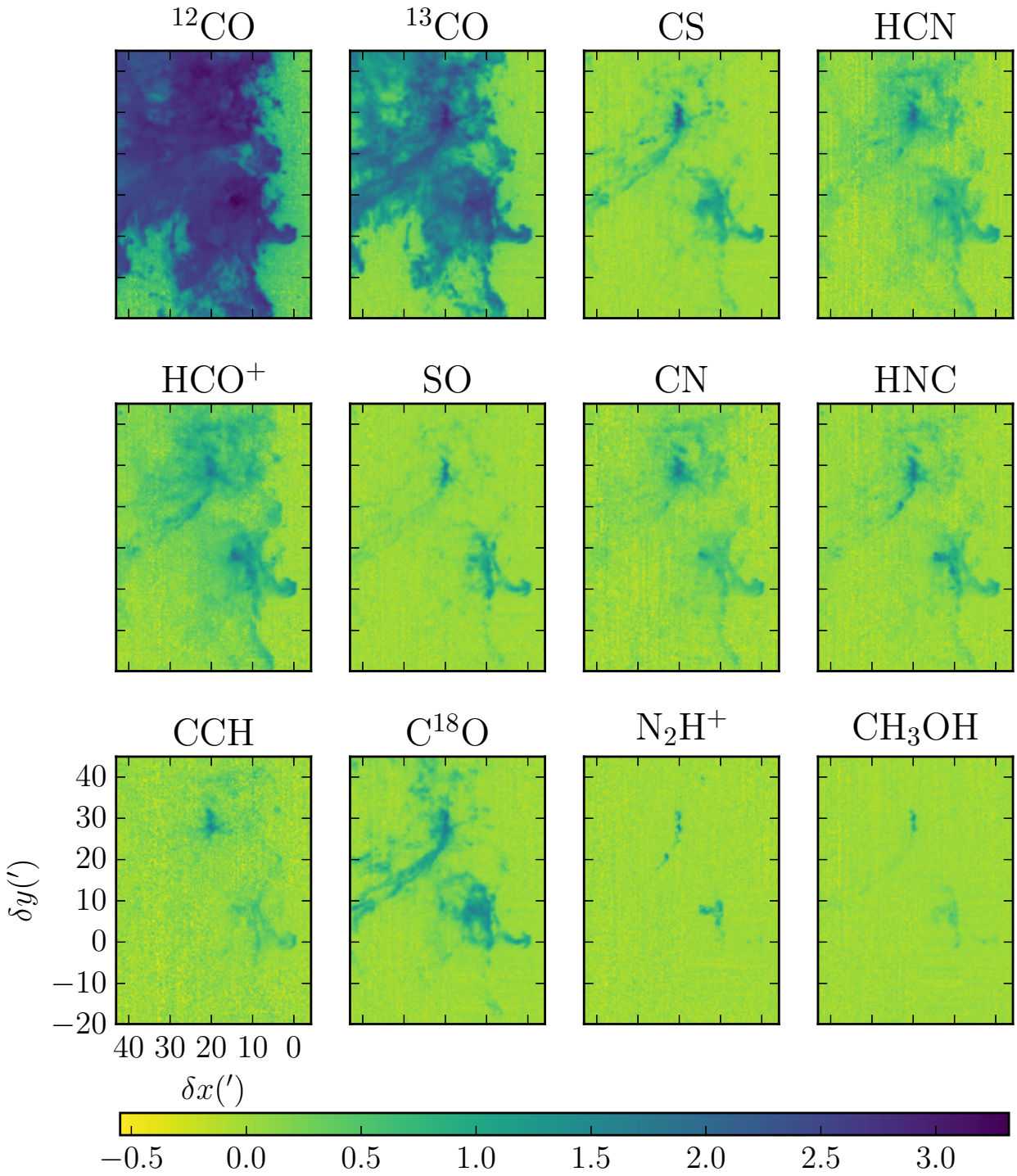


Fig. 4. Maps of molecular emission after reparametrization by the asinh function.

the regions of large positive values of PC2. It thus highlight further chemical variations inside dense cores. This component is completely dominated by opposite contributions of CH_3OH and N_2H^+ and thus traces variations in the ratio of these two lines. The component map seems to highlight smaller-sized cores embedded in some of the clumps revealed by PC2, and thus probably highlights the chemistry of the densest cores (larger N_2H^+ to CH_3OH ratios).

The fifth principal component shows positive contributions of sulfur species (CS and SO), and $C^{18}O$, and negative contributions of ^{12}CO, CCH, and CH_3OH. Its large positive values highlight larger scale regions embedding the dense clouds shown by PC2 (but with negative values where PC2 is very large), and this PC could thus trace the chemistry of moderately dense gas.

The sixth principal component shows negative contributions of HCN, HCO^+, and CN, which can all originate in photochemistry, and positive contributions from CCH, $C^{18}O$, ^{13}CO, and SO. Although the latter are usually associated with a larger amount of shielded gas, CCH can also be bright in UV illuminated regions. The component map shows a wide blue region around NGC 2024, similar to the large warm dust region seen in the dust temperature map of the region (Schneider et al. 2013).

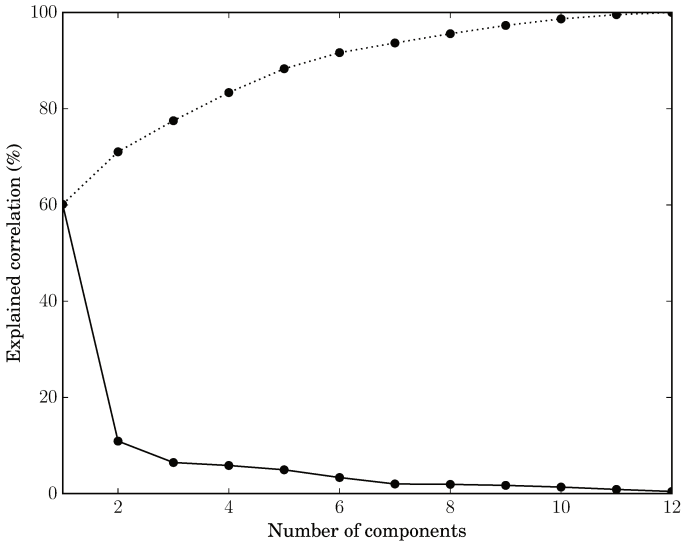


Fig. 5. Percentage of the explained correlation as a function of the number of components in the principal component analysis, dotted line: cumulative percentage.

This region could also be related to the radiation field, but trace a different aspect from PC3, that is characterized by lower CCH intensities relative to the other lines.

The remaining components are more difficult to interpret, but tend to describe opposite variations in pairs of lines that varied together in previous PCs. PCs 7, 9 and 10 display opposite variations in pairs of lines of the group HCN, HNC, CN and HCO^+, whose variations were correlated in the previous PCs in which they had large weights (PC1 and 6). PC8 shows anticorrelated variations of SO and $C^{18}O$, and its component map shows a striking spatial pattern: negative values (high SO/$C^{18}O$ ratios) in the Horsehead, the molecular gas at the base of the Horsehead, and the small scale clumps in NGC 2024; and positive values (low SO/$C^{18}O$ ratios) in a dense filament stretching away from NGC 2024. PC11 is strongly dominated by CS, and thus shows specific variations of CS, mostly uncorrelated with the other lines (somewhat anticorrelated with $C^{18}O$), and that were not described by the previous PCs. Its component map shows small scale spots of positive values, mostly surrounding NGC 2024 and NGC 2023. The fact that it appears so late in the decomposition can be explained by the small size of the highlighted region, having little weight in the correlation matrix. PC12 is completely constrained by orthogonality to the previous PCs and is thus only an artifact of previous PCs.

4.3. Studying the effect of noise

The noisy nature of our data can have two kinds of effects. It can first induce variability in our results (the results would vary for a different realization of the random noise). We verified the stability of our results by using a bootstrapping method. Bootstrapping is a method of choice to compute uncertainties on an estimator (here the PCA components) when the distribution of estimator values cannot be assumed to follow a simple distribution (Feigelson & Babu 2012). The idea of bootstrapping is to use a Monte Carlo method to create new resampled datasets of the same size as the original dataset by sampling with replacement from the original dataset. We constructed 5000 such bootstrapped datasets and ran the PCA algorithm on each. To avoid overestimating the uncertainty, and because the PCA is invariant through the change of sign of the PCs, we ensured that the signs were consistent before computing the distribution of the PC coefficients. The results are presented both for the eigenspectra in Fig. 6, and the correlation wheels in Fig. 7.

The PC coefficients appear overall very stable, the variances are completely negligible for PC 1 and 2, and very small for PC 3 to 6. Only PCs 7 and 8 show higher variability. This can be understood as PCA results are particularly sensitive to noise when two PCs correspond to very close eigenvalues, and PCs 7 and 8 have the closest eigenvalues with respectively 2% and 1.9% of the total correlation. Indeed, PCs with equal eigenvalues are degenerate in the sense that any basis of the subspace they define satisfies the definition of PCA. As a result, when eigenvalues are not exactly equal but very close, the noise can result in a random rotation of this group of PCs inside their subspace. Our results, which focus on the first few PCs, are thus unaffected by noise variability.

The second possible effect of noise is to bias the results. Principal component analysis is unbiased if the noise is spherical (i.e., has equal variance in all directions) in the final dataset on which PCA is applied (i.e., after standardization in our correlation-based variant). In this case, the noise can only hide the lowest PCs (that describe variations smaller than the noise level) and make them degenerate. In our case however, the noise levels on the different molecular lines are initially close but not equal (variations by a factor of three at most, as demonstrated in Table 1). The non-linear reparametrization keeps these relative variances. Finally, the last normalization step (by the standard deviation) results in final noise variances that are proportional to the ratios of noise standard deviation to total standard deviation of the reparametrized intensities. These differ by up to a factor of 14.8 between the lines, and possible biases may be present in our results, giving higher weight to the lines with the largest ratios of noise variance to total variance. However, it was not possible with the PCA method to avoid giving higher weight to the brightest lines (which led us to use the correlation-based PCA), and at the same time ensure equal noise on all variables. We note that previous PCA studies of molecular clouds were less concerned by noise-induced bias as they only used lines that were clearly detected in all pixels. We chose to perform the PCA on the full region, which led to the identification of two PCs associated with dense core chemistry.

5. Correlation of the principal component maps with independently measured physical parameters maps

In the previous section, we combined two sources of information to interpret the main principal components: 1) astrochemistry, which teaches us that some molecules trace certain physical conditions, and 2) Orion B is an extremely well-studied source, implying that its spatial structure is well known. For instance, the molecular cloud is known to be illuminated by well-defined young massive stars (discussion in Pety et al. 2016). This allowed us to infer a link between the first three principal components and physical parameters such as the column density, the volume density, and UV illumination. In this section, we will quantitatively assert these potential relations by studying the correlation of each component map with a set of independently measured maps of physical parameters.

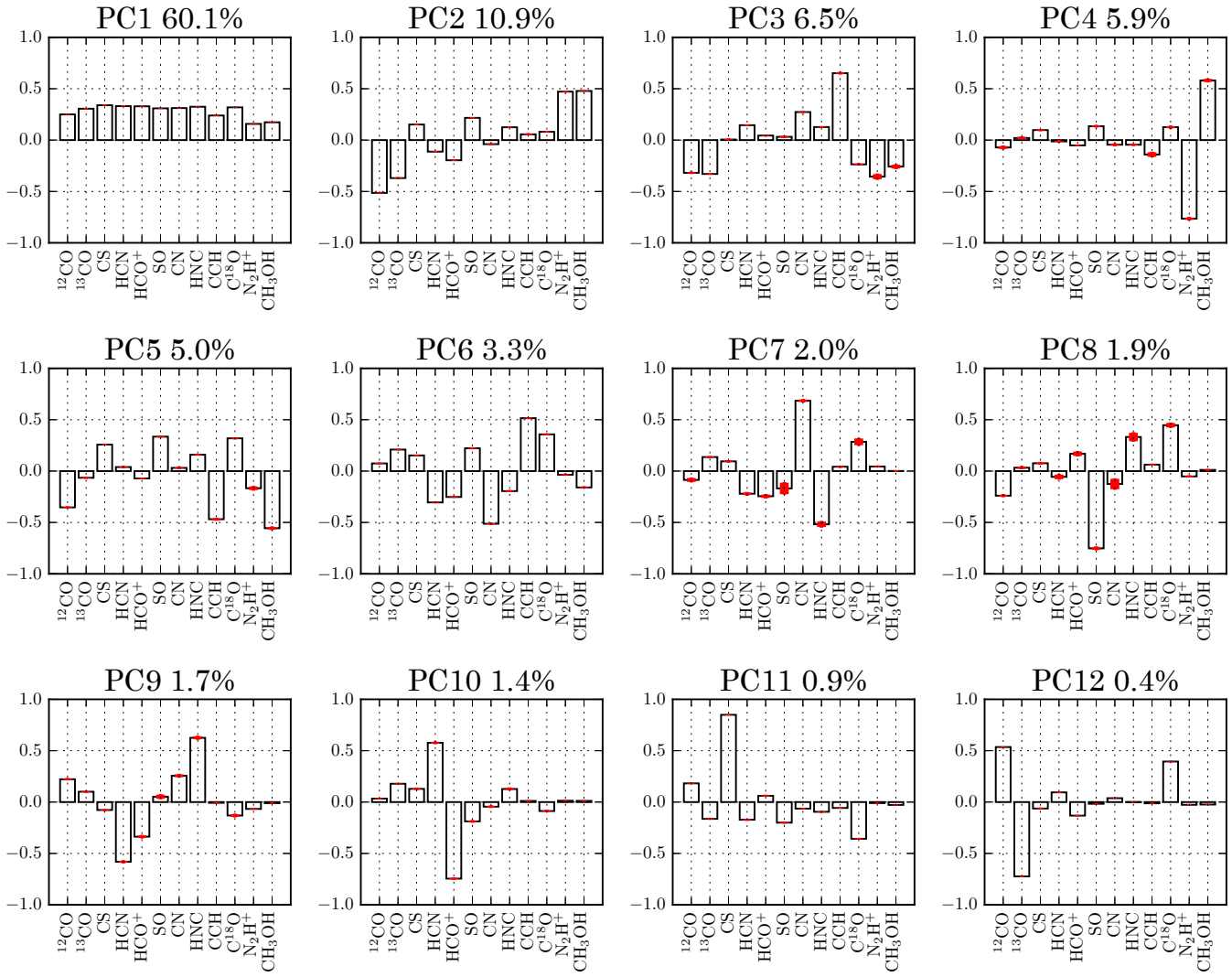


Fig. 6. Bar plots showing the contribution of each line intensity to each principal component (with the fraction of the total correlation accounted for by each PC given as a percentage). The uncertainties (standard deviations) shown in red are obtained by bootstrapping as described in Sect. 4.3.

5.1. Independent measure of the physical parameters

The goal of this section is to find the principal component that is best associated to each of the physical parameters, not to assign an absolute physical meaning to some of the components. It is therefore not necessary to have absolute values of the independently measured physical parameter maps. Only the relative variation of each physical parameter is required to compute the correlation coefficient. Figure 9 shows the different maps of the physical parameters that we will correlate with the first three principal components. This section describes how these maps were obtained.

5.1.1. Column density

The dust column density map is from the Hershel Gould Belt Survey (PI: P. Andre) Orion B map (André et al. 2010; Schneider et al. 2013)[2]. This map was obtained by fitting the far infrared spectral energy distribution by greybodies. We applied a logarithmic scaling to the data to reduce the dynamical range. The resulting map is plotted in the left panel of Fig. 9.

[2] http://www.herschel.fr/cea/gouldbelt/en/

5.1.2. Volume density

Volume density is a difficult quantity to measure because one needs both a mass estimate and an associated volume. Density is thus dependent on the scale that it is computed at. We used the catalog of cores identified and characterized in Kirk et al. (2016) and computed masses from each cloud's 850 μm flux using their equation three. To do this, we assumed a common temperature of 17 K for all clouds. From this mass and their observed size estimates we computed a volume density for each of the dense cores in our observed field of view. In this case correlation could not be carried out over the full map but we correlated the density measured for each core with the value of the principal components measured in the nearest pixel. The data is shown as a scatter plot in the middle panel of Fig. 9.

5.1.3. UV radiation field

We computed the UV radiation field by using the fact that PAH emissivity is roughly constant per unit H and unit radiation field (Draine & Li 2007). In practice, we used the WISE (Meisner & Finkbeiner 2014) 12 μm maps divided by the column density, and clipped to a maximum value of 10^{22} cm^{-2}.

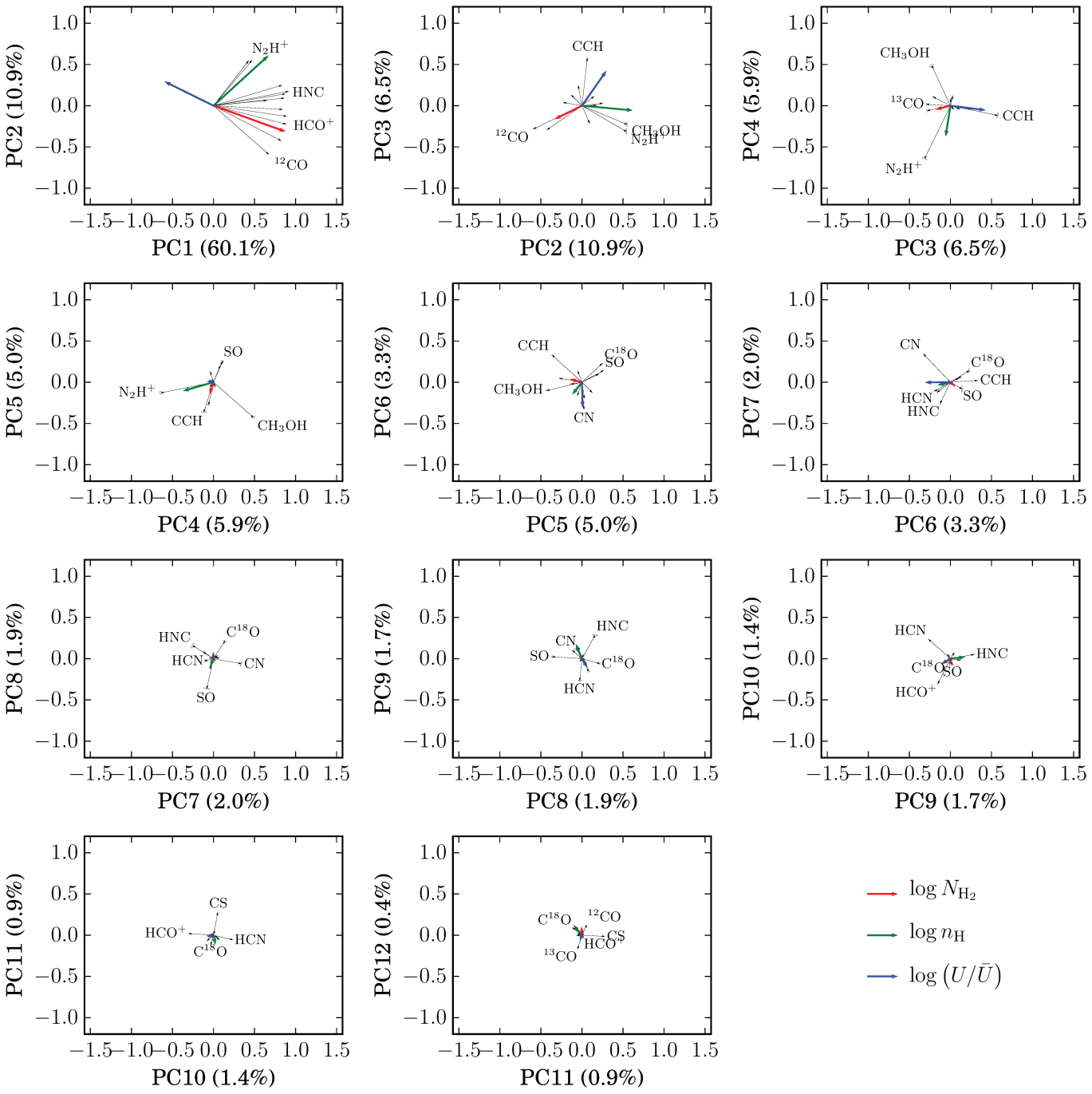


Fig. 7. Correlation wheels, showing the initial line intensities as vectors having as coordinates their correlation coefficients to each PC, represented in the planes of successive pairs of PCs. Uncertainties from our bootstraping analysis (see Sect. 4.3) are presented as thin black contours around the arrows' heads (isodensity contours containing 68% of the distribution). Also represented in colored arrows are the correlations of our independent physical parameters with the PCs (red: $\log N_{H_2}$, green: $\log n_H$, blue: $\log(U/\bar{U})$).

We do not claim to have an absolute value of the UV radiation field but a quantity that should be proportional to it. The quantity $\log(U/\bar{U})$ where $\bar{U}$ is the mean value of U is shown in the rightmost panel of Fig. 9.

The proper way to compute the UV radiation field from PAH emission would be to divide by the volume density but as we discussed in the previous paragraph, it is not possible to get a full map of volume density. We chose to use column density as a proxy for volume density even though it entails strong constrains on the spacial distribution of the gas along the line of sight. Since we are interested in relative variation of density and not absolute values it is sufficient to assume that the matter is clustered into clouds that are of similar spatial extents.

5.2. Correlation of principal component maps with physical parameters

We computed the Spearman's rank correlation coefficient between each pair of principal component maps and physical parameters maps. We used Spearman's rank correlation instead of the Pearson linear correlation coefficient because the potential relations between the principal components and the physical

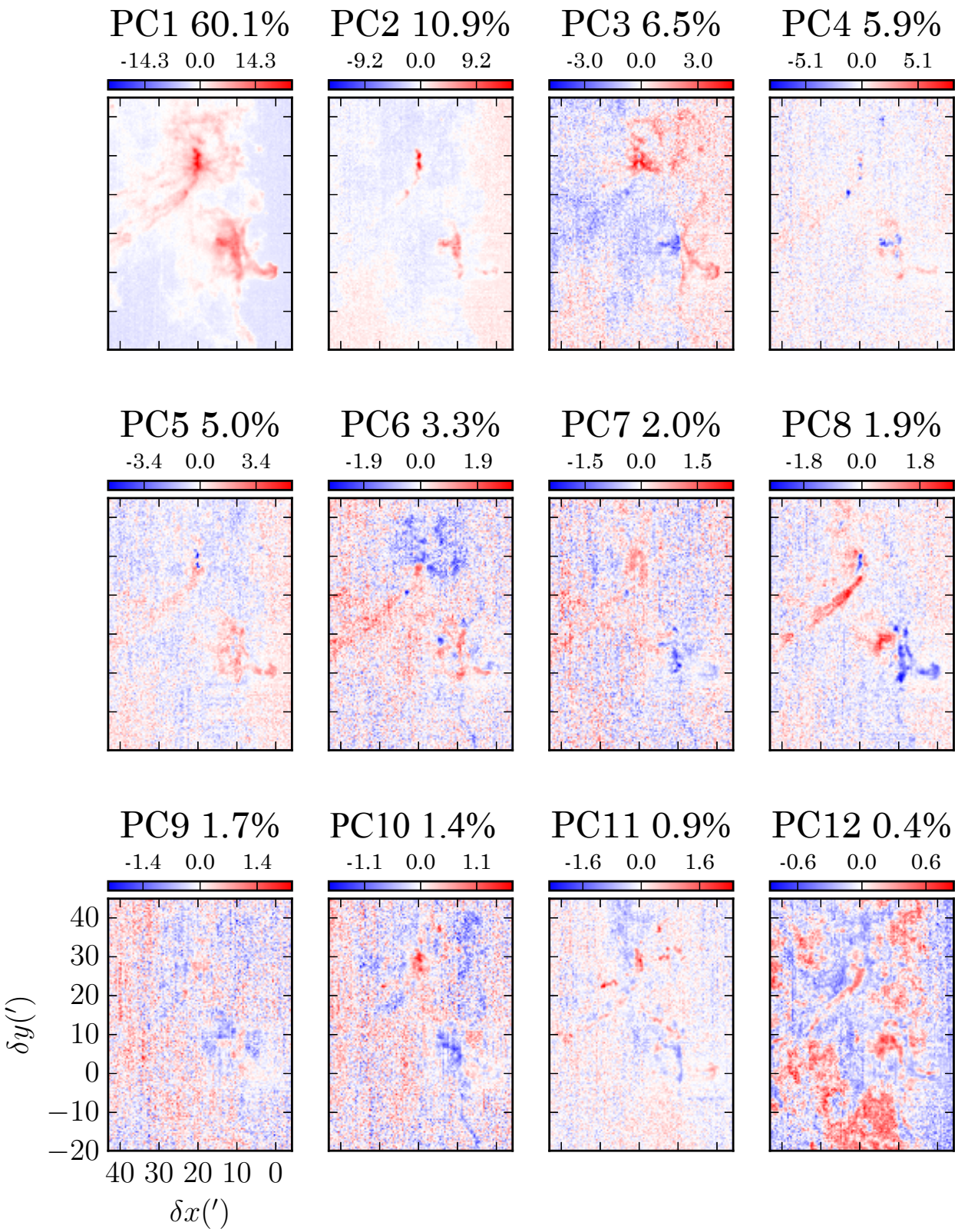


Fig. 8. Principal component maps. These maps represent the value of each observed pixel when they are projected in the space of the principal components.

parameters are most certainly non linear in nature. The rank coefficient used is only sensitive to the ordering of the values and is thus not affected by the possible non-linearities of the correlation. Table 2 summarizes all theses values and Fig. 10 shows the scatter plots for the most significant correlations discussed in the next paragraphs. An alternative way of exploring the correlations between the independent physical parameters and the PCs is to represent the correlation between each physical parameters and the PCs in the correlation wheels of Fig. 7.

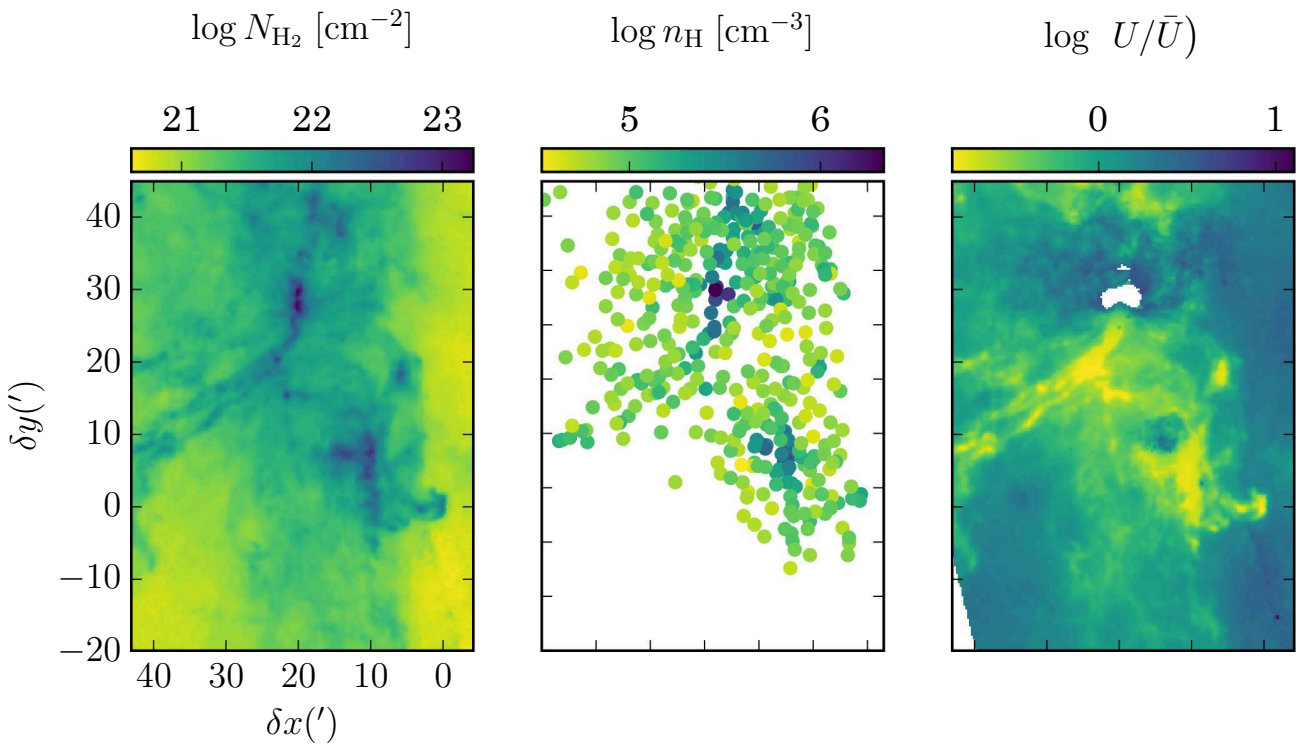


Fig. 9. Maps of the independently measured physical parameters, H_2 column density (*left*), volumic density (*middle*), UV illumination (*right*).

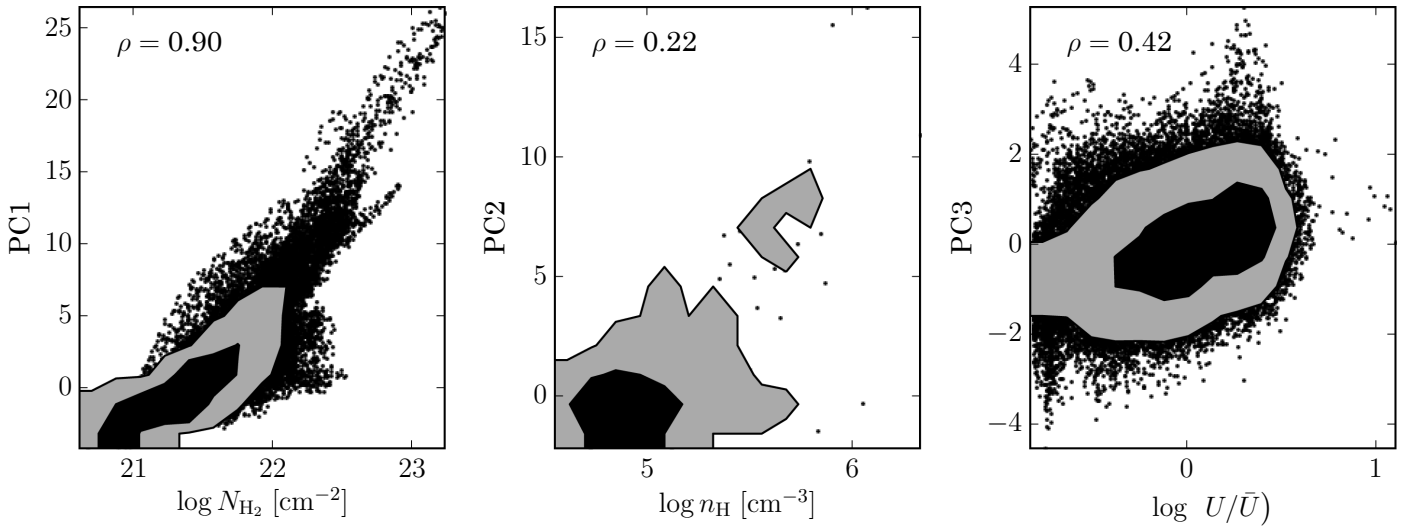


Fig. 10. Scatter plots of the first three principal components with the independent physical parameters. Contours in black and gray correspond to 68% and 95% of the samples respectively.

For this analysis, it must be kept in mind that while the principal components are necessarily uncorrelated, the physical parameters considered here are correlated: $N(H_2)$ is an integral of n_H along the line of sight and the two are thus strongly correlated, whereas U is inversely proportional to $N(H_2)$ by construction and they are thus anticorrelated. As a result, the principal components will tend to represent the uncorrelated part of the variations of the underlying physical parameters.

Column density: The component map showing the highest correlation coefficient with $N(H_2)$ is PC1. Spearman's rank correlation coefficient is extremely high at 0.90, and the scatter plot (Fig. 10, left panel) shows a strongly linear relation between PC1 and $\log(N(H_2))$.

Because PC1 is the first PC (axis of largest variation), it is unaffected by the decorrelation constraint that affects the other PCs. This first principal component can thus be interpreted as a global measure of total column density, as suspected in our previous discussion. Since n_H and U are positively and negatively correlated with $N(H_2)$, respectively, these physical parameters exhibit relatively strong positive and negative correlations with PC1.

Volume density: The PC most correlated to n_H is also PC1, due to the large correlation between $N(H_2)$ and n_H. The next

Table 2. Spearman's rank correlation coefficient between the principal components and the physical parameters.

	$\log N_{H_2}$	$\log n_H$	$\log(U/\bar{U})$
PC1	0.90	0.43	−0.66
PC2	−0.57	0.22	0.43
PC3	−0.20	0.06	0.42
PC4	−0.01	−0.23	−0.04
PC5	−0.16	0.02	0.09
PC6	0.04	−0.07	−0.26
PC7	−0.02	−0.03	0.00
PC8	−0.02	0.05	0.12
PC9	−0.02	0.06	−0.11
PC10	−0.06	−0.10	−0.07
PC11	−0.03	−0.13	−0.04
PC12	0.04	0.11	−0.05

principal components most correlated with our limited sample of volume density measurements are PC2, which shows a Spearman's rank correlation coefficient of 0.22, and PC4, with a Spearman's rank correlation coefficient of −0.23.

As was discussed in Sect. 4, PC2 and PC4 both trace chemical differences typical of dense cores. PC2 and PC4 can thus be interpreted as indicator of the presence of dense cores. We note that this comparison was only done with a limited sample of rather dense clouds. We can thus only say that PC2 traces increased density among dense clouds. Because of the opposite sign of the correlation of the density with PC4, negative values of this PC probably trace an even higher density regime. As noted before, the behavior of PC2 in less dense region is probably anticorrelated with density, and these PCs are thus only indicative of density in the high density regime.

Radiation field: For the radiation field, the most correlated PC is again PC1 (negative correlation) because high column density tends to result in highly shielded gas. Not considering PC1 and PC2, the third principal component shows the highest correlation with our estimation of the radiation field, with a Spearman's rank coefficient of 0.42. It thus describes the part of the radiation field variations that are not correlated with the cloud column density. As a result, PC3 highlights the part of the cloud where specific sources cause increased illumination. A strong positive correlation (0.43) with PC2 is also found, which is most likely an artifact due to the positive values of PC2 in the diffuse regions surrounding the molecular cloud (where most of the lines involved in PC2 are undetected, making PC2 irrelevant). PC6 also has a significant correlation with the radiation field (−0.26). These results thus confirm our previous discussion of PC3 and PC6.

These results can be inferred graphically from the correlation wheels of Fig. 7, in which the colored arrows tracing the location of $N(H_2)$ n_H, and U in the PC space have a significant size only for the first four PCs. Furthermore, each arrow is roughly aligned with one of the PCs: PC1 for $N(H_2)$, PC2 for n_H and PC3 for U.

6. Discussion

6.1. Comparison with other works

PCA has been extensively used in astronomy as a multivariale analysis tool starting from the work of Deeming (1964) on the classification of stellar spectra. Its use for the study of molecular maps of the ISM is more recent, starting with the work of Ungerechts & Thaddeus (1987). Notable studies include Neufeld et al. (2007), Lo et al. (2009), Melnick et al. (2011), and Jones et al. (2012).

We first discuss the common points between these studies. On a technical aspect, all these studies apply only subtraction by mean and normalization by variance, and do not attempt to introduce a non linear reparametrization of the observed intensities. The effect of noise is considered by limiting the number of observed lines to the set of brightest tracers (Lo et al. 2009) and masking regions of low emission (Jones et al. 2012). All of these studies identify the utility of PCA as a means of studying the correlations between molecular lines by studying the commonality between tracers and their variation, and as a tool to identify regions interesting for further study. With the notable exception of Ungerechts & Thaddeus (1987), rarely a discussion is made relating the principal components with the underlying physical parameters of the ISM. However, specific correlations or anticorrelations are often discussed in a chemical view or by invoking opacity effects (Lo et al. 2009).

Ungerechts et al. (1997) present a dataset of 360 spatial points in 32 lines of 20 chemical species including isotopologues toward the Orion A molecular cloud with the 14 m FCRAO telescope. Using PCA they show that the chemical abundances of most species stay similar for the Orion ridge, and that the main differences stand up for the BN-KL region. They note that their first three PCs contain 80% of the observed correlation and they use their component maps mainly to identify regions for further astrochemical study. They nevertheless discuss that data mostly lie in a 3D space spanned by the first three PCs because the molecular emission probably depends on three physical parameters of the ISM, namely the column density, volumic density, and gas temperature. Melnick et al. (2011) compared the distribution of the ground-state transition of water vapor with that of the ground state transition of N_2H^+, CCH, HCN, CN, and $^{13}CO(5-4)$. Water vapor is found to best correlate with species like $^{13}CO(5-4)$ and CN, tracing the cloud surface up to a few magnitudes of extinction, and is poorly correlated with N_2H^+ tracing the shielded regions. Using MOPRA, Jones et al. (2012) have mapped the central molecular zone (CMZ) near the center of the Galaxy in 20 spectral lines in the 85.3 to 93.3 GHz range. They performed a PCA analysis using the strongest eight lines (HCN, HCO^+, HNC, HNCO, N_2H^+, SiO, CH_3CN, and HC_3N) in the restricted area around SgrB2 and SgrA, where the N_2H^+ line is stronger than 10 K km s^{-1}. The analysis recovers the overall similarity of the line maps. The main differences are found in the SgrA and SgrB2 cores between the bright lines HCN, HNC, HCO^+, and the other species, and is attributed by Jones et al. (2012) to a difference in opacity. The other PCA components reveal specific regions where CH_3CN, HNCO, and SiO abundances are enhanced, possibly due to shocks or hot cores. Lo et al. (2009) studied the G333 molecular cloud with MOPRA. The PCA is performed on eight molecular lines with high S/N ratio (^{13}CO, $C^{18}O$, CS, HCO^+, HCN, HNC, N_2H^+, and CCH). The PCA analysis reveals differences between the regions traced by CCH and N_2H^+. In star forming regions it also reveals an anticorrelation between ^{13}CO, $C^{18}O$, and N_2H^+, and between N_2H^+ and HCO^+. PCA is also used by Neufeld et al. (2007) to separate the different regions impacted by supernovae shock waves.

While the analysis discussed in the previous paragraph used integrated line intensities, PCA has also been used on spectral line profiles as a mean to extract information on the

spatial properties of the turbulence (Heyer & Peter Schloerb 1997; Roman-Duval et al. 2011; Brunt & Heyer 2013), to study line absorption depth (Neufeld et al. 2015), or to measure cloud properties (Rosolowsky & Leroy 2006). To our knowledge no PCA analysis takes into account the full velocity profile of the molecular emission at every spatial position. Further inquiry on this subject is required as it can add a further dimension, namely the shape of the line profiles, to study the emission correlations.

6.2. Non-linearities and multiple physical regimes

Two important properties of PCA must be kept in mind. The first is that it is a linear method. It distinguishes the axes of variations in the dataset as linear combinations of the initial variables. Thus, non-linear (approximate) relations between the variables cannot be properly captured. In this case, a single relation could be described by several PCs, one describing the best linear approximation, and additional PCs describing directions in which the non-linear relation deviates from linearity. Using our non-linear transform, equivalent to a logarithm for high S/N values, alleviates part of the problem because it allows us to describe power-law relations. However, other non-linearities (such as saturation for the ^{12}CO line) are still not captured. Non-linear extensions of PCA exist (kernel-PCA, neural network-based dimensionality reduction such as the self-organizing maps), but their results tend to be harder to interpret.

The second property is that PCA is based on the global correlation matrix of the data. If different physical regimes are present in the dataset, each with different relations between the variables, PCA will provide a global (linear) approximation including the different regimes. Depending on the fraction of samples (pixels) representing the different regimes, it may give more weight to some regimes than others, neglect some regimes, or mainly represent one of the regimes.

6.3. Reduction of dimensionality

PCA is often used as a dimensionality-reduction tool, by keeping only a subset of the PCs that account for a sufficiently large fraction of the variance in the dataset. We saw that PC1 to 6 explain more than 90% of the correlation structure of the data. Moreover, PC6 defines a transition between several PCs with similar levels, and others with different levels (5% for PC3−4−5, 1−2% for the PCs after 6). The projection on the first six PC thus define a six-dimensional hyperplane in which the data is approximately embedded. However, we saw some striking spatial features appearing in later PCs, indicating meaningful axes of variations, such as PCs 8 and 11. The pattern of small scale spots shown by PC11 (corresponding to overbright CS) is particularly interesting, and its late apparition in the decomposition could simply be a consequence of the small fraction of pixels concerned. Thus, even PCs with lower fractions of explained correlation can contain important information, such as specific variations occurring in small regions only.

6.4. A synthetic view of Orion B

Using the physical interpretation of the principal components derived from the previous section it is possible to derive a synthetic view of the Orion B cloud rendered through a color image (see Fig. 11). The principal components 1 (column density), 2 (volume density) and 3 (UV illumination) are used in the following way: the column density is used to encode the luminosity, and the volume density and UV illumination are combined orthogonally to define a color,

$$\text{hue} = \text{atan}(uv, \text{density}). \quad (2)$$

In this way, it is possible to identify, by color only, the physical properties associated with every line of sight.

Most of the region is composed of low density gas either obscured (green) or UV illuminated (yellow). Notable features are the moderately dense (orange) photodissociation regions that are present as the surface of pillars (e.g. around the Horsehead nebula) and as globules surrounding the NGC 2024 massive H II region in the upper part of the map. A sharp illumination gradient is visible at the base of the neck of the Horse with transition from illuminated (yellow) to shielded (green) gas.

Concerning the dust lane in front of NGC 2024, there is a clear sharp frontier between the northern and southern part, the north being strongly UV illuminated (yellow, orange, and red), the southern part much more obscured (cyan and green). The variation in the density of dense cores are visible with transitions from moderately dense (cyan) to higher density (dark blue) gas.

7. Conclusion

To study the correlations between maps of the emission of 12 bright lines belonging to the 3 mm band over the south-western edge of the Orion B molecular cloud, we applied the PCA to these data. Before this analysis, we applied a non-linear transformation that is close to linear around zero and is equivalent to a logarithmic transform at large values. The goal of this non-linear transform is two-fold.

Firstly, although ratios of brightness temperatures are easier to interpret, PCA assumes that the relations in the input data set are linear. Applying the logarithm to the input data allows us to transform ratios of brightness temperatures into subtractions well adapted to a linear analysis.

Secondly, signal is only detected on a line-dependent subset of the field of view. Applying the logarithm to noisy brightness temperatures centered around zero is mathematically ill-defined. Having a linear transform around zero solves this problem. We tuned the transition value between the linear and logarithmic value that is typically eight times the typical noise value of the dataset. We showed that the results are not very sensitive to this value.

The PCA delivers a set of maps that are a linear combination of the input brightness temperatures, taking into account their (anti-)correlations. Although PCA does not use the spatial information of the input dataset, the output maps expose well-defined structures. We thus limited our analysis to the first few principal components that expose the largest correlations present in the initial dataset. The analysis of these correlations allowed us to propose links between the first three components and physical parameters, in this case the column density, volume density, and UV radiation field. We quantified these links by computing the correlation coefficients of these principal components with independent measurements of the column density, volume density, and UV illumination. The first principal component is highly correlated to the column density measured from the dust extinction and has positive contributions from all molecules, as has been noted in Pety et al. (2016). The third principal component is well correlated to our estimation of the UV illumination, with positive contributions from CCH, CN and anticorrelations with N_2H^+ and CH_3OH. The second principal component is correlated with the volume density in the dense cores having

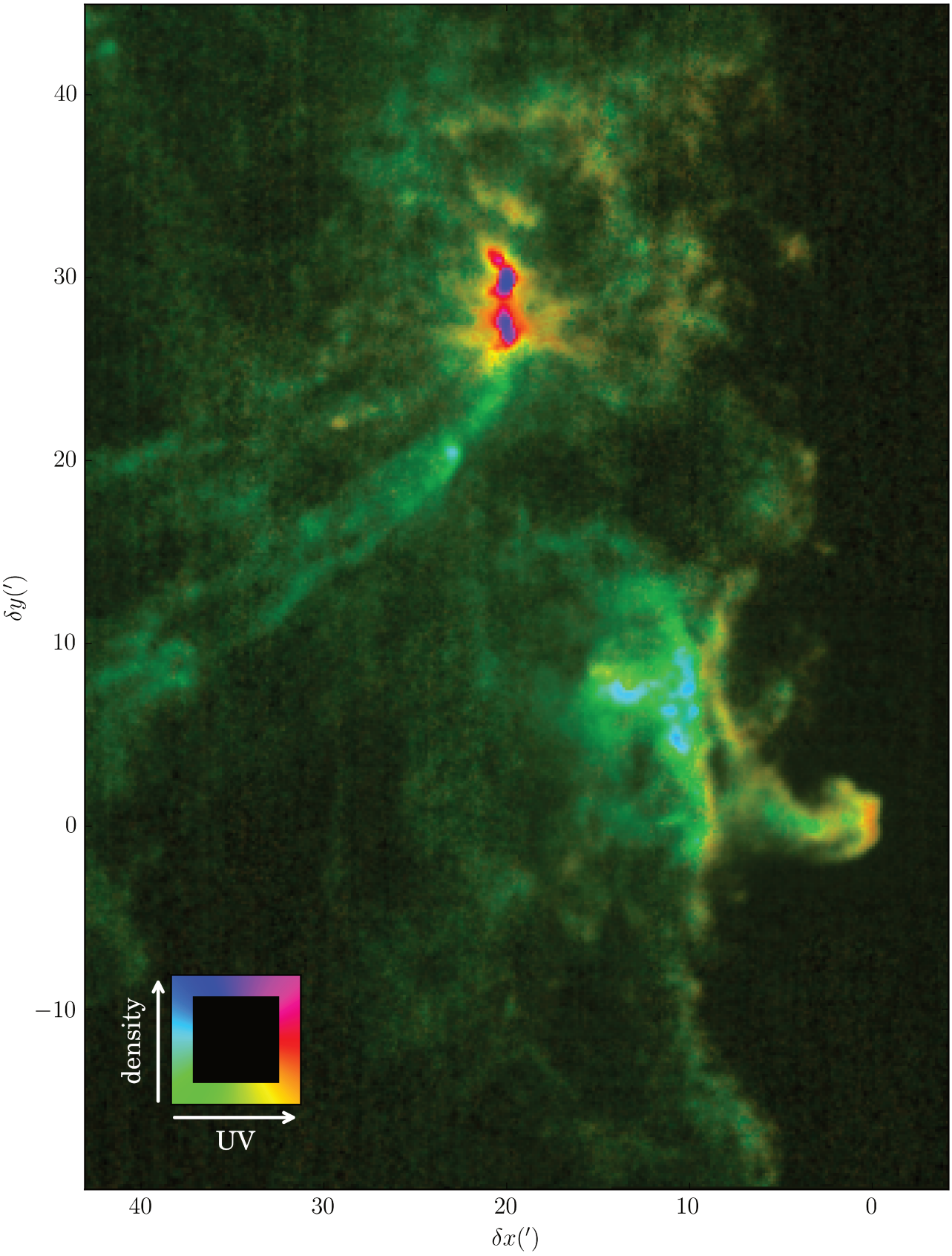


Fig. 11. Synthetic view of the Orion B molecular cloud. In this colormap, the intensity of each pixel is encoded by PC1 (column density) and the hue is encoded by the angle of the vector constructed using two orthogonal components PC2 (volume density) and PC3 (UV radiation field). It is possible to identify limiting cases. Magenta: dense PDR, yellow: diffuse PDR, green: diffuse non illuminated, blue: dense non illuminated.

a combined positive contribution from N_2H^+ and CH_3OH and a negative contribution from ^{12}CO and ^{13}CO.

The possibility of linking linear combinations of the brightness temperatures of a set of 3 mm lines to physical parameters such as the column density, volume density, or UV illumination opens an interesting avenue to analyze the large spectro-imaging data sets that (sub)-mm radioastronomy starts to produce. As PCA analysis only works on the brightness temperatures independent of their spatial relations, it also offers an easy possibility to compare with large grids of detailed 1D models of photo-dissociation regions. In future papers, we will continue to explore this with more advanced decomposition techniques that may take into account missing values, noise effects, or non-linear relations in the input dataset.

Acknowledgements. This work was supported by the French program Physique et Chimie du Milieu Interstellaire (PCMI) funded by the Conseil National de la Recherche Scientifique (CNRS) and Centre National d'Études Spatiales (CNES). We thank the CIAS for its hospitality during the two workshops devoted to this project. P.G. thanks ERC starting grant (3DICE, grant agreement 336474) for funding during this work. P.G.'s current postdoctoral position is funded by the INSU/CNRS. NRAO is operated by Associated Universities Inc. under contract with the National Science Foundation. We thank P. Andre and N. Schneider to kindly give us access to the *Herschel* Gould Belt Survey data. This research has made use of data from the *Herschel* Gould Belt survey (HGBS) project (http://gouldbelt-herschel.cea.fr). The HGBS is a *Herschel* Key Programme jointly carried out by SPIRE Specialist Astronomy Group 3 (SAG 3), scientists of several institutes in the PACS Consortium (CEA Saclay, INAF-IFSI Rome and INAF-Arcetri, KU Leuven, MPIA Heidelberg), and scientists of the *Herschel* Science Center (HSC). We than the anonymous referee for his/her constructive comments.

References

Agúndez, M., & Wakelam, V. 2013, Chem. Rev., 113, 8710
André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
Brunt, C. M., & Heyer, M. H. 2013, MNRAS, 433, 117
Deeming, T. J. 1964, MNRAS, 127, 493
Draine, B. T., & Li, A. 2007, ApJ, 657, 810
Feigelson, E. D., & Babu, G. J. 2012, Modern Statistical Methods for Astronomy (Cambridge: Cambridge University Press)
Heyer, M. H., & Peter Schloerb, F. 1997, ApJ, 475, 173
Hotelling, H. 1933, J. Educ. Psych., 24
Ilin, A., & Raiko, T. 2010, J. Mach. Learn. Res., 11, 1957
Ivezić, Ž., Connolly, A., Vanderplas, J., & Gray, A. 2014, Statistics, Data Mining and Machine Learning in Astronomy (Princeton University Press)
Jolliffe, I. 2002, Principal Component Analysis (New York: Springer Verlag)
Jones, P. A., Burton, M. G., Cunningham, M. R., et al. 2012, MNRAS, 419, 2961
Kirk, H., Di Francesco, J., Johnstone, D., et al. 2016, ApJ, 817, 167
Le Bourlot, J., Le Petit, F., Pinto, C., Roueff, E., & Roy, F. 2012, A&A, 541, A76
Lo, N., Cunningham, M. R., Jones, P. A., et al. 2009, MNRAS, 395, 1021
Meisner, A. M., & Finkbeiner, D. P. 2014, ApJ, 781, 5
Melnick, G. J., Tolls, V., Snell, R. L., et al. 2011, ApJ, 727, 13
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Neufeld, D. A., Hollenbach, D. J., Kaufman, M. J., et al. 2007, ApJ, 664, 890
Neufeld, D. A., Godard, B., Gerin, M., et al. 2015, A&A, 577, A49
Orkisz, J. H., Pety, J., Gerin, M., et al. 2017, A&A, 599, A99
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Lear. Res., 12, 2825
Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98
Roman-Duval, J., Federrath, C., Brunt, C., et al. 2011, ApJ, 740, 120
Rosolowsky, E., & Leroy, A. 2006, PASP, 118, 590
Scargle, J. D., Norris, J. P., Jackson, B., & Chiang, J. 2013, ApJ, 764, 167
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Ungerechts, H., & Thaddeus, P. 1987, ApJS, 63, 645
Ungerechts, H., Bergin, E. A., Goldsmith, P. F., et al. 1997, ApJ, 482, 245
Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47
Ward-Thompson, D., Nutter, D., Bontemps, S., Whitworth, A., & Attwood, R. 2006, MNRAS, 369, 1201

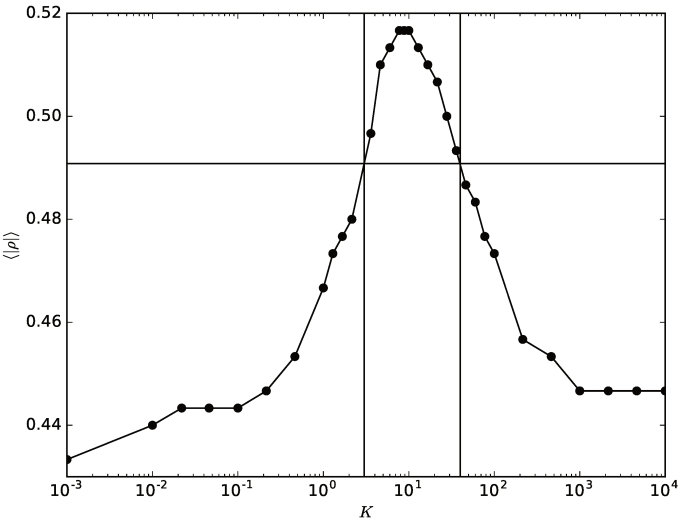


Fig. A.1. Variation of $\langle|\rho|\rangle$, the mean of the absolute values of the Spearman's correlation coefficients as a function of K, with $a = K\text{median}(\sigma)$. The optimal value is found for $K = 8$.

Appendix A: Optimal value of *a* in the asinh reparametrisation

The only free parameter in the asinh reparametrization is a, the parameter which marks the boundary between the linear and logarithmic regimes of the asinh function (see Fig. 3). As shown in Table 1 the noise across different lines is similar and we express a as the product of a constant factor K by the median noise 0.08 K. The quantity we choose to maximize is the mean of the absolute value of the correlation coefficient of the principal components with the physical maps $N(H_2)$, U, and n_H, we note this quantity $\langle|\rho|\rangle$. Figure A.1 shows the evolution of this quantity with increasing values of K, a maximum value of $\langle|\rho|\rangle$ around $K = 8$ although an acceptable range of K values (reduction of $\langle|\rho|\rangle$) by less than 5%) spans values from 3 to 40.

A&A 610, A12 (2018)
DOI: 10.1051/0004-6361/201731833

Astronomy & Astrophysics

Clustering the Orion B giant molecular cloud based on its molecular emission⋆

Emeric Bron[1,2], Chloé Daudon[3], Jérôme Pety[4,3], François Levrier[3], Maryvonne Gerin[3], Pierre Gratier[5], Jan H. Orkisz[6,4,3], Viviana Guzman[7], Sébastien Bardeau[4], Javier R. Goicoechea[1], Harvey Liszt[8], Karin Öberg[9], Nicolas Peretto[10], Albrecht Sievers[4], and Pascal Tremblin[11]

[1] ICMM, Consejo Superior de Investigaciones Cientificas (CSIC), 28049 Madrid, Spain
[2] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, 92190 Meudon, France
e-mail: emeric.bron@obspm.fr
[3] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, École normale supérieure, 75005 Paris, France
[4] IRAM, 300 rue de la Piscine, 38406 Saint Martin d'Hères, France
[5] Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[6] Univ. Grenoble Alpes, IRAM, 38000 Grenoble, France
[7] Joint ALMA Observatory (JAO), Alonso de Cordova 3107 Vitacura, Santiago de Chile, Chile
[8] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[9] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[10] School of Physics and Astronomy, Cardiff University, Queen's buildings, Cardiff CF24 3AA, UK
[11] Maison de la Simulation, CEA-CNRS-INRIA-UPS-UVSQ, USR 3441, Centre d'étude de Saclay, 91191 Gif-Sur-Yvette, France

Received 25 August 2017 / Accepted 2 October 2017

ABSTRACT

Context. Previous attempts at segmenting molecular line maps of molecular clouds have focused on using position-position-velocity data cubes of a single molecular line to separate the spatial components of the cloud. In contrast, wide field spectral imaging over a large spectral bandwidth in the (sub)mm domain now allows one to combine multiple molecular tracers to understand the different physical and chemical phases that constitute giant molecular clouds (GMCs).
Aims. We aim at using multiple tracers (sensitive to different physical processes and conditions) to segment a molecular cloud into physically/chemically similar regions (rather than spatially connected components), thus disentangling the different physical/chemical phases present in the cloud.
Methods. We use a machine learning clustering method, namely the Meanshift algorithm, to cluster pixels with similar molecular emission, ignoring spatial information. Clusters are defined around each maximum of the multidimensional probability density function (PDF) of the line integrated intensities. Simple radiative transfer models were used to interpret the astrophysical information uncovered by the clustering analysis.
Results. A clustering analysis based only on the $J = 1$–0 lines of three isotopologues of CO proves sufficient to reveal distinct density/column density regimes ($n_{\rm H} \sim 100\,{\rm cm^{-3}}$, $\sim$500 cm^{-3}, and >1000 cm^{-3}), closely related to the usual definitions of diffuse, translucent and high-column-density regions. Adding two UV-sensitive tracers, the $J = 1$–0 line of HCO^+ and the $N = 1$–0 line of CN, allows us to distinguish two clearly distinct chemical regimes, characteristic of UV-illuminated and UV-shielded gas. The UV-illuminated regime shows overbright HCO^+ and CN emission, which we relate to a photochemical enrichment effect. We also find a tail of high CN/HCO^+ intensity ratio in UV-illuminated regions. Finer distinctions in density classes ($n_{\rm H} \sim 7 \times 10^3$ cm^{-3}, $\sim 4 \times 10^4$ cm^{-3}) for the densest regions are also identified, likely related to the higher critical density of the CN and HCO^+ (1–0) lines. These distinctions are only possible because the high-density regions are spatially resolved.
Conclusions. Molecules are versatile tracers of GMCs because their line intensities bear the signature of the physics and chemistry at play in the gas. The association of simultaneous multi-line, wide-field mapping and powerful machine learning methods such as the Meanshift clustering algorithm reveals how to decode the complex information available in these molecular tracers.

Key words. astrochemistry – ISM: molecules – ISM: clouds – ISM: structure – methods: statistical – ISM: individual objects: Orion B

1. Introduction

The interstellar medium (ISM) is made of several physical/chemical phases: dense vs. diffuse gas, hot vs. cold gas, ionised, atomic, or molecular gas, far-UV (FUV)-illuminated vs. FUV-shielded gas, and gravitationally bound vs. free-floating gas. The ISM molecular composition is particularly sensitive to the changes that affect the gas and dust when they cycle between these different phases. Wide-field mapping of the line emission of many molecules sensitive to different physical processes could thus enable one to segment giant molecular clouds (GMCs) into regions belonging to distinct physical/chemical

⋆ Data products associated with this paper are available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/610/A12 and at http://www.iram.fr/~pety/ORION-B

phases, so that we can then study the physics and chemistry of these regions in detail. Moreover, understanding the prevalence of these different phases inside a given molecular cloud, and learning how to disentangle their relative contributions to each molecular tracer is of interest to interpret the spatially unresolved molecular emission in extragalactic studies.

The advent of wide-band high-spectral-resolution spectrometers associated to future multi-beam receivers in the (sub-)millimetre domain will enable radio-astronomers to easily map the emission of many lines from tens of species over tens of square degrees on the sky. The pioneer ORION-B project (Outstanding Radio-Imaging of OrioN-B, PI: J. Pety and M. Gerin) currently uses the IRAM-30m/EMIR spectrometer to image about 4.5 square degrees of the Southern part of the Orion B molecular cloud at typical spectral resolutions of 0.6 km s^{-1} and an angular resolution of 26″ (i.e.∼50 mpc or ∼10^4 AU at the distance of Orion B: 400 pc, Menten et al. 2007; Schlafly et al. 2014) and a typical sensitivity of 0.1 K over almost all of the 3 mm atmospheric window.

This paper is part of the first series of papers based on the already acquired dataset that covers about 1 square degree of the Orion B molecular cloud surrounding the Horsehead nebula, NGC 2023, and NGC 2024, in the 84−116 GHz frequency range. Pety et al. (2017) introduce the molecular anatomy of the Orion B GMC, including relationships between line intensities and gas column density or FUV radiation field, and correlations between selected lines and line ratios. They obtain a dust-traced gas mass that is less than approximately one third of the CO-traced mass, using the standard $X_{\rm CO}$ conversion factor. The presence of overluminous CO can be traced back to the dependence of the CO intensity on the gas kinetic temperature, which in turn is affected by the FUV illumination (photo-electric heating). While most lines show some dependence on the UV radiation field, CN and C_2H are found to be the most sensitive. Moreover, dense cloud cores are almost exclusively traced by N_2H^+. Other traditional high-density tracers, such as HCN(1–0) or HCO^+(1–0), are also easily detected in extended translucent regions at a typical density of about 500 H cm^{-3}. Gratier et al. (2017) propose a first multi-line approach applying principal component analysis (PCA; Jolliffe 2002) on 12 of the brightest lines (integrated over a narrow velocity range) to reveal the pattern of correlations between the different tracers. This approach emphasises three clear trends: 1) the line intensities are well correlated with the column density, that is, the more matter along the line of sight, the brighter the lines; 2) CCH, CN, HCN are correlated with the FUV (<13.6 eV) irradiation (while N_2H^+ and the CO isotopologues are anti-correlated); and 3) the PCA method confirms the known anticorrelation between N_2H^+ and CO in dense cores. Finally, the ^{13}CO position-position-velocity cube has been used by Orkisz et al. (2017) to show that solenoidal motions clearly dominate over the observed field of view, in agreement with the low star formation efficiency measured in Orion B (Lada 1992; Carpenter 2000; Megeath et al. 2016).

In this paper, we take a further step to characterise the different ISM phases from a multi-line wide-field dataset. The basic idea is similar to remote sensing in Earth studies (e.g. Inglada et al. 2017) which tries to classify environments (forests, deserts, mountains, oceans, etc.), based on the dominant colour they emit. In other words, we wish to segment the dataset into a small (yet unknown) number of classes that have a well defined physical or chemical meaning, based on their molecular emission. This goal requires the use of data-mining techniques in order to go beyond a tracer-by-tracer analysis, and take advantage of the full information hidden in the joint variations of the different tracers. Classification techniques are divided into two categories. Supervised ones use known examples of the desired classes to learn how to automatically classify new observations. They thus require a priori independent knowledge of the physical or chemical properties of the different ISM phases. These approaches will be explored in future papers. In contrast, clustering, which is an unsupervised technique, aims to reveal how the data points naturally group themselves into distinct clusters of points with similar properties, hinting at the existence of different physical or chemical regimes. This is the approach that we adopt in this paper.

Traditional segmentation approaches in GMC studies typically segment the map of emission of a single tracer into constitutive clumps (Stutzki & Guesten 1990; Williams et al. 1994; Rosolowsky & Leroy 2006; Colombo et al. 2015). These methods use in one way or another the topology (contiguity) of the emission in the position-position-velocity space, sometimes associated with additional physical properties such as the virial state. Their goal is thus to separate the spatial components of a GMC. In contrast, we here propose to work on multi-dimensional probability density distributions (PDFs) of the line integrated intensities. The PDF shape can indeed show distinct components, which can reveal distinct physical/chemical regimes, and which we want to automatically separate. For instance, in the Hertzprung-Russell (HR) diagram, different branches in the colour vs. magnitude plot correspond to different stages of stellar evolution (main sequence, giant branch, etc.). However, while recognising structure by eye is possible in two-dimensional (2D) datasets, direct visualisation of the data becomes difficult in higher dimension. Simple 2D projections for each pair of line intensities do not necessary reveal all the existing structure, and clustering algorithms become necessary. In the case of the ORION-B dataset, Gratier et al. (2017) show that understanding the physics and chemistry underlying the extended molecular line emission requires a multi-dimensional analysis of the data. To our knowledge, clustering a GMC based on its (multi-molecule) molecular emission similarity rather than spatial (or PPV) contiguity has never been done before.

This paper is organised as follows. In Sect. 2, we present the data used in our analysis. We then explain the clustering algorithm that we chose to use in Sect. 3. This clustering method is first applied to the most widely observed lines in the millimetre wave domain, that is, the $J = 1 \rightarrow 0$ lines of three CO isotopologues (^{12}CO, ^{13}CO, and C^{18}O) in Sect. 4. Section 5 then discusses the additional results obtained when adding the ground-state transitions of HCO^+ and CN to the analysis, whose intensities are known to be related to the FUV illumination. We discuss the benefits and limits of the method in Sect. 6. We present our conclusions in Sect. 7.

2. Data

2.1. IRAM-30m observations

Pety et al. (2017) present in detail the acquisition and data reduction of the dataset used in this study. In short, the data were acquired at the IRAM-30m telescope by the ORION-B project from August 2013 to November 2014. The frequency range from 84 to 116 GHz was completely sampled at 200 kHz spectral resolution. Such a large bandwidth allowed us to image over 20 chemical species at a median sensitivity of 0.1 K (main-beam temperature) per channel. As opposed to several small bandwidth mappings, the spectral lines in this survey are observed in only two tunings covering 16 GHz each. They are thus well

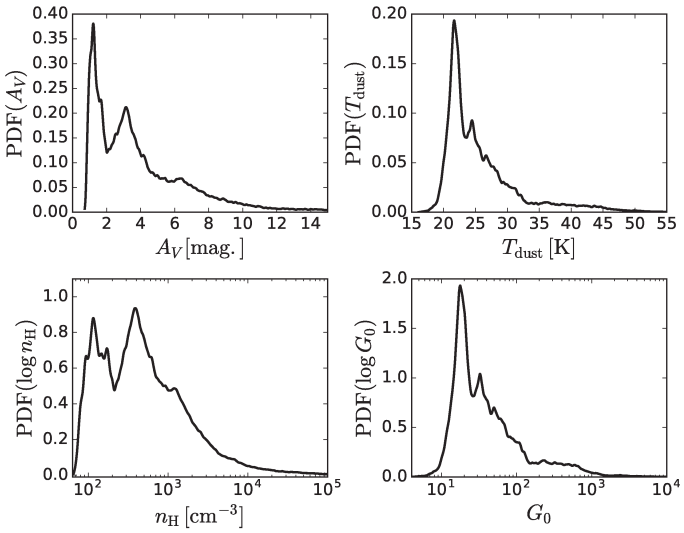


Fig. 1. PDFs of the dust visual extinction (*top left*), the dust temperature (*top right*), the deduced approximate gas volume density (*bottom left*), and the deduced FUV illumination (*bottom right*) in the observed field of view.

inter-calibrated, which gives an unprecedented spectral accuracy for such a large field of view. The intensity dynamic range reaches ~720.

The field of view presented covers 0.8° × 1.1° towards the Orion B molecular cloud part that contains the Horsehead nebula, and the HII regions NGC 2023, NGC 2024, IC 434, and IC 435. The angular resolution ranges from 22.5 to 30.5″. The position-position-velocity cubes of each line were smoothed at a common angular resolution of 31″ to avoid resolution effects during the comparison. At a distance of 400 pc (Menten et al. 2007), the sampled linear scales range from ~50 mpc to ~8 pc.

The observations provided a position-position-frequency cube of 315 × 420 × 160 000 pixels, each pixel covering 9″ × 9″ × 0.5 km s^{-1} (Nyquist sampling at the highest frequency, i.e. ^{12}CO(1–0) at 115.27 GHz).

We here study maps of line integrated intensities. Lines are detected over quite different velocity ranges. Using a large velocity range would artificially increase the noise for most of the lines, while adapting the velocity range to each line could bias the comparisons. We thus focus on a narrow velocity range where the bulk of the gas emits. While Gratier et al. (2017) computed for each line the mean of three 0.5 km s^{-1} velocity channels around 10.5 km s^{-1}, we use here a more common radioastronomical approach, which is to integrate the line intensity profile. We here integrate over the [9, 12 km s^{-1}] velocity interval where the main velocity component of the Orion B cloud sits (see Sect. 2.5 of Pety et al. 2017). To first order, both sets of maps are proportional to one another (the respective velocity ranges differ slightly).

2.2. Herschel observations

In order to get independent constraints on the physical conditions in the Orion B cloud, we use the dust continuum observations from the *Herschel* Gould Belt Survey (André et al. 2010; Schneider et al. 2013) and from the *Planck* satellite (Planck Collaboration 2011). The fit of the spectral energy distribution by Lombardi et al. (2014) gives us access to the spatial distributions of the dust opacity at 850 μm and of the dust temperature. As in Pety et al. (2017), we converted $\tau_{850\,\mu m}$ to visual extinctions using $A_V = 2.7 \times 10^4\, \tau_{850}$ mag. The top panels of Fig. 1 show the PDF of the dust visual extinction and temperature.

The A_V PDF shows three distinct peaks, indicating that the field of view samples three different regimes: A_V = 1–2, 2–6, and ≥6. These regimes are consistent with the usual distinction between diffuse, translucent, and high-column-density regions (Snow & McCall 2006). As in Pety et al. (2017), we use $N_H/A_V = 1.8 \times 10^{21}$ cm^{-2}/mag as conversion factor between visual extinction and hydrogen column density: $N_H = N_{HI} + 2N_{H_2}$. In addition, we propose a conversion from the column density to an approximate volume density map. The procedure is discussed in detail in Appendix A. In summary, we assume a rough isotropy of the cloud (similar dimensions along the line of sight and in the plane of the sky) to deduce an estimate of the average hydrogen density along each line of sight as follows. For a given column density value x, we consider the region where $N_H \geq x$. We then estimate the line-of-sight depth l of this region from its plane-of-the-sky surface S as $l \simeq \sqrt{S}$ (using our isotropy assumption). We finally assign the approximate volume density $n_H = x/l$ to the pixels where $N_H = x$.

The resulting approximate volume density PDF is shown in the bottom-left panel of Fig. 1. The three distinct A_V regimes correspond to three volume density regimes: one low-density peak close to 10^2 cm^{-3} corresponds to diffuse gas, a second peak covering the range 300–800 cm^{-3} is associated to the translucent gas, and a third smaller peak slightly above 10^3 cm^{-3} with a long tail extending up to a few 10^6 cm^{-3} corresponds to denser gas. The values found after our conversion are consistent with the usual orders of magnitude for diffuse, translucent, and denser gas. More quantitatively, Appendix A shows that our estimation of the volume density is valid in a statistical way with a bias of a factor of 3 at most and a typical scatter of one order of magnitude, when compared with volume density estimates from the literature that make different hypotheses. The deduced values of n_H are rough estimates that should not be trusted beyond order-of-magnitude comparisons. However, this method reproduces the observed range of densities fairly well, indicating that the shape of the PDF is also approximately correct.

The T_{dust} PDF shows a less marked multi-peak structure with a sharp first peak at ~22 K, a small secondary peak at ~25 K and a shallow third peak at ~27 K. A first steep tail extends up to ~33 K, followed by a second flatter tail (reaching values up to 100 K). These two tails are indicative of highly FUV-illuminated regions. Pety et al. (2017) converted the dust temperature map into an approximate map of the FUV radiation field G_0 in units of the Habing interstellar standard radiation field (ISRF; Habing 1968), using the simple approximation of Hollenbach et al. (1991)

$$G_0 = \left(\frac{T_{dust}}{12.2\,\mathrm{K}}\right)^5 . \quad (1)$$

Shimajiri et al. (2017) compared this estimation with another estimation directly using the far infra-red intensities at 70 and 100 μm. Both estimates agree within 30%. The PDF of log(G_0) is shown as the bottom right panel of Fig. 1 and is very similar to the dust temperature PDF (as the conversion is a simple power law). As mentioned before, the main peak is close to 20 times the ISRF while the tail extends up to several thousand times the ISRF. As for our estimate of n_H, the deduced values of G_0 should only be trusted at order-of-magnitude levels.

3. The Meanshift clustering method

From a mathematical viewpoint, the data are a set of points characterised by their two sky coordinates and the integrated intensities of D molecular lines. The dataset thus lies in a space of $D + 2$ dimensions. As discussed before, we aim to cluster the datapoints based on their molecular emission only, and not their spatial proximity. For clarity, we thus separate the data space into two parts: 1) the usual position space; and 2) the line space of dimension D spanned by the molecular line intensities. Clustering will only be done in the line space, ignoring the location of the pixels in the position space. After a non-exhaustive discussion of methods that segment the data based on their multi-dimensional PDF, we describe the Meanshift algorithm and the implementation used in this paper.

3.1. The two families of PDF-based clustering methods

We are interested in clustering methods that aim to separate components in the (multi-dimensional) PDF of the data. Two families of such methods can be defined. The first family assumes that the data PDF can naturally be decomposed into components of some given functional form, for example, Gaussian functions, controlled by free parameters. These methods are thus called parametric approaches. The complete dataset is assumed to be a mixture of several components, with the same functional form but different values for the control parameters. These algorithms are thus usually called mixture models (see Bishop 2006, Chap. 9). This approach has two main drawbacks. First, the number of free parameters tends to increase quickly with the dimension of the problem, resulting in a difficult and often degenerate optimisation. To alleviate this problem, restrictions are imposed on the free parameters. For instance, assuming Gaussian components and forcing them to all have the same scalar covariance matrix (thus forcing spherically symmetric clusters of equal size) yields the K-means algorithm, one of the most used clustering algorithms. Second, the assumption of a functional PDF form is a strong a priori that can bias the clustering when this form is inadequate for the studied data. More flexible functional forms reduce this problem but result in more free parameters. In this family, a compromise has to be made between the flexibility of the assumed functional PDF form and the number of free parameters.

The second family takes a data-driven approach, by defining clusters around local maxima of the data PDF. Each cluster is thus a region of high density in the line space, separated from the other clusters by regions of lower density. This definition has two advantages: 1) It allows to capture any shape of the PDF of the clusters (possibly curved and elongated); and 2) The number of clusters is determined automatically from the data. Data clusters must however create a maximum in the PDF to be detected. A small group of datapoints blended in the tail of another more common cluster might thus not be detected. The two most famous algorithms in this family take a different approach to finding the high-density regions in the line space. First, the DBSCAN algorithm (Ester et al. 1996) uses a graph-based approach to find high-density regions, but it assumes a similar density of points inside all clusters. Second, the Meanshift algorithm (Comaniciu & Meer 2002) searches for the maxima of the data PDF using a kernel-based approach. We choose to use the Meanshift approach, as it can detect clusters that have both different sizes and different densities. Moreover, its direct link to the data PDF eases the interpretation of the clusters. The following section describes this algorithm in detail.

3.2. The Meanshift algorithm

3.2.1. General description

The Meanshift algorithm (see Comaniciu & Meer 2002 for more details) associates each data point to the closest local maximum of some empirical estimate of the PDF. The algorithm iteratively climbs up the slope of the PDF starting from each of the datapoints. The set of datapoints converging to the same PDF maximum constitutes a cluster.

The algorithm is based on the same concept as the Kernel Density Estimate method (Rosenblatt 1956; Parzen 1962), which estimates the PDF of a random variable (here, the intensities) from one sample realisation. The kernel density estimator for a given set of N D-dimensional data points $\{\boldsymbol{x}_i\}_{1\leq i\leq N}$ is

$$f_{h,K}(\boldsymbol{x}) = \frac{C_h}{N}\sum_{i=1}^{N} K\left(\frac{\mathrm{dist}(\boldsymbol{x},\boldsymbol{x}_i)^2}{h^2}\right), \tag{2}$$

where $\mathrm{dist}(\boldsymbol{x},\boldsymbol{x}_i)$ is a distance in the line space between a given vector $\boldsymbol{x}$ and the vector $\boldsymbol{x}_i$ associated to the ith datapoint (both are vectors of line intensities), K is the smoothing kernel (a non-negative decreasing function of $\mathbb{R}^+$), h is the bandwidth of the smoothing and C_h a normalisation constant. The kernel often has a finite support $[0, 1]$, so that the estimation of the PDF at $\boldsymbol{x}$ is only based on the datapoints that are closer to $\boldsymbol{x}$ than the bandwidth h.

The Meanshift algorithm avoids the estimation of the PDF itself by directly estimating the PDF gradient with the same kernel smoothing approach. By taking the gradient of Eq. (2) in the case of an Euclidean distance, and noting $G(x) = -K'(x)$, which is a new kernel, we obtain

$$\boldsymbol{\nabla} f_{h,K}(\boldsymbol{x}) \propto f_{h,G}(\boldsymbol{x})\, \boldsymbol{m}_{h,G}(\boldsymbol{x}), \tag{3}$$

$$\text{where} \quad \boldsymbol{m}_{h,G}(\boldsymbol{x}) = \frac{\sum_{i=1}^{N} \boldsymbol{x}_i\, G\left(\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|_2^2}{h^2}\right)}{\sum_{i=1}^{n} G\left(\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|_2^2}{h^2}\right)} - \boldsymbol{x}, \tag{4}$$

which is called the Mean Shift vector as it gives the shift from the current position $\boldsymbol{x}$ to the mean of the datapoints weighted by the kernel G centered on $\boldsymbol{x}$. Equation (3) indicates that this Mean Shift vector gives an estimate of the relative gradient (the local PDF gradient divided by the local PDF).

The following iterative algorithm is applied, starting from each of the datapoints:

- Compute the Mean Shift vector $\boldsymbol{m}_{h,G}(\boldsymbol{x})$ at the current estimate $\boldsymbol{x}$ of the searched local maxima.
- Modify the current estimate by shifting it by the Mean Shift vector.

This algorithm converges to points where the PDF gradient estimate is zero, and that usually are local maxima due to its hill climbing nature. Convergence points that are closer to each other than the bandwidth h are then merged, and clusters are defined as the sets of datapoints that have converged to the same extremum. Data points lying close to local minima can sometimes stay stuck due to the associated null gradient, but the resulting unwanted clusters can easily be recognised by the very small number of datapoints they contain, and removed by assigning their datapoints to the closest clusters.

3.2.2. The FAMS implementation

In this article, we used the Fast Adaptive Mean Shift (FAMS) code described in Georgescu et al. (2003), and provided by the authors on their webpage[1]. This code implements several modifications to the Meanshift principle in order to improve the quality of the gradient estimate in low-density regions.

The quality of the PDF gradient estimate depends on the number of datapoints present inside the smoothing kernel. Consequently, using a fixed bandwidth results in insufficiently sampled gradient estimates in the tails of the data PDF. This leads to many artificial maxima being found in these tails, unless the bandwidth is increased. But the central parts of the PDF then risk being overly smoothed and significant maxima can be lost. To alleviate this problem, Comaniciu et al. (2001) proposed an adaptive-bandwidth Meanshift algorithm, in which a different bandwidth h_i is assigned to each data point $\boldsymbol{x}_i$ (h is replaced by h_i in Eq. (4)). The Adaptive Meanshift Algorithm uses the simplest method to choose the bandwidth for each datapoint: taking the distance to the kth-nearest neighbour to the datapoint as the bandwidth, so that there is always roughly the same number of datapoints inside the kernel. This number of neighbours k becomes a parameter of the method to be adjusted by the user (instead of the bandwidth value in the classical Meanshift algorithm). In other words, there is a trade-off between the sampling variance and the smoothing factor to estimate the PDF: an adaptive smoothing is applied so that low-density regions are smoothed on a larger scale than high-density regions. All estimates of the PDF gradient are then equally well sampled.

Second, the FAMS implementation uses a L_1 distance ($\mathrm{dist}(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{D} |x_i - y_i|$) instead of the usual Euclidean distance ($\mathrm{dist}(\boldsymbol{x},\boldsymbol{y}) = \sqrt{\sum_{i=1}^{D}(x_i - y_i)^2}$). This allows for an additional optimisation in high dimension (Georgescu et al. 2003); but it is only an approximation of the Meanshift algorithm, as Eq. (4) assumes a Euclidean distance. Finally, the kernel used in this implementation is $G(x) = (1 - x)^2$.

3.2.3. In practice

The main control parameter of the Meanshift algorithm is either the bandwidth value h when using the fixed bandwidth implementation or the number of nearest neighbours k when using the adaptative bandwidth version. This parameter controls the scale on which the PDF is smoothed in both cases. We use here the adaptive bandwidth version. In each studied case, k was varied from a few hundred to a few thousand, and adjusted to increase or decrease the number of clusters found. Any cluster that contains a number of pixels that is very small compared to k is merged with the closest cluster. This can happen when the Meanshift algorithm is stuck into a local minimum of the PDF.

The smoothing part uses a spherical kernel, so that the applied smoothing is isotropic, that is, it has the same absolute bandwidth in all directions. It is thus necessary to ensure that the variability of the dataset along the different dimensions (that is the intensity dynamic of each line) is comparable. Without a linear rescaling, either the variations of the faint lines would be smoothed out by a bandwidth adapted to the bright lines or the bright lines would drive the segmentation into many small clusters when using a bandwidth adapted to the faint lines. We thus standardised the dataset (we ensured that the intensity PDF of each line has a unit standard deviation) before applying the Meanshift algorithm. Additional non-linear transformations applied to the dataset before clustering it with Meanshift would in general modify the number and positions of PDF maxima and thus affect the results. As we wish to check the amount of physical/chemical information encoded in the line intensities, we chose to only linearly standardise the data.

The clustering of our $\sim 10^5$ data points with $D = 5$ and $k = 500-2000$, typically takes 10 to 20 h of computation on a single standard CPU in 2017. Taking into account that our data has some redundancy (Nyquist sampling), we tested the method on a decimated dataset before getting the final results on the full dataset.

Comaniciu et al. (2001) showed that adding spatial coordinates in addition to intensities when clustering images may smooth the resulting clusters. We did not use the spatial information present in our dataset, as our focus is on grouping pixels where the intensities are similar rather that pixels belonging to the same spatial structure. We will however use the spatial coherence of the clusters found as a consistency check of the results because some amount of physical/chemical similarity is expected between neighbouring pixels.

When discussing the results, we visualise the 2D PDFs of pairs of lines (comparing the contributions of the obtained clusters) using kernel density estimation (Rosenblatt 1956; Parzen 1962) from the `scikit-learn` Python package (Pedregosa et al. 2011), with an Epanechnikov kernel ($\propto 1 - x^2$, optimal in terms of mean squared error, Epanechnikov 1969). This implementation uses a fixed bandwidth. This is however only used as a visualisation tool, and is independent of our Meanshift clustering analysis.

4. The CO isotopologue emission enables us to separate the diffuse, translucent, and denser gas regimes

The ^{12}CO, ^{13}CO, and C^{18}O $J = 1$–0 lines are amongst the most observed radio lines in molecular clouds. At constant elemental ratios of the carbon isotopes, the naive chemical interpretation suggests that the relative abundances of these three species should be identical in all lines of sight of a GMC. Moreover, the critical densities of these three lines for collisional excitation with H_2 are similar ($\sim 2 \times 10^3$ cm^{-3}), implying similar excitation conditions. Differences in optical depths should therefore be the main factor governing the intensity ratios. In this section, we ask whether or not clustering the intensities of these lines can distinguish physical regimes known to happen in the studied field of view. We thus first apply the Meanshift algorithm to a dataset consisting of the maps of the $J = 1$–0 lines of ^{12}CO, ^{13}CO, and C^{18}O only ($N = 141\,050$, $D = 3$).

After several trials for the number of neighbours in the adaptive kernel width, we chose a compromise between avoiding picking up sampling fluctuations in the PDF as artificial maxima and smoothing out physical maxima of interest (see Appendix B for a discussion of this choice). We settled on $k = 1900$ neighbours. In order to understand and interpret the clustering, we present in the following the spatial distribution of the clusters as well as PDFs of 1D or 2D projections of the data (that is, PDFs of single lines or pairs of lines).

4.1. Spatial distribution

We find nine clusters, whose spatial distribution is shown in Fig. 2. For the following discussion, we name these clusters

[1] http://coewww.rutgers.edu/riul/research/code/AMS/index.html

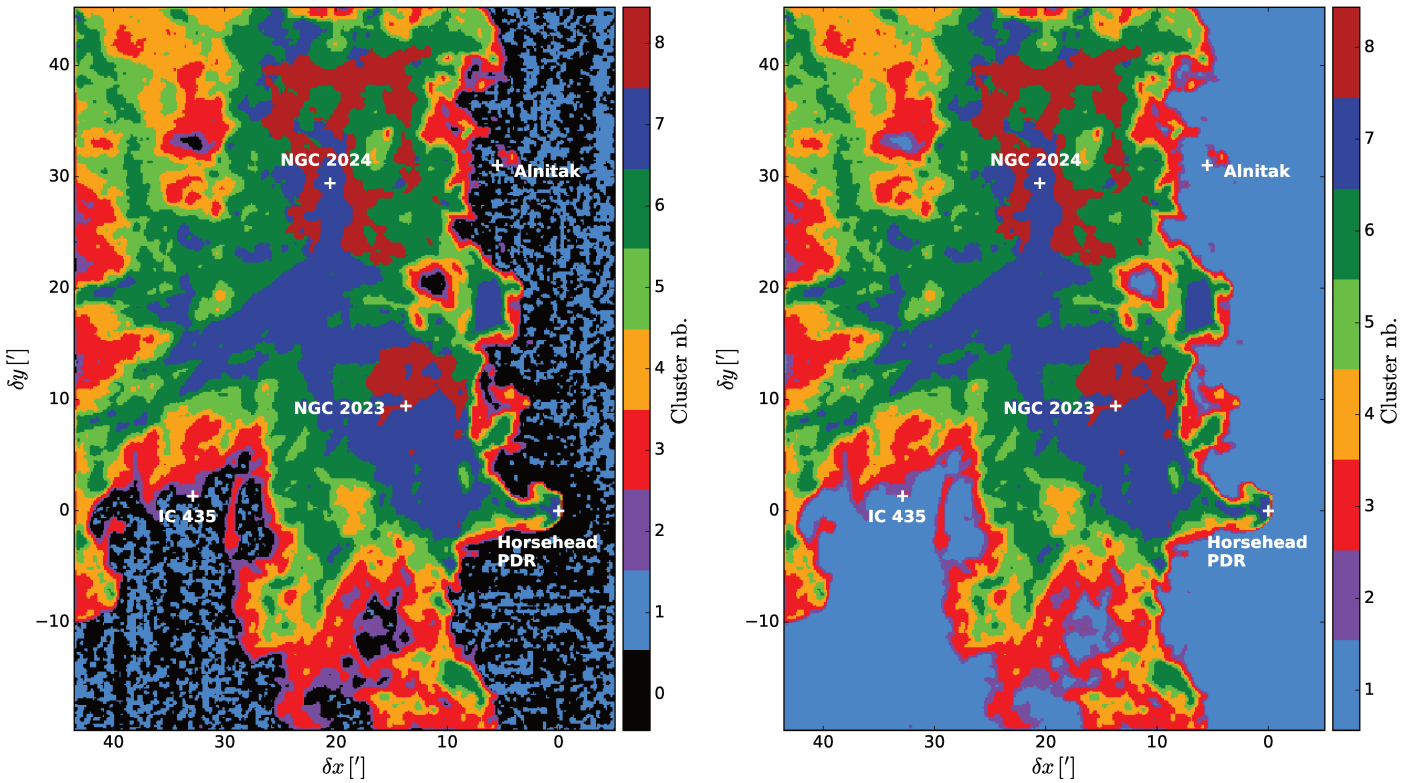


Fig. 2. Map of the clusters based on the ^{12}CO, ^{13}CO, and C^{18}O (1–0) line integrated intensities. White crosses mark the positions of remarkable regions. *Left*: raw results. *Right*: clusters CO-0 and CO-1 are merged into a single cluster named 1.

CO-0 to CO-8. While our use of the Meanshift clustering does not take into account the spatial contiguity, they show very consistent spatial distributions, except for clusters CO-0 and CO-1 which share the outer region of the cloud and display a noise-like pattern. A visual inspection of the CO isotopologue PDFs shows that this distinction most likely comes from noise properties. We thus merge cluster CO-0 with cluster CO-1. We call the new cluster CO-1 and represent it in light blue. The resulting cluster map is shown on the right panel of Fig. 2.

Clusters CO-1 to CO-7 show a nested pattern highlighting successive layers from the surface to the inner parts of the cloud. These clusters are present both on the eastern and western sides that surround the inner region of the imaged field of view. However, an asymmetry between the two sides is visible in the much smaller thickness of the surface layers highlighted by the transition from clusters CO-2 to CO-5 on the western FUV-illuminated side. This is a consequence of the much steeper intensity gradients for all three CO lines on this side of the cloud. In other words, high FUV illumination has a much stronger impact on the CO intensity gradients than on the intensities themselves. Cluster CO-8 differs as it only appears towards the two HII regions, NGC 2023 and NGC 2024, which are embedded in the south-western part of Orion B.

One way to check the quality of the clustering is to compare the spatial distributions of the original line intensities with the line intensities averaged in each cluster. This is somehow a test of the ability of the clustering method to compress the information contained in the line intensity maps while retaining the most important aspects. Taking the mean of the line intensities for each cluster conserves the total flux in the output images. Figure 3 presents such a comparison. The mean intensities (and other characteristic intensity values) of each line in each cluster are listed in Table C.1. Most of the ^{12}CO and ^{13}CO spatial features are preserved in the clustered images, while the representation of the C^{18}O image is not as good: some spatial features appear and others disappear. This comparison also highlights that cluster CO-8 is characterised by an increased ^{12}CO intensity compared to its surrounding. We can note that the Horsehead pillar and the other dense clumps that emerge from the IC 434 HII region belong to cluster CO-7 which is associated with relatively dense gas (as discussed in the following subsections).

4.2. Projected PDFs

Figure 4 shows the ^{12}CO 1D PDF computed for the full dataset and for each individual cluster. The full dataset PDF is clearly multi-peaked with a main peak at low intensities (~2 K km s^{-1}), two other peaks around 37 and 60 K km s^{-1} separated by a plateau, and several minor peaks, one of them corresponding to very bright ^{12}CO around 90 K km s^{-1}. In contrast, the 1D PDF of ^{13}CO and C^{18}O (not shown here) are mostly mono-modal (only one clear peak). A good correspondence between the maxima of the ^{12}CO PDF and most of the clusters can be seen. This indicates that ^{12}CO plays a major role in the definition of the clusters. Cluster CO-1 corresponds to the highest and narrowest, low-intensity peak, clusters CO-3 and CO-4 constitute the 37 K km s^{-1} peak, cluster CO-5 contributes to the plateau between the 37 and the 60 K km s^{-1} peaks, cluster CO-6 and CO-7 corresponds to the 60 K km s^{-1} peak, and cluster CO-8 corresponds to the clear bump in the high-intensity tail of the PDF (around 90 K km s^{-1}).

However, significant overlap between the clusters can be seen, and two clusters share the 60 K km s^{-1} peak. Both facts highlight the influence of the other two isotopologues on the clustering. To understand the role of ^{13}CO and C^{18}O in the clustering, Fig. 5 shows the 2D PDFs of ^{13}CO vs. ^{12}CO, and

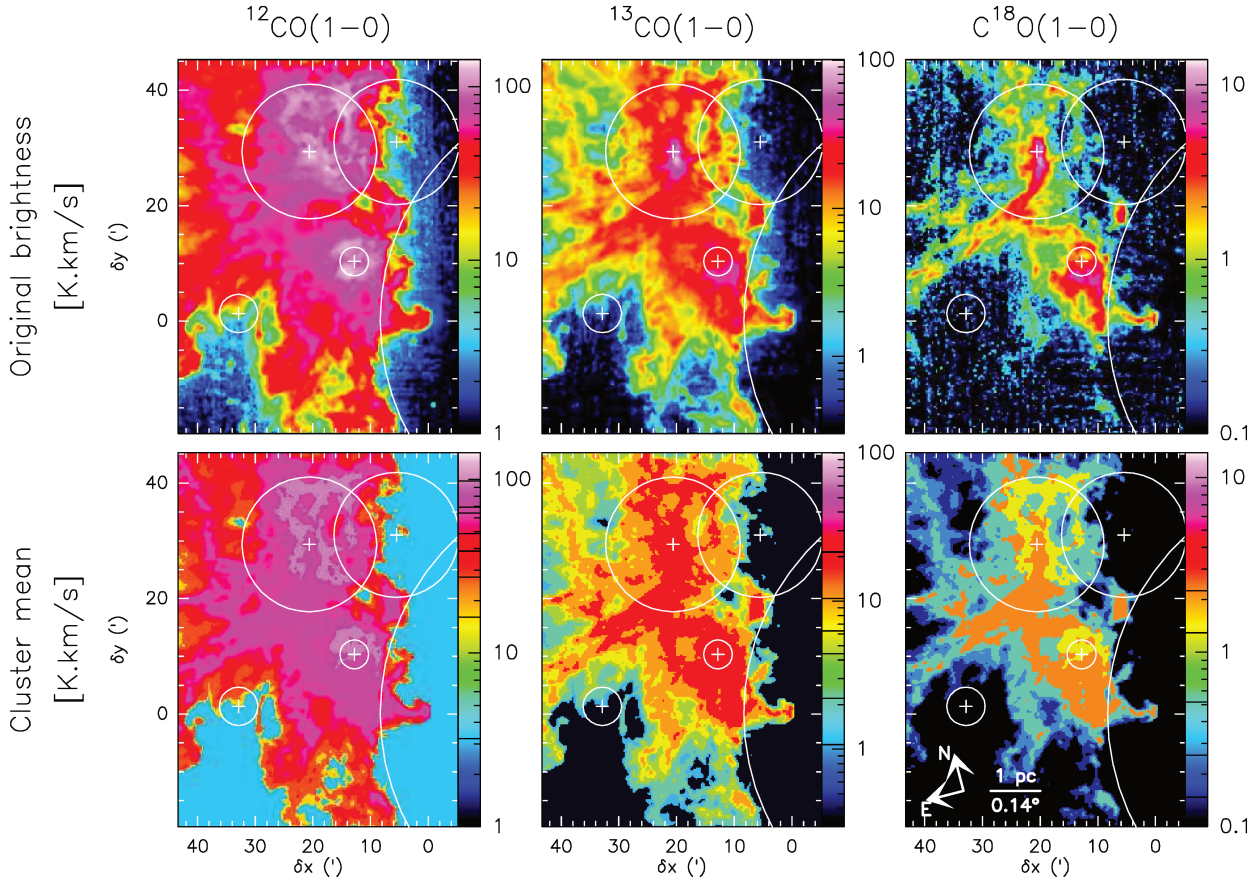


Fig. 3. Comparison of the original intensity spatial distribution (*top row*) with the mean intensity computed for each cluster separately (*bottom row*). The colour scales are identical for the top and bottom rows, but they differ from one column to another. The levels shown on the bottom colour lookup table represents the mean values of the cluster intensities. The clusters were defined using the ^{12}CO, ^{13}CO, and C^{18}O (1–0) lines. The circles show the typical extensions of the HII regions and the crosses show the position of the associated exciting stars (see Pety et al. 2017, for details).

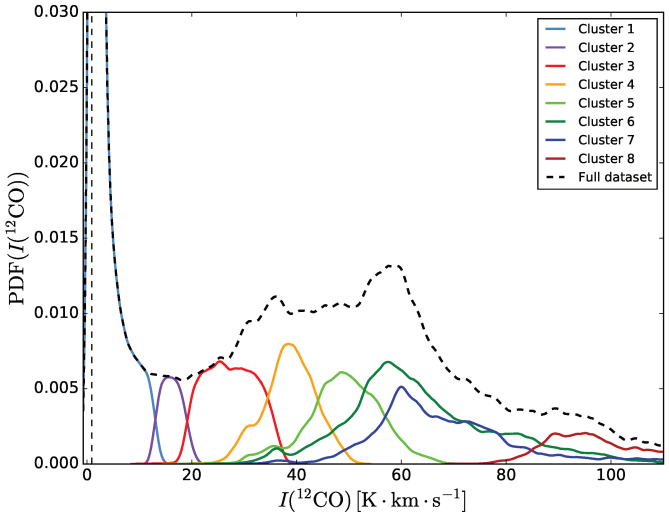


Fig. 4. PDF of the ^{12}CO J = 1–0 line intensity, comparing the PDF of the total dataset (dashed) to the contributions of the different clusters (solid contours coloured according to the colour coding of clusters in Fig. 2). The thin vertical dashed line shows the median 4σ noise level.

^{13}CO vs. C^{18}O as contour plots. In each case, the PDF of the full dataset is shown as black dotted contours, while the PDFs of the different clusters are coloured according to the cluster colours presented in Fig. 2. All clusters have clear separations in the ^{13}CO vs. ^{12}CO plane, so that the overlap seen in Fig. 4 is only a projection effect. Clusters CO-1 to CO-5 clearly follow the ridgeline (the line connecting the cluster maxima, similar to the ridgeline of a mountain ridge connecting the summits) of the 2D PDF. The maxima to which they are associated are small bumps along this ridge. The separations of the basins of attraction of each of these maxima thus lie roughly orthogonally to the direction of this ridge. This is probably why the shape of CO-5 cluster looks like an anti-correlation. Clusters CO-2 to CO-5 are associated to relatively small fluctuations of the PDF along the ridge line. While statistically significant, these fluctuations might be too weak to be each attributed a physical meaning: these four clusters might thus represent a single physical category. In the ^{13}CO vs. C^{18}O space, the PDFs of clusters CO-1 to CO-5 are nearly indistinguishable because they lie below the C^{18}O detection limit. We thus grouped them into a single PDF (grey contours) for better readability.

While clusters CO-6 and CO-7 are undistinguishable on the ^{12}CO 1D PDF, they are clearly separated by their ^{13}CO intensities (cluster CO-7 having $I_{^{13}\mathrm{CO}} \gtrapprox 13$ K km s^{-1}). Cluster CO-8 is distinguished both by its high ^{12}CO intensity and by higher ^{13}CO/C^{18}O ratios at similar C^{18}O intensity than lines of sight belonging to cluster CO-7. Clusters CO-7 and CO-8 thus correspond to a separation of the ^{13}CO vs. ^{12}CO ridge line into two distinct ones at high intensities. Cluster CO-6 is an intermediate cluster that probably lies around the intersection of these two

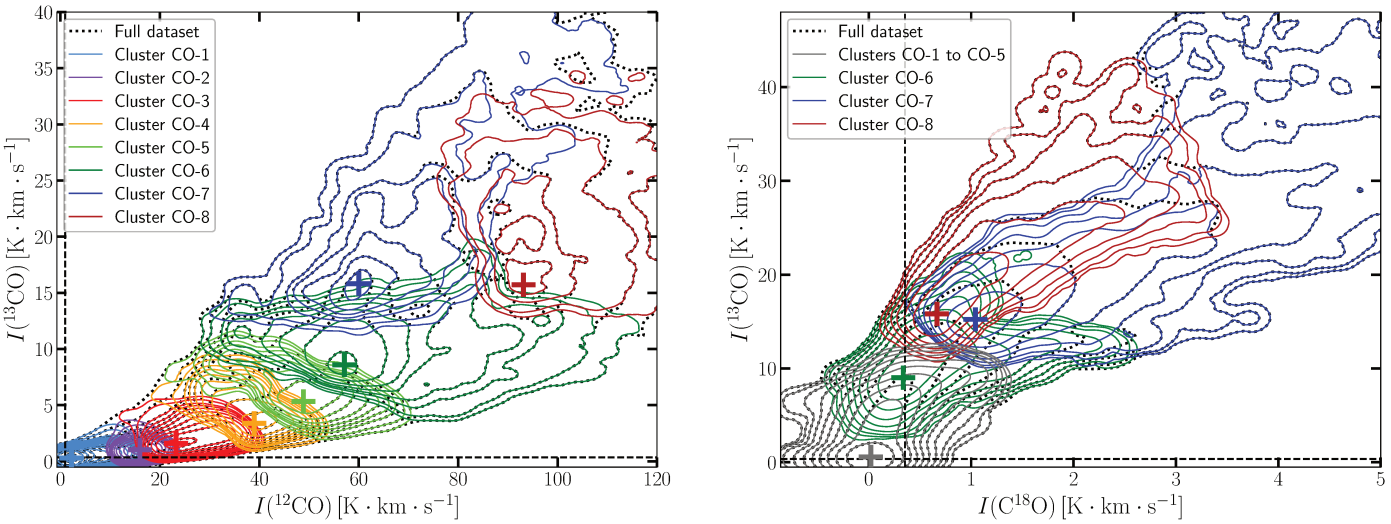


Fig. 5. Contour plot of the 2D PDFs of ^{13}CO vs. ^{12}CO (*left*), and ^{13}CO vs. C^{18}O (*right*). The PDFs of the total dataset are shown as black dotted contours, while the PDFs of the individual clusters are shown as solid contours coloured according to the colour coding of clusters in Fig. 2. On *the right panel* clusters CO-1 to CO-5 have been grouped (grey contours) for better readability. The thin vertical and horizontal dashed lines show the median 4σ noise levels, while the coloured crosses show the positions of the PDF maxima for each cluster.

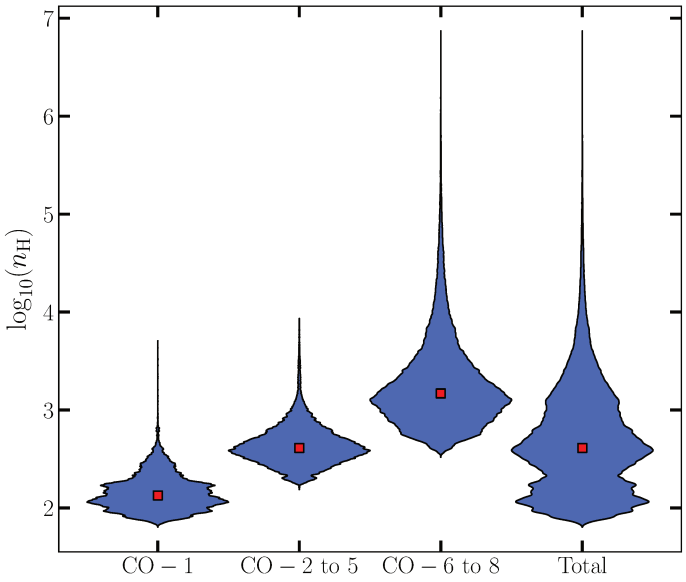


Fig. 6. Violin plot showing the PDF of the approximate gas volume density n_{H}, comparing the contributions of the three groups of CO clusters discussed in the text (CO-1, CO-2 to 5, and CO-6 to 8) to the PDF of the total data set.

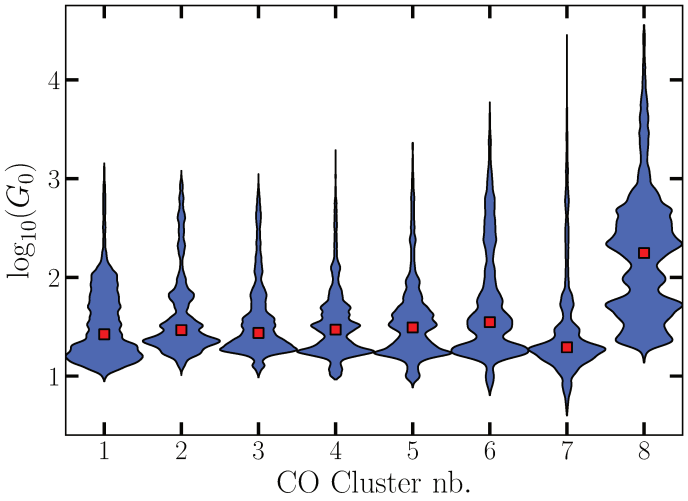


Fig. 7. Violin plot showing the PDF of $\log_{10}(G_0)$ for each CO cluster (blue profiles) and the median value in each cluster (red squares).

ridge lines. This explains why it is so extended along the ^{12}CO axis and narrow along the ^{13}CO axis. We discuss these cluster shapes in relation to local thermodynamic equilibrium (LTE) calculations in Sect. 4.4.

4.3. Relation with volume density and FUV illumination

Figure 6 shows how the clusters contribute to the PDF of the approximate volume density (see Sect. 2.2) in the form of a violin plot: for each group of clusters, the blue profiles show the volume density PDF (normalised to an identical width) and the median values are shown as red squares. We find a close correspondance between our clusters and the three peaks of the density PDF discussed in Sect. 2.2: Cluster CO-1 corresponds to diffuse gas ($n_{\mathrm{H}} \sim 100\,\mathrm{cm}^{-3}$), the group of clusters CO-2 to CO-5 are associated to translucent gas ($n_{\mathrm{H}} \sim 500\,\mathrm{cm}^{-3}$), while clusters CO-6, 7, and 8 correspond to denser gas ($n_{\mathrm{H}} \gtrsim 1000\,\mathrm{cm}^{-3}$). The CO clusters thus reveal underlying density regimes.

Figure 7 compares the distributions of the FUV illumination for the different clusters. Clusters CO-1 to CO-6 have similar median values of $G_0 \sim 30$–35. Moreover, a higher G_0 wing is present for all these clusters. In contrast, cluster CO-7 has a significantly lower value of $G_0 \sim 20$. This cluster thus tracks gas relatively shielded from the FUV illumination. In addition, the high-G_0 wing is negligible for this cluster. Finally, cluster CO-8 has a much larger median value of $G_0 \sim 180$, and its PDF has two broad components at typical values of $G_0 \sim 50$ and 300, consistent with the presence of the NGC 2023 and NGC 2024 HII regions.

In summary, clustering of the CO isotopologues allows us: 1) to distinguish three different regimes of column/volume density (diffuse, translucent, and higher density); and 2) to start distinguishing FUV-illuminated from FUV-shielded gas; but only for relatively dense gas.

4.4. Interpretation: nested CO isotopologues and higher [^{13}CO]/[C^{18}O] abundance ratios in FUV-illuminated dense gas

We wish to understand the physical and chemical processes that determine the variations in line intensities detected by the clustering method. We thus make the simplest possible radiative transfer model that will allow us to match the observed line intensities and ratios of the CO isotopologue lines.We then interpret the astrophysical information uncovered by the clustering of the CO isotopologues.

4.4.1. Modeling principles

The typical density over the studied field of view is $\sim$300 $H_2\,cm^{-3}$, and the density of at least 25% of the field (Pety et al. 2017) is larger than 2×10^3 $H_2\,cm^{-3}$ (the typical critical density of CO J = 1–0). We thus chose to use LTE models, as we are mainly interested by the denser parts of the field of view. The clustering analysis taught us that the CO isotopologue intensities trace different ranges of hydrogen column density, and consequently of approximate volume density (see Sect. 2.2). We have thus chosen to model intensity curves for several fixed values of the CO isotopologue column densities, varying the kinetic temperature.

Figure 3 indicates that the structure of the dense inner parts of the cloud is well delineated by the C^{18}O (1–0) emission, still visible in the ^{13}CO (1–0) emission, and mostly hidden in the ^{12}CO (1–0) emission. The usual interpretation is that the ^{12}CO line is so optically thick that the outer, less dense layers along the line of sight can already produce a saturated ^{12}CO emission. This means that the ^{13}CO and C^{18}O (1–0) lines would be sensitive to denser, cooler gas more deeply embedded along the line of sight while the ^{12}CO (1–0) line would to first order be mostly sensitive to the foreground, more FUV-illuminated and thus warmer gas. We thus propose to use two different values for the excitation temperature: a high value for ^{12}CO and a lower value for ^{13}CO and C^{18}O. This in turn implies that the model ^{12}CO column density will be a lower limit to the total CO column density as it only represents the warm gas.

The detailed parametrisation of our modelling is described in Appendix D.

4.4.2. Observations and modeled curves

Figure 8 shows the modeled curves over the joint histograms of the ^{13}CO vs. ^{12}CO emission (left column), of the ^{13}CO vs. C^{18}O emission (middle column), and of the ^{13}CO/C^{18}O vs. ^{12}CO/^{13}CO line ratios (right column). The ratio vs. ratio histograms allow us to check how the models take care of the covariations of the three studied CO lines.

The first row presents the observations for the full field of view, while the next three rows present the observations for different sets of CO clusters (from 1 to 6, the 7th one, and the 8th one, respectively). The sets of input parameters described above each row were chosen to deliver the best visual match between the modelled curves and the three associated histograms.

In all cases, only the lines of sight where the isotopologue lines considered have intensities above 4σ are used to compute the histogram. For each ^{13}CO opacity (i.e. along each white curve), the ^{13}CO and C^{18}O excitation temperature increases clockwise and counter-clockwise for the ^{13}CO vs. ^{12}CO, and ^{13}CO vs. C^{18}O histograms, respectively. On the ratio vs. ratio histograms (right column), the ^{13}CO and C^{18}O excitation temperature increases from left to right.

On the line vs. line histograms (left and middle columns), the higher the ^{13}CO opacity, the more opened the corresponding model curve. In contrast, the dependency on the column density is reduced in the ratio vs. ratio histograms, as indicated by the fact that all curves for different ^{13}CO opacities almost overlap. This is linked to the fact the line intensity is proportional to the column density to lowest order. And therefore, line ratios remove this trend.

4.4.3. Global results

The first row of Fig. 8 shows the best visual match between observations and models for the full field of view. The FWHM of the lines (2 km s^{-1}) is the median value measured over the field of view on the 10.5 km s^{-1} main component of the ^{13}CO and C^{18}O (1–0) lines. The line emission of the modelled curves is integrated over 3 km s^{-1} as in the observations.

The range of ^{13}CO (1–0) opacities runs from optically thin lines (minimum: 0.03) to moderately saturated lines (maximum: 2.5). The C^{18}O (1–0) line is always optically thin as expected from the fact that C^{18}O (1–0) shows an excellent correlation with the visual extinction (Pety et al. 2017). The ^{12}CO (1–0) line is almost always optically thick.

Kinetic temperatures of up to $\sim$100 K are required to explain the low intensity part of the ^{13}CO vs. ^{12}CO histogram. The [^{13}CO]/[C^{18}O] abundance ratio is larger than the expected elemental ratio value of $\sim$8 (Wilson & Rood 1994). Both inferences can be explained by the significant FUV illumination in the observed field of view, with contributions from external and embedded HII regions. The joint histogram of ^{13}CO vs. ^{12}CO intensities, especially the range of ^{13}CO intensities at a given ^{12}CO intensity, can only be explained if ^{12}CO and ^{13}CO have different kinetic temperatures. In contrast, the observation space can be understood with similar kinetic temperatures of the gas that emits the ^{13}CO and C^{18}O (1–0) lines.

4.4.4. Results per CO clusters

In order to better understand how the CO clustering can distinguish different regimes of density and FUV illuminations, we now discuss the input parameters that deliver the best visual match between the modelled curves and the three histograms for three different subsets of the CO clusters.

Starting with the histograms computed for clusters CO-1 to CO-6, shown on the second rows, we obtain the same set of parameters as for the entire field of view (first row). There are only two exceptions. First, we need relatively low ^{13}CO opacities (from 0.03 to 0.5), confirming that we deal with low-opacity lines of sight. The horizontal edge between clusters CO-6 and CO-7/8 in the ^{13}CO vs. ^{12}CO histogram closely follows a constant column density curve, confirming that it separates two column density regimes (the cut corresponds to a ^{13}CO opacity of $\sim$0.5). Second, the minimum kinetic temperature is slightly lower than for the global fit as we are less constrained by the lower edge of the ^{13}CO vs. C^{18}O histogram. It is unclear whether this fact is significant.

The third and fourth rows show our best matches for clusters CO-7 and CO-8, respectively. In both cases, only the high ^{13}CO opacity (0.65 to 2.5) curves are displayed. This confirms that we are in the high-column-density regime. In both cases, we need to restrict the ^{13}CO kinetic temperature range from 11–12 to 20 K.

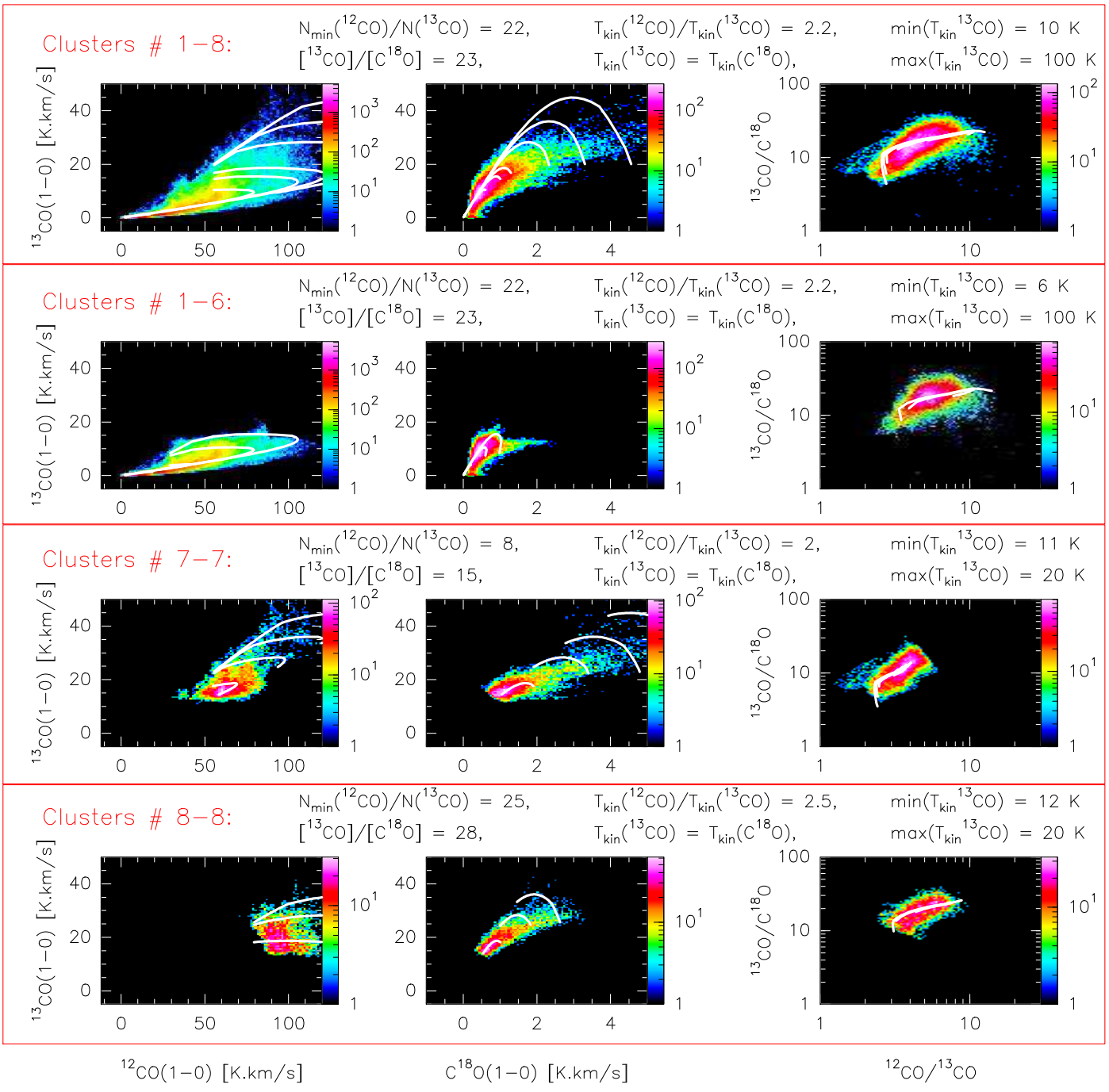


Fig. 8. LTE radiative transfer models for the three main CO isotopologues. The four rows show the best match between the observations and the models for, *from top to bottom*, the full field of view studied here, clusters 1 to 6, and the 7th and 8th clusters. The control parameters of the family of models are written on top of each associated row. The *left and middle columns* show the joint histogram of the (1–0) lines of ^{13}CO vs. ^{12}CO, and of ^{13}CO vs. C^{18}O. The *right column* shows the joint histogram of the ^{13}CO/C^{18}O vs. ^{12}CO/^{13}CO intensity ratios. The colour look-up tables show the number of lines of sight that fall within a given bin of the histogram. The white curves present the LTE intensity variations as a function of the ^{13}CO kinetic temperature for different fixed ^{13}CO opacities (0.03, 0.1, 0.3, 0.5, 0.65, 1.2, 1.75, 2.5).

The other parameters differentiate the two clusters. First, the ^{12}CO/^{13}CO kinetic temperature ratio is higher in cluster CO-8 than in cluster CO-7. This confirms the idea that the outer layers of the CO-8 cluster are more exposed to the FUV illumination than those of the CO-7 cluster. Finally, the [^{13}CO]/[C^{18}O] abundance ratio is much closer to the expected elemental abundance ratios for cluster CO-7 than for cluster CO-8. This is consistent with the idea that most of the gas in cluster CO-7 is well shielded, in complete contrast with the gas in cluster CO-8, as discussed in the following section.

4.4.5. Discussion

In all our models, we need to distinguish the kinetic temperature of the gas that emits the ^{12}CO (1–0) line on the one hand, and the ^{13}CO and C^{18}O (1–0) lines on the other. The effect is the most pronounced in cluster CO-8 that is highly FUV illuminated and then in clusters CO-1 to CO-6 that contain diffuse and translucent gas. This implies that the ^{12}CO and ^{13}CO/C^{18}O emissions have different spatial extents along the line of sight (as ^{12}CO emission quickly saturates and thus only traces a limited surface

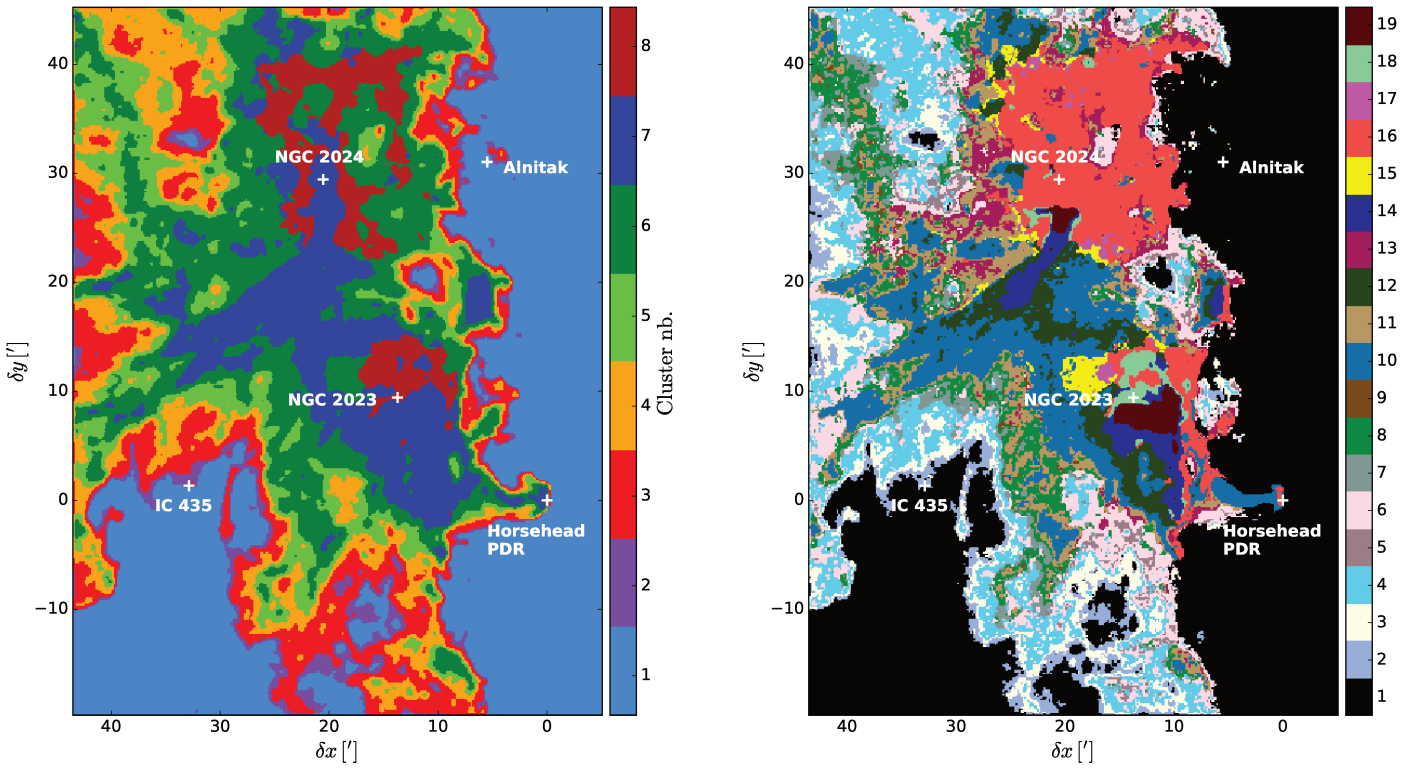


Fig. 9. Comparison of the clusters obtained based on the CO isotopologues alone (*left*) and the clusters obtained when associating the CO isotopologues to HCO^+ and CN (*right*). In both cases, the clusters have been ordered by increasing ^{12}CO (1–0) mean intensity. White crosses mark the positions of remarkable regions.

layer). It is thus impossible to use the flux ratio of these lines to try to infer the $[^{12}CO]/[^{13}CO]$ and $[^{12}CO]/[C^{18}O]$ abundance ratios.

In the high column/volume density clusters, ^{12}CO is very optically thick (the saturation of the ^{12}CO vs. A_V relation is noticeable starting from $A_V \sim 5$ as shown by Pety et al. 2017), and the variations of integrated intensities are largely caused by variations in the CO excitation temperature, which is close to the gas kinetic temperature. Cluster CO-8 shows significantly higher ^{12}CO intensities (~90 K km s^{-1}) than clusters CO-6 and CO-7 (~60 K km s^{-1}) and thus traces a warmer dense gas regime. This is consistent with its location around the NGC 2024 and NGC 2023 star forming regions.

Cluster CO-8 is also distinguished from cluster CO-7 by higher $^{13}CO/C^{18}O$ intensity ratios: the median ratio is ~18 in cluster CO-8, while it is ~11 in cluster CO-7. This difference is a sign of FUV illumination as the ^{13}CO abundance in PDRs is tightly coupled to ^{12}CO by the isotopic fractionation reaction (Langer et al. 1984)

$$^{12}CO + {}^{13}C^+ \rightarrow {}^{13}CO + {}^{12}C^+ + 35\,K, \quad (5)$$

which, at the gas temperatures of PDRs, does not favour any enrichment, but ensures a strong coupling between the abundances of ^{12}CO and ^{13}CO. In contrast, $C^{18}O$ in FUV-illuminated regions is formed separately from pure carbon chemistry followed by reactions of small hydrocarbons such as CH, CH_2 or C_2H with ^{18}O. As a result, ^{13}CO indirectly benefits from ^{12}CO self-shielding while $C^{18}O$ is easily dissociated. This might explain the large $[^{13}CO]/[C^{18}O]$ abundance ratio compared to the value expected from elemental abundances.

Clusters CO-7 and CO-8 thus highlight a separation of the global intensity PDF into two different tails at high column density that correspond to warm illuminated dense regions around massive star forming regions (cluster CO-8) and shielded dense gas (cluster CO-7).

5. Adding HCO^+ and CN to get a better clustering of high-density and high-FUV-illumination regimes

Clustering ability is limited by the information contained in the tracers input to the algorithm. Our first application of the Meanshift clustering algorithm to the CO isotopologues proved its ability to reveal several distinct density regimes. It also hinted at a first distinction of FUV illumination regimes. However, using only the three CO isotopologues is insufficient to clearly distinguish FUV illumination effects. We thus now include in the clustering analysis the HCO^+ and CN (1–0) maps together with the three CO isotopologues maps. Indeed, Pety et al. (2017) and Gratier et al. (2017) have shown that HCO^+ and CN were sensitive to FUV illumination. We chose CN rather than small hydrocarbons (C_2H or c-C_3H_2) as the latter are detected at a lower signal-to-noise ratio (S/N), making the clustering noisier. For simplicity, we only used the brightest hyperfine component of the CN (1–0) transition. In addition, these two lines have significantly higher critical densities (~2×10^5 cm^{-3} for HCO^+ and ~2×10^6 cm^{-3} for CN).

We used in this case the adaptive bandwidth method with 425 neighbours. This number is a compromise between eliminating artificial clusters coming from sampling fluctuations of the PDF and retaining sufficiently fine cluster subdivisions to find the interesting physical distinctions.

5.1. Resulting spatial distribution

Figure 9 compares the clusters obtained based on the CO isotopologues alone with the clusters obtained by adding CN and

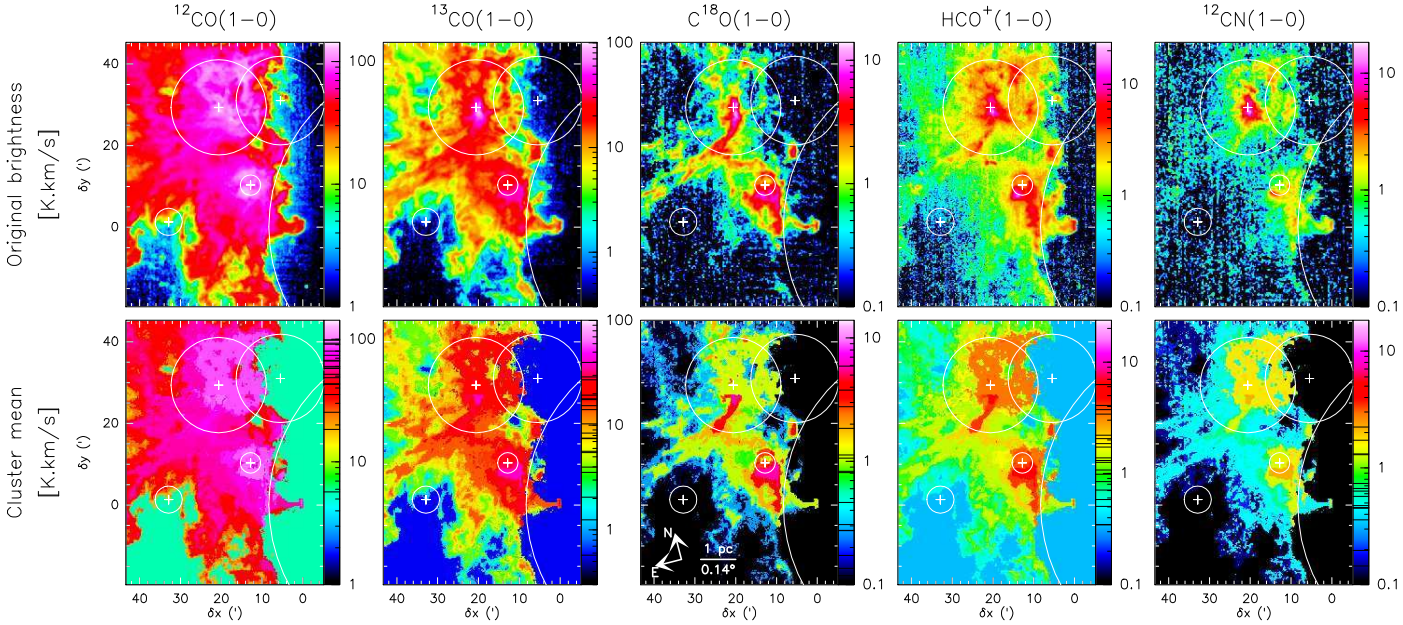


Fig. 10. Same as Fig. 3, except that clusters were defined on the ^{12}CO, ^{13}CO, C^{18}O, CN, and HCO$^+$ (1–0) lines.

HCO$^+$. Several striking facts are visible. First, the CO clustering brought 8 clusters while the new clustering identifies 19 clusters. The increase in number of clusters is related to the finer physics we wish to reveal through the increase of the dimension of the intensity space from 3 to 5 lines, at a constant number of pixels. We name the new clusters FUV-1 to FUV-19. For convenience they have been numbered by order of increasing mean ^{12}CO intensity.

Second, the spatial edges of the FUV clusters that appear from the north-eastern to the south-western corners appear noisier. Indeed, mostly translucent gas is present in these regions, implying that CN and HCO$^+$ are barely detected there. Third, a clear East-West asymmetry is now seen in the distributions of the clusters. Some clusters, such as FUV-2, 3, 4, 7, 8 or 11, appear mostly on regions less exposed to FUV illumination, while cluster FUV-16 is clearly associated with the NGC 2024, NGC 2023, and IC 434 HII regions. In the previous clustering, cluster CO-8 does not tag the PDRs associated with IC 434 (the Horsehead PDR, for instance). The western edges of the CO-3 to CO-5 clusters are mostly merged now in the cluster FUV-1 that mainly contains diffuse gas. For instance, the envelope of the Horsehead nebula has been merged into FUV-1, only leaving the less familiar silhouette of the denser parts of the Horsehead visible.

Figure 10 compares the spatial distributions of the original line intensities with those of the line intensities averaged per cluster. Comparing with Fig. 3, we see that the FUV clustering reproduces the ^{12}CO (1–0) faint intensity regimes ($\leq$5 K km s^{-1}) less well, but it much better samples the ^{12}CO (1–0) high intensity regime ($\geq$50 K km s^{-1}) and the ^{13}CO and C^{18}O median intensity regimes (between 10 and 50 K km s^{-1}, and between 2 and 6 K km s^{-1}, respectively). The better sampling of the median- to high-intensity regimes of the CO isotopologue (1–0) lines is linked to the detection of several clusters at relatively high HCO$^+$ integrated intensity ($\geq$2 K km s^{-1}). This is particularly clear on the C^{18}O and HCO$^+$ compressed maps that emphasise dense regions extending south of NGC 2024 and surrounding NGC 2023. In contrast, the high-CN-integrated-intensity end ($\geq$2 K km s^{-1}) is not well sampled by the new clustering.

In summary, this clustering seems to provide most of the FUV illumination contrast between east and west. On the one hand, the FUV clustering thus provides a better data compression in the inner dense parts and in the FUV-illuminated regions; on the other, the shapes of the Horsehead and of the western illuminated edge are less well reproduced.

While the 19 clusters are statistically significant, interpreting all of them is difficult. Indeed, clusters with extreme behaviour have relatively clear physico-chemical interpretations but they are separated in the line space by clusters with intermediate properties that reflect subtler, second-order distinctions. The major physical distinctions brought forward by this clustering are thus best discussed in terms of groups of clusters. Moreover, we wish to understand the first-order roles of the HCO$^+$ and CN (1–0) lines in the classification. We thus group the clusters in two ways. We group together clusters with similar most probable intensities of HCO$^+$ or CN, respectively, ordered by increasing values. Our goal is to keep the minimum number of groups needed to visualise the physico-chemical regimes first brought forward by each line.

5.2. HCO$^+$-based grouping and high-density regimes

We first constitute groups based on the HCO$^+$ intensities in each cluster following the numerical recipe explained in the previous section.

5.2.1. One-dimensional PDFs and spatial distribution

The left panels of Fig. 11 show the HCO$^+$ PDF of each individual cluster, and the way we grouped them: the clusters whose most probable values (1D-PDF peaks) gather at similar integrated intensities are grouped. We end up with seven groups, named HCO$^+$-1 to HCO$^+$-7, whose mean integrated intensities regularly increase from 0.3 to 6.3 K km s^{-1}. Cluster FUV-1 alone constitutes group HCO$^+$-1 as it mostly traces diffuse gas surrounding the molecular cloud. We nevertheless note that it has a wing between 1 and 3 K km s^{-1} that corresponds, for

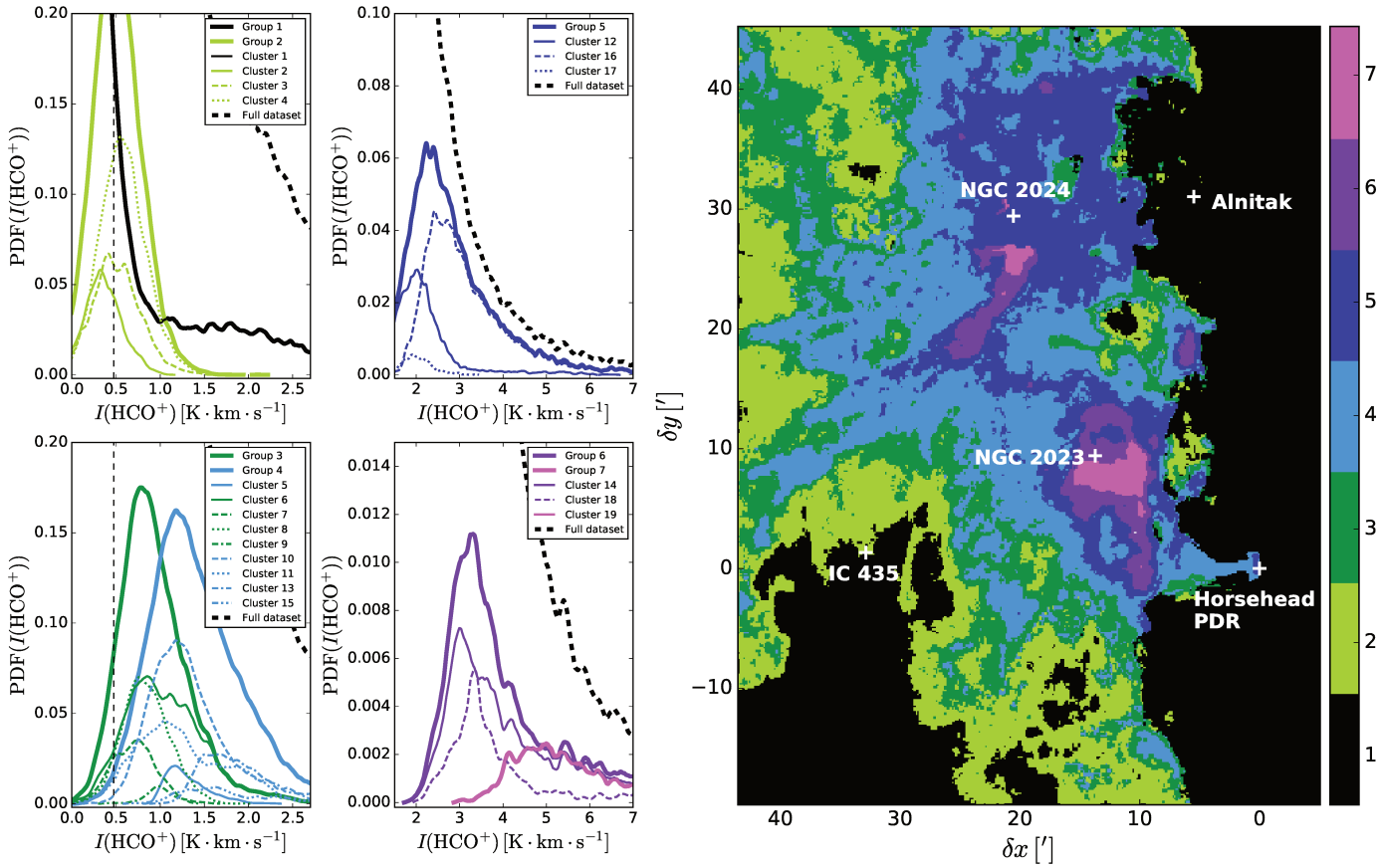


Fig. 11. *Left*: 1D PDF of the HCO+ (1–0) line intensity, comparing the full dataset PDF (thick dashed black line), the contribution of each of the groups defined in Sect. 5.2 (coloured thick lines), and the contribution of each individual cluster (thin coloured lines). The cluster contributions are coloured according to the group to which they belong. For readability, we have separated groups HCO^+-1 to HCO^+-4 and their constitutive clusters (*left panels*) and groups HCO^+-5 to HCO^+-7 and their constitutive clusters (*right panels*). *Right*: map of the 7 groups resulting from the grouping of consecutive clusters described in the text (Sect. 5.2).

instance, to the Horsehead envelope. Group HCO^+-2 contains clusters FUV-2, 3 and 4, group HCO^+-3 clusters FUV-6, 7, 8, and 9, group HCO^+-4 clusters FUV-5, 10, 11, 13, and 15, group HCO^+-5 clusters FUV-12, 16, and 17, group HCO^+-6 clusters FUV-14 and 18, and finally, group HCO^+-7 contains only cluster FUV-19. The characteristic intensity values (median, mean, standard deviation) of the lines in each group are listed in Table C.2.

The spatial distribution of these groups is displayed as the right panel of Fig. 11. We see that the resulting groups have a faint mean HCO^+ intensity in the outer part of the cloud and that this intensity increases towards the densest parts. In comparison to the CO clustering, the CO-6 to CO-8 bright clusters are now distributed over groups HCO^+-3 to HCO^+-7, and clusters CO-1 to CO-5 (faint CO intensity) are distributed over the groups HCO^+-1 and HCO^+-2. This suggests that the HCO^+ (1–0) line is better at discriminating higher-density regimes than the CO (1–0) lines, even though about half of the HCO^+ flux over the observed field of view is coming from diffuse and translucent regions (Pety et al. 2017). Finally, the boundaries of some groups are close to the CO clustering results, indicating that CO isotopologues still play an important role in defining some of the groups. For instance, the boundary between the HCO^+-4 and HCO^+-5 groups is similar to the boundary between clusters CO-7/8 and CO-6.

5.2.2. Two-dimensional PDFs

To further understand the relative roles of the CO isotopologues and HCO^+, we show in Fig. 12 the 2D-PDFs of ^{13}CO vs. $C^{18}O$ (left) and HCO^+ vs. $C^{18}O$ (right), which we found to be the most informative among the possible pairs of lines. The first striking impression is that the groups overlap considerably in both 2D PDFs. However, some groups clearly separate in one of the 2D PDFs but not in the other. For instance, while groups HCO^+-4 and 6 strongly overlap in the (HCO^+ vs. $C^{18}O$) PDF, they are cleanly separated in the (^{13}CO vs. $C^{18}O$) PDF. Finding the right 2D projection to reveal cluster separations quickly becomes impossible. Moreover, this 2D projection might not even exist when the clusters are not linearly separable; for example, when one cluster is completely surrounded by another one. We have to rely on the Meanshift algorithm to reveal information about the morphology of the complete PDFs (maxima and their associated basins of attraction) that we cannot otherwise directly visualise.

The groups form a sequence that mostly follows a single trend with increasing line intensities. This is better visualised when trying to connect the crosses that represent the group mean intensities. Group HCO^+-5 only partially follows this trend: a part of it (actually mostly cluster FUV-16) is overluminous in HCO^+ at constant $C^{18}O$ intensity. This is linked to the sensitivity

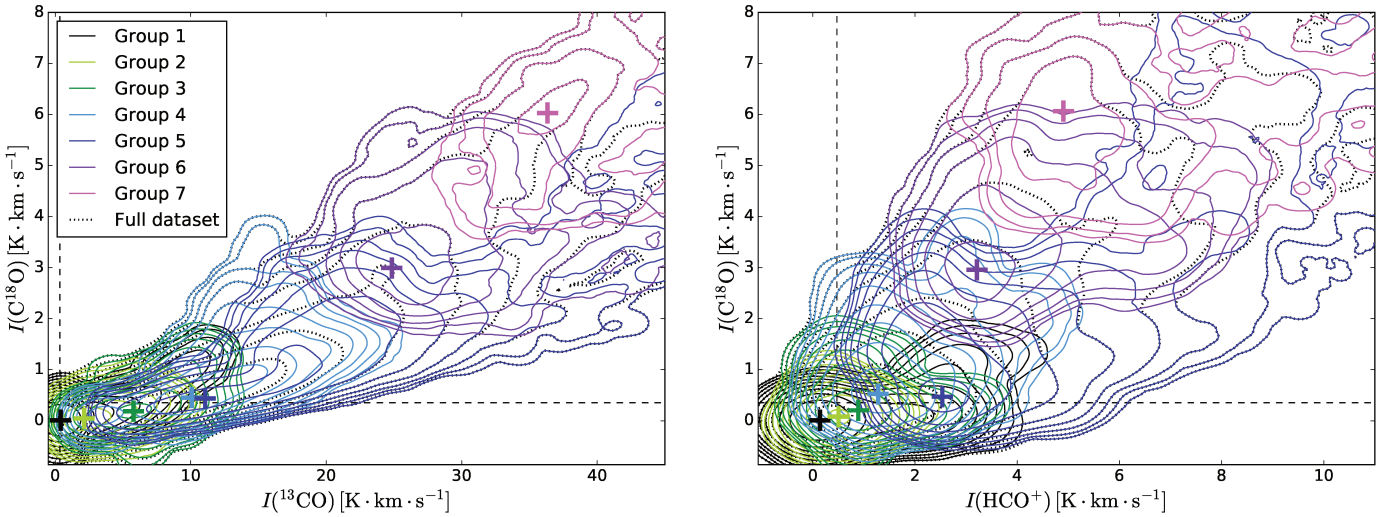


Fig. 12. Contour plot of the 2D PDFs of $C^{18}O$ vs. ^{13}CO (*left*) and $C^{18}O$ vs. HCO^+ (*right*). The PDFs of the total dataset are shown as black dotted contours. The contributions of the 7 groups resulting from the grouping discussed in the text are shown in contours coloured according to Fig. 11 (right). In addition, the PDF maximum of each group is shown as a cross with the same colour as the group.

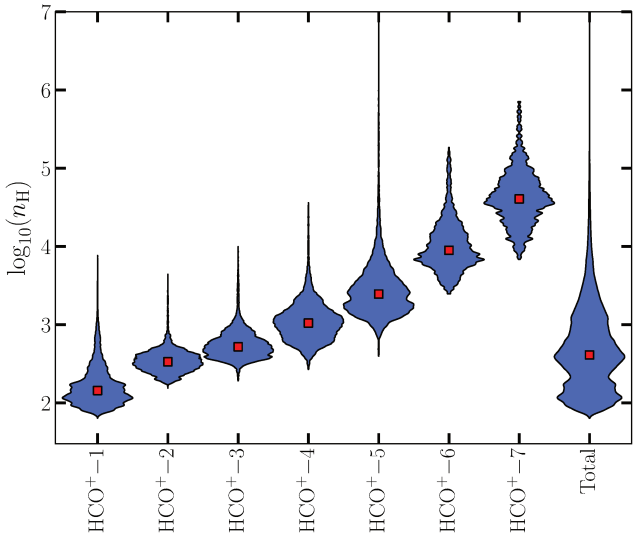


Fig. 13. Violin plots showing the PDF (blue profiles) and median values (red squares) of the approximate volume density n_H for each HCO^+-group, and for the full map (label "Total").

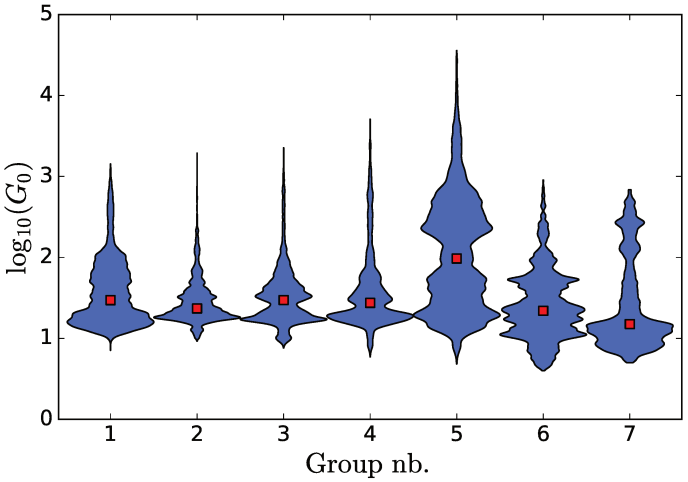


Fig. 14. Violin plots showing the PDF (blue profiles) and median values (red squares) of the FUV illumination G_0 for each HCO^+-group.

of the HCO^+ (1–0) line to the FUV illumination (cf. the following section).

5.2.3. Link with volume density and FUV illumination

As shown in Gratier et al. (2017), the main underlying parameter contributing to intensity variations across our maps is the gas column density. The single trend highlighted here could thus to first order be associated with column density variations and thus approximate volume density variations as discussed in Sect. 2.2. Figure 13 shows the PDF and median value of the volume density in each group (violin plots). We indeed see that the HCO^+-groups correspond to increasing ranges of volume densities. On the one hand, the first five HCO^+-groups contribute to the three main peaks of the approximate volume density PDF: the HCO^+-1 group corresponds to the diffuse gas peak, the HCO^+-2 and 3 groups dominate the translucent gas peak, and the HCO^+-4 and 5 groups contribute to the denser gas peak. On the other hand, groups HCO^+-6 and 7 are located in the high-density tails.

While part of the distinction is linked to the underlying existence of the three density regimes, the distinction between groups HCO^+-6 and 7 hints at the existence of higher-density regimes. Their rarity makes them only barely noticeable as bumps in the density PDF. This distinction is probably the result of an excitation effect. Indeed, their typical volume densities (probably underestimated as they are averaged along the line of sight) are 7×10^3, and 4×10^4 cm^{-3}, respectively. These values approach the critical density of HCO^+ for collisional excitation with H_2 ($\sim 10^5$ cm^{-3}). We thus probably experience a transition from a weak excitation regime (Liszt & Pety 2016) towards a regime closer to the usual thermalised excitation.

Figure 14 shows the PDF and median value of the FUV illumination for each group. In contrast to the density, no clear separation of the groups in terms of FUV illumination is visible, except for the fact that HCO^+-5 has a much broader G_0 distribution than the other groups. The HCO^+ grouping thus does not cleanly capture distinctions only related to FUV illumination,

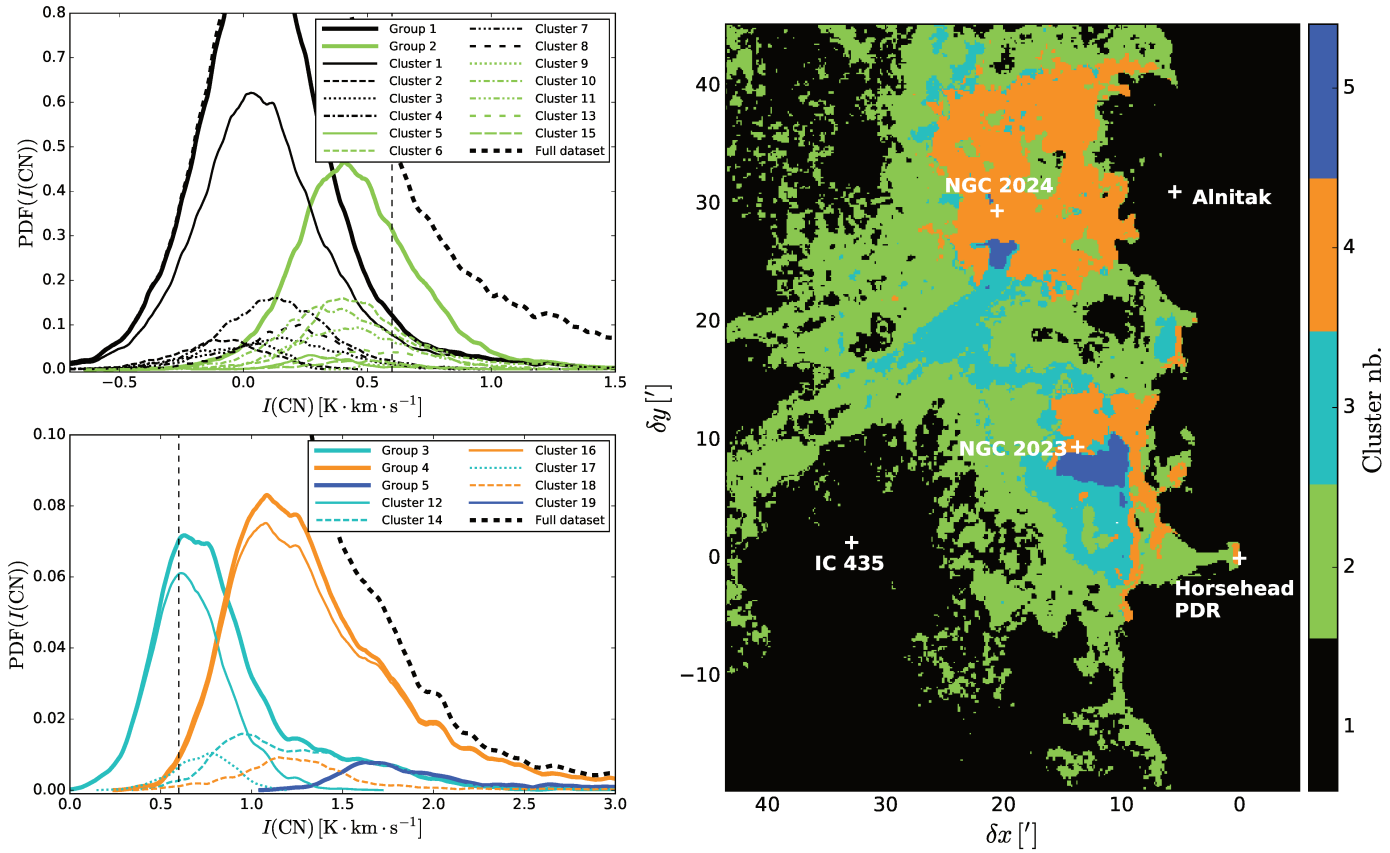


Fig. 15. *Left*: 1D PDF of CN (1–0) line intensity, comparing the full dataset PDF (thick dashed black line), the contribution of each of the groups defined in Sect. 5.3 (coloured thick lines), and the contribution of each individual cluster (thin coloured lines). The contributions of clusters are coloured according to the group to which they belong. For readability, we have separated groups CN-1 and CN-2 and their constitutive clusters (*top panel*) and groups CN-3 to CN-5 and their constitutive clusters (*bottom panel*). *Right*: map of the 5 groups resulting from the grouping of consecutive clusters described in the text (Sect. 5.3).

even though the complex behaviour of HCO^+-5 in the 2D PDFs of the line intensities is likely related to the presence of a mixture of FUV illumination (varying by more than one order of magnitude) in this group. This is not a property of the initial 19 clusters but of the grouping, as is shown in the following section.

5.3. CN-based grouping and FUV illumination regimes

We now present the second grouping of the clusters, based on their CN intensities.

5.3.1. One-dimensional PDFs and spatial distribution

The left panels of Fig. 15 show how the CN PDF of the individual clusters contributes to the PDF of their group. After sorting them by increasing CN mean intensity, we merged the first six clusters into group CN-1 (clusters FUV-1, 2, 3, 4, 7, 8) because they correspond to regions where CN is not detected. We then merged the next seven clusters into group CN-2 (clusters FUV-5, 6, 9, 10, 11, 13, 15). The CN line is barely detected in these clusters. Groups CN-3 and CN-4 gather the next three (FUV-12, 14 and 17) and two clusters (FUV-16 and 18), respectively. The last cluster (FUV-19) is significantly brighter in CN. It thus has its own group CN-5. The characteristic intensity values (median, mean, standard deviation) of the lines in each group are listed in Table C.3.

The most striking distinction revealed by the resulting spatial distribution (see the right panel of Fig. 15) is the separation of the central regions of the cloud between groups CN-3 and CN-4; contrary to the previous cases, this separation does not show a nested pattern. The CN-3 group covers dense regions in the inner parts of the cloud. The CN-4 group appears towards the interfaces between the molecular cloud and the NGC 2024, NGC 2023, and IC 434 HII regions. This distinction thus seems to separate FUV-shielded dense gas (group CN-3) and FUV-illuminated dense gas (group CN-4). Group CN-5 highlights smaller regions at the interface between groups CN-3 and CN-4. Groups CN-1 and 2 represent outer regions of the cloud.

5.3.2. Two-dimensional PDFs

Figure 16 shows the 2D PDFs of CN vs. $C^{18}O$ (left panel), and CN vs. HCO^+ (right panel); it compares the full dataset PDF (dashed line) with the contributions of our five groups (solid lines). Two distinct trends in the tail of the PDF are obvious on the 2D PDF of CN vs. HCO^+: a low CN/HCO^+ ratio corresponding to group CN-5, and a high CN/HCO^+ ratio corresponding to group CN-4. The other groups lie in a low-intensity region where the two trends are blended.

This dual trend is already seen on the CN vs. $C^{18}O$ 2D-PDF: for groups CN-2, CN-3, and CN-5, CN increases very slowly with $C^{18}O$, in a mostly linear way. In contrast, group CN-4 has higher CN intensities than the other groups, with the CN

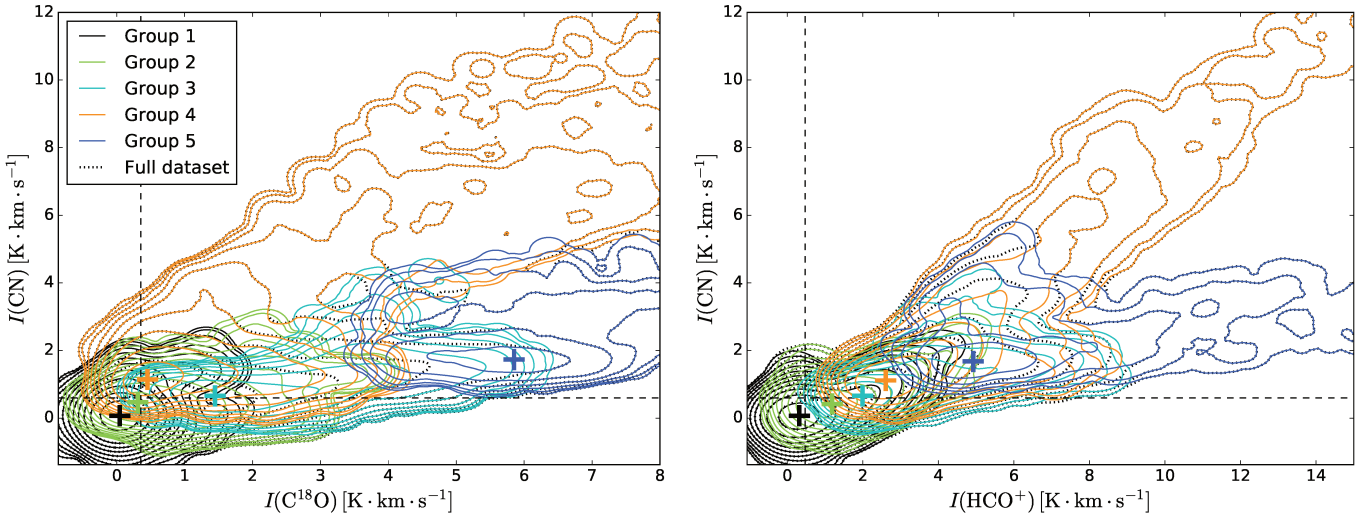


Fig. 16. Contour plot of the 2D PDFs of CN vs. $C^{18}O$ (*left*) and CN vs. HCO^+ (*right*). The PDFs of the total dataset are shown as black dotted contours. The contribution of the 5 groups resulting from the grouping discussed in the text are shown coloured according to Fig. 15 (right). In addition, the PDF maximum of each group is shown as a cross with the same colour as the group.

intensity increasing much faster with the $C^{18}O$ one; although this trend has a larger scatter.

Finally, while there is a large overlap region between groups CN-4 and CN-5 in the PDF of CN vs. HCO^+, these two groups are clearly separated in the PDF of CN vs. $C^{18}O$. Group CN-4 is thus observationally distinguished by an overly bright CN emission.

5.3.3. Link with FUV illumination (and volume density)

As a species easily detected in diffuse clouds, CN is a good tracer of FUV-illuminated gas (Snow & McCall 2006). We thus interpret the dual trend as a separation between regions where the photo-chemistry is active and regions of dense FUV-shielded molecular gas. The spatial consistency of groups CN-3 with regions of dense FUV-shielded gas and CN-4 with FUV-illuminated gas strengthens this interpretation, all the more so that spatial information is not used in the clustering analysis.

More quantitatively, Fig. 17 shows the approximate volume densities (left panel) and FUV illuminations (right panel) found for the different groups. Groups CN-3 and CN-4 have very similar volume densities (median densities of $\sim 3 \times 10^3$ and $\sim 4 \times 10^3$ cm^{-3}, respectively), while the CN-1, CN-2, and CN-5 correspond to distinct ranges of volume densities (respective median densities of $\sim 2 \times 10^2$, $\sim 8 \times 10^2$, and $\sim 4 \times 10^4$ cm^{-3}). In contrast, the G_0 distributions show that group CN-4 clearly has higher FUV illumination (median $G_0 \sim 210$) than all other groups (median G_0 between 15 and 30), in particular groups CN-3 and CN-5 (median $G_0 \sim 20$ and $G_0 \sim 15$, respectively). This confirms our interpretation that group CN-4 corresponds to FUV-illuminated (relatively) dense gas.

Groups CN-1 and CN-2 have lower approximate volume densities than groups CN-3 and CN-4. However the distribution of $T_{\rm dust}$ and therefore G_0 in groups CN-1 and CN-2 overlaps with that of both groups CN-3 and CN-4. This shows that it is more difficult to separate the influence of the radiation field for low-density regions. This is related to the fact that the envelope of the Horsehead nebula has been merged into group CN-1, leaving only the less familiar silhouette of the denser parts of the Horsehead visible.

The highest-density group (CN-5) is found only in the immediate vicinity of the two star-forming regions NGC 2023 and NGC 2024 and their HII regions. This might be a signature of compression of the molecular gas by the expansion of the HII regions: Tremblin et al. (2014) have indeed found this process to cause bimodality in the column density PDF (on spatial scales of a few pc). The gas kinematics in these two regions was in addition found to be dominated by compressive (rather than solenoidal) motions by Orkisz et al. (2017), in contrast to the rest of the field of view.

5.3.4. Interpretation: enrichment of HCO^+ and CN with respect to $C^{18}O$ in FUV-illuminated gas

The PDFs of HCO^+ vs. $C^{18}O$ and CN vs. $C^{18}O$ show a dual regime at high $C^{18}O$ (1–0) intensity (cf. Fig. 12, right panel, and Fig. 16, left panel), with one regime where HCO^+ and CN are overluminous relative to $C^{18}O$. This latter regime is clearly associated with high FUV illumination at high volume density. We here check whether this is the sign of a chemical enrichment of HCO^+ and CN in FUV-illuminated regions, using a non-LTE code (RADEX, van der Tak et al. 2007) to estimate the column densities of these species.

The modeling details and the derived column densities are described in Appendix E. For simplicity, we only model the radiative transfer for typical conditions in each group (median gas volume density, median kinetic temperature, and median line integrated intensities). As an estimate of the kinetic temperature, we take a combination of the dust temperature and the ^{12}CO excitation temperature as in the Appendix A of Orkisz et al. (2017): we take the ^{12}CO excitation temperature when it is above 60 K, and the maximum of the dust temperature and ^{12}CO excitation temperature otherwise.

Figure 18 shows the behaviour of the derived abundance ratios $[HCO^+]/[C^{18}O]$, and $[CN]/[C^{18}O]$ as a function of the $G_0/n_{\rm H}$ parameter, which is expected to be the dominant controlling parameter of the physics and chemistry in a PDR (Hollenbach & Tielens 1997). An increase by about two orders of magnitude of the HCO^+ and CN abundances relative to $C^{18}O$ can be seen when increasing $G_0/n_{\rm H}$. The abundances

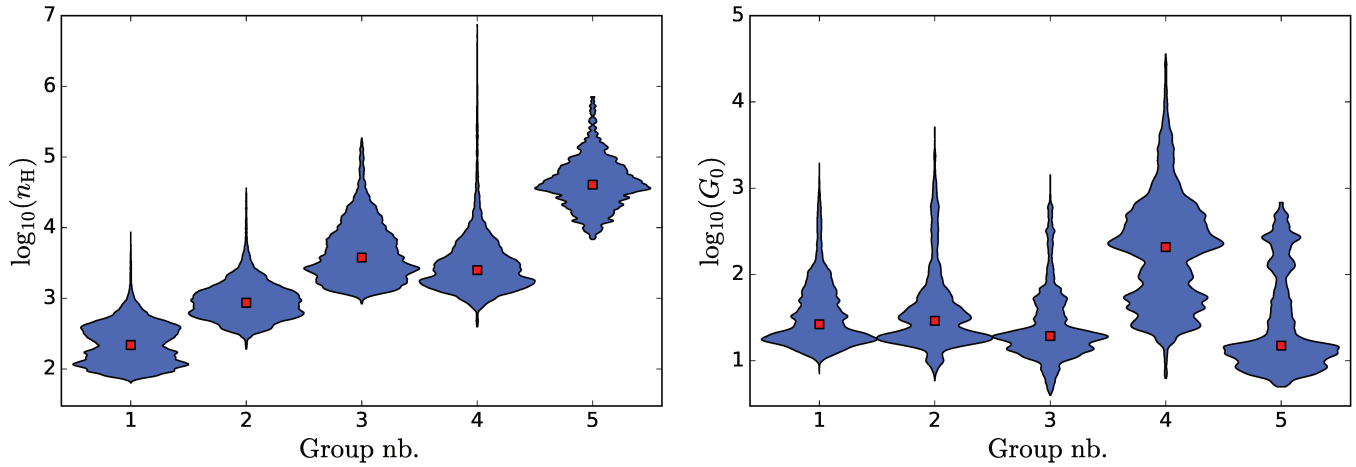


Fig. 17. Violin plots showing the PDF (blue profiles) and median values (red squares) of the approximate $n_{\rm H}$ (*left panel*) and G_0 (*right panel*) in each of the groups CN-1 to 5.

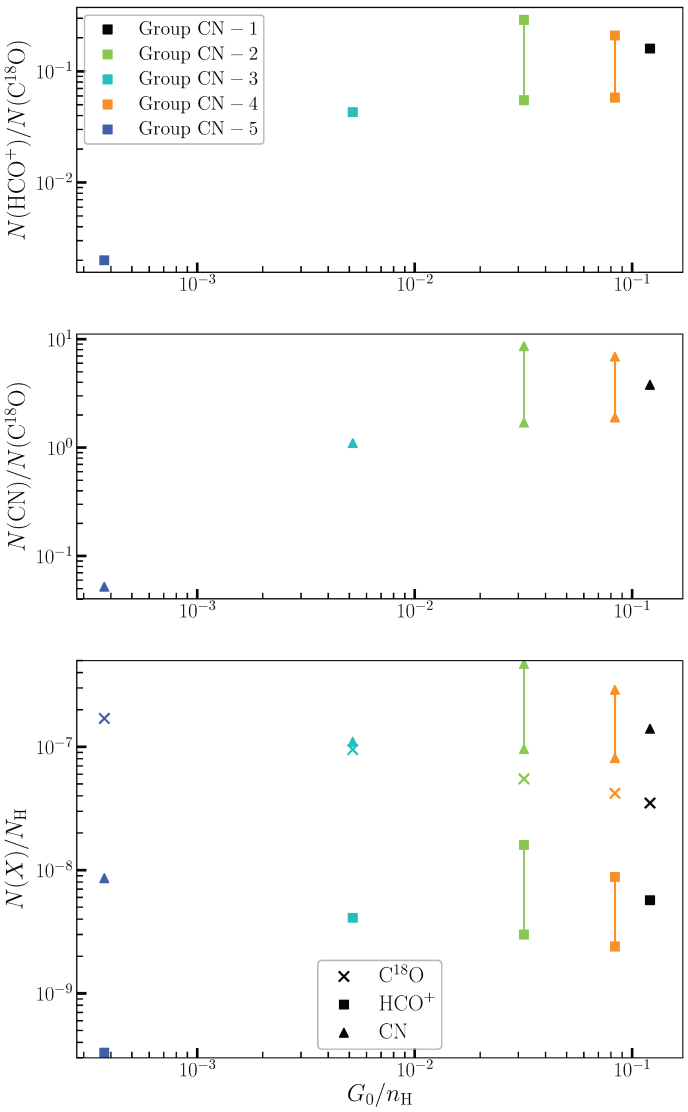


Fig. 18. Column density ratios $N(\mathrm{HCO^+})/N(\mathrm{C^{18}O})$ (*top panel*) and $N(\mathrm{CN})/N(\mathrm{C^{18}O})$ (*middle panel*) as a function of the $G_0/n_{\rm H}$ ratio for the five CN groups. *The lower panel* gives fractional abundances for each of the three species. A range of values is given for CN and $\mathrm{HCO^+}$ when the ionisation fraction is uncertain (with the limits being $x_{\rm e} = 0$ and $x_{\rm e} = 1.4 \times 10^{-4}$).

relative to $\mathrm{C^{18}O}$ seem to reach a constant value at high $G_0/n_{\rm H}$ values in both cases.

Fractional abundances for each of the three species are also computed relative to the total (dust-derived) column density $N_{\rm H}$. These abundances are shown in the lower panel of Fig. 18 as a function of $G_0/n_{\rm H}$. Both the $\mathrm{HCO^+}$ and CN abundances steeply increase with $G_0/n_{\rm H}$ with $G_0/n_{\rm H}$ at first (up to $G_0/n_{\rm H} \sim 10^{-2}$), before saturating at a nearly constant value up to $G_0/n_{\rm H} \sim 10^{-1}$. The $\mathrm{C^{18}O}$ abundance smoothly decreases with increasing $G_0/n_{\rm H}$ over the full range of values. This marked difference of chemical behaviour therefore explains the difference in line intensities identified by the MeanShift algorithm. The abundances of CN and $\mathrm{HCO^+}$ remain at a high but nearly constant value in strongly or mildly FUV-illuminated regions, while the $\mathrm{C^{18}O}$ abundance decreases with increasing FUV illumination. The overbright $\mathrm{HCO^+}$ and CN regime found in FUV-illuminated regions is thus caused by a combination of photochemical enrichment in CN and $\mathrm{HCO^+}$ and photodissociation of $\mathrm{C^{18}O}$.

6. The Meanshift algorithm, an interesting clustering method: biases and data requirements

We first present why and how clustering and PCA are complementary. We then discuss the effects (noise, sampling, dimensionality) that can alter our results.

6.1. On the complementarity of clustering and principal component analyses

We chose here to use a clustering approach in order to analyse the structure of the multi-dimensional PDF of several line intensities, based on the idea that this structure can reveal interesting insights into the physics and chemistry at play. As soon as we use a dataset with more than two dimensions, visualising the structure of its PDF becomes difficult, and specific methods must be used. Each of these methods is usually focused on highlighting a particular kind of structure, and applying different methods to the same dataset thus provides complementary results.

For instance, clustering provides a complementary approach to PCA. On the one hand, PCA highlights the non-sphericity of the data by revealing the axes of strong covariance or correlation. However, PCA cannot capture non-linear patterns of co-variations between the intensities. Moreover, PCA highlights variations around a centre of the dataset (usually the mean),

which might not be relevant if the data points are gathered in several natural clusters, with different centres (that is, if the PDF is multimodal). On the other hand, clustering algorithms aim at revealing any grouping of the data points in different regions of intensity space, that is, to reveal multimodality in the (multi-dimensional) PDF of the data.

By applying PCA to this dataset, Gratier et al. (2017) showed that column density, volume density, and FUV illumination are some of the underlying parameters controlling the intensity variations. And they listed the tracers that are the most affected by each of these parameters. The clustering analysis we have performed in the present study reveals, in addition, a multimodality of the line intensity PDF with modes related to the density and the FUV illumination. In the case of the CO clustering, the modes of the CO isotopologue PDF are directly related to modes of the column/volume density PDF, thus revealing the existence of distinct density regimes in the Orion B cloud. In our clustering analysis including HCO^+ and CN, the transition to a photon-dominated chemistry leads to a separate mode corresponding to dense PDRs. In this case, clustering reveals a transition between two different physical/chemical regimes.

Moreover, both approaches can be used as data compression methods, in order to reduce the volume of data before applying some other very time-consuming data analysis. PCA compresses the dataset by reducing the number of variables characterising each data point, while clustering can be used as a segmentation method, discretising the possible values of each variable. Finally, these two approaches could be combined in future work. One possibility would be to decompose a multi-line PDF into separate components with simpler structure before applying PCA to each of these components. Alternatively, a PCA analysis could be performed first to eliminate irrelevant components of the data (e.g. noise), followed by a clustering analysis restricted to the relevant features deduced by the PCA.

6.2. What is the impact of noise?

The effect of the measurement noise present in our line-intensity maps on the results of the Meanshift algorithm can be understood in two ways: 1) Its effect on the data PDF, used by the Meanshift algorithm to define clusters; and 2) its effect on the attribution of a given pixel to one of the clusters.

Assuming an identical noise rms, σ, on all datapoints (a good approximation for this dataset, see Pety et al. 2017), the addition of Gaussian noise to the true variables replaces the underlying PDF by an observed PDF that is the convolution of the underlying PDF with a Gaussian of standard deviation σ. This is equivalent to a Gaussian smoothing of the PDF. This effect will mostly tend to merge some maxima if their separation is too small compared to the smoothing scale (that is, the noise level), rather than creating artificial maxima. The smoothing effect can also slightly shift the position of the extrema. But the existence of the clusters will be unaffected as long as their PDF maxima are well separated compared to the smoothing scale.

Noise also alters the boundaries between the clusters. If the true intensity values of a given pixel place it close enough to a boundary between clusters in the line space (typically closer than the noise level), adding noise can move this pixel across the border, and thus change the cluster to which it belongs. As a result, noise on the line intensities tends to make the spatial boundaries between clusters appear noisy (not forming a regular curve on the map). This effect is more pronounced in regions of the map where intensity gradients are small. In this case, pixels relatively far away from the cluster spatial boundary can still be close to the cluster boundary in the line space and thus be transferred to another cluster. On the contrary, if intensity gradients are steep at the spatial boundary, even pixels located just one or two pixels away from the spatial boundary can be far enough from the cluster boundary in the line space so that noise is unlikely to transfer them to another cluster. This effect is at play when we include CN and HCO^+ which have low S/Ns on significant fractions of the map. Relatively noisy boundaries can be seen on the eastern side of the cloud while the boundaries on the western edge remain sharp (cf. Fig. 9). The intensity gradients are indeed much steeper on the western side of the cloud than on the eastern side.

6.3. What is the impact of limited sampling (field of view)?

The Meanshift algorithm estimates the PDF gradient in the line space in order to find the PDF maxima, and it needs to estimate this gradient from a finite sample (the observed dataset). There are two different aspects here. First, the observed field of view may be biased towards some values of the parameters that control the physics or the chemistry. Pety et al. (2017) showed that this is the case here, as the studied field of view has a large FUV illumination compared to the ISRF because it includes several HII regions. In the ORION-B project, we will increase the observed field of view towards regions of lower FUV illumination to circumvent this limitation. In the meantime, our clustering analysis must be interpreted with this limitation in mind.

Second, the gradient estimate can be affected by sampling noise: a different dataset drawn from the same underlying PDF (corresponding to this specific field of view) would yield slightly different gradient estimates, and thus converge towards slightly different maxima. We chose to use an adaptive bandwidth (so that the kernel always includes the same number of datapoints), rather than a fixed bandwidth (which would give better sampled gradient estimates close to the major PDF peaks than in the tails) to reduce the impact of this effect. This ensures that the sampling noise is similar for all gradient estimations, and it avoids finding sampling-noise-induced artificial maxima in the tail of the PDF. Having at least a few hundred datapoints in the kernel generally ensures that the sampling noise has negligible effect.

However, implicitly increasing the kernel bandwidth in low-density regions of the PDF means a decreased capacity to resolve small-scale features in the PDF. As a result, PDF maxima corresponding to a small number of pixels (compared to the kernel size) might be smoothed out unless these pixels have intensities widely different from all other pixels. This means that our analysis is likely to miss specific physical or chemical regimes if they occur on too small a region of the map. This was the case in our tests where we found that dense cores are difficult to capture as a cluster with the Meanshift algorithm even when including specific dense core chemical tracers such as N_2H^+.

6.4. Choice of the number of molecular lines included in the analysis

In contrast to our PCA study (Gratier et al. 2017), we limited the clustering analysis to a moderate number of lines (5 at most). This choice was driven by several considerations.

The first reason is practical. As discussed in Sect. 6.2, while including low S/N line maps may change the total number of clusters, the presence of such low-S/N data always degrades the quality of the cluster boundaries. Well defined clusters with clear interfaces are only obtained in the regions where all lines have high S/N. A consequence of using several lower-brightness lines

at relatively constant noise level (our observing case) is therefore a reduction of the usable pixels to the regions where all the lines have high S/N, that is the regions of highest column densities. This requirement of significant S/N for all line maps limits the usefulness of the Meanshift algorithm when applying it to a large set of lines of varying S/N. We thus restricted our study to lines showing extended emission with high S/N on a large fraction of the map.

The second reason is more fundamental. In this paper, we try to understand the physical and chemical processes that regulate the intensities of the lines used in the clustering algorithm. The clustering of the CO isotopologue ground-state line maps, complemented by the HCO^+ and CN lines, showed two trends. First, the number of significant clusters increases with the number of lines because we add lines that exhibit different sensitivities to the physical or chemical processes at work. Second, the interpretation of a large number of clusters is difficult because the associated clusters have less data points implying a lower statistical significance of the trends. Moreover, it is difficult to get a good appreciation of the full distribution of the data over which the Meanshift algorithm operates in dimensions larger than two through standard 2D PDFs.

Our experience is thus that it is better to start clustering in low dimension to understand the sensitivity of the different lines to the many underlying physical and chemical processes at play in the ISM. It will then be possible to cluster higher-dimension data to get a finer segmentation that will depend on the underlying properties that need to be emphasised.

7. Conclusions

In this paper, we present a segmentation of the Orion B molecular cloud into regions of similar molecular emission, in order to reveal the different physical and chemical phases constitutive of molecular clouds. We have applied the Meanshift algorithm, a PDF-based (unsupervised) clustering algorithm defining clusters around the maxima of the PDF, to the (high-dimensional) multi-line PDF of our dataset. This is the first application of a clustering analysis based on molecular emission properties only (and not spatial proximity between pixels) to ISM data.

We first applied the clustering analysis to the maps of the three main CO isotopologue lines only. While the clustering did not take the spatial distribution of the CO emission into account, it highlighted a nested pattern from the outer edges to the innermost parts of the Orion B cloud. Comparison with an approximate volume density map showed that the clusters have increasing typical volume densities with significant grouping at densities of 100, 500, $>1000\,\mathrm{cm}^{-3}$. The CO isotopologue maps alone were thus found to be sufficient to reveal the existence of the diffuse, translucent and high-column-density regimes. Simple LTE radiative transfer modelling implies that the gas emitting the ^{12}CO (1–0) line is more extended than the gas emitting the ^{13}CO/C^{18}O (1–0) lines. It is thus impossible to use the flux ratio of these lines to try to infer the [^{12}CO]/[^{13}CO] and [^{12}CO]/[C^{18}O] abundance ratios.

In the densest regime, an additional separation of the PDF in two distinct tails was found, which we could associate to FUV-illumination effects. Comparison with LTE radiative transfer models shows that this distinction is related to the presence of a warmer ^{12}CO-traced surface layer and higher than usual [^{13}CO]/[C^{18}O] ratios, which can both be explained by the increased FUV illumination caused by the nearby HII regions. It however proved insufficient to get a satisfactory separation of the FUV illuminated regions.

We thus added two FUV-sensitive tracers (the (1–0) lines of HCO^+ and CN) to the CO isotopologues, and performed a second clustering analysis. This analysis revealed a similar separation into increasing density regimes, but captured finer distinctions at higher density ($n_\mathrm{H} \sim 10^4$ and $5\times10^4\,\mathrm{cm}^{-3}$) due to the high critical density of the added tracers. Pety et al. (2017) have however shown that about half of the HCO^+ flux over the observed field of view is coming from diffuse and translucent regions, implying that the use of the HCO^+ line intensity as a tracer of high density gas ($\gtrsim10^4\,\mathrm{cm}^{-3}$) in unresolved GMC observations is questionable.

Moreover, the clustering also revealed the existence of another clear separation of the data at high column density. On the one hand, part of the data presents a CN and HCO^+ (1–0) emission that is overly bright with respect to the C^{18}O (1–0). This data also shows a high CN/HCO^+ intensity ratio. The associated lines of sight form the dense PDR regions around the star forming regions NGC 2023 and NGC 2024, and on the Orion B western edge illuminated by σ Ori (including the Horsehead PDR). On the other hand, other high-column-density lines of sight have low CN/C^{18}O, HCO^+/C^{18}O, and CN/HCO^+ intensity ratios. These lines of sight correspond to the FUV-shielded, dense regions in the inner parts of Orion B. Non-LTE models show that this distinction is related to a clear increase of the [HCO^+]/[C^{18}O] and [CN]/[C^{18}O] abundance ratios with G_0/n_H.

Our clustering analysis based on the (1–0) lines of the CO isotopologues, HCO^+ and CN, thus managed to both capture finer density categories in the densest regions, and to reveal the existence of two distinct chemical phases (characterised by different abundance ratios) corresponding to FUV-induced photochemistry and shielded-gas chemistry. This exposes the wealth of physical and chemical information that can be inferred from molecular tracers when powerful statistical methods (as the Meanshift algorithm) are applied to large amounts of data. One of the next steps in the ORION-B project is to stack the spectra inside each of the clusters found here, to better characterise the molecular content of each regime using the whole information available in the 3 mm band.

Acknowledgements. The authors thank the referee for his constructive comments, useful suggestions, and his very fast refereeing. E.B. post-doctoral position during this work was funded by the ERC grant ERC-2013-Syg-610256-NANOCOSMOS. This work was in part supported by the Programme National "Physique et Chimie du Milieu Interstellaire" (PCMI) of CNRS/INSU with INC/INP, and co-funded by CEA and CNES. We thank the CIAS for its hospitality during the three workshops devoted to this project. JRG thanks the Spanish MINECO for funding support through grant AYA2012-32032. NRAO is operated by Associated Universities Inc. under contract with the National Science Foundation. This paper is based on observations carried out at the IRAM-30 m single-dish telescope. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain). This research also used data from the *Herschel* Gould Belt survey (HGBS) project (http://gouldbelt-herschel.cea.fr). The HGBS is a *Herschel* Key Programme jointly carried out by SPIRE Specialist Astronomy Group 3 (SAG 3), scientists of several institutes in the PACS Consortium (CEA Saclay, INAF-IFSI Rome and INAF-Arcetri, KU Leuven, MPIA Heidelberg), and scientists of the *Herschel* Science Center (HSC).

References

Allison, A. C., & Dalgarno, A. 1971, A&A, 13, 331
André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
Bishop, C. 2006, Pattern Recognition and Machine Learning (New York: Springer-Verlag)
Black, J. H., & van Dishoeck, E. F. 1991, ApJ, 369, L9
Carpenter, J. M. 2000, AJ, 120, 3139
Colombo, D., Rosolowsky, E., Ginsburg, A., Duarte-Cabral, A., & Hughes, A. 2015, MNRAS, 454, 2067

Comaniciu, D., & Meer, P. 2002, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 603
Comaniciu, D., Ramesh, V., & Meer, P. 2001, in Proc. 8th Intl. Conf. on Computer Vision, 438
Epanechnikov, V. A. 1969, Theory of Probability and Its Applications, 14, 153
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Knowledge Discovery and Data Mining Conference (AAAI Press), 226
Faure, A., & Tennyson, J. 2001, MNRAS, 325, 443
Feigelson, E. D., & Babu, G. J. 2012, Modern Statistical Methods for Astronomy (Cambridge: Cambridge University Press)
Flower, D. R. 1999, MNRAS, 305, 651
Fuente, A., García-Burillo, S., Usero, A., et al. 2008, A&A, 492, 675
Georgescu, B., Shimshoni, I., & Meer, P. 2003, in Computer Vision, Proc. Ninth IEEE International Conference, 456
Goldsmith, P. F., & Kauffmann, J. 2017, ApJ, 841, 25
Gratier, P., Bron, E., Gerin, M., et al. 2017, A&A, 599, A100
Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421
Hollenbach, D. J., Takahashi, T., & Tielens, A. G. G. M. 1991, ApJ, 377, 192
Hollenbach, D. J., & Tielens, A. G. G. M. 1997, ARA&A, 35, 179
Inglada, J., Vincent, A., Arias, M., et al. 2017, Remote Sensing, 9, 95
Jolliffe, I. 2002, Principal Component Analysis, Springer Series in Statistics (New York: Springer-Verlag)
Kirk, H., Di Francesco, J., Johnstone, D., et al. 2016, ApJ, 817, 167
Krčo, M., & Goldsmith, P. F. 2016, ApJ, 822, 10
Lada, E. A. 1992, ApJ, 393, L25
Langer, W. D., Graedel, T. E., Frerking, M. A., & Armentrout, P. B. 1984, ApJ, 277, 581
Lique, F., Spielfiedel, A., Feautrier, N., et al. 2010, J. Chem. Phys., 132, 024303
Liszt, H. S. 2012, A&A, 538, A27
Liszt, H. S., & Pety, J. 2016, ApJ, 823, 124
Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, A&A, 566, A45
Martin-Zaïdi, C., Deleuil, M., Le Bourlot, J., et al. 2008, A&A, 484, 225
McCauley, P. I., Mangum, J. G., & Wootten, A. 2011, ApJ, 742, 58
Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, AJ, 151, 5
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Orkisz, J. H., Pety, J., Gerin, M., et al. 2017, A&A, 599, A99
Parzen, E. 1962, Ann. Math. Statist., 33, 1065
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Machine Learning Res., 12, 2825
Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A1
Rosenblatt, M. 1956, Ann. Math. Statist., 27, 832
Rosolowsky, E., & Leroy, A. 2006, PASP, 118, 590
Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, ApJ, 786, 29
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Shimajiri, Y., André, P., Braine, J., et al. 2017, A&A, 604, A74
Snow, T. P., & McCall, B. J. 2006, ARA&A, 44, 367
Sobolev, V. V. 1960, Soviet Ast., 4, 1
Stutzki, J., & Guesten, R. 1990, ApJ, 356, 513
Tremblin, P., Schneider, N., Minier, V., et al. 2014, A&A, 564, A106
van der Tak, F. F. S., Black, J. H., Schöier, F. L., Jansen, D. J., & van Dishoeck, E. F. 2007, A&A, 468, 627
Ward-Thompson, D., Nutter, D., Bontemps, S., Whitworth, A., & Attwood, R. 2006, MNRAS, 369, 1201
Williams, J. P., de Geus, E. J., & Blitz, L. 1994, ApJ, 428, 693
Wilson, T. L., & Rood, R. 1994, ARA&A, 32, 191
Yang, B., Stancil, P. C., Balakrishnan, N., & Forrey, R. C. 2010, ApJ, 718, 1062

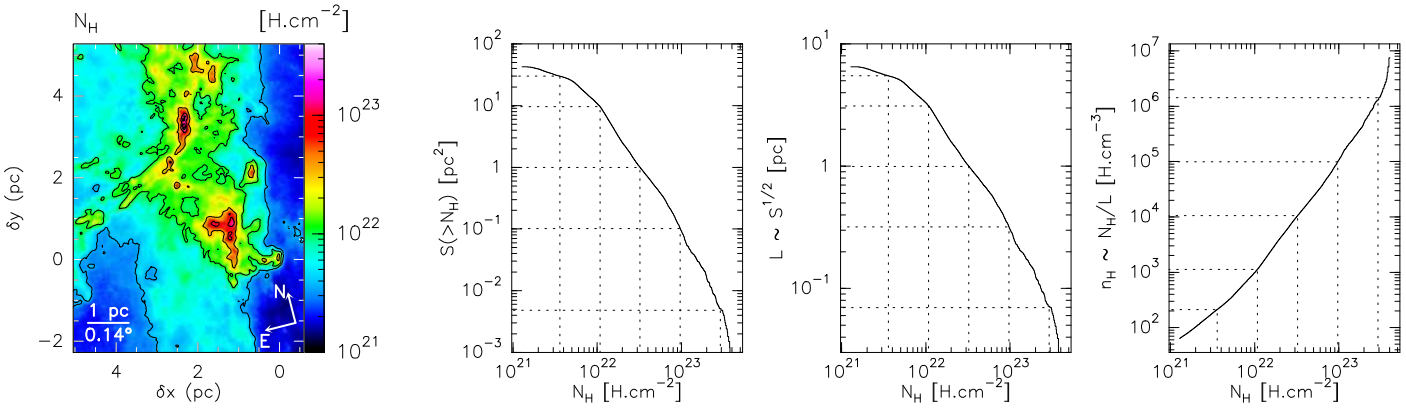


Fig. A.1. Steps to statistically derive an approximate volume density from the column density under the assumptions of isotropy and of nested distributions of increasing density. *Panel a*: spatial distribution of the column density of gas as deduced from the spectral energy distribution of the dust continuum emission. *Panel b*: surface area of the observed field of view that has a column density larger than a lower limit. *Panel c*: typical depth (square root of the previous surface) associated with a given column density. *Panel d*: typical volume density associated with a given column density. The contour levels in *panel a* and the dashed vertical lines in *panels b to d* take the column density values corresponding to a visual extinction of 2, 6, 18, 54, and 162 mag.

Appendix A: Column density and approximate volume density

In this section, we explore whether the assumptions of isotropy and of a nested distribution of volume density (density smoothly increasing from the outer regions to the inner parts of the cloud) can be used to derive an approximate volume density from the spatial distribution of the column density. After a discussion of the spatial and statistical distributions of the column density, we use the above assumptions to derive an approximate volume density. We then compare the derived values to previously published measures of the density in this field of view to understand the limits of the method and to estimate its accuracy.

Figure A.1 shows the spatial (panel a) and statistical (panel b) distributions of the gas column density deduced from the dust continuum emission (see Sect. 2.2). The targeted region exhibits a large range of column densities. There is slightly more than a factor of 100 between the minimum and maximum values. A hypothesis of nearly constant gas volume density would imply that the high-column-density regions should be about one hundred times deeper than the low-column-density ones, which would require an unrealistic cloud geometry. Moreover, the spatial distribution of the column density shows a nested pattern. Indeed the highest-column-density contours are surrounded by the smaller-column-density ones. This is also easily seen on the column density PDF (Fig. 1), which shows a faint tail at high column density. Higher-column-density regions must thus be less extended along the line of sight (according to our isotropy hypothesis), and as a consequence be associated with higher volume densities, at least in the statistical sense.

Assuming the simplest possible hypotheses about the spatial distribution of volume density, that is the hypothesis of no privileged direction and nested increasing volume density, we estimate the typical lengthscale l of regions with a given column density value x as the square root of the projected surface area where $N_{\rm H} \geq x$. Pixels with $N_{\rm H}$ are then given the volume density $n_{\rm H} = N_{\rm H}/l$. This procedure is illustrated in the last three panels of Fig. A.1. This reasoning assumes a one to one correspondence between a column density and the approximate volume density. We only expect this relationship to hold in a statistical sense: it may not be valid pixel by pixel but we expect it to correctly represent the range of volume densities at a given column density.

In order to estimate the accuracy of this estimate, we gathered volume density estimations from the literature at positions within our field of view, and derived from a variety of different methods. From the catalogue of cores of Kirk et al. (2016) (SCUBA dust emission observations), we derived core masses from their Eq. (3) (using the background-substracted fluxes), and we complemented the mass value with their measured size to yield volume densities. We completed this sample with a few density estimates at particular positions. The volume density of the diffuse foreground gas of HD 38087 was derived by Martin-Zaïdi et al. (2008) from H_2 UV absorption lines through the use of PDR models. The volume densities for two positions in NGC2024 was inferred from H_2CO emission lines (McCauley et al. 2011). And the volume densities in the B33-SMM1 and B33-SMM2 cores in the Horsehead were derived from dust emission observations (Ward-Thompson et al. 2006). Figure A.2 shows our volume density estimates compared to the ones from the literature. In addition, the red points and their error bars show the average of our density estimate in log-spaced bins of the literature density values.

On average, our density estimate is within a factor of 3 of the densities from the literature, with a typical scatter of one order of magnitude. For the diffuse medium data point of Martin-Zaïdi et al. (2008), we find a density ~10 times larger. This is consistent with the typical scatter, but could also come from a bias of our estimate at low densities, as our lengthscale estimate is limited to the size of our field of view while diffuse medium might span larger scales. We note that our estimate is however qualitatively correct by predicting a diffuse-medium-like density at this position.

As an additional check, we consider a simple analytical example: a spherically symmetrical cloud whose volume density profile is

$$n(r) = n_0 \left(1 + \frac{r^2}{r_0^2}\right)^{-\frac{1+\alpha}{2}} \quad \text{with} \quad \alpha \geq 1, \tag{A.1}$$

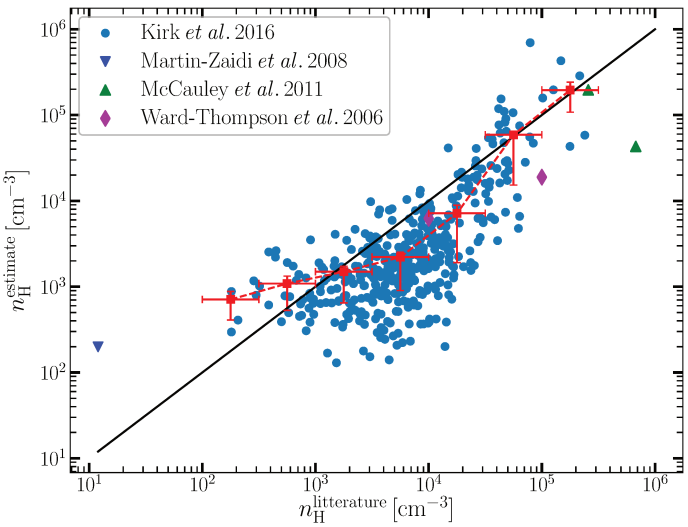


Fig. A.2. Approximate estimation of the volume density as a function of independent estimations from the literature. The solid black line shows a perfect one-to-one relationship, while the red points and their error bars show the bin averaged relationship. The horizontal and vertical error bars indicates the bin size and the interquartile interval, respectively.

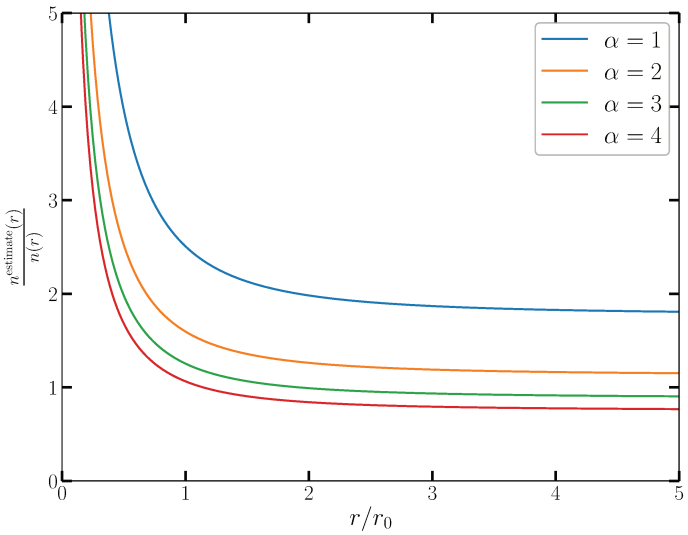


Fig. A.3. Approximate density estimate divided by the true density for different analytical spherical density profiles as a function of distance to the centre.

(used for instance in Krčo & Goldsmith 2016). This corresponds to an observed column density profile

$$N(r) = \frac{n_0 r_0 \Gamma(\frac{\alpha}{2}) \sqrt{\pi}}{\Gamma(\frac{\alpha+1}{2})} \left(1 + \frac{r^2}{r_0^2}\right)^{-\frac{\alpha}{2}}. \tag{A.2}$$

Our simple estimate of the volume density would yield at a distance r from the cloud centre

$$n^{\rm estimate}(r) = \frac{N(r)}{\sqrt{\pi r^2}}. \tag{A.3}$$

The ratio of the two values,

$$\frac{n^{\rm estimate}(r)}{n(r)} = \frac{r_0}{r} \sqrt{1 + \frac{r^2}{r_0^2}} \frac{\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{\alpha+1}{2})}, \tag{A.4}$$

is shown on Fig. A.3 for different values of α. Our estimation ranges from 0.8 to 2.5 times the true density value for all estimation scales larger than the typical scale r_0 and it diverges when the estimation scale is much smaller than r_0. This comes from the fact that the surface estimate describing the scale of the inner region goes to zero close to the centre. However, the spatial resolution of our data (60 mpc) avoids the regime $r \ll r_0$. Indeed, our comparison to literature values does not show any dramatic overestimation for the high-column-density pixels.

In conclusion, the method proposed here provides a reasonable estimate in a statistical sense with a bias of a factor 3 at most and a typical scatter of one order of magnitude. It can thus be trusted for order-of-magnitude comparisons. Moreover, this estimation of the volume density map is completely independent from the clustering analysis presented in this paper, and is only used to help in the interpretation of the clustering analysis.

Appendix B: Impact of the number of neighbours on the number of clusters

In the Adaptive Meanshift algorithm, $k_{\rm neighbors}$ controls the size of the adaptive bandwidth of the smoothing kernel: for each datapoint the bandwidth is automatically adjusted to include its $k_{\rm neighbors}$ nearest neighbours. As a result, this parameter controls the degree of smoothing applied to the PDF gradient estimate, while ensuring equal sampling in the peaks and tails of the PDF.

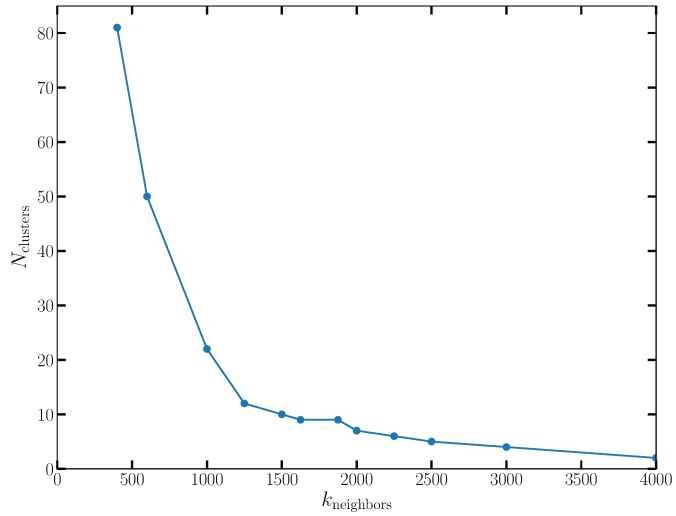


Fig. B.1. Variation of the number of clusters found $N_{\rm clusters}$ with the number of neighbours $k_{\rm neighbors}$ in the adaptive kernel of the Meanshift algorithm, when applied to the three CO isotope lines only.

Most of the time, increasing $k_{\rm neighbors}$ will result in a decrease in the number of clusters found by the algorithm as local maxima are merged by the increased smoothing. As discussed in Sect. 6.3, too low a value of $k_{\rm neighbors}$ results in finding artificial maxima caused by the sampling noise. A very large value will smooth the data PDF to a single peak with the shape of the kernel. When decreasing $k_{\rm neighbors}$ from this large value, the most well separated maxima of the PDF will be distinguished first, then maxima that are weaker or close to the highest maxima.

The first maxima that appear are thus likely to correspond to the strongest physical distinctions, and the following ones to subtler and subtler distinctions (until artificial maxima caused by sampling noise start to appear). As a result, the choice of $k_{\rm neighbors}$

is mostly a choice of the level of detail we want in our analysis, as long as it is large enough to get rid of sampling noise effects.

Figure B.1 shows the variation of the number of clusters found as a function of $k_{\rm neighbors}$, for the clustering analysis of the three CO isotopologue lines (cf. Sect. 4). We see a flat plateau close to 10 clusters starting from $k_{\rm neighbors} > 1200$, indicating that these clusters are more strongly separated than the ones appearing at lower values of $k_{\rm neighbors}$ (they do not disappear with increased smoothing until much larger smoothing). We thus chose a value corresponding to this plateau.

Appendix C: Characteristic intensities in the clusters

Table C.1 lists characteristic intensity values of the three CO isotopologues for the clusters derived from CO isotopologues only. The values given correspond to the mean, median, standard deviation inside each cluster, as well as the most represented intensity of the cluster (i.e., the PDF local maxima). The cluster's most represented intensity is generally fainter than the cluster median and mean. This is a natural consequence of the asymmetry of the PDFs that have extended high intensity tails.

Similarly, Table C.2 and C.3 give the characteristic intensity values (median, mean and standard deviation) for the groups of clusters HCO^+-1 to 7 and CN-1 to 5.

Appendix D: Details of the CO LTE modeling

Here, we describe the details of the modeling that allows us to derive the physical and chemical conditions discussed in Sect. 4.4.

D.1. Parametrization

The intensity integrated over the line profile is defined as

$$I_{ij} = T_{ij}^{\rm B}\, w_{ij}, \tag{D.1}$$

with ij = 12, 13, and 18 for the ^{12}CO, ^{13}CO, and C^{18}O (1–0) lines, respectively. In this equation, $T_{ij}^{\rm B}$ is the LTE intensity that is defined as

$$T_{ij}^{\rm B} = T_{ij}^{0}\left[1-\exp\left(-\tau_{ij}\right)\right]\left\{\frac{1}{\exp\left(\frac{T_{ij}^{0}}{T_{ij}^{\rm exc}}\right)-1} - \frac{1}{\exp\left(\frac{T_{ij}^{0}}{T_{\rm cmb}}\right)-1}\right\}, \tag{D.2}$$

where τ_{ij} is the opacity at the line center, $T_{\rm cmb}$ = 2.73 K is the cosmic microwave background temperature, and $T_{ij}^{0} = h\nu/k_{\rm B}$ (=5.53 K for ^{12}CO (1–0), 5.29 K for ^{13}CO, and 5.27 K for C^{18}O).

Assuming LTE, we will parametrize the modeled curves with increasing kinetic temperature ($T^{\rm kin}$) at constant ^{13}CO opacity

$$T_{13}^{\rm exc} = T_{18}^{\rm exc} = T_{\rm min}^{\rm kin} + (T_{\rm max}^{\rm kin} - T_{\rm min}^{\rm kin})\,\frac{i-1}{n-1} \quad \text{with} \quad 1 \le i \le n. \tag{D.3}$$

Moreover, we will use the additional freedom to have a higher excitation temperature for ^{12}CO, i.e.,

$$\frac{T_{12}^{\rm exc}}{T_{13}^{\rm exc}} = \text{cste} \ge 1. \tag{D.4}$$

The variation of the opacity with the temperature is computed using the opacity at 20 K as reference, i.e.,

$$\tau_{ij} = \tau_{ij}^{20\,\rm K}\left\{\frac{20\,\rm K}{T_{ij}^{\rm exc}}\right\}\left\{\frac{1-\exp\left(-\frac{T_{ij}^{0}}{T_{ij}^{\rm exc}}\right)}{1-\exp\left(-\frac{T_{ij}^{0}}{20\,\rm K}\right)}\right\} \tag{D.5}$$

$$\text{with} \quad \frac{\tau_{12}}{\tau_{13}} = \frac{N_{\rm min}(^{12}\rm CO)}{N(^{13}\rm CO)}, \tag{D.6}$$

$$\text{and} \quad \frac{\tau_{13}}{\tau_{18}} = \frac{N(^{13}\rm CO)}{N(\rm C^{18}O)} = \frac{[^{13}\rm CO]}{[\rm C^{18}O]}, \tag{D.7}$$

where N(CO) is the column density of each CO isotopologue (in fact, a lower limit for ^{12}CO), and [CO] is the abundance relative to H_2 of each CO isotopologue. Only the $I(^{13}\rm CO)/I(\rm C^{18}O)$ can be interpreted in terms of relative abundances because the emission of these isotopologues is probably co-spatial, while the ^{12}CO line is mainly sensitive to the foreground part of the emission.

The second factor on the right-hand side of Eq. (D.1) takes care of the integration over a Gaussian line profile, including the opacity broadening term

$$w_{ij} = FWHM\,\sqrt{\frac{1}{\log 2}\,\log\left\{\frac{\tau_{ij}}{\log\left[\frac{2}{1+\exp(-\tau_{ij})}\right]}\right\}}, \tag{D.8}$$

where FWHM is the measured linewidth of the Gaussian profile for an optically thin line. As we limited the range of velocity over which we integrate the CO lines to Δv, we saturate w_{ij} as follows

$$w_{ij} = \Delta v \quad \text{where} \quad w_{ij} > \Delta v. \tag{D.9}$$

D.2. Impact of each input parameter

Here, we describe the specific influence that each input parameter has on the curves in the histograms of Fig. 8. At constant ^{13}CO opacity, a change of the line FWHM has an homothetic effect on the modeled curves in the intensity vs. intensity histograms: the higher the FWHM, the larger the amplitude of the curve. Limiting the interval of velocity over which we integrate quickly leads to a saturation of the ^{13}CO integrated intensities. This is the reason why the upper edge of the ^{13}CO vs. ^{12}CO histogram is relatively sharp. This effect is less obvious for the ^{12}CO line because the opacity is so large that the line is already saturated. Saturation of the ^{13}CO emission also explains the range of observed values of CO isotopologue ratios (bottom left panel). When both ^{12}CO and ^{13}CO get saturated the intensity ratio is mainly controlled by the ratio of the excitation temperatures, with a modest influence of the opacity broadening term.

The minimum kinetic temperature sets the lower edge of the histogram of ^{13}CO vs. C^{18}O. Using a lower temperature would result in curves that go beyond the observed minimum ^{13}CO intensity for each given C^{18}O intensity. The maximum kinetic temperature controls the regions of low intensities at low column density for all three isotopologues. In other words, we do not populate correctly the low intensity part of the histogram when the maximum temperature is too low.

Table C.1. Characteristic line intensities in each CO cluster: PDF maximum of the cluster, median value, mean value, and standard deviation.

	Cluster	CO-1	CO-2	CO-3	CO-4	CO-5	CO-6	CO-7	CO-8	Global
Nb. of pixels		47 440	4943	14 352	13 336	14 109	24 769	15 625	6476	141 050
^{12}CO	center	1.3	13.1	20.8	37.7	45.4	56.7	59.6	93.5	
	median	2.1	16.1	27.3	38.8	49.2	61.5	66.0	96.1	35.5
	mean	3.2	16.1	27.4	38.6	48.9	64.1	69.3	97.9	36.4
	std	2.9	1.9	4.8	5.0	7.2	14.6	15.6	10.5	30.3
^{13}CO	center	0.2	0.8	1.3	3.1	5.0	7.8	13.8	15.2	
	median	0.3	1.0	2.1	4.2	6.1	10.4	18.1	20.8	3.6
	mean	0.4	1.1	2.2	4.4	6.3	10.4	21.5	21.5	6.6
	std	0.3	0.4	0.8	1.4	1.6	2.3	9.7	5.2	8.2
C^{18}O	center	0.0	0.0	0.0	0.1	0.1	0.3	0.8	0.6	
	median	0.0	0.0	0.1	0.1	0.2	0.5	1.6	1.1	0.2
	mean	0.0	0.0	0.1	0.1	0.3	0.5	2.3	1.3	0.4
	std	0.1	0.1	0.1	0.2	0.2	0.3	1.7	0.6	0.9

Notes. All values are in K km s^{-1}.

Table C.2. Characteristic line intensities in each HCO$^+$ group: median value, mean value, and standard deviation.

	Group	HCO$^+$-1	HCO$^+$-2	HCO$^+$-3	HCO$^+$-4	HCO$^+$-5	HCO$^+$-6	HCO$^+$-7	Global
Nb. of pixels		53 049	21 998	20 541	25 681	15 885	2926	970	141 050
^{12}CO	median	2.3	30.7	47.8	60.1	85.0	78.9	95.2	35.5
	mean	5.7	29.4	47.2	60.2	84.2	81.6	98.7	36.4
	std	8.7	8.2	8.3	11.2	17.2	16.2	15.0	30.3
^{13}CO	median	0.3	2.5	6.1	11.7	17.7	27.1	38.0	3.6
	mean	0.6	2.9	6.2	11.7	18.3	27.8	39.7	6.6
	std	1.0	1.7	2.1	3.7	9.2	4.2	6.6	8.2
C^{18}O	median	0.02	0.07	0.2	0.6	1.1	3.3	5.8	0.2
	mean	0.02	0.08	0.3	0.7	1.3	3.5	5.9	0.4
	std	0.2	0.2	0.2	0.5	1.2	0.8	1.1	0.9
HCO$^+$	median	0.2	0.5	0.9	1.3	2.6	3.6	5.4	0.7
	mean	0.3	0.5	0.9	1.4	3.0	4.0	6.3	1.1
	std	0.6	0.3	0.4	0.5	1.6	1.3	2.5	1.3
CN	median	0.07	0.09	0.3	0.4	1.1	1.2	1.9	0.2
	mean	0.1	0.09	0.3	0.5	1.4	1.3	2.2	0.4
	std	0.3	0.3	0.3	0.3	1.3	0.5	0.8	0.7

Notes. All values are in K km s^{-1}.

The $N(^{12}\mathrm{CO})/N(^{13}\mathrm{CO})$ column density ratio controls the observed lower edge of the associated intensity vs. intensity histogram. A too low value underestimates the ^{13}CO intensity at constant ^{12}CO intensity and vice versa. In a similar way, the $[^{13}\mathrm{CO}]/[\mathrm{C}^{18}\mathrm{O}]$ abundance ratio controls the observed upper edge of the associated intensity vs. intensity histogram.

Finally, the ratio of the $T^{\mathrm{exc}}_{^{12}\mathrm{CO}}/T^{\mathrm{exc}}_{^{13}\mathrm{CO}}$ excitation temperatures controls the ^{13}CO "width" of the curves. A higher $T^{\mathrm{exc}}_{^{12}\mathrm{CO}}/T^{\mathrm{exc}}_{^{13}\mathrm{CO}}$ implies both a lower and higher ^{13}CO intensity (at low and high $T^{\mathrm{exc}}_{^{13}\mathrm{CO}}$ value) for the same ^{12}CO intensity. In other words, a higher $T^{\mathrm{exc}}_{^{12}\mathrm{CO}}/T^{\mathrm{exc}}_{^{13}\mathrm{CO}}$ would less well describe both the lower and upper edge of the ^{13}CO vs. ^{12}CO histogram if all other parameters stay constant. A higher $T^{\mathrm{exc}}_{^{12}\mathrm{CO}}/T^{\mathrm{exc}}_{^{13}\mathrm{CO}}$ also implies a higher slope of the $^{13}\mathrm{CO}/\mathrm{C}^{18}\mathrm{O}$ vs. $^{12}\mathrm{CO}/^{13}\mathrm{CO}$ curves.

Appendix E: Details of the CN, HCO$^+$, and C^{18}O RADEX modeling

Here, we describe the detail of our modeling approach for deriving the HCO$^+$ and CN abundances and abundance ratios discussed in Sect. 5.3.4.

E.1. RADEX non-LTE radiative transfer models

The radiative transfer modeling is subtle for two reasons. First, the critical densities of C^{18}O, HCO$^+$, and CN (for collisional excitation with H_2) differ by two to three orders of magnitude ($\sim 2\times10^3$, 2×10^5, and $\sim 2\times10^6$ cm^{-3}, respectively). The emission of C^{18}O is mostly thermalized in all groups (except CN-1). In contrast, the excitation of HCO$^+$, and CN is subthermal. We

Table C.3. Characteristic line intensities in each CN group: median value, mean value, and standard deviation.

	Group	CN-1	CN-2	CN-3	CN-4	CN-5	Global
Nb. of pixels		84 631	36 638	6991	11 820	970	141 050
^{12}CO	median	7.6	56.2	68.6	91.9	95.2	35.5
	mean	17.0	55.4	70.7	91.5	98.7	36.4
	std	17.8	12.9	12.0	14.8	15.0	30.3
^{13}CO	median	0.6	9.9	20.6	17.2	38.0	3.6
	mean	1.9	9.9	21.5	18.8	39.7	6.6
	std	2.4	4.3	4.9	10.9	6.6	8.2
C^{18}O	median	0.05	0.5	2.0	0.8	5.8	0.2
	mean	0.06	0.6	2.3	1.3	5.9	0.4
	std	0.2	0.5	1.1	1.4	1.1	0.9
HCO$^+$	median	0.3	1.2	2.3	3.0	5.4	0.7
	mean	0.4	1.3	2.7	3.4	6.3	1.1
	std	0.5	0.5	1.3	1.7	2.5	1.3
CN	median	0.08	0.4	0.8	1.3	1.9	0.2
	mean	0.1	0.5	0.9	1.7	2.2	0.4
	std	0.3	0.3	0.4	1.4	0.8	0.7

Notes. All values are in K km s^{-1}.

Table E.1. Median gas volume density, gas temperature, C^{18}O, HCO$^+$, and CN integrated intensities and associated column densities derived from RADEX models for each CN group.

	Median n_H [cm^{-3}]	Median T [K]	Median N_H [cm^{-2}]	Median I(C^{18}O) [K km s^{-1}]	Median I(HCO$^+$) [K km s^{-1}]	Median I(CN) [K km s^{-1}]	N(C^{18}O) [cm^{-2}]	N(HCO$^+$) [cm^{-2}]	N(CN) [cm^{-2}]
1	2.2×10^2	24	3.7×10^{21}	4.9×10^{-2}	3.3×10^{-1}	7.9×10^{-2}	1.3×10^{14}	2.1×10^{13b}	5.0×10^{14b}
2	8.6×10^2	28	9.3×10^{21}	4.7×10^{-1}	1.2	4.4×10^{-1}	5.1×10^{14}	2.8×10^{13b} –1.5×10^{14a}	8.9×10^{14b} –4.4×10^{15a}
3	3.8×10^3	29	1.9×10^{22}	2.0	2.3	7.6×10^{-1}	1.8×10^{15}	7.7×10^{13a}	2.0×10^{15a}
4	2.5×10^3	42	1.6×10^{22}	8.3×10^{-1}	3.0	1.3	6.7×10^{14}	3.9×10^{13b} –1.4×10^{14a}	1.3×10^{15b} –4.6×10^{15a}
5	4.1×10^4	39	6.6×10^{22}	5.8	5.4	1.9	1.1×10^{16}	2.2×10^{13a}	5.7×10^{14a}

Notes. [(a)] RADEX result for $x_e = 0$. [(b)] RADEX result for $x_e = 1.4 \times 10^{-4}$.

thus use a non-LTE radiative transfer approach. Second, collisions with electrons can dominate the excitation of high dipole moment species such as CN and HCO$^+$ in regions where the electronic fraction is high enough (Black & van Dishoeck 1991; Liszt 2012; Goldsmith & Kauffmann 2017). This effect can be important for the regions of low visual extinction, e.g., groups CN-1 and to a lesser extent CN-2, and for the highly FUV-illuminated PDRs of group CN-4.

We used the RADEX code (van der Tak et al. 2007) which uses the escape probability approach of Sobolev (1960) to compute the non-LTE level populations and the emission from a region of given temperature and volume density of collision partners within a given velocity interval. The position-position-velocity data suggest to use a linewidth of 1 km s^{-1}. The cross-section coefficients for collisional excitation are obtained from the LAMDA database[2]. The coefficient data for collisional excitation of C^{18}O, HCO$^+$, and CN with H_2 were computed by Yang et al. (2010), Flower (1999), and Lique et al. (2010), respectively. The data for the excitation by electrons were computed by Faure & Tennyson (2001), Fuente et al. (2008) for HCO$^+$, and Allison & Dalgarno (1971) for CN.

[2] http://home.strw.leidenuniv.nl/~moldata/

For each CN group, we ran RADEX models for the median values of the volume density and temperature corresponding to each group, and we adjusted the column densities of C^{18}O, HCO$^+$, and CN to best reproduce the median line intensities. As electrons may be important for the collisional excitation of high dipole moment molecules, we used two hypotheses for the electron fraction: (1) The electron fraction x_e is 0 and only H_2 contributes to the excitation; (2) The electron fraction is set by the ionisation of all carbon atoms and $x_e = 1.4 \times 10^{-4}$. Case (2) was computed only for groups where this hypothesis can be relevant : CN-1, 2 and 4. For the diffuse medium of group CN-1, case (2) is a good hypothesis and we take only the corresponding column density values. In group CN-2 and CN-4 however, the electronic fraction is more uncertain, and we give a range of values corresponding to the extreme cases (1) and (2). We assumed a negligible electronic fraction in the other groups.

E.2. Uncertainties

We estimated the uncertainties on the median line intensities by two different methods. First, we used a perturbative Monte Carlo approach. We produced 1000 perturbed datasets (where Gaussian noise is added to each pixel's intensities according to the local noise rms). We then computed the median intensities

inside each group. And we finally took the standard deviation of these results as the uncertainty on the median intensity. Second, we used a bootstrapping method (Feigelson & Babu 2012). We produced 1000 bootstrapped datasets of the same size as the original one (drawing with replacement from the initial dataset). And we took the standard deviation of the medians of these datasets. Both approaches give similar uncertainties. The relative (1σ) uncertainties are below 2% for all groups and for the three molecular lines considered here.

Due to the large number of pixels in each group, the median values are highly statistically significant, even in group CN-1 where the three lines stay undetected in most pixels. However, potential unknown biases in the integrated intensities (baseline distortion, etc.) are not taken into account. As a result, the median values for group CN-1 are less reliable than for the other groups. Our estimates of the volume density and kinetic temperature are likely to suffer from global biases. We thus did not try to estimate a noise-associated uncertainty for these quantities.

E.3. Column densities

Figure E.1 illustrates how the column densities were derived from the radiative transfer models. Table E.1 lists the median input data and the output results. For $C^{18}O$, the curves of all groups are roughly superimposed (thermalized emission) and the derived column densities are mostly proportional to the median intensities of the groups. In contrast, the intensity vs. column density relations are strongly dependent on the median density of the group for HCO^+ and CN. For HCO^+ and CN, the excitation is subthermal, and closer to the weak excitation regime of Liszt & Pety (2016). As a result, the intensity-column density relations are strongly dependent on the median volume density of the group (in the weak excitation regime, the intensity is proportional to the product of column density and volume density).

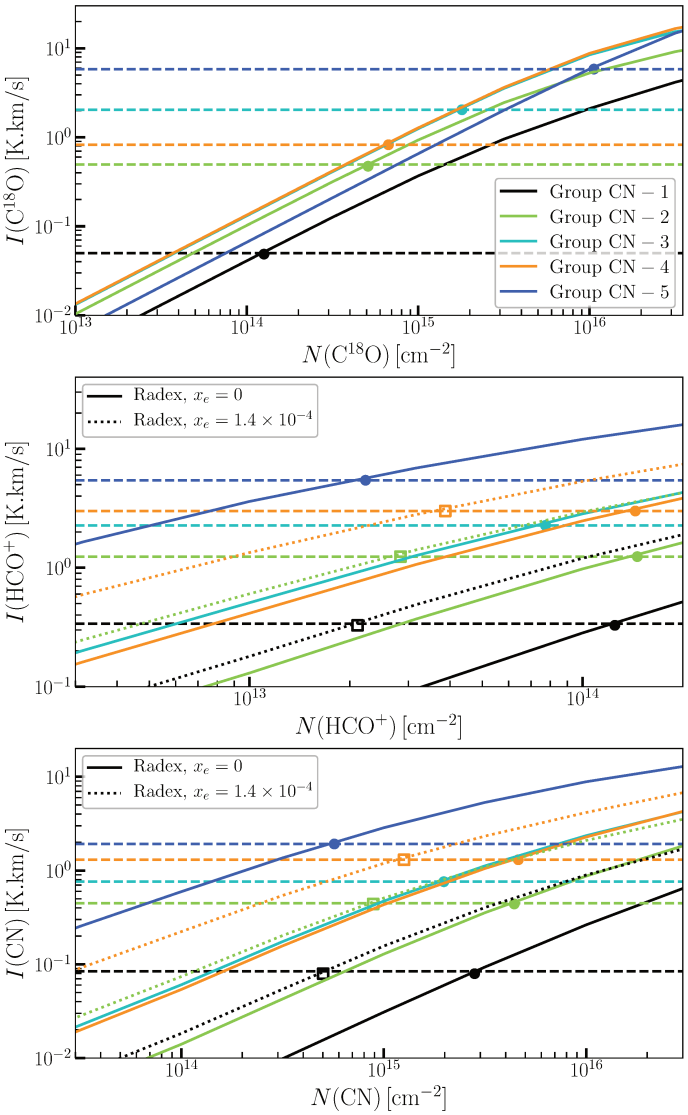


Fig. E.1. Line intensity as a function of column density for non-LTE radiative transfer models at the median gas volume density and temperature for each CN group. The full and dotted lines assume a ionisation fraction of 0 and 1.4×10^{-4}, respectively. The horizontal dashed lines represent the median observed intensity computed for each group. The symbols show the best column density for each group according to the RADEX models, with full circles for models with $x_e = 0$ and open squares for $x_e = 1.4 \times 10^{-4}$.

Bibliography

Aalto, S., Spaans, M., Wiedner, M. C., & Hüttemeister, S. 2007, A&A, 464, 193 (Cited on page 135)

Abergel, A., Bernard, J. P., Boulanger, F., et al. 2002, A&A, 389, 239 (Cited on page 114)

Adam, R., Adane, A., Ade, P. A. R., et al. 2018, A&A, 609, A115 (Cited on page 93)

Agúndez, M. & Wakelam, V. 2013, Chemical Reviews, 113, 8710 (Cited on page 141)

Akbari, V., Doulgeris, A. P., & Eltoft, T. 2014, IEEE Transactions on Geoscience and Remote Sensing, 52, 3729 (Cited on page 24)

Aladro, R., Martín, S., Riquelme, D., et al. 2015, A&A, 579, A101 (Cited on page 133)

Allison, A. C. & Dalgarno, A. 1971, A&A, 13, 331 (Cited on page 180)

Alves, J., Lombardi, M., & Lada, C. J. 2017, A&A, 606, L2 (Cited on page 65)

André, P. 2017, Comptes Rendus Geoscience, 349, 187 (Cited on page 70)

André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, Protostars and Planets VI, 27 (Cited on pages 68, 74, 75, 79, and 90)

André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102 (Cited on pages 25, 50, 68, 79, 80, 88, 113, 148, and 158)

Anthony-Twarog, B. J. 1982, AJ, 87, 1213 (Cited on page 112)

Arzoumanian, D., André, P., Didelon, P., et al. 2011, A&A, 529, L6 (Cited on pages 68, 69, 73, 76, 79, 83, 84, 85, 87, 88, and 95)

Arzoumanian, D., André, P., Peretto, N., & Könyves, V. 2013, A&A, 553, A119 (Cited on pages 70, 73, 76, 79, 85, 89, and 91)

Aylward, S. R. & Bullitt, E. 2002, IEEE transactions on medical imaging, 21, 61 (Cited on page 95)

Baan, W. A., Henkel, C., Loenen, A. F., Baudry, A., & Wiklind, T. 2008, A&A, 477, 747 (Cited on page 135)

Bally, J. 2008, Overview of the Orion Complex, ed. B. Reipurth, 459 (Cited on page 15)

Bally, J., Langer, W. D., Stark, A. A., & Wilson, R. W. 1987, ApJ, 312, L45 (Cited on pages 15 and 67)

Barnard, E. E. 1913, ApJ, 38 (Cited on page 9)

Barnard, E. E. 1919, ApJ, 49 (Cited on page 9)

Beck, M. R., Scarlata, C., Fortson, L. F., et al. 2018, MNRAS, 476, 5516 (Cited on page 24)

Bergin, E. A. & Tafalla, M. 2007, ARA&A, 45, 339 (Cited on page 116)

Beuther, H., Meidt, S., Schinnerer, E., Paladino, R., & Leroy, A. 2017, A&A, 597, A85 (Cited on page 23)

Bigiel, F., Bolatto, A. D., Leroy, A. K., et al. 2010a, ApJ, 725, 1159 (Cited on pages 21 and 26)

Bigiel, F., Leroy, A., Seibert, M., et al. 2010b, ApJ, 720, L31 (Cited on pages 21 and 25)

Bigiel, F., Leroy, A., Walter, F., et al. 2010c, AJ, 140, 1194 (Cited on page 21)

Bigiel, F., Leroy, A. K., Jiménez-Donaire, M. J., et al. 2016, ApJ, 822, L26 (Cited on pages 20, 21, 27, 111, and 125)

Bigiel, F., Leroy, A. K., Walter, F., et al. 2011, ApJ, 730, L13 (Cited on page 21)

Bik, A., Lenorzer, A., Kaper, L., et al. 2003, A&A, 404, 249 (Cited on pages 16 and 111)

Bishop, C. 2006, Pattern Recognition and Machine Learning (Springer-Verlag New York) (Cited on page 159)

Black, J. H. & van Dishoeck, E. F. 1991, ApJ, 369, L9 (Cited on page 180)

Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132 (Cited on pages 80 and 118)

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207 (Cited on pages 25, 89, and 118)

Bouy, H. & Alves, J. 2015, A&A, 584, A26 (Cited on page 15)

Bron, E., Daudon, C., Pety, J., et al. 2018, A&A, 610, A12 (Cited on pages x, 16, 33, 34, 65, 80, 88, 93, 104, 106, and 109)

Brunt, C. M. & Federrath, C. 2014, MNRAS, 442, 1451 (Cited on pages xiii, 46, 47, 50, 51, 52, 53, 57, 58, 60, 64, and 65)

Brunt, C. M., Federrath, C., & Price, D. J. 2010, MNRAS, 403, 1507 (Cited on pages 47, 50, 51, 52, 53, and 57)

Brunt, C. M. & Heyer, M. H. 2013, MNRAS, 433, 117 (Cited on page 153)

Buhl, D. & Snyder, L. E. 1970, Nature, 228, 267 (Cited on page 10)

Burkert, A. & Hartmann, L. 2004, ApJ, 616, 288 (Cited on page 78)

Burkhart, B., Lazarian, A., Goodman, A., & Rosolowsky, E. 2013, ApJ, 770, 141 (Cited on page 107)

Busquet, G., Zhang, Q., Palau, A., et al. 2013, ApJ, 764, L26 (Cited on pages 69, 70, 73, and 79)

Capitanio, L., Lallement, R., Vergely, J. L., Elyajouri, M., & Monreal-Ibero, A. 2017, A&A, 606, A65 (Cited on pages 65 and 106)

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245 (Cited on pages 80 and 114)

Carpenter, J. M. 2000, AJ, 120, 3139 (Cited on pages 59, 60, 89, 90, and 157)

Carter, M., Lazareff, B., Maier, D., et al. 2012, A&A, 538, A89 (Cited on page 19)

Catalano, A., Adam, R., Ade, P. A. R., et al. 2018, Journal of Low Temperature Physics (Cited on page 93)

Chernin, L. M. 1996, ApJ, 460, 711 (Cited on page 116)

Cheung, A. C., Rank, D. M., Townes, C. H., Thornton, D. D., & Welch, W. J. 1968, Physical Review Letters, 21, 1701 (Cited on page 10)

Chira, R.-A., Kainulainen, J., Ibáñez-Mejía, J. C., Henning, T., & Mac Low, M.-M. 2018, A&A, 610, A62 (Cited on pages 68 and 69)

Clarke, S. D., Whitworth, A. P., Duarte-Cabral, A., & Hubber, D. A. 2017, MNRAS, 468, 2489 (Cited on pages 70, 72, 75, and 76)

Clarke, S. D., Whitworth, A. P., & Hubber, D. A. 2016, MNRAS, 458, 319 (Cited on pages 70, 75, and 90)

Colombo, D., Meidt, S. E., Schinnerer, E., et al. 2014, ApJ, 784, 4 (Cited on page 23)

Colombo, D., Rosolowsky, E., Ginsburg, A., Duarte-Cabral, A., & Hughes, A. 2015, MNRAS, 454, 2067 (Cited on pages 24, 33, and 157)

Comaniciu, D. & Meer, P. 2002, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 603 (Cited on page 159)

Comaniciu, D., Ramesh, V., & Meer, P. 2001, in in Proc. 8th Intl. Conf. on Computer Vision, 438–445 (Cited on page 160)

Cooley, J. W. & Tukey, J. W. 1964, Math. Comp., 19, 297 (Cited on page 54)

Crutcher, R. M., Roberts, D. A., Troland, T. H., & Goss, W. M. 1999, ApJ, 515, 275 (Cited on page 41)

Deeming, T. J. 1964, MNRAS, 127, 493 (Cited on page 152)

Dickman, R. L. 1978, ApJS, 37, 407 (Cited on page 51)

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, MNRAS, 476, 3661 (Cited on page 24)

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Cited on pages 9, 12, and 49)

Draine, B. T., Dale, D. A., Bendo, G., et al. 2007, ApJ, 663, 866 (Cited on page 148)

Duarte-Cabral, A. & Dobbs, C. L. 2017, MNRAS, 470, 4261 (Cited on page 67)

Edgeworth, K. E. 1946, MNRAS, 106, 470 (Cited on pages 67, 72, and 73)

Epanechnikov, V. A. 1969, Theory of Probability & Its Applications, 14, 153 (Cited on page 160)

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Knowledge Discovery and Data Mining Conference (AAAI Press), 226–231 (Cited on page 159)

Ewen, H. I. & Purcell, E. M. 1951, Nature, 168, 356 (Cited on page 10)

Fabricius, C., Bastian, U., Portell, J., & others. 2016, ArXiv e-prints (Cited on page 112)

Falgarone, E., Panis, J.-F., Heithausen, A., et al. 1998, A&A, 331, 669 (Cited on page 67)

Falgarone, E. & Phillips, T. G. 1990, ApJ, 359, 344 (Cited on pages 13 and 41)

Falgarone, E. & Puget, J.-L. 1995, A&A, 293, 840 (Cited on page 41)

Faure, A. & Tennyson, J. 2001, MNRAS, 325, 443 (Cited on page 180)

Federrath, C. 2013, MNRAS, 436, 1245 (Cited on page 55)

Federrath, C. 2016, MNRAS, 457, 375 (Cited on pages 68, 72, 73, 78, 79, 87, and 91)

Federrath, C., Chabrier, G., Schober, J., et al. 2011, Physical Review Letters, 107, 114504 (Cited on pages 52, 53, and 57)

Federrath, C. & Klessen, R. S. 2012, ApJ, 761, 156 (Cited on pages 44, 45, 49, 50, 60, and 110)

Federrath, C. & Klessen, R. S. 2013, ApJ, 763, 51 (Cited on pages 50, 59, 89, 90, and 110)

Federrath, C., Klessen, R. S., & Schmidt, W. 2008, ApJ, 688, L79 (Cited on pages 41, 42, 46, and 53)

Federrath, C., Klessen, R. S., & Schmidt, W. 2009, ApJ, 692, 364 (Cited on page 49)

Federrath, C., Roman-Duval, J., Klessen, R. S., Schmidt, W., & Mac Low, M.-M. 2010, A&A, 512, A81 (Cited on pages 42, 43, 44, 45, and 50)

Feigelson, E. D. & Babu, G. J. 2012, Modern Statistical Methods for Astronomy (Cited on pages 147 and 181)

Fiege, J. D. & Pudritz, R. E. 2000a, MNRAS, 311, 85 (Cited on page 69)

Fiege, J. D. & Pudritz, R. E. 2000b, MNRAS, 311, 105 (Cited on page 73)

Fiege, J. D. & Pudritz, R. E. 2000c, ApJ, 534, 291 (Cited on page 74)

Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, in BAAS, Vol. 1, Bulletin of the American Astronomical Society, 240 (Cited on page 11)

Fischer, W. J., Megeath, S. T., Stutz, A. M., et al. 2013, Astronomische Nachrichten, 334, 53 (Cited on page 85)

Fischera, J. & Martin, P. G. 2012, A&A, 542, A77 (Cited on page 74)

Florack, L. M., ter Haar Romeny, B. M., Koenderink, J. J., & Viergever, M. A. 1992, Image and Vision Computing, 10, 376 , information Processing in Medical Imaging (Cited on page 94)

Flower, D. R. 1999, MNRAS, 305, 651 (Cited on page 180)

Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. 1998, in Medical Image Computing and Computer-Assisted Intervention (Springer Berlin Heidelberg), 130–137 (Cited on pages 94, 95, and 97)

Frerking, M. A., Langer, W. D., & Wilson, R. W. 1982, ApJ, 262, 590 (Cited on page 116)

Friesen, R. K., Pineda, J. E., co-PIs, et al. 2017, ApJ, 843, 63 (Cited on page 79)

Fuente, A., García-Burillo, S., Gerin, M., et al. 2005, ApJ, 619, L155 (Cited on page 135)

Fuente, A., García-Burillo, S., Usero, A., et al. 2008, A&A, 492, 675 (Cited on page 180)

Fuente, A., Goicoechea, J. R., Pety, J., et al. 2017, ApJ, 851, L49 (Cited on page 17)

Fukui, Y. & Kawamura, A. 2010, ARA&A, 48, 547 (Cited on page 67)

Furlan, E., Fischer, W. J., Ali, B., et al. 2016, ApJS, 224, 5 (Cited on pages 85 and 86)

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016a, A&A, 595, A2 (Cited on page 112)

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016b, A&A, 595, A1 (Cited on page 112)

Gaida, M., Ungerechts, H., & Winnewisser, G. 1984, A&A, 137, 17 (Cited on page 67)

Gaze, V. F. & Shajn, G. A. 1952, Izvestiya Ordena Trudovogo Krasnogo Znameni Krymskoj Astrofizicheskoj Observatorii, 8, 17 (Cited on page 67)

Geen, S., Hennebelle, P., Tremblin, P., & Rosdahl, J. 2015, MNRAS, 454, 4484 (Cited on page 59)

Georgescu, B., Shimshoni, I., & Meer, P. 2003, in Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, 456–463 vol.1 (Cited on page 160)

Gerin, M., Ruaud, M., Goicoechea, J. R., et al. 2015, A&A, 573, A30 (Cited on page 80)

Gibson, S. J., Taylor, A. R., Higgs, L. A., & Dewdney, P. E. 2000, ApJ, 540, 851 (Cited on page 67)

Girichidis, P., Federrath, C., Banerjee, R., & Klessen, R. S. 2011, MNRAS, 413, 2741 (Cited on pages 44 and 45)

Goicoechea, J. R., Teyssier, D., Etxaluze, M., et al. 2015, ApJ, 812, 75 (Cited on page 132)

Goldsmith, P. F. 2001, ApJ, 557, 736 (Cited on pages 62 and 80)

Goldsmith, P. F., Heyer, M., Narayanan, G., et al. 2008, ApJ, 680, 428 (Cited on pages 67, 69, 70, and 73)

Goldsmith, P. F. & Kauffmann, J. 2017, ApJ, 841, 25 (Cited on page 180)

Gómez, G. C. & Vázquez-Semadeni, E. 2014, ApJ, 791, 124 (Cited on pages 68, 72, and 73)

Goodman, A. A., Barranco, J. A., Wilner, D. J., & Heyer, M. H. 1998, ApJ, 504, 223 (Cited on pages 13 and 40)

Goodman, A. A., Pineda, J. E., & Schnee, S. L. 2009, ApJ, 692, 91 (Cited on pages 43, 44, and 45)

Graninger, D. M., Herbst, E., Öberg, K. I., & Vasyunin, A. I. 2014, ApJ, 787, 74 (Cited on page 135)

Gratier, P., Bron, E., Gerin, M., et al. 2017, A&A, 599, A100 (Cited on pages x, 29, 50, 80, 93, 95, 104, 106, 109, 111, 123, 135, 157, 158, 166, 169, and 173)

Gratier, P., Pety, J., Guzmán, V., et al. 2013, A&A, 557, A101 (Cited on page 17)

Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2015, ApJ, 810, 25 (Cited on page 117)

Groves, B. A., Schinnerer, E., Leroy, A., et al. 2015, ApJ, 799, 96 (Cited on pages 21 and 25)

Guzmán, V., Pety, J., Goicoechea, J. R., Gerin, M., & Roueff, E. 2011, A&A, 534, A49 (Cited on page 17)

Guzmán, V., Pety, J., Gratier, P., et al. 2012a, A&A, 543, L1 (Cited on page 17)

Guzmán, V., Roueff, E., Gauss, J., et al. 2012b, A&A, 548, A94 (Cited on page 17)

Guzmán, V. V. 2013, PhD thesis (Cited on pages viii, 17, and 106)

Guzmán, V. V., Goicoechea, J. R., Pety, J., et al. 2013, A&A, 560, A73 (Cited on page 17)

Guzmán, V. V., Pety, J., Goicoechea, J. R., et al. 2015, ApJ, 800, L33 (Cited on pages 17 and 135)

Guzmán, V. V., Pety, J., Gratier, P., et al. 2014, Faraday Discussions, 168, 103 (Cited on page 17)

Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421 (Cited on pages 26, 113, 114, 116, 119, 121, 131, 132, and 158)

Hacar, A., Kainulainen, J., Tafalla, M., Beuther, H., & Alves, J. 2016, A&A, 587, A97 (Cited on pages 69 and 79)

Hacar, A. & Tafalla, M. 2011, A&A, 533, A34 (Cited on pages 68, 69, and 70)

Hacar, A., Tafalla, M., Forbrich, J., et al. 2018, A&A, 610, A77 (Cited on pages 70, 79, 90, and 92)

Hacar, A., Tafalla, M., Kauffmann, J., & Kovács, A. 2013, A&A, 554, A55 (Cited on pages 70, 71, 73, 74, 76, 79, 85, 90, 92, 93, and 100)

Hartmann, J. 1904, ApJ, 19 (Cited on page 9)

Hartmann, L. 2002, ApJ, 578, 914 (Cited on pages 67, 68, 70, 72, 73, and 74)

Hartmann, L. & Burkert, A. 2007, ApJ, 654, 988 (Cited on page 79)

Hatchell, J., Richer, J. S., Fuller, G. A., et al. 2005, A&A, 440, 151 (Cited on pages 67, 68, 69, 70, and 91)

Helmholtz, H. 1858, J. für die reine Angewandte Mathematik, 55, 25 (Cited on page 40)

Hennebelle, P. 2013, A&A, 556, A153 (Cited on pages 40, 72, 73, 78, 91, and 110)

Hennebelle, P., Banerjee, R., Vázquez-Semadeni, E., Klessen, R. S., & Audit, E. 2008, A&A, 486, L43 (Cited on pages 72 and 73)

Hennebelle, P. & Chabrier, G. 2011, ApJ, 743, L29 (Cited on page 110)

Hennebelle, P. & Falgarone, E. 2012, A&A Rev., 20, 55 (Cited on pages 39, 46, 49, 50, 51, and 57)

Hennemann, M., Motte, F., Schneider, N., et al. 2012, A&A, 543, L3 (Cited on pages 68 and 69)

Henshaw, J. D., Jiménez-Serra, I., Longmore, S. N., et al. 2017, MNRAS, 464, L31 (Cited on page 68)

Henshaw, J. D., Longmore, S. N., Kruijssen, J. M. D., et al. 2016, MNRAS, 457, 2675 (Cited on pages 22 and 33)

Herrera, C. N., Boulanger, F., Nesvadba, N. P. H., & Falgarone, E. 2012, A&A, 538, L9 (Cited on page 22)

Herschel, W. 1785, Philosophical Transactions of the Royal Society of London Series I, 75, 213 (Cited on page 9)

Hester, J. J., Stone, J. M., Scowen, P. A., et al. 1996, ApJ, 456, 225 (Cited on page 46)

Heyer, M. & Dame, T. M. 2015, ARA&A, 53, 583 (Cited on page 132)

Heyer, M. H. & Peter Schloerb, F. 1997, ApJ, 475, 173 (Cited on page 153)

Heyer, M. H., Vrba, F. J., Snell, R. L., et al. 1987, ApJ, 321, 855 (Cited on page 67)

Heyer, M. H., Williams, J. P., & Brunt, C. M. 2006, ApJ, 643, 956 (Cited on page 43)

Hily-Blant, P., Falgarone, E., & Pety, J. 2008, A&A, 481, 367 (Cited on pages 43, 45, and 49)

Hily-Blant, P., Teyssier, D., Philipp, S., & Güsten, R. 2005, A&A, 440, 909 (Cited on pages 16 and 56)

Hollenbach, D. J., Takahashi, T., & Tielens, A. G. G. M. 1991, ApJ, 377, 192 (Cited on pages 113 and 158)

Hollenbach, D. J. & Tielens, A. G. G. M. 1997, ARA&A, 35, 179 (Cited on page 171)

Hotelling, H. 1933, J. Educ. Psych., 24 (Cited on pages 29, 30, and 143)

Hughes, A., Meidt, S. E., Colombo, D., et al. 2013, ApJ, 779, 46 (Cited on page 110)

Hummel, C. A., Rivinius, T., Nieva, M.-F., et al. 2013, A&A, 554, A52 (Cited on page 112)

Iffrig, O. & Hennebelle, P. 2015, A&A, 576, A95 (Cited on page 13)

Ilin, A. & Raiko, T. 2010, J. Mach. Learn. Res., 11, 1957 (Cited on page 144)

Inglada, J., Vincent, A., Arias, M., et al. 2017, Remote Sensing, 9, 95 (Cited on pages 24 and 157)

Inoue, T. & Fukui, Y. 2013, ApJ, 774, L31 (Cited on pages 72 and 73)

Inoue, T., Hennebelle, P., Fukui, Y., et al. 2018, PASJ, 70, S53 (Cited on page 72)

Inutsuka, S.-I. & Miyama, S. M. 1992, ApJ, 388, 392 (Cited on pages 73 and 90)

Inutsuka, S.-i. & Miyama, S. M. 1997, ApJ, 480, 681 (Cited on pages 73 and 74)

Ishida, E. E. O. & de Souza, R. S. 2013, MNRAS, 430, 509 (Cited on page 24)

Ivezić, Ž., Connolly, A., Vanderplas, J., & Gray, A. 2014, Statistics, Data Mining and Machine Learning in Astronomy (Princeton University Press) (Cited on page 144)

Jackson, J. M., Finn, S. C., Chambers, E. T., Rathborne, J. M., & Simon, R. 2010, ApJ, 719, L185 (Cited on page 67)

Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1 (Cited on page 74)

Jimenez, R., Heavens, A. F., Panter, B., & Tojeiro, R. 2008, in American Institute of Physics Conference Series, Vol. 1082, American Institute of Physics Conference Series, ed. C. A. L. Bailer-Jones, 92–98 (Cited on page 24)

Jiménez-Donaire, M. J., Bigiel, F., Leroy, A. K., et al. 2017a, MNRAS, 466, 49 (Cited on page 22)

Jiménez-Donaire, M. J., Cormier, D., Bigiel, F., et al. 2017b, ApJ, 836, L29 (Cited on page 22)

Johnstone, D. & Bally, J. 1999, ApJ, 510, L49 (Cited on pages 67 and 69)

Jolliffe, I. 2002, Principal component analysis (New York: Springer Verlag) (Cited on pages 29, 142, and 157)

Jones, P. A., Burton, M. G., Cunningham, M. R., et al. 2012, MNRAS, 419, 2961 (Cited on pages 142 and 152)

Juvela, M., Ristorcelli, I., Pagani, L., et al. 2012, A&A, 541, A12 (Cited on pages 68 and 69)

Kainulainen, J., Beuther, H., Henning, T., & Plume, R. 2009, A&A, 508, L35 (Cited on page 43)

Kainulainen, J., Federrath, C., & Henning, T. 2013a, A&A, 553, L8 (Cited on pages 45, 46, and 65)

Kainulainen, J., Hacar, A., Alves, J., et al. 2016, A&A, 586, A27 (Cited on pages 69, 77, 84, 85, 88, 89, and 90)

Kainulainen, J., Ragan, S. E., Henning, T., & Stutz, A. 2013b, A&A, 557, A120 (Cited on pages 74 and 75)

Kainulainen, J., Stutz, A. M., Stanke, T., et al. 2017, A&A, 600, A141 (Cited on pages 70, 74, and 88)

Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775 (Cited on page 118)

Kandel, D., Lazarian, A., & Pogosyan, D. 2016, MNRAS, 461, 1227 (Cited on page 65)

Kennicutt, Jr., R. C., Armus, L., Bendo, G., et al. 2003, PASP, 115, 928 (Cited on page 20)

Kim, C.-G. & Ostriker, E. C. 2015, ApJ, 802, 99 (Cited on page 49)

Kim, C.-G., Ostriker, E. C., & Kim, W.-T. 2013, ApJ, 776, 1 (Cited on page 110)

Kirk, H., Di Francesco, J., Johnstone, D., et al. 2016, ApJ, 817, 167 (Cited on pages 32, 51, 56, 148, and 176)

Kirk, H., Klassen, M., Pudritz, R., & Pillsworth, S. 2015, ApJ, 802, 75 (Cited on pages 68, 69, 70, 79, and 88)

Kirk, H., Myers, P. C., Bourke, T. L., et al. 2013, ApJ, 766, 115 (Cited on pages 69, 70, 75, and 79)

Klein, B., Hochgürtel, S., Krämer, I., et al. 2012, A&A, 542, L3 (Cited on page 19)

Klemperer, W. 1970, Nature, 227, 1230 (Cited on page 10)

Klessen, R. S. 2004, in Revista Mexicana de Astronomia y Astrofisica, vol. 27, Vol. 22, Revista Mexicana de Astronomia y Astrofisica Conference Series, ed. G. Garcia-Segura, G. Tenorio-Tagle, J. Franco, & H. W. Yorke, 8–12 (Cited on page 72)

Koch, E. W. & Rosolowsky, E. W. 2015, MNRAS, 452, 3435 (Cited on pages 95 and 100)

Koch, E. W., Ward, C. G., Offner, S., Loeppky, J. L., & Rosolowsky, E. W. 2017, MNRAS, 471, 1506 (Cited on page 107)

Koda, J., Scoville, N., Sawada, T., et al. 2009, ApJ, 700, L132 (Cited on page 67)

Koenderink, J. J. 1984, Biological cybernetics, 50, 363 (Cited on page 94)

Kolmogorov, A. 1941, Akademiia Nauk SSSR Doklady, 30, 301 (Cited on page 38)

Kong, S., Arce, H. G., Feddersen, J. R., et al. 2018, ArXiv e-prints (Cited on page 79)

Könyves, V., André, P., Men'shchikov, A., et al. 2015, A&A, 584, A91 (Cited on pages 68, 74, and 79)

Kramer, C., Stutzki, J., & Winnewisser, G. 1996, A&A, 307, 915 (Cited on page 50)

Kreckel, K., Blanc, G. A., Schinnerer, E., et al. 2016, ApJ, 827, 103 (Cited on page 21)

Kritsuk, A. G., Norman, M. L., Padoan, P., & Wagner, R. 2007, ApJ, 665, 416 (Cited on pages 41 and 42)

Krumholz, M. R. & McKee, C. F. 2005, ApJ, 630, 250 (Cited on pages 40 and 44)

Krčo, M. & Goldsmith, P. F. 2016, ApJ, 822, 10 (Cited on page 177)

Lada, C. J., Forbrich, J., Lombardi, M., & Alves, J. F. 2012, ApJ, 745, 190 (Cited on pages 111 and 125)

Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687 (Cited on pages 44, 59, 90, 111, and 135)

Lada, C. J., Lombardi, M., Roman-Zuniga, C., Forbrich, J., & Alves, J. F. 2013, ApJ, 778, 133 (Cited on pages 60, 111, and 125)

Lada, E. A. 1992, ApJ, 393, L25 (Cited on pages 59, 60, 89, 90, and 157)

Lallement, R., Capitanio, L., Ruiz-Dern, L., et al. 2018, ArXiv e-prints (Cited on pages 65 and 106)

Lallement, R., Vergely, J.-L., Valette, B., et al. 2014, A&A, 561, A91 (Cited on page 117)

Langer, W. D., Graedel, T. E., Frerking, M. A., & Armentrout, P. B. 1984, ApJ, 277, 581 (Cited on page 166)

Larson, R. B. 1981, MNRAS, 194, 809 (Cited on pages 49 and 70)

Larson, R. B. 1985, MNRAS, 214, 379 (Cited on page 73)

Lazarian, A. 2005, in American Institute of Physics Conference Series, Vol. 784, Magnetic Fields in the Universe: From Laboratory and Stars to Primordial Structures., ed. E. M. de Gouveia dal Pino, G. Lugones, & A. Lazarian, 42–53 (Cited on page 13)

Le Bourlot, J., Le Petit, F., Pinto, C., Roueff, E., & Roy, F. 2012, A&A, 541, A76 (Cited on page 141)

Lefloch, B. & Lazareff, B. 1994, A&A, 289, 559 (Cited on page 16)

Lequeux, J., Falgarone, E., & Ryter, C. 2004, The Interstellar Medium, Astronomy and Astrophysics Library (Springer Berlin Heidelberg) (Cited on pages 9 and 12)

Leroy, A. K., Hughes, A., Schruba, A., et al. 2016, ArXiv e-prints (Cited on page 111)

Leroy, A. K., Usero, A., Schruba, A., et al. 2017, ApJ, 835, 217 (Cited on page 22)

Leroy, A. K., Walter, F., Bigiel, F., et al. 2009, AJ, 137, 4670 (Cited on pages 20, 21, and 67)

Leroy, A. K., Walter, F., Sandstrom, K., et al. 2013, AJ, 146, 19 (Cited on page 21)

Lindegren, L., Lammers, U., Bastian, U., et al. 2016, ArXiv e-prints (Cited on page 112)

Lique, F., Spielfiedel, A., Feautrier, N., et al. 2010, J. Chem. Phys., 132, 024303 (Cited on page 180)

Lis, D. C., Pety, J., Phillips, T. G., & Falgarone, E. 1996, ApJ, 463, 623 (Cited on page 41)

Liszt, H. 2014, ApJ, 783, 17 (Cited on page 80)

Liszt, H. & Lucas, R. 2001, A&A, 370, 576 (Cited on page 128)

Liszt, H. S. & Pety, J. 2012, A&A, 541, A58 (Cited on pages 110, 132, and 180)

Liszt, H. S. & Pety, J. 2016, ApJ, 823, 124 (Cited on pages 25, 28, 36, 50, 56, 128, 133, 135, 169, and 181)

Lo, N., Cunningham, M. R., Jones, P. A., et al. 2009, MNRAS, 395, 1021 (Cited on pages 142 and 152)

Lomax, O., Whitworth, A. P., & Hubber, D. A. 2015, MNRAS, 449, 662 (Cited on page 58)

Lombardi, M., Alves, J., & Lada, C. J. 2011, A&A, 535, A16 (Cited on page 15)

Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, A&A, 566, A45 (Cited on pages 15, 16, 25, 26, 32, 50, 80, 81, 113, 114, 115, and 158)

Low, F. J., Beintema, D. A., Gautier, T. N., et al. 1984, ApJ, 278, L19 (Cited on page 67)

Lucas, R. & Liszt, H. 1996, A&A, 307, 237 (Cited on page 128)

Lucas, R. & Liszt, H. S. 2000, A&A, 358, 1069 (Cited on page 128)

Mac Low, M.-M. 1999, ApJ, 524, 169 (Cited on pages 39, 40, and 49)

Maddalena, R. J., Morris, M., Moscowitz, J., & Thaddeus, P. 1986, ApJ, 303, 375 (Cited on page 15)

Mangum, J. G., Emerson, D. T., & Greisen, E. W. 2007, A&A, 474, 679 (Cited on page 112)

Mangum, J. G. & Shirley, Y. L. 2015, PASP, 127, 266 (Cited on pages 25, 51, and 80)

Marka, C., Pety, J., Bardeau, S., & Sievers, A. 2017, Observational examples of spectral line calibration at the 30m telescope with MRTCAL and MIRA, Tech. Rep. 2017-1, IRAM (Cited on page 19)

Marsh, K. A., Griffin, M. J., Palmeirim, P., et al. 2014, MNRAS, 439, 3683 (Cited on page 131)

Marsh, K. A., Whitworth, A. P., & Lomax, O. 2015, MNRAS, 454, 4282 (Cited on pages 25 and 65)

Marsh, K. A., Whitworth, A. P., Lomax, O., et al. 2017, MNRAS, 471, 2730 (Cited on pages 25, 65, and 106)

Martin, P. G., Blagrave, K. P. M., Lockman, F. J., et al. 2015, ApJ, 809, 153 (Cited on pages 54 and 55)

Martin-Zaïdi, C., Deleuil, M., Le Bourlot, J., et al. 2008, A&A, 484, 225 (Cited on page 176)

McCauley, P. I., Mangum, J. G., & Wootten, A. 2011, ApJ, 742, 58 (Cited on page 176)

McKee, C. F. & Ostriker, E. C. 2007, ARA&A, 45, 565 (Cited on page 13)

McKee, C. F. & Ostriker, J. P. 1977, ApJ, 218, 148 (Cited on page 11)

Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, AJ, 144, 192 (Cited on pages 59 and 60)

Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, AJ, 151, 5 (Cited on pages 48, 59, 60, 85, 86, 89, 90, 135, and 157)

Meidt, S. E., Schinnerer, E., García-Burillo, S., et al. 2013, ApJ, 779, 45 (Cited on pages 23 and 110)

Meier, D. S. & Turner, J. L. 2012, ApJ, 755, 104 (Cited on page 133)

Meier, D. S., Walter, F., Bolatto, A. D., et al. 2015, ApJ, 801, 63 (Cited on page 133)

Meisner, A. M. & Finkbeiner, D. P. 2014, ApJ, 781, 5 (Cited on pages 32 and 148)

Melnick, G. J., Tolls, V., Snell, R. L., et al. 2011, ApJ, 727, 13 (Cited on pages 142 and 152)

Men'shchikov, A. 2013, A&A, 560, A63 (Cited on page 100)

Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515 (Cited on pages 15, 50, 79, 115, 117, 142, 157, and 158)

Messier, C. 1781, Connoissance des Temps ou des Mouvements Célestes, for 1784, 227 (Cited on page 9)

Miville-Deschênes, M.-A., Duc, P.-A., Marleau, F., et al. 2016, A&A, 593, A4 (Cited on page 51)

Miville-Deschênes, M.-A., Martin, P. G., Abergel, A., et al. 2010, A&A, 518, L104 (Cited on page 67)

Miyama, S. M., Narita, S., & Hayashi, C. 1987, Progress of Theoretical Physics, 78, 1051 (Cited on pages 72 and 73)

Molina, F. Z., Glover, S. C. O., Federrath, C., & Klessen, R. S. 2012, MNRAS, 423, 2680 (Cited on page 41)

Molinari, S., Swinyard, B., Bally, J., et al. 2010, A&A, 518, L100 (Cited on pages 74 and 79)

Moretti, M. I., Hatzidimitriou, D., Karampelas, A., et al. 2018, MNRAS, 477, 2664 (Cited on page 24)

Mouschovias, T. C. 1991, ApJ, 373, 169 (Cited on page 13)

Muller, C. A. & Oort, J. H. 1951, Nature, 168, 357 (Cited on page 10)

Myers, P. C. 2009, ApJ, 700, 1609 (Cited on pages 72, 73, and 75)

Myers, P. C. 2011, ApJ, 735, 82 (Cited on pages 69, 75, and 76)

Müller, H. S., Schlöder, F., Stutzki, J., & Winnewisser, G. 2005, Journal of Molecular Structure, 742, 215 (Cited on page 80)

Nagai, T., Inutsuka, S.-i., & Miyama, S. M. 1998, ApJ, 506, 306 (Cited on pages 72, 73, 79, and 90)

Nakamura, F. & Li, Z.-Y. 2008, ApJ, 687, 354 (Cited on pages 72 and 73)

Narayanan, G., Heyer, M. H., Brunt, C., et al. 2008, ApJS, 177, 341 (Cited on page 87)

Neufeld, D. A., Godard, B., Gerin, M., et al. 2015, A&A, 577, A49 (Cited on page 153)

Neufeld, D. A., Hollenbach, D. J., Kaufman, M. J., et al. 2007, ApJ, 664, 890 (Cited on pages 142 and 152)

Nishimura, Y., Shimonishi, T., Watanabe, Y., et al. 2016, ApJ, 818, 161 (Cited on page 133)

Nolan, C. A., Federrath, C., & Sutherland, R. S. 2015, MNRAS, 451, 1380 (Cited on page 60)

Orkisz, J. H., Peretto, N., Pety, J., Gerin, M., & Levrier, F. 2018, A&A, submitted (Cited on pages xv, 68, 77, and 100)

Orkisz, J. H., Pety, J., Gerin, M., et al. 2017, A&A, 599, A99 (Cited on pages xiii, 47, 48, 64, 86, 89, 90, 91, 109, 111, 131, 142, 157, and 171)

Ossenkopf, V. & Mac Low, M.-M. 2002, A&A, 390, 307 (Cited on page 43)

Ostriker, J. 1964, ApJ, 140, 1056 (Cited on pages 74 and 85)

Padoan, P., Federrath, C., Chabrier, G., et al. 2014, Protostars and Planets VI, 77 (Cited on pages 49 and 60)

Padoan, P., Juvela, M., Goodman, A. A., & Nordlund, Å. 2001, ApJ, 553, 227 (Cited on pages 72, 73, and 78)

Padoan, P. & Nordlund, Å. 1999, ApJ, 526, 279 (Cited on page 41)

Padoan, P., Nordlund, A., & Jones, B. J. T. 1997, MNRAS, 288, 145 (Cited on page 41)

Palmeirim, P., André, P., Kirk, J., et al. 2013, A&A, 550, A38 (Cited on pages 68, 70, 73, and 100)

Panopoulou, G. V., Psaradaki, I., Skalidis, R., Tassis, K., & Andrews, J. J. 2017, MNRAS, 466, 2529 (Cited on pages 77, 79, 84, 86, and 87)

Panopoulou, G. V., Tassis, K., Goldsmith, P. F., & Heyer, M. H. 2014, MNRAS, 444, 2507 (Cited on pages 68, 69, 70, 82, 83, 84, 87, 88, and 92)

Parvathi, V. S., Sofia, U. J., Murthy, J., & Babu, B. R. S. 2012, ApJ, 760, 36 (Cited on page 80)

Parzen, E. 1962, Ann. Math. Statist., 33, 1065 (Cited on pages 159 and 160)

Paubert, G. 2000, Multi-beam Correlator, Tech. rep., IRAM (Cited on page 19)

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825 (Cited on pages 144 and 160)

Penzias, A. A. & Burrus, C. A. 1973, ARA&A, 11, 51 (Cited on page 113)

Peratt, A. L. 2015, Physics of the Plasma Universe (Springer New York), 299–309 (Cited on page 67)

Peretto, N., André, P., Könyves, V., et al. 2012, A&A, 541, A63 (Cited on pages 69, 72, 73, and 78)

Peretto, N., Fuller, G. A., André, P., et al. 2014, A&A, 561, A83 (Cited on pages 40, 68, 70, 75, 79, 85, 90, and 92)

Peretto, N., Fuller, G. A., Duarte-Cabral, A., et al. 2013, A&A, 555, A112 (Cited on pages 70, 74, 75, and 79)

Peretto, N., Hennebelle, P., & André, P. 2007, A&A, 464, 983 (Cited on pages 70 and 79)

Pety, J. 2005, in SF2A-2005: Semaine de l'Astrophysique Francaise, ed. F. Casoli, T. Contini, J. M. Hameury, & L. Pagani, 721–722 (Cited on pages 19 and 113)

Pety, J., Bardeau, S., & Reynier, E. 2009, IRAM-30m EMIR time/sensitivity estimator, Tech. rep., iRAM Memo 2009-1 (Cited on page 112)

Pety, J. & Falgarone, É. 2000, A&A, 356, 279 (Cited on pages 49, 72, 73, and 78)

Pety, J., Gratier, P., Guzmán, V., et al. 2012, A&A, 548, A68 (Cited on page 17)

Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98 (Cited on pages ix, 17, 19, 25, 26, 27, 28, 29, 32, 35, 36, 38, 50, 51, 56, 62, 65, 79, 80, 81, 89, 103, 104, 109, 142, 144, 145, 147, 153, 157, 158, 162, 164, 166, 168, 173, and 174)

Pety, J., Schinnerer, E., Leroy, A. K., et al. 2013, ApJ, 779, 43 (Cited on pages 22 and 132)

Pety, J., Teyssier, D., Fossé, D., et al. 2005, A&A, 435, 885 (Cited on page 135)

Planck and Fermi Collaborations, Ade, P. A. R., Aghanim, N., et al. 2015, A&A, 582, A31 (Cited on page 132)

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A1 (Cited on pages 113 and 158)

Planck Collaboration Int. XIX. 2015, A&A, 576, A104 (Cited on page 91)

Planck Collaboration Int. XX. 2015, A&A, 576, A105 (Cited on page 91)

Planck Collaboration Int. XXXII. 2016, A&A, 586, A135 (Cited on page 94)

Planck Collaboration Int. XXXIV. 2016, A&A, 586, A137 (Cited on page 91)

Planck Collaboration XIX. 2011, A&A, 536, A19 (Cited on pages 25, 80, and 132)

Plewa, P. M. 2018, MNRAS, 476, 3974 (Cited on page 24)

Polychroni, D., Schisano, E., Elia, D., et al. 2013, ApJ, 777, L33 (Cited on pages 74 and 79)

Porth, O., Komissarov, S. S., & Keppens, R. 2014, MNRAS, 443, 547 (Cited on page 46)

Reipurth, B. & Bouchet, P. 1984, A&A, 137, L1 (Cited on page 16)

Renaud, F., Bournaud, F., Kraljic, K., & Duc, P.-A. 2014, MNRAS, 442, L33 (Cited on pages 44, 46, and 50)

Reynolds, O. 1883, Proceedings of the Royal Society of London Series I, 35, 84 (Cited on pages 37 and 38)

Richer, J. S. 1990, MNRAS, 245, 24P (Cited on page 116)

Rieke, G. H. & Lebofsky, M. J. 1985, ApJ, 288, 618 (Cited on pages 80 and 114)

Ripple, F., Heyer, M. H., Gutermuth, R., Snell, R. L., & Brunt, C. M. 2013, MNRAS, 431, 1296 (Cited on pages 50 and 62)

Rohlfs, K. & Wilson, T. L. 2004, Tools of radio astronomy (Cited on page 62)

Roman-Duval, J., Federrath, C., Brunt, C., et al. 2011, ApJ, 740, 120 (Cited on page 153)

Rosenblatt, M. 1956, Ann. Math. Statist., 27, 832 (Cited on pages 159 and 160)

Rosolowsky, E. & Leroy, A. 2006, PASP, 118, 590 (Cited on pages 33, 153, and 157)

Roy, A., André, P., Arzoumanian, D., et al. 2015, A&A, 584, A111 (Cited on pages 70, 75, and 100)

Russell, H. N. 1914, Popular Astronomy, 22, 275 (Cited on page 24)

Ryden, B. & Pogge, R. 2015, Interstellar and Intergalactic Medium, Ohio State Graduate Astrophysics Series (The Ohio State University) (Cited on pages 9 and 12)

Saintonge, A., Catinella, B., Cortese, L., et al. 2016, MNRAS, 462, 1749 (Cited on page 21)

Saintonge, A., Kauffmann, G., Kramer, C., et al. 2011a, MNRAS, 415, 32 (Cited on pages 20 and 21)

Saintonge, A., Kauffmann, G., Wang, J., et al. 2011b, MNRAS, 415, 61 (Cited on page 21)

Salas, P., Galaz, G., Salter, D., et al. 2014, ApJ, 797, 134 (Cited on page 133)

Salji, C. J., Richer, J. S., Buckle, J. V., et al. 2015, MNRAS, 449, 1782 (Cited on pages 68, 77, 94, 95, and 100)

Sandell, G., Avery, L. W., Baas, F., et al. 1999, ApJ, 519, 236 (Cited on page 116)

Scargle, J. D., Norris, J. P., Jackson, B., & Chiang, J. 2013, ApJ, 764, 167 (Cited on page 144)

Schaefer, G. H., Hummel, C. A., Gies, D. R., et al. 2016, AJ, 152, 213 (Cited on pages 79 and 112)

Schilke, P., Walmsley, C. M., Pineau Des Forets, G., et al. 1992, A&A, 256, 595 (Cited on page 135)

Schinnerer, E., Meidt, S. E., Colombo, D., et al. 2017, ApJ, 836, 62 (Cited on pages 23 and 67)

Schinnerer, E., Meidt, S. E., Pety, J., et al. 2013, ApJ, 779, 42 (Cited on pages 20, 23, and 132)

Schisano, E., Rygl, K. L. J., Molinari, S., et al. 2014, ApJ, 791, 27 (Cited on pages 69, 77, 83, and 100)

Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, ApJ, 786, 29 (Cited on pages 15, 50, 79, 115, and 157)

Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17 (Cited on pages 25, 50, 57, 58, 62, 80, 113, 128, 146, 148, and 158)

Schneider, N., Csengeri, T., Bontemps, S., et al. 2010, A&A, 520, A49 (Cited on page 70)

Schneider, S. & Elmegreen, B. G. 1979, ApJS, 41, 87 (Cited on pages 67, 70, 73, and 78)

Schulz, N. 2012, The Formation and Early Evolution of Stars: From Dust to Stars and Planets, Astronomy and Astrophysics Library (Springer Berlin Heidelberg) (Cited on page 9)

Schulz-Dubois, E. O. & Rehberg, I. 1981, Applied Physics, 24, 323 (Cited on page 90)

She, Z.-S. & Leveque, E. 1994, Phys. Rev. Lett., 72, 336 (Cited on page 72)

Shetty, R., Kauffmann, J., Schnee, S., Goodman, A. A., & Ercolano, B. 2009, ApJ, 696, 2234 (Cited on pages 25 and 131)

Shimajiri, Y., André, P., Braine, J., et al. 2017, ArXiv e-prints (Cited on page 158)

Shirley, Y. L. 2015, PASP, 127, 299 (Cited on pages 27, 28, and 133)

Smith, R. J., Glover, S. C. O., Clark, P. C., Klessen, R. S., & Springel, V. 2014a, MNRAS, 441, 1628 (Cited on page 25)

Smith, R. J., Glover, S. C. O., & Klessen, R. S. 2014b, MNRAS, 445, 2900 (Cited on pages 68, 70, 73, 76, 77, and 78)

Smith, R. J., Glover, S. C. O., Klessen, R. S., & Fuller, G. A. 2016, MNRAS, 455, 3640 (Cited on pages 70, 71, 73, 76, and 91)

Snow, T. P. & McCall, B. J. 2006, ARA&A, 44, 367 (Cited on pages 158 and 171)

Snyder, L. E., Buhl, D., Zuckerman, B., & Palmer, P. 1969, Physical Review Letters, 22, 679 (Cited on page 10)

Sobolev, V. V. 1960, Soviet Ast., 4, 1 (Cited on page 180)

Sofia, U. J., Lauroesch, J. T., Meyer, D. M., & Cartledge, S. I. B. 2004, ApJ, 605, 272 (Cited on page 80)

Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730 (Cited on pages 89 and 118)

Sousbie, T. 2011, MNRAS, 414, 350 (Cited on pages 88 and 100)

Spitzer, L. 1978, Physical processes in the interstellar medium (Cited on page 118)

Spitzer, L. 1998, Physical Processes in the Interstellar Medium, 335 (Cited on page 90)

Stutz, A. M. & Gould, A. 2016, A&A, 590, A2 (Cited on page 68)

Stutz, A. M. & Kainulainen, J. 2015, A&A, 577, L6 (Cited on pages 69 and 88)

Stutz, A. M., Tobin, J. J., Stanke, T., et al. 2013, ApJ, 767, 36 (Cited on page 92)

Stutzki, J. & Guesten, R. 1990, ApJ, 356, 513 (Cited on page 157)

Tafalla, M. & Hacar, A. 2015, A&A, 574, A104 (Cited on pages 70, 74, and 76)

Takahashi, S., Ho, P. T. P., Teixeira, P. S., Zapata, L. A., & Su, Y.-N. 2013, ApJ, 763, 57 (Cited on pages 70, 74, and 90)

Teixeira, P. S., Takahashi, S., Zapata, L. A., & Ho, P. T. P. 2016, A&A, 587, A47 (Cited on pages 70, 74, 88, 89, and 90)

Tremblin, P., Anderson, L. D., Didelon, P., et al. 2014a, A&A, 568, A4 (Cited on pages 43, 44, 59, 60, and 100)

Tremblin, P., Audit, E., Minier, V., & Schneider, N. 2012, A&A, 538, A31 (Cited on page 16)

Tremblin, P., Schneider, N., Minier, V., et al. 2014b, A&A, 564, A106 (Cited on pages 60 and 171)

Trumpler, R. J. 1930, PASP, 42, 214 (Cited on page 9)

Ungerechts, H., Bergin, E. A., Goldsmith, P. F., et al. 1997, ApJ, 482, 245 (Cited on pages 31, 142, and 152)

van der Tak, F. F. S., Black, J. H., Schöier, F. L., Jansen, D. J., & van Dishoeck, E. F. 2007, A&A, 468, 627 (Cited on pages 36, 171, and 180)

van Leeuwen, F. 2007, A&A, 474, 653 (Cited on page 112)

Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47 –54 (Cited on page 144)

von Weizsäcker, C. F. 1951, ApJ, 114, 165 (Cited on page 38)

Walch, S. & Naab, T. 2015, MNRAS, 451, 2757 (Cited on page 40)

Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563 (Cited on page 20)

Ward-Thompson, D., Kirk, J. M., André, P., et al. 2010, A&A, 518, L92 (Cited on page 78)

Ward-Thompson, D., Nutter, D., Bontemps, S., Whitworth, A., & Attwood, R. 2006, MNRAS, 369, 1201 (Cited on pages 59, 145, and 176)

Watanabe, Y., Sakai, N., Sorai, K., & Yamamoto, S. 2014, ApJ, 788, 4 (Cited on page 133)

Williams, G. M., Peretto, N., Avison, A., Duarte-Cabral, A., & Fuller, G. A. 2018, ArXiv e-prints (Cited on pages 69, 76, and 79)

Williams, J. P., de Geus, E. J., & Blitz, L. 1994, ApJ, 428, 693 (Cited on pages 24, 33, and 157)

Wilson, R. W., Jefferts, K. B., & Penzias, A. A. 1970, ApJ, 161, L43 (Cited on page 10)

Wilson, T. L. & Rood, R. 1994, ARA&A, 32, 191 (Cited on pages 51, 80, and 164)

Wiseman, J. J. & Ho, P. T. P. 1998, ApJ, 502, 676 (Cited on pages 67 and 69)

Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191 (Cited on pages 110 and 132)

Wyrzykowski, Ł. & Belokurov, V. 2008, in American Institute of Physics Conference Series, Vol. 1082, American Institute of Physics Conference Series, ed. C. A. L. Bailer-Jones, 201–206 (Cited on page 24)

Yang, B., Stancil, P. C., Balakrishnan, N., & Forrey, R. C. 2010, ApJ, 718, 1062 (Cited on page 180)

Young, J. S., Xie, S., Tacconi, L., et al. 1995, ApJS, 98, 219 (Cited on page 20)

Ysard, N., Jones, A. P., Demyk, K., Boutéraon, T., & Koehler, M. 2018, accepted by A&A (Cited on page 25)

Zamora-Avilés, M., Ballesteros-Paredes, J., & Hartmann, L. W. 2017, MNRAS, 472, 647 (Cited on page 76)

Zhang, T. Y. & Suen, C. Y. 1984, Communications of the ACM, 27, 236 (Cited on page 95)

Jan Orkisz

Comprendre la structure des nuages moléculaires : imagerie hyperspectrale d'Orion B

La dernière génération de récepteurs radio, dotés à la fois d'une grande bande passante et d'une haute résolution, fait de toute observation radio-astronomique une étude spectroscopique. Dans le cas de l'imagerie à grand champ du milieu interstellaire, cela génère de très gros jeux de données en grande dimension. Une telle abondance de données fournit de nouveaux outils de diagnostic, mais pose simultanément de nouveaux défis en termes de traitement et d'analyse des données - il est donc nécessaire de développer de nouvelles méthodes. C'est l'objectif du projet ORION-B, qui observe 5 degrés carrés du nuage moléculaire Orion B, soit près de la moitié de la surface du nuage, dans toute la bande à 3 mm. Les observations ont produit des cartes résolues en vitesse de l'émission de dizaines de traceurs moléculaires, qui sont analysées avec des approches statistiques d'apprentissage automatique.

L'accès à des cartes résolues spatialement pour de nombreuses espèces chimiques nous permet d'identifier les traceurs les plus caractéristiques de telles ou telles régions, qui peuvent avoir des densités ou éclairements UV différents. Réciproquement, étudier les données dans l'espace "chimique" multi-dimensionnel permet de segmenter le nuage moléculaire en régions caractérisées par une émission moléculaire similaire, et de quantifier les corrélations les plus importantes entre différents traceurs moléculaires, et entre les traceurs et des quantités physiques telles que la densité ou la température des poussières.

La grande plage dynamique spatiale et spectrale des données fournit également une vue détaillée de la cinématique et de la dynamique du nuage moléculaire, en particulier en ce qui concerne son activité de formation stellaire. Dans cette thèse, je caractérise statistiquement la nature des mouvements du nuages, et je quantifie la fraction de quantité de mouvement contenue dans les modes compressifs et solénoïdaux (rotatifs) de la turbulence. Le nuage est dominé par des mouvements solénoïdaux, tandis que les modes compressifs sont concentrés dans deux régions de formation stellaire. J'aborde aussi la formation stellaire d'un point de vue morphologique, en identifiant et caractérisant les nombreux filaments présents dans le nuage. Ceux-ci s'avèrent avoir de faibles densités, et être très stables face à l'effondrement gravitationnel - mais ils montrent cependant des signes de fragmentation radiale et longitudinale. Tous ces résultats soulignent le rôle du forçage compressif et des filaments denses dans le processus de formation stellaire. Ils sont également cohérents avec la très faible efficacité de formation stellaire d'Orion B, mais suggèrent aussi que le nuage pourrait devenir plus actif à l'avenir.

Understanding the structure of molecular clouds: Multi-line wide-field imaging of Orion B

The new generation of wide-bandwidth high-resolution receivers turns almost any radio observation into a spectral survey. In the case of wide-field imaging of the interstellar medium, large, highly multi-dimensional datasets are generated. Such a wealth of data simultaneously provides new diagnostic tools and creates new challenges in terms of data processing and analysis - new methods therefore need to be developed. This is the goal of the ORION-B project, which is observing 5 square degrees of the Orion B molecular cloud, or about half of the cloud's surface, over the entire 3 mm band. The observations have provided velocity-resolved maps of the emission of tens of molecular tracers, which are analysed using statistical, machine learning approaches.

Having access to spatially resolved maps from many molecular species enables us to identify the tracers most typical of given regions, with different gas densities and illuminations. Conversely, the multi-dimensional data in chemical space allow us to segment the molecular cloud into typical regions based on their molecular emission, and to quantify the most meaningful correlations of different molecular tracers with each other and with physical quantities such as density or dust temperature.

The large spatial and spectral dynamical range of the dataset also provides a detailed view of the kinematics and dynamics of the molecular cloud, in particular in relation with its star formation activity. In this thesis, I statistically characterize the nature of the motions in the cloud, and quantify the amount of momentum in the compressive and solenoidal (rotational) modes of turbulence. The cloud is dominated by solenoidal motions, with the compressive modes being concentrated in two star-forming regions. I also approach star formation from a morphological point of view, by identifying and characterizing the numerous filaments present in the cloud. These prove to have rather low densities, and to be very stable against gravitational collapse - and yet they show signs of longitudinal and radial fragmentation. All these results highlight the role of compressive forcing and dense filamentary structures in the star formation process. They are also consistent with the overall very low star formation efficiency of the Orion B cloud, while suggesting that it might host more triggered star formation in the future.