A&A, 687, A140 (2024)
https://doi.org/10.1051/0004-6361/202348565

Astronomy & Astrophysics

Emergence of high-mass stars in complex fiber networks (EMERGE)

I. Early ALMA Survey: Observations and massive data reduction

A. Hacar[1], A. Socci[1], F. Bonanomi[1], D. Petry[2], M. Tafalla[3], D. Harsono[4], J. Forbrich[5], J. Alves[1], J. Grossschedl[1], J. R. Goicoechea[6], J. Pety[7,8], A. Burkert[9,10,11], and G.X. Li[12]

[1] Department of Astrophysics, University of Vienna, Türkenschanzstrasse 17, 1180 Vienna, Austria
e-mail: alvaro.hacar@univie.ac.at
[2] European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany
[3] Observatorio Astronómico Nacional (IGN), Alfonso XII 3, 28014 Madrid, Spain
[4] Institute of Astronomy, Department of Physics, National Tsing Hua University, Hsinchu, Taiwan
[5] University of Hertfordshire, Centre for Astrophysics Research, College Lane, Hatfield AL10 9AB, UK
[6] Instituto de Física Fundamental (CSIC), c/ Serrano 121–123, 28006 Madrid, Spain
[7] IRAM, 300 rue de la Piscine, 38406 Saint Martin d'Hères, France
[8] LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, 75014 Paris, France
[9] Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 Munich, Germany
[10] Excellence Cluster ORIGINS, Boltzmannstrasse 2, 85748 Garching, Germany
[11] Max-Planck Institute for Extraterrestrial Physics, Giessenbachstrasse 1, 85748 Garching, Germany
[12] South-Western Institute for Astronomy Research, Yunnan University, Kunming 650500 Yunnan, PR China

Received 10 November 2023 / Accepted 6 March 2024

ABSTRACT

Context. Recent molecular surveys have revealed the rich gas organization of sonic-like filaments at small scales (so-called fibers) in all types of environments prior to the formation of low- and high-mass stars. These fibers form at the end of the turbulent cascade and are identified as the fine substructure within the hierarchical nature of the gas in the interstellar medium (ISM).
Aims. Isolated fibers provide the subsonic conditions for the formation of low-mass stars. This paper introduces the Emergence of high-mass stars in complex fiber networks (EMERGE) project, which investigates whether complex fiber arrangements (networks) can also explain the origin of high-mass stars and clusters.
Methods. We analyzed the EMERGE Early ALMA Survey including seven star-forming regions in Orion (OMC-1,2,3, and 4 South, LDN 1641N, NGC 2023, and the Flame Nebula) that were homogeneously surveyed in three molecular lines (N_2H^+ J = 1–0, HNC J = 1–0, and HC_3N J = 10–9) and in the 3 mm continuum using a combination of interferometric ALMA mosaics and IRAM-30 m single-dish (SD) maps, together with a series of *Herschel*, *Spitzer*, and WISE archival data. We also developed a systematic data reduction framework allowing the massive data processing of ALMA observations.
Results. We obtained independent continuum maps and spectral cubes for all our targets and molecular lines at different (SD and interferometric) resolutions, and we explored multiple data combination techniques. Based on our low-resolution (SD) observations (30″ or ∼12 000 au), we describe the global properties of our sample, which covers a wide range of physical conditions, including low- (OMC-4 South and NGC 2023), intermediate (OMC-2, OMC-3, and LDN 1641N), and high-mass (OMC-1 and Flame Nebula) star-forming regions in different evolutionary stages. The comparison between our single-dish maps and ancillary YSO catalogs denotes N_2H^+ (1–0) as the best proxy for the dense, star-forming gas in our targets, which show a constant star formation efficiency and a fast time evolution of $\lesssim$1 Myr. While apparently clumpy and filamentary in our SD data, all targets show a much more complex fibrous substructure at the enhanced resolution of our combined ALMA+IRAM-30 m maps (4.″5 or ∼2000 au). A large number of filamentary features at subparsec scales are clearly recognized in the high-density gas ($\gtrsim 10^5$ cm^{-3}) that is traced by N_2H^+ (1–0) directly connected to the formation of individual protostars. Surprisingly, this complex gas organization appears to extend farther into the more diffuse gas (∼10^3–10^4 cm^{-3}) traced by HNC (1–0).
Conclusions. This paper presents the EMERGE Early ALMA Survey, which includes a first data release of continuum maps and spectral products for this project that are to be analysed in future papers of this series. A first look at these results illustrates the need of advanced data combination techniques between high-resolution interferometric (ALMA) and high-sensitivity, low-resolution single-dish (IRAM-30 m) datasets to investigate the intrinsic multiscale, gas structure of the ISM.

Key words. stars: formation – ISM: clouds – ISM: kinematics and dynamics – submillimeter: ISM

1. Introduction

A classical observational dichotomy separates the origin of solar-like stars from their high-mass counterparts. Low-mass stars (few $M_\odot$; e.g., HL Tau) can be found in isolation or in small groups in quiescent, low-mass environments such as Taurus. In contrast, high-mass stars (>10 $M_\odot$; e.g., θ^1 Ori C in the ONC) are typically found within clusters (which also form low-mass stars) and under the influence of intense stellar feedback (i.e., UV radiation, winds, or outflows) in giant molecular clouds such as

Orion (e.g., Pabst et al. 2019). These findings led to the proposal of independent formation scenarios for low- (Shu et al. 1987) and high-mass (Bonnell et al. 2001; McKee & Tan 2003) stars. Based on similar (spherical) initial conditions (either a low-mass, dense core or a massive clump, respectively), these models have dominated the star formation field in the past decades.

A series of observational studies have highlighted the direct link between the filamentary structure of the interstellar medium (ISM) and the initial conditions for star-formation (see André et al. 2014; Hacar et al. 2023; Pineda et al. 2023, for recent reviews). According to large-scale *Herschel* far-IR (FIR) dust continuum emission surveys, most dense cores and young stellar objects (YSO) are preferentially formed in association with dense filaments in both nearby (André et al. 2010) and more distant Galactic Plane (Molinari et al. 2010) molecular clouds. Low-mass cores are found along parsec-scale filaments (Hartmann et al. 2001) and likely form out of their internal gravitational fragmentation (Schneider & Elmegreen 1979; Inutsuka & Miyama 1997). On the other hand, most of the dense clumps harboring high-mass stars and clusters are found at the junction of massive filaments, forming the so-called hub-filament systems (HFS; Myers 2009; Kumar et al. 2020). Filaments provide a preferential orientation with an additional mass reservoir and could funnel large amounts of material toward cores and clumps (e.g., Peretto et al. 2013). The connection between the previous star formation models and the new filamentary conditions of the ISM remains debated, however, particularly in the case of high-mass stars (e.g., Motte et al. 2018).

The analysis of the gas kinematics of many of the above filamentary clouds has revealed a high level of gas organization prior to the formation of stars. Single-dish (low spatial resolution) observations demonstrated that many of the parsec-size filaments detected by *Herschel* in nearby low-mass clouds (e.g., B213; Palmeirim et al. 2013) are actually collections of velocity-coherent filaments at smaller scales, which were identified by their internal velocity dispersion close to the sound speed and usually referred to as fibers (Hacar et al. 2013; André et al. 2014; Pineda et al. 2023). Sharing similar kinematic properties but smaller, further ALMA (high spatial resolution) observations resolved a similar fiber structure at sub-parsec scales in massive star-forming regions (Hacar et al. 2018). Since their discovery, filamentary structures of different lengths (observed to have between ~0.1 and ~7 pc) but similar sonic internal velocity dispersion than those identified in fibers have been systematically identified in all types of low-mass clouds (Arzoumanian et al. 2013; Fehér et al. 2016; Kainulainen et al. 2016; Hacar et al. 2016), clusters (Fernández-López et al. 2014; Hacar et al. 2017b), high-mass star-forming regions (Treviño-Morales et al. 2019; Cao et al. 2022), and infrared dark clouds (IRDCs; Henshaw et al. 2014; Chen et al. 2019), to cite a few examples. While first identified from the gas velocity of low-density tracers and potentially affected by line-of-sight confusion (see Zamora-Avilés et al. 2017; Clarke et al. 2018), the subsequent detection of fibers in high-density tracers such as N_2H^+ (e.g., Chen et al. 2019; Shimajiri et al. 2019), NH_3 (Monsch et al. 2018; Sokolov et al. 2019), $H^{13}CO^+$, and NH_2D^+ (e.g., Cao et al. 2022) guarantees a good correspondence with the true dense gas structure within clouds in large statistical samples. Systematically produced in hydro- and turbulent simulations (Smith et al. 2014; Moeckel & Burkert 2015; Clarke et al. 2017; Li & Klein 2019), these "filaments within filaments" appear as the first subsonic structures created at the end of the turbulent cascade as part of the intrinsic hierarchical structure of gas in the ISM (see Hacar et al. 2023, for a discussion).

Fibers could potentially connect the formation of low- and high-mass stars (see Hacar et al. 2018, for a discussion). Individual fibers promote the formation of small chains of low-mass Jeans-like cores via quasi-static fragmentation (Schneider & Elmegreen 1979; Hacar & Tafalla 2011; Tafalla & Hacar 2015), which if associated in large groups, could become self-gravitating and might agglomerate a high number of cores and stars (Hacar et al. 2017a,b). In parallel, interactions between fibers could also form super-Jeans cores via merging or collision (Clarke et al. 2017; Hoemann et al. 2021), while at the same time, they might provide large mass reservoirs favoring the enhanced accretion rates needed to form high-mass stars (Bernasconi & Maeder 1996; Behrend & Maeder 2001). Massive fiber arrangements could therefore simultaneously generate stellar clusters and high-mass stars (Hacar et al. 2018). A similar scenario has been proposed from the merging of parsec-scale filaments (Kumar et al. 2020, 2022). Given these promising results, this new fiber scenario for star formation deserves further exploration.

This paper introduces the Emergence of high-mass stars in complex fiber networks (EMERGE) project[1]. As its working hypothesis, EMERGE aims to investigate whether both high-mass stars and clusters could be created as emergent phenomena in densely populated fiber systems. Rather than by distinct processes, EMERGE will explore whether the formation of low- and high-mass stars together with clusters arises as a natural consequence of the global properties of these fiber systems showing a continuum spectrum of masses as a function of the network density.

EMERGE will systematically analyze the substructure, internal interactions, and dynamical evolution of these filamentary systems throughout the Milky Way. These massive regions are typically located at kpc distances, and resolution and sensitivity constraints have so far limited their analysis to individual targets or small samples. EMERGE will survey a large, homogeneous ALMA sample of massive fiber networks extracted from its public archive using a novel combination of data mining techniques (Ahmadi & Hacar 2023) and massive data processing (van Terwisga et al. 2022). In this first work (hereafter Paper I), we present the EMERGE Early ALMA Survey (see Sect. 2), which includes seven star-forming regions in Orion that were systematically analyzed using a combination of high spatial resolution resolution ALMA (interferometric) mosaics plus large-scale IRAM-30 m (single-dish) observations at 3 mm (N_2H^+, HNC, HC_3N, HCN, and 3 mm continuum). Designed as a proof of concept of this project, Paper I introduces a standard set of procedures for data combination and massive data reduction. In a series of accompanying papers, we will investigate the effects of interferometric filtering on ISM observations carried out with ALMA (Bonanomi et al. 2024, Paper II) and present the analysis of the physical properties of dense fibers identified in this first EMERGE sample (Socci et al. 2024, Paper III). Future EMERGE papers will extend this analysis to other regions.

2. The EMERGE Early ALMA Survey

With our EMERGE early ALMA survey we aim to investigate the internal gas structure and physical conditions for star formation in different environments of Orion (Menten et al. 2007; Großschedl et al. 2018, D~400 pc; see also Sect. 5). This first survey sampled a total of seven prototypical star-forming regions: OMC-1, OMC-2, OMC-3, OMC-4 South, and LDN 1641N in

[1] EMERGE Project website: https://emerge.univie.ac.at/

Table 1. EMERGE early ALMA survey: targets.

Target	Region	D$^{(1)}$ (pc)	$M_{\rm tot}{}^{(2)}$ ($M_\odot$)	$M_{\rm dense}{}^{(3)}$ ($M_\odot$)	$f_{\rm dense}^{(4)}$	$\langle T_{\rm K}\rangle^{(5)}$ (K)	O-star?	P$^{(6)}$	P+D$^{(6)}$	P/D
OMC-1	Orion A	400	1032$^{(\star)}$	582.2	0.56	33.7	Yes	36	423	0.09
OMC-2	Orion A	400	581.0	372.8	0.64	26.3	No	33	142	0.30
OMC-3	Orion A	400	643.2	351.1	0.55	25.4	No	26	102	0.34
OMC-4 South	Orion A	400	469.4	73.6	0.16	20.3	?	11	59	0.23
LDN 1641N	Orion A	400	478.6	174.3	0.36	25.4	No	13	51	0.34
NGC 2023	Orion B	423	330.6	70.7	0.21	21.9	No	6	8	0.75
Flame Nebula	Orion B	423	287.8$^{(\star)}$	94.1	0.33	28.8	Yes	21	141	0.18

Notes. Physical parameters derived from low-resolution (*Herschel* and/or IRAM-30 m) maps within an area of 1.5×1.5 pc^2 (~700×700 arcsec^2) around the phase center of our ALMA maps (see coordinates in Table 2): (1) Adopted cloud distances; (2) Total mass, M(total), derived from *Herschel* column density maps (Lombardi et al. 2014); (3) Total dense gas mass, M(dense), derived from N_2H^+(1–0); (4) Fraction of dense gas (i.e. $f = M_{\rm tot}/M_{\rm dense}$; (5) Mean gas kinetic temperature, $\langle T_{\rm K}\rangle$; (6) Number of Protostars (P) and Disk stars (D) identified by *Spitzer* (Megeath et al. 2012) (see also Fig. 7). $^{(\star)}$ Note how the mass contribution in stars could significantly increase the mass load of high-mass, clustered regions such as OMC-1 and Flame Nebula.

Orion A, and NGC 2023 and Flame Nebula (or NGC 2024) in Orion B (see Table 1). We designed our ALMA survey to cover both low- and high-mass star-forming regions, which show different levels of stellar activity, mass, and evolution. We show the location, column density maps, and stellar populations in these regions in Figs. 1 and 2.

Previous stellar surveys have classified these regions as function of their clustering properties and star-formation activity (see Peterson & Megeath 2008; Bally 2008). OMC-1 and Flame Nebula are identified as the two most massive embedded clusters in Orion showing the highest peak stellar densities in this complex (>1500 stars pc^{-2}) including several high-mass stars. Although less highly concentrated, OMC-2, OMC-3, and LDN 1641 N contain large embedded populations of YSOs that form smaller groups. In contrast, OMC-4 South and NGC 2023 only form a handfull of low-mass sources mainly in isolation (see Megeath et al. 2016, for a full comparison).

Our source selection also covers a wide range of environments. OMC-1 and Flame Nebula are exposed to intense extreme-UV (EUV) ionizing radiation (producing HII regions) and to strong far-UV (FUV) dissociating radiation (creating neutral photodissociated regions, PDRs) produced by the O-type θ^1 Ori C (O6Vp) and NGC 2024 IRS-2**b** (O8V) stars, respectively (see Bally 2008). Although less severe, the stellar winds and UV radiation of nearby clusters and high-mass stars seem to influence the OMC-3 (NGC 1977 and 42 Orionis; Peterson & Megeath 2008), OMC-4 South (NGC 1980 and ι Ori; Alves & Bouy 2012), and NGC 2023 (σ Ori; Brown et al. 1994) regions as denoted in previous velocity-resolved observations of the FIR [CII] 158μm fine-structure line (see Pabst et al. 2017, 2020). Despite active in star formation as denoted by a large population of young outflows (e.g., Stanke & Williams 2007; Sato et al. 2023), OMC-2 and LDN 1641N appear to be more pristine regions and therefore become good candidates to also investigate unperturbed gas.

Most of our targets have been systematically surveyed at long wavelengths, particularly at low spatial resolution ($\theta \gtrsim 15''$). The Orion complex has been mapped using large-scale FIR (36″; Lombardi et al. 2014) plus (sub-) millimeter (submm) (14″; Stanke et al. 2022) continuum surveys as well as in some CO transitions (11″–14″; Berné et al. 2010; Goicoechea et al. 2020; Stanke et al. 2022). In the Orion A cloud, the OMC clouds (OMC-1, 2, 3, and 4 South; see Fig. 1) are part of the famous Integral Shaped Filament (ISF; Bally et al. 1987), the most massive filamentary cloud in the solar neighborhood, and the one containing the most massive cluster within the local kiloparsec, the Orion Nebula Cluster (see Peterson & Megeath 2008, for a review). The ISF has been mapped at sub-mm wavelengths in continuum (14″; Johnstone & Bally 1999), by molecular surveys (~30″; Kauffmann et al. 2017; Brinkmann et al. 2020) and dedicated observations of both diffuse (Shimajiri et al. 2014; Stanke et al. 2022) and dense molecular tracers (~30″; Tatematsu et al. 2008; Hacar et al. 2017a), to cite few examples (for more information see Peterson & Megeath 2008; Hacar et al. 2018, and references therein). Away from this main star-forming site, LDN 1641N has received less attention, although it has been covered in some large-scale surveys in Orion A ($\gtrsim$ 15″; e.g., Nishimura et al. 2015; Mairs et al. 2016). NGC 2023 and Flame Nebula have been observed as part of cloud-size studies in Orion B (see Fig. 2) both in single-line observations ($\gtrsim$19″; Lada et al. 1991; Stanke et al. 2022) and unbiased molecular surveys ($\gtrsim$22″; Pety et al. 2017; Orkisz et al. 2019; Santa-Maria et al. 2023).

The clouds in our sample have been classified as filamentary in previous low spatial resolution studies (e.g., Johnstone & Bally 1999; Orkisz et al. 2019; Könyves et al. 2020; Gaudel et al. 2023). Nonetheless, a close inspection of Figs. 1 and 2 reveals a variety of cloud morphology. Regions such as OMC-2, OMC-3, and Flame Nebula appear elongated and to follow the large-scale filamentary structure of their parental cloud (Johnstone & Bally 1999). The OMC-1 region is identified as a hub-filament System structure that is characteristic of massive clumps (Rodriguez-Franco et al. 1992). On the other hand, OMC-4 South and NGC 2023 show a more irregular and diffuse organization, while LDN 1641N exhibits a cometary shape (Mairs et al. 2016; Kirk et al. 2016).

Compared to the large number of low spatial resolution observations, the study of these targets at higher spatial resolution ($\theta \leq 10''$) is more limited. Only the targets in Orion A (OMC-1, 2, 3, and 4 South, and LDN 1641N) have been systematically observed using large-scale CO maps (~7″; Kong et al. 2018; Suri et al. 2019) but a similar counterpart study of Orion B is lacking. Local surveys mainly focused the northern part of the ISF region (OMC-1, 2, and 3) mapped the FIR (~8″; Schuller et al. 2021) or millimeter (~4.″5; Takahashi et al. 2013) continuum. On the other hand, interferometric observations of

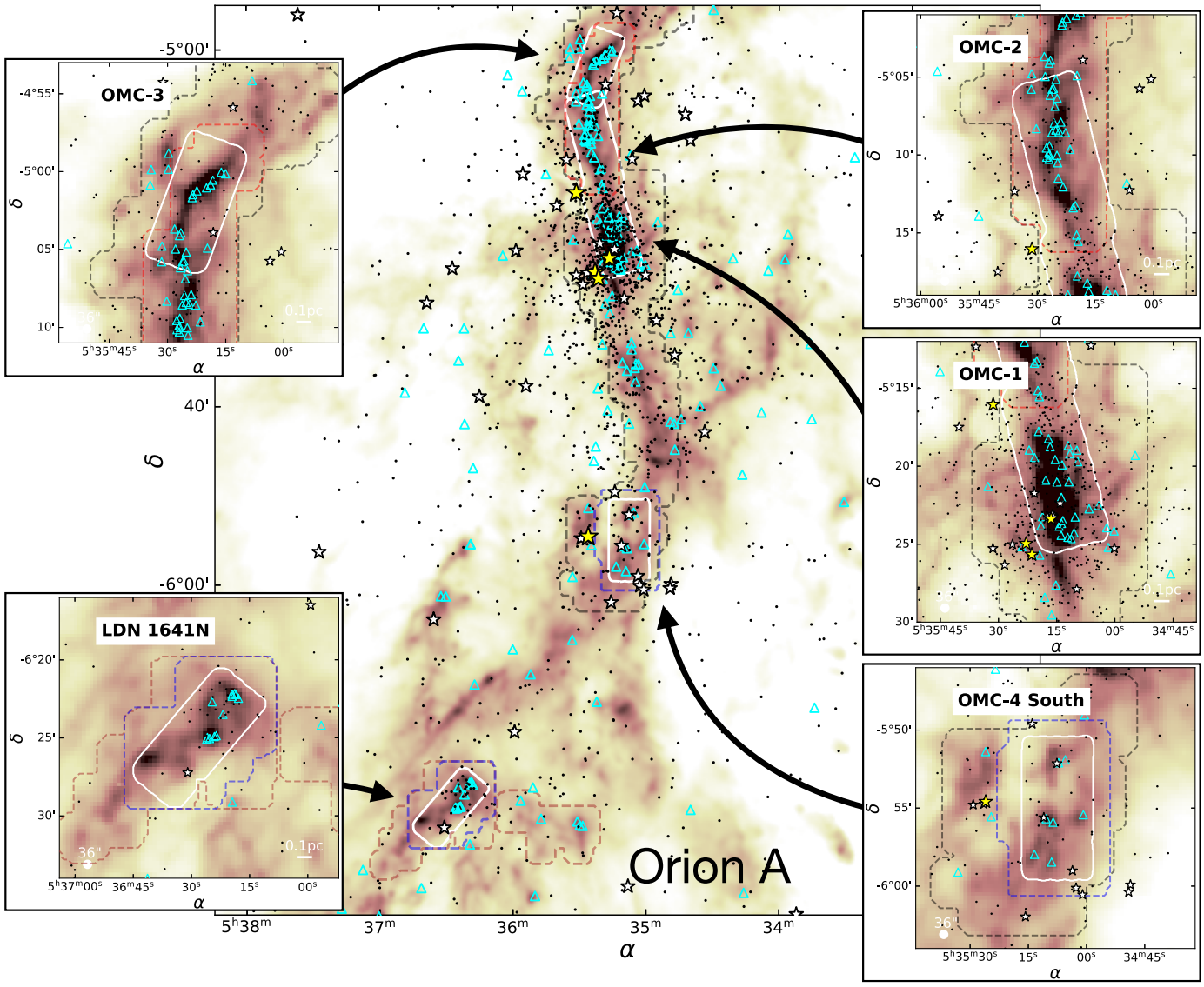

Fig. 1. EMERGE Early ALMA Survey in Orion A (central panel), including zoom-ins of the OMC-1, OMC-2, OMC-3, OMC-4 South, and LDN 1641N regions (see subpanels), shown over the total gas column density derived from *Herschel* observations (background image at 36 arcsec resolution; Lombardi et al. 2014). The footprints of our ALMA mosaics (white lines) as well as our IRAM-30 m observations (colored lines; see color-code in Table 3) are indicated in all panels together with the positions of O (yellow stars) and B stars (white stars) at the distance of Orion. The positions of young protostars (cyan triangles) and disk stars (black dots) (Megeath et al. 2012) and the corresponding *Herschel* beam size and 0.1 pc scale bar are marked in the individual zoom-ins.

high-density tracers (>10^4 cm^{-3}) have been restricted to regions of specific interest (~4″; e.g., OMC-1 by Wiseman & Ho 1998; Hacar et al. 2018), sections of these regions (~3–5″; e.g., parts of OMC-2 by Kainulainen et al. 2017 or Flame Nebula by Shimajiri et al. 2023), and/or peculiar targets within these fields (~1–2″; e.g., the head of LDN 1641N by Stanke & Williams 2007, OMC-2 FIR-4 by Chahine et al. 2022).

We surveyed our EMERGE Early ALMA sample with a homogeneous set of high spatial resolution, high sensitivity ALMA observations (4.″5) in combination with additional large-scale, low spatial resolution, single-dish (30″) maps (Sect. 3) for which we also created a new data reduction framework (Sect. 4). We investigated the gas properties in these regions (star formation, column density, temperature, etc.) using a dedicated suite of molecular tracers (N_2H^+, HNC, and HC_3N) and continuum maps (Sect. 5). By selecting targets within the same star-forming complex, we carried out a direct comparison of these regions at similar resolution while avoiding additional variations depending on their local Galactic environment. As primary goal of our study, we systematically investigate the physical conditions in both low- and high-mass star-forming regions from parsec scales down to 2000 au resolutions (see also Papers II and III).

3. Observations

3.1. Interferometric ALMA observations

We surveyed the molecular and continuum emission of the OMC-3, OMC-4 South, LDN 1641, NGC 2023 and Flame Nebula regions using the Atacama Large Millimeter Array (ALMA) (project ID: 2019.1.00641.S; PI: Hacar) in Chajnantor (Chile) during ALMA Cycle-7. We summarize the main parameters of our ALMA observations in Table 2. We used independent

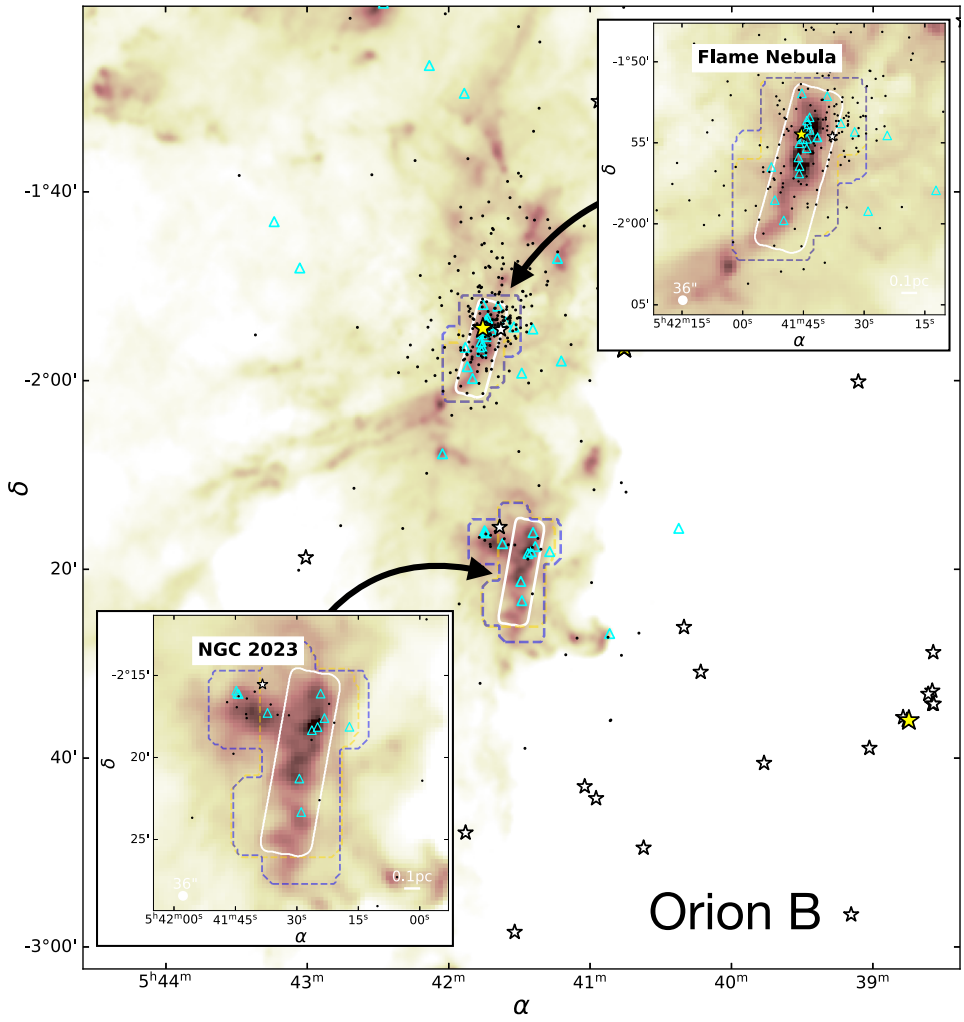

Fig. 2. Similar to Fig. 1 but for the case of Orion B with zoom-ins of NGC 2023 and Flame Nebula (NGC 2024).

Table 2. EMERGE early ALMA survey: ALMA fields.

Field	Phase center α(J2000)	Phase center δ(J2000)	Map size (arcsec^2)	No. of pointings	Spectral windows	Proj. ID
OMC-1$^{(\#)}$	$05^h35^m27^s.2$	$-05°09'42''$	$220'' \times 600''$	148	N_2H^+, 3 mm-Cont.	2015.1.00669.S
OMC-2$^{(\#)}$	$05^h35^m14^s.2$	$-05°22'21''$	$220'' \times 600''$	148	N_2H^+, 3 mm-Cont.	2015.1.00669.S
OMC-3	$05^h35^m22^s.5$	$-05°02'00''$	$210'' \times 500''$	124	N_2H^+, HNC, HC_3N, 3 mm-Cont.	2019.1.00641.S
OMC-4 South	$05^h35^m07^s.5$	$-05°55'00''$	$230'' \times 540''$	149	N_2H^+, HNC, HC_3N, 3 mm-Cont.	2019.1.00641.S
LDN 1641N	$05^h36^m27^s.0$	$-06°25'00''$	$210'' \times 540''$	132	N_2H^+, HNC, HC_3N, 3 mm-Cont.	2019.1.00641.S
NGC 2023	$05^h41^m29^s.0$	$-02°20'20''$	$150'' \times 650''$	123	N_2H^+, HNC, HC_3N, 3 mm-Cont.	2019.1.00641.S
Flame Nebula	$05^h41^m46^s.0$	$-01°56'37''$	$150'' \times 600''$	111	N_2H^+, HNC, HC_3N, 3 mm-Cont.	2019.1.00641.S

Notes. $^{(\#)}$ Observations presented in Hacar et al. (2018).

large-scale ALMA mosaics, containing between 111 and 149 pointings each, with typical map sizes of $\sim 200 \times 600$ arcsec^2 (or $\sim 0.4 \times 1.2$ pc^2 at the distance of Orion). The footprints of the individual ALMA fields are shown in Figs. 1 and 2 (white contours) superposed on the *Herschel* total gas column density maps of these regions (background image; Lombardi et al. 2014). Our choice for the phase center, size, and orientation of our maps follows the distribution of the N_2H^+ (1–0) integrated emission reported in previous single-dish surveys of the Orion A (Hacar et al. 2017a) and Orion B (Pety et al. 2017) clouds.

We mapped the molecular and continuum emission of all of our targets with the ALMA-12m array (only) in its most compact configuration (C43-1), and with baselines between 15 and $\sim$310 m, achieving a native resolution of $\theta \sim 3.5$ arcsec. All fields, each observed as individual scheduling blocks (SB), were covered at least twice and were observed for a total of 2–3 h per

Table 3. EMERGE early ALMA survey: observational setup.

Species	Frequency	$E_u^{(1)}$	$n_{eff}^{(2)}$	ALMA		IRAM-30 m$^{(3)}$		ALMA+IRAM	
				Proj.ID	δv	Proj.ID	δv	ID	δv
	(GHz)	(K)	(cm^{-3})		(km s^{-1})		(km s^{-1})		(km s^{-1})
N_2H^+ (J = 1–0)	93.173764	4.47	1.0×10^4	2015.1.00669.S	0.098	032-13	0.06	narrow	0.10
				2019.1.00641.S	0.114	034-16	0.06	narrow	0.15
						120-22	0.06	narrow	0.15
HNC (J = 1–0)	90.663564	4.53	3.7×10^3	2019.1.00641.S	0.233	032-13	0.66	broad	0.66
						034-16	0.66	broad	0.66
						060-22	0.06	narrow	0.25
						133-22	0.06	narrow	0.25
HC_3N (J = 10–9)	90.979023	24.1	4.3×10^5	2019.1.00641.S	0.233	032-13	0.66	broad	0.66
						034-16	0.66	broad	0.66
						060-22	0.16	narrow	0.25
						133-22	0.16	narrow	0.25
HCN (J = 1–0)	88.631602	4.25	8.4×10^3	–$^{(\star)}$	–$^{(\star)}$	032-13	0.66	broad	–
						034-16	0.66	broad	–
						133-22	0.66	broad	–
3 mm-Cont.	93, 104	–	–	2015.1.00669.S	6.0 and 5.4	–$^{(\star)}$	–$^{(\star)}$	–	Cont.$^{(\star\star)}$
3 mm-Cont.	93.2, 91.2	–	–	2019.1.00641.S	3.71	–$^{(\star)}$	–$^{(\star)}$	–	Cont.$^{(\star\star)}$

Notes. (1) Values for the upper energy levels (E_u), without hyperfine structure, are taken from the Leiden Atomic and Molecular Database (LAMDA; Schöier et al. 2005; van der Tak et al. 2020). (2) Effective excitation density at 10 K (Shirley 2015). (3) The footprints of the different IRAM-30 m data are color-coded in Figs. 1 and 2: 032-12 (black), 034-16 (brown), 120-20 (gold), 133-22 (blue), and 060-22 (red). $^{(\star)}$ No data available. $^{(\star\star)}$ ALMA-alone maps.

field with a minimum of 45 antennas under average weather conditions (precipitable water vapor PWV=2.5–5 mm). Following standard procedures the pointing, bandpass, and flux calibrations were carried out at the beginning of each observing run, for which phase calibration was obtained every 10–15 min using different sources.

Our observations were carried out in Band 3 using a single spectral setup that simultaneously combine three narrow line plus two broad continuum spectral windows (SPWs) simultaneously. We include a detailed description of the different observational setup in Table 3. In our new ALMA observations we used a narrow SPW to map the N_2H^+ (1–0) (93.17 GHz) emission at high spectral resolution (δv = 0.114 km s^{-1}). We set independent SPWs for both HNC (1–0) (90.66 GHz) and HC_3N (10–9) (90.97 GHz) lines at intermediate spectral resolution (δv = 0.233 km s^{-1}). Moreover, we set two additional broadband, 1.875 GHz wide SPWs to map the continuum emission at 93.2 and 91.2 GHz.

To complete our survey we added similar ALMA Cycle-3 mosaics of the OMC-1 and OMC-2 regions in Band 3 (project ID: 2015.1.00669.S; PI: Hacar). The description of these previous datasets is presented in Hacar et al. (2018; see also Tables 2 and 3). Independent N_2H^+ (1–0) maps were carried out along OMC-1 and OMC-2 at high spectral resolutions (δv = 0.098 km s^{-1}), together with two 1.875 GHz wide, broad continuum bands at 93 and 104 GHz, respectively (see also van Terwisga et al. 2019). Unfortunately, no HNC (1–0) or HC_3N (10–9) observations were possible at the time due to the limitations of the data rates during Cycle-3. No ALMA map is therefore available for these transitions in OMC-1 or OMC-2.

When needed (e.g., during data combination), we assumed standard flux conversions, that is, $\left(\frac{T_{mb}}{1\,K}\right) = \left(\frac{S_\nu}{1\,Jy}\right)\left[13.6\left(\frac{300\,GHz}{\nu}\right)^2\left(\frac{1\,arcsec}{\theta}\right)^2\right]$ (e.g., 6.96 and 0.16 K Jy^{-1} for beams of 4.″5 and 30″, respectively; see ALMA technical handbook, Remijan et al. 2019). The reduction of our ALMA observations, including data combination with additional single-dish observations (Sect. 3.2), is discussed in Sect. 4.

3.2. Single-dish IRAM-30 m observations

We complemented our ALMA survey with additional single-dish observations at 3 mm carried out at the 30-m Institute Radioastronomie Millimetric telescope (IRAM-30 m) in Granada (Spain) in multiple observing campaigns between 2013 and 2023. We mapped the molecular emission of all our ALMA targets using independent large-scale IRAM-30 m mosaics, all observed with the Eight Mixer Receiver (EMIR) connected to both the VErsatile SPectrometer Array (VESPA) and Fast Fourier Transform Spectrometer (FTS) backends. We summarize these observations in Table 3. We observed all our fields with two independent spectral setups. In a first set of observations (project IDs: 032-13, 034-16, and 120-20) we obtained large-scale maps of N_2H^+ (1–0) (93.17 GHz) using VESPA at high-spectral resolution (δv = 0.06 km s^{-1}). Simultaneously, we mapped the emission of HNC (1–0) (90.66 GHz), HCN (1–0) (88.63 GHz), and HC_3N (10–9) (90.97 GHz), among other lines, using multiple FTS units in broad configuration at low spectral resolution (δv = 0.66 km s^{-1}). In a second set of maps (proj. IDs: 060-22, 133-22) we then focused on the local HNC (1–0) (90.66 GHz) and HC_3N (10–9) (90.97 GHz) emission and reobserved our ALMA fields using smaller mosaics at higher spectral resolution with VESPA (δv = 0.06 km s^{-1}) and FTS in narrow configuration (δv = 0.16 km s^{-1}).

Each large single-dish mosaic was constructed by several individual tiles, typically with sizes of 200 × 200 arcsec^2, observed multiple times in orthogonal directions using on-the-fly (OTF), position switching (PSw) observations. The footprints of

our low- and high-resolution IRAM-30 m maps is again shown in Figs. 1 and 2. We carried out our data reduction using the GILDAS-CLASS software (Pety 2005; Gildas Team 2013) using standard procedures for OTF observations. We reduced each tile independently and applied facility-provided efficiencies to calibrate our spectra into main-beam temperature scale ($T_{\rm mb}$). For each of our target lines, we combined our calibrated spectra, observed at a native resolution ∼27 arcsec, into a regular grid and convolved them into a final resolution of θ = 30 arcsec in order to obtain individual Nyquist-sampled cubes. As final data products of our single-dish observations we obtained individual N_2H^+ (1–0), HNC (1–0), HCN (1–0), and HC_3N (10–9) datacubes (including their entire hyperfine structure) for each of the sources in our survey (OMC-1, 2, 3, and 4 South, LDN 1641N, NGC 2023, and Flame Nebula). In the case of HNC and HC_3N, we reduced these cubes at low and high spectral resolution.

Following Hacar et al. (2020), we combined our low spectral resolution (broad) HCN (1–0) and HNC (1–0) observations to investigate the gas temperature of the gas in our survey. We obtained individual integrated-intensity maps of each of these tracers (I(HCN) and I(HNC)) adjusted to the velocity range of each target seen in HCN. Both HCN and HNC present bright emission that usually follows the column density distribution of our clouds and extends beyond the coverage of our maps (see Pety et al. 2017; Hacar et al. 2020; Tafalla et al. 2023). We considered only those positions with I(HNC) ≥ 2 K km s^{-1}, however, and then converted the resulting I(HCN)/I(HNC) intensity ratios into the corresponding gas kinetic temperature $T_{\rm K}$(HCN/HNC) per pixel using the empirical correlations defined by Hacar et al. (2020). The I(HCN)/I(HNC) intensity ratios provide a good proxy of the gas kinetic temperature at column densities $N(H_2) > 10^{22}$ cm^{-2} and within an optimal temperature range between ∼15 and 40 K, with only minor contamination in local regions that are affected by outflows (e.g., Ori BN/KL in OMC-1). Using this method, we then obtained large-scale kinetic temperature $T_{\rm K}$(HCN/HNC) maps (hereafter $T_{\rm K}$ maps) for all the sources in our sample (see Sect. 5.1).

3.3. Complementary infrared data

We investigated the connection between stars and gas in our different ALMA fields. In addition to our millimeter observations, our project benefits from a wealth of ancillary, public surveys in Orion at multiple wavelengths. In particular, we used the total H_2 column density maps obtained by Lombardi et al. (2014) from the FIR *Herschel* observations of the entire Orion A and B clouds. We converted these maps, originally measuring the dust opacity at 350 μm ($\tau_{350\mu m}$), first into equivalent K-band extinction values (A_K) using the conversion values provided by Lombardi et al, and later into total H_2 column density maps assuming a standard reddening law (A_K/A_V=0.112; Rieke & Lebofsky 1985) and dust-to-gas conversion factor ($N(H_2)/A_V$ = 0.93×10^{21} cm^{-2} mag^{-1}; Bohlin et al. 1978). We therefore obtained total H_2 column density maps ($N(H_2)$) of all our sources at 36″ resolution (Sect. 5.1).

For illustrative purposes, we also made use of archival 3.4 μm emission images obtained by the Wide-field Infrared Survey Explorer (WISE) (Wright et al. 2010). Together with the detection of most of the young stars in our targets, the WISE 3.4 μm band includes prominent polycyclic aromatic hydrocarbons (PAH) features, such as the C-H stretching band of FUV-pumped PAH (e.g., Chown et al. 2024), highlighting the effects of external FUV radiation onto the dust grains. Compared to the equivalent 12 μm images, the WISE 3.4 μm band provides a wide dynamic range in intensity without undesired saturation effects in bright regions such as OMC-1. These WISE images offer a complementary view to the molecular gas traced by our (sub-)millimeter maps.

The YSOs in both Orion A and B clouds have been systematically surveyed by Megeath et al. (2012) using *Spitzer* observations. From the analysis of their mid-IR colors, Megeath et al. classified these sources as protostars (P) or pre-main-sequence stars with disks (D), typically corresponding to Class 0/I and Class II objects (Greene et al. 1994), respectively. We adopted the Megeath et al. (2012) catalog as reference for our study. We refined this catalog by including additional young objects (protostars and flat spectrum, all included as P objects in our catalog) reported in follow-up *Herschel* observations (Furlan et al. 2016; Stutz et al. 2013). Overall, we expect a high homogeneity and completeness of this catalog in describing the low-mass YSO population within our ALMA fields with the possible exception of the OMC-1 and Flame Nebula due to their bright nebulae and crowded populations (see also Megeath et al. 2016; Großschedl et al. 2019, for a detailed discussion). The lack of large-scale, systematic surveys of Class III objects in Orion B (e.g., in X-rays) limits the extension of this analysis to older populations. In a separate catalog for massive stars, we collected sources in Orion that were classified as spectral types O and B using Simbad (Wenger et al. 2000).

4. Massive data reduction of ALMA observations

The nature and volume of our EMERGE Early ALMA Survey, including mosaics of a large number of target areas (5+2) combining different molecular (three) and continuum (one) maps, sometimes in multiple spectral configurations (two), requires the development of automatic data reduction techniques. The large number of potential data products plus the use of our single-dish data using different data combination techniques (see below) make the manual and customized reduction of each target not only highly inefficient but also costly in time. The optimization of this process and the production of homogeneous datasets is of paramount interest for our study.

In this section, we introduce a data reduction framework in which the standardization of the calibration and imaging steps allows the full automation of the ALMA data reduction process shortening the effective processing time by more than an order of magnitude. First, describe the calibration and imaging steps for individual observing setups in Sect. 4.1. Second, we describe how these standardized processes can easily be parallelized using a supercomputer in Sect. 4.2. Third, we quantify the quality of our reductions using advance quality assessments in Sect. 4.3. Finally, we present the resulting molecular line cubes and continuum maps, including when possible combined ALMA+IRAM-30 m observations, as final data products of our survey in Sect. 4.4.

4.1. Calibration, imaging, and data combination

We present the workflow of our data reduction scheme in Fig. 3 (lower panel). We exemplify this reduction process for the data obtained in the OMC-3 region from the original ALMA scheduling block (SB) containing raw visibilities including continuum and lines to the final data products in FITS format in both interferometry-alone and combined maps and cubes. Overall, the full data reduction process is divided into two stages: (1) the calibration of the raw visibilities, and (2) the imaging (or

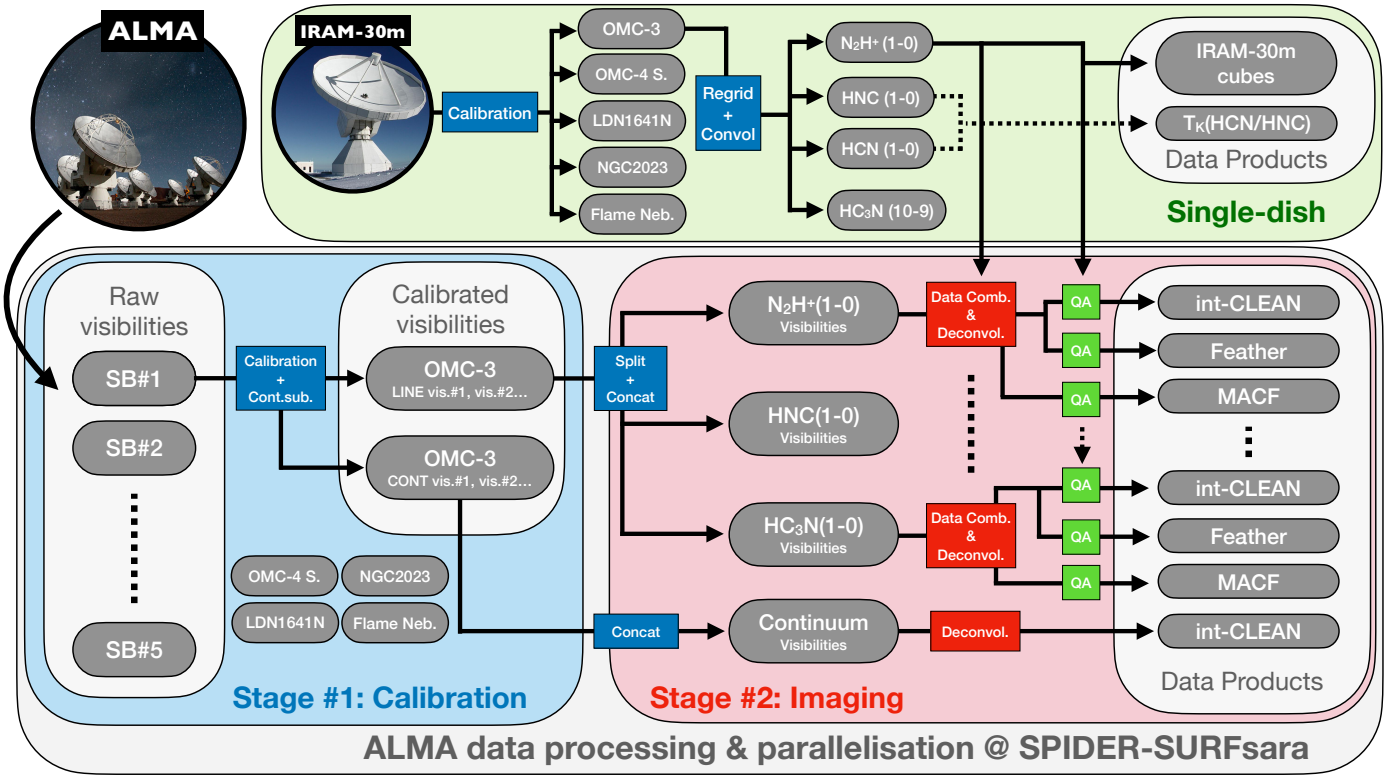

Fig. 3. EMERGE Early Science data reduction workflow for our single-dish, IRAM-30 m *(upper panel)* and interferometric ALMA (*lower panel*) observations. In the case of our ALMA reductions, both calibration (stage #1; light blue) and imaging (stage #2; light red) processes are carried out at the SPIDER-SURFsara supercomputer using parallel processing.

deconvolution) process. We carried out both calibration and imaging steps using the Common Astronomy Software Applications (CASA) package (CASA Team 2022; Emonts et al. 2022). We independently discuss each of these steps below.

As part of stage #1, we applied standard calibration to the raw visibilities in order to prepare our data for imaging (Fig. 3, bottom panel, left part). Different executions of the same SB lead to different measurement sets (MS) including multiple SPWs (lines + continuum). We recreated the ALMA pipeline calibration of each MS file applying facility-provided scripts (*scriptforPI*) using CASA versions 4.7.2 or 5.4.0 depending on the target. We then applied a continuum subtraction for all science SPWs in each calibrated MS file and created individual files for both continuum and continuum-subtracted line visibilities.

In stage #2, we independently imaged each line and continuum mosaic (Fig. 3, bottom panel, right part). We first split (only for line SPWs) and then concatenated the previous calibrated visibilities producing single visibility files for each molecular species and continuum dataset per field. Each ALMA mosaic was then imaged using the CLEAN deconvolution algorithm (Högbom 1974; Clark 1980) with the task `clean` in CASA version 5.7[2]. When possible, we applied data combination to include our IRAM-30 m data (Sect. 3.1) as short-spacing information (see below). To facilitate its implementation, we homogenized most of the `clean` parameters in our reductions. Our line and continuum mosaics were deconvolved adopting the phase centers indicated in Table 2 with a `mapsize` of 1600″×1600″ with a `cellsize` of 0.″45, and reduced using standard `clark` plus `briggs` weighting schemes with `robust` parameter equal to 0.5 using $5{\times}10^4$ `iterations`. As main differences, we set thresholds of 1 mJy for maps in continuum reduced in `msf` mode and of 10 mJy per channel for our line cubes obtained in `velocity` mode. In the case of the line cubes, we also optimized the spectral resolution (see last column in Table 3) and number of channels for the deconvolution depending on the species and resolution (narrow and broad) of our interferometric and single-dish data (see Table 3).

Plunkett et al. (2023) recently highlighted the role of data combination in the analysis of extended sources such as our ALMA targets. Due to the so-called short-spacing problem (Wilner & Welch 1994), the lack of sensitivity at short-baselines leads to scale-dependent filtering effects in both continuum and line observations (e.g., Pety et al. 2013). As shown in previous works (e.g., Leroy et al. 2021), these issues can critically affect the fidelity of interferometric observations. We explored the effects of data combination using three independent reductions for each of our molecular line mosaics: int-CLEAN, Feather, and MACF. A detailed description of these data combination methods and their implementation can be found in Plunkett et al. (2023).

In a first and basic reduction, we obtained an interferometric-only map by deconvolving the ALMA-only visibilities (int-CLEAN). Second, we used the task `feather` (Cotton 2017) as standard data combination CASA method to combine our previous int-CLEAN maps with our IRAM-30 m observations (Feather). Third, and finally, we applied a model-assisted

[2] We note that the ALMA task `tclean` showed unstable results when using a `startmodel` option during data combination in CASA version 5.7. These issues are not present in `clean` which motivated our preference for this latter task. Newer versions of CASA seem to have fixed this problem and `tclean` will be used in future EMERGE data reductions.

CLEAN and Feather combination (MACF) which introduces the single-dish information both as initial cleaning model and feathering image. In contrast to the line observations, no single-dish information is available in continuum. As a result, we were only able to produce interferometric-only (int-CLEAN) continuum maps for our sources. Additional data combination techniques (see Plunkett et al. 2023, for a review) will be explored in future papers of this series.

The deconvolution of our ALMA observations produces images with typical beam sizes of about $\theta_0 \sim 3\farcs5$. We applied a primary beam (PB) correction to all our cleaned images and masked all emission channels (voxels) with PB values lower than 0.5. To increase the stability of our maps, we also applied a small convolution after deconvolution (and before feathering) and resampled the resulting cubes every half beam, leading into Nyquist-sampled maps (i.e. $\theta/2$) and cubes with a final resolution of $\theta = 4\farcs5$.

Our specific choices for the above stage #2 parameters aim to define single parameter values that can homogeneously be applied to our entire sample. Spatial parameters such as the `mapsize` are automatically defined to cover the largest dimensions of our dedicated IRAM-30 m data around the corresponding ALMA fields. By design, we also adopted a single `cellsize` value of $\theta/10$ in order to correctly oversample the native beamsize θ_0 (i.e. `cellsize`$< \theta_0/5$) while at the same time facilitate the Nyquist resampling of our data products at their final resolution (final pixel size = 5× original `cellsize`). Other cleaning parameters were chosen as a compromise between the maximization of the quality of our maps (see quality assessments in Sect. 4.3) and the simultaneous minimization of the total computing time per field. Additional tests using larger number of iterations or lower thresholds produced no significant differences on the final data products while having a direct impact on the deconvolution timescales when applied to the large number of targets and lines included in our sample (76 individual reductions; see below). Dedicated reductions might produce better results for individual observations at the expense of longer computing times and human intervention per field. In contrast, our EMERGE Early ALMA Survey leverages a large statistical sample of targets and lines that are homogeneously reduced on short timescales (see Sect. 4.2). Similar optimization strategies appear of critical importance for future ALMA archival studies.

4.2. Parallelization on a high-performance computer

The standardization of the calibration and imaging steps (Sect. 4.1) makes our data reduction process easy to parallelize into separate CASA jobs. As shown in Fig. 3, we also took advantage of the original ALMA data structure. Each of our ALMA fields was stored in a single SB that could be recreated separately. After an early split, the imaging process of the resulting calibrated continuum and line visibilities could then be then run independently in an automatic way. Moreover, each of these modular steps could be carried out using dedicated CASA versions.

We executed parallel runs of both calibration and imaging steps for our entire ALMA sample in the SPIDER Data Processing Platform[3], which is part of the SURF[4] Data Processing facilities in Amsterdam[5]. The high performance and flexibility of SURF-SPIDER permitted us to run a large number of CASA instances in parallel. First, we simultaneously recreated all our calibrated visibilities at once (stage #1). Second, we simultaneously carried out all individual line and continuum deconvolutions (including data combination) (stage #2). Each CASA run was assigned to five CPUs providing a total memory of 40 GB per job to optimize the CPU and queuing times. Additional tests indicated no performance improvement when scaling our reductions with a larger number of cores[6] or with CASA executions in parallel mode (`mpicasa`) in the SURF-SPIDER architecture. This approach is similar to the parallel reduction method introduced by van Terwisga et al. (2022) for the massive data processing of ALMA single-pointing continuum maps, this time, extended to mosaics and spectral cubes.

The efficiency of our data reduction process strongly benefits from the use of high-performance computers such as SURF-SPIDER. Our full ALMA survey consists of a total of 76 individual data products including all molecular line and continuum maps and data combination techniques of all our ALMA fields. The processing time per individual target line (including the three data combination methods) is about 30–60 h of computing time depending on the spectral resolution and queuing time. On the other hand, with the implementation of our parallel data processing the entire EMERGE Early ALMA Sample can be reduced in $\lesssim$72 h (including overheads), that is, at least an order of magnitude less time than the standard (linear) reduction. The application of this parallel methodology to in massive ALMA archive research will be presented in a future work.

4.3. Quality assessments

The use of automatic reductions (see Sect. 4.2) requires a careful evaluation of the quality of our data products. To homogeneously quantify the goodness of our reduction and combination processes, we have developed several of the quality assessment (QA) metrics described by Plunkett et al. (2023). We focus on the comparison of the flux recovered in our spectral cubes (N_2H^+, HNC, HC_3N) using different methods (int-CLEAN, Feather, and MACF) with respect the total flux present in our single-dish, IRAM-30 m observations (see Fig. 3). A detailed discussion of our QA results can be found in Appendix A.

Our analysis demonstrates that the addition of the the short-spacing information significantly improves the image fidelity of our ALMA observations (Appendix A.1; see also Paper II). Similar to the results discussed in Plunkett et al. (2023), our int-CLEAN reductions exhibit flux losses of more than 70% of the flux detected in the single-dish data due to interferometric filtering. In contrast, the implementation of data combination techniques such as Feather and MACF improves the flux recovery

[3] https://spiderdocs.readthedocs.io/en/latest/
[4] https://www.surf.nl/
[5] The EMERGE Early Science Survey was executed as part of the projects "Advance ALMA data reduction using SPIDER" (PI: A. Hacar) and "Advanced ALMA data reduction" (PI: A. Ahmadi) carried out at SURF-SPIDER in collaboration with the Allegro Dutch ARC node in Leiden.
[6] We remark here that the use of a higher number of cores (CPUs) indeed reduces the execution time per reduction although this does not significantly improve the overall performance of the sample processing. Our tests indicate that the reduction process inversely scales with the number of cores. However, this performance scaling becomes sub-linear with more than 5 cores and usually saturates at ∼10 cores. The request of >5 cores makes the reduction process not only more expensive in computational resources but also more difficult to schedule by the queuing system in large numbers. Our choice is therefore justified in order to optimize the use of computational resources while reducing the execution time (queuing + reduction) for the entire sample.

to ≳90%. The lack of short-spacing information also has a dramatic impact on the gas kinematics at all scales (Appendix A.3). Different velocity components are selectively (and largely unpredictably) filtered in velocity, altering the derived parameters such as line intensities and ratios, centroid velocities, and linewidths. Our QAs highlight the need of data combination for the analysis of regions with complex and extended emission similar to those included in our sample.

While both the Feather and MACF methods produce satisfactory reductions, MACF at high spectral resolution (narrow) shows the most stable results in our assessments (total flux recovery, flux per intensity bin, and flux per channel) in most of our fields and spectral setups (see Appendix A.2 for a full discussion). For a homogeneous analysis, and to facilitate the comparison with the OMC-1 and OMC-2 results (Hacar et al. 2018), we adopted our MACF reductions as final science products in all our fields (see Sect. 4.4) and will use them during the analysis of our EMERGE sample (hereafter referred to as ALMA+IRAM-30 m data). On the other hand, no single-dish data were available to complement our continuum observations. Our ALMA Band-3 continuum int-CLEAN maps should therefore be treated with appropriate caution.

4.4. Final data products and data release 1

Our final EMERGE Early ALMA Survey consists of 76 ALMA datasets, including spectral and 3 mm continuum maps and multiple data combinations, in the OMC-3, OMC-4 South, LDN 1641N, NGC 2023, and Flame Nebula star-forming regions, all with a final resolution of 4.″5 or ∼2000 au at the distance of Orion. Our sample also includes a total of 30 IRAM-30 m-only spectral maps of the same line tracers (see Sect. 3.2) at 30″resolution. Moreover, we obtained the T_K (30″; Sect. 3.2) and derived a set of total column density maps at high spatial resolution (4.″5) using a new technique that is to be presented in Hacar et al (in prep.) in all these targets (five times each) as high-level data products. Our new observations are also complemented by similar ALMA (only continuum and N_2H^+) plus IRAM-30 m observations in OMC-1 and OMC-2 (Hacar et al. 2018; van Terwisga et al. 2019).

All our fully reduced, ALMA plus IRAM-30 m data products will be included in different data releases (DR) that are accessible at a dedicated page on our website[7]. Accompanying this paper, we have released all integrated-intensity maps of our ALMA and IRAM-30 m observations in FITS format as part of the EMERGE data release 1 (DR1). Future DRs will include all spectral cubes and column density maps of our sample.

5. Sample properties

In this section we present the main physical properties of the target sample observed in our EMERGE Early ALMA Survey based on the analysis of their integrated-intensity and continuum maps at both (low-) single-dish (Sect. 5.1) and (high-) interferometric (Sect. 5.2) resolutions. To illustrate our results (e.g., emission distribution, and gas structure) we use OMC-3 as representative showcase of our dataset (e.g., Fig. 4). Similar plots for other regions can be found in Appendix B. To facilitate the comparison between the targets, we display all regions within the same plotting ranges in each of the figure panels. Our ALMA targets cover a wide range of physical conditions in terms of mass, structure, stellar content, and evolution (see below). We quantify and summarize some of these properties in Table 1. Additional analyses using the full spectral information will be presented in future papers of this series (e.g., gas kinematics; see Paper III).

In our calculations, we adopted standard distances to our targets (see Table 1). For sources in Orion A (OMC-1, 2, 3, and 4 South, and LDN 1641N), we assumed a typical distance of 400 pc as a compromise between the VLBI measurements of the ONC (414±7; Menten et al. 2007) and the *Gaia* estimates for the head of the ISF (393±13; Großschedl et al. 2018). On the other hand, we use a distance of 423 pc for Orion B (NGC 2023 and Flame Nebula) in agreement with the *Gaia* (423±21; Zucker et al. 2019) and VLBI (∼420 pc; Kounkel et al. 2017) results. Uncertainties of ±10–20 pc (or ∼5% of the cloud distance; see references above) are expected from these values, but produce only minor effects on our mass and size calculations. For a typical distance of 400 pc, the corresponding physical resolution of our data would then be 0.058 pc (or 12 000 au) for our IRAM-30 m observations (30″) and 0.009 pc (or 1800 au) for our ALMA maps (4.″5).

5.1. Low-resolution observations

5.1.1. Gas column density and temperature

While limited to only seven targets, our sample covers an extensive range of physical conditions and environments. In addition to the overall description of our survey (Sect. 2), we quantified additional evolutionary properties of our sources considering all positions covered by our IRAM-30 m (broad) observations within a representative area of 1.5×1.5 pc^2 from the center of our maps (∼700×700 arcsec^2; roughly similar to the zoom-in maps in Figs. 1 and 2).

The wide range of stellar activity in our sample is reflected in the populations of protostars (P) plus disk (D) YSOs (Table 1). Our survey covers almost two orders of magnitude in average stellar density (i.e. (P+D)/area or simply P+D) from the highly clustered OMC-1 region to the more diffuse NGC 2023, which forms stars almost in isolation. We note here that the stellar density is largely underestimated in the case of OMC-1 and Flame Nebula since our stellar counting does not include any of the optical stars that have been identified in these clouds or the bright emission of their nebulae, which rapidly increases the incompleteness of IR surveys (Megeath et al. 2016).

The diversity of our sample becomes apparent in the distribution of the column density probability distribution functions (N-PDF) displayed in Fig. 5 (left panel)[8]. Large amounts of gas at $N(H_2) > 10^{23}$ cm^{-2} are seen in regions forming O-type stars (OMC-1 and Flame Nebula), close to the theoretical predictions for the formation of high-mass stars (i.e. $N(H_2) = 10^{23.4}$ cm^{-2}; Krumholz & McKee 2008). In contrast, low-mass and diffuse regions (e.g., OMC-4 South or NGC 2023) reach maximum column densities of $N(H_2) \sim 10^{23}$ cm^{-2} in only a few locations. Likewise, the high-end of the N-PDF distributions in active regions with higher fractions of dense gas (OMC-1, 2, and 3, LDN 1641N, and Flame Nebula) shows shallower slopes than those that are found in more diffuse clouds (OMC-4 South and NGC 2023) in agreement with previous works in some of these clouds (Stutz & Kainulainen 2015).

[7] https://emerge.univie.ac.at/results/data/

[8] We note that the *Herschel* maps used in Lombardi et al. (2014) show several saturated pixels at the brightest positions in both the OMC-1 and Flame Nebula fields. The derived $N(H_2)$ values at these positions should be then taken with caution (see secondary peak in the N-PDF of OMC-1 in Fig. 5, top left panel).

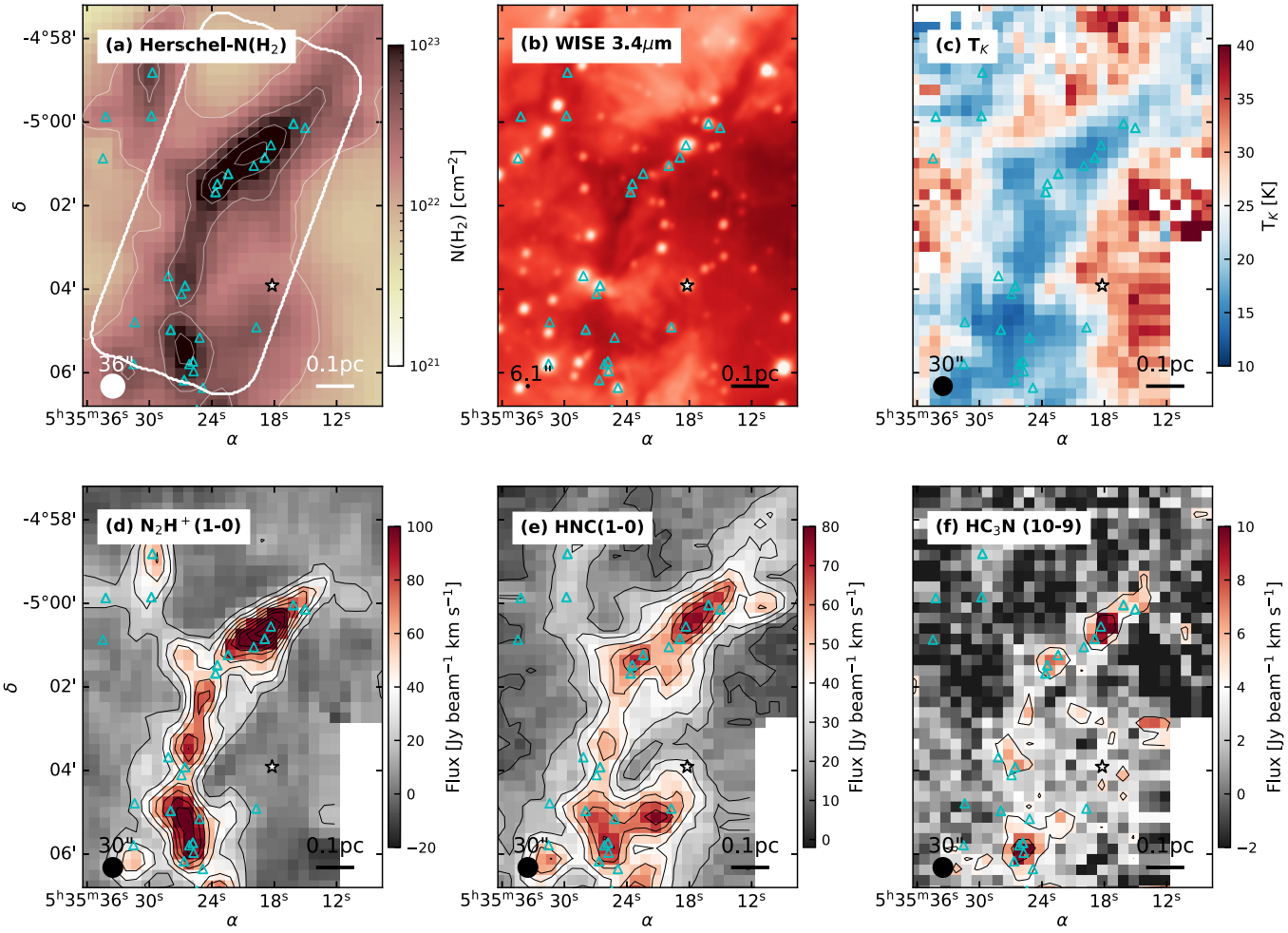

Fig. 4. Low-resolution observations in OMC-3. From left to right and from top to bottom we plot (*a*) *Herschel*, total gas column density, $N(H_2)$ (36″); (*b*) WISE 3.4 μm emission (6.1″); (*c*) gas kinetic temperature, T_K, (*d*) N_2H^+ (1–0), (*e*) HNC (1–0), and (*f*) HC_3N (10–9) integrated intensity maps obtained in our IRAM-30 m (broad) observations (30″). The symbols are similar to those in Fig. 1. The corresponding beam size and scale bar are indicated in the lower corner of each panel. For reference, panel a includes contours at $N(H_2)$ = [15, 50, 100, 150, 200] $\times 10^{21}$ cm^{-2} and the footprint of our ALMA observations shown in Fig. 8. Similar plots for all the other regions can be found in Appendix B.

Our gas kinetic temperature T_K maps (see Fig. 4c) illustrate the thermal evolution of gas during the star formation process. Well shielded high column density regions with $N(H_2) > 2 \times 10^{22}$ cm^{-2} usually show cold temperatures of $\lesssim$25 K. In contrast, lower column density material is usually found at higher temperatures up to ∼40 K, which is heated by external radiation and results in a positive thermal gradient toward the edge of the cloud. Local heating effects are seen to be associated with some locations with strong star formation activity (e.g., compare the positions of the WISE sources with our T_K maps around OMC-2 FIR-2 in Figs. B.2b and B.2c). More prominently, regions under the influence of high-mass stars (OMC-1 and Flame Nebula) show large areas of both low- and high column density gas at temperatures above 40 K that is coincident with the bright emission in the WISE 3.4μm maps (see Figs. B.1 and B.6, respectively). These results demonstrate that low-resolution HCN/HNC measurements are sensitive enough to describe the thermal properties across a wide range of column densities and environments (see Hacar et al. 2020, for a discussion).

Feedback also shapes the overall temperature distribution within these clouds, as shown in the T_K histograms (or temperature PDF, T_K-PDF) in Fig. 5 (right panels), and it changes the average gas temperature in these regions (see Table 1). Most of the gas in low-mass star-forming regions (e.g., OMC-3 or OMC-4 South) is found at temperatures between 10–30 K with mean values of ∼25 K. In contrast, regions with embedded massive stars show higher average values of $\gtrsim$30 K with significant amounts of gas at $T_K > 45$ K (see the tails of the T_K-PDF in OMC-1 and Flame Nebula). Interestingly, LDN 1641N shows a significant fraction of low column density gas above 40 K surrounding the cloud (see the external heating in Fig. B.4c). The origin of this warm component is unclear, although the cometary shape of this cloud that points toward the NGC 1980 cluster might suggest a potential interaction with ι Ori and NGC 1980.

5.1.2. Molecular emission and density-selective tracers

We describe the global emission properties of our sample using our low spatial resolution observations shown in Fig. 4 (lower panels). To do this, we took advantage of the extended spatial coverage of our (broad spectral resolution) IRAM-30 m maps at low spectral resolution (see Sect. 3.2). Together with the

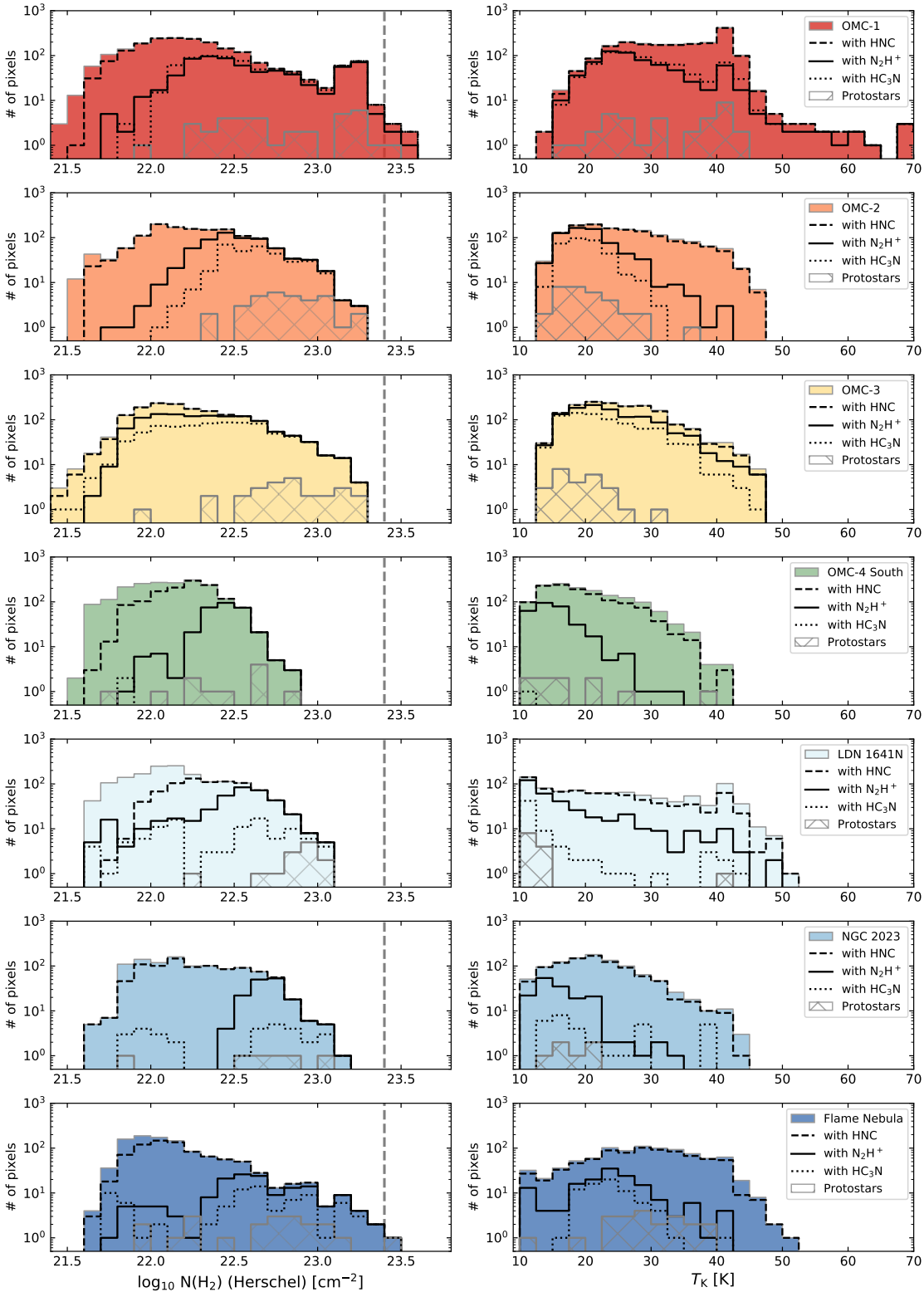

Fig. 5. Histograms of the (*left*) gas column density $N(H_2)$ and (*right*) gas kinetic temperature T_K distributions detected at low spatial resolution (*Herschel* or IRAM-30 m) within the central 1.5×1.5 pc^2 (~700×700 $arcsec^2$) of all regions part of the EMERGE Early ALMA Survey (see also Tables 1 and 2). *From top to bottom*: OMC-1, OMC-2, OMC-3, OMC-4 South, LDN 1641N, NGC 2023, and Flame Nebula. Positions showing detected emission above the first contour in Fig. 4 (and Figs. B.1–B.6) of HNC (black dashed lines), N_2H^+ (black solid lines), and HC_3N (black dotted lines), as well as including protostars (hatched gray areas) are highlighted in all histograms. A vertical gray dashed line in the $N(H_2)$ histograms indicates the minimum column density of $10^{23.4}$ cm^{-2} (or 1 g cm^{-2}) needed to form a high-mass star (Krumholz & McKee 2008).

$N(H_2)$, WISE, and T_K maps (panels a-c) that we presented in Sect. 5.1.1, we display the N_2H^+ (1–0), HNC (1–0), and HC_3N (10–9) integrated-intensity maps in Fig. 4 (panels d–f) in OMC-3. We obtained each of these molecular maps by adapting the range of integrated velocities to the corresponding cloud and transition. Similar plots for other regions can be found in Figs. B.1–B.6.

The emission of all molecular tracers is clearly detected in all sources (panels d–f), except for the HC_3N emission in OMC-4 South (see Fig. B.3). The overall emission properties in our maps are similar to the results reported in previous large-scale surveys in Orion (Pety et al. 2017; Kauffmann et al. 2017). N_2H^+ exhibits the brightest integrated-intensity values, which are typically associated with regions of high column density above $N(H_2) > 2 \times 10^{22}$ cm^{-2} (or $A_V \gtrsim 20$ mag). HNC (as well as HCN, not shown) displays the most extended emission in our sample. The HNC emission reaches down to column densities of $N(H_2) \sim 10^{21}$ cm^{-2} (or few A_V), and in many cases, extends over the limits of our IRAM-30 m maps. On the other hand, the HC_3N emission is usually clumpier, located in smaller areas, and comparatively weaker, although it shows bright emission features that re closely related to the position of some (but not all) young protostars. The relative contributions of the different tracers become apparent in the $N(H_2)$ and T_K histograms in Fig. 5, where we highlight pixels with N_2H^+ (solid black line), HNC (dashed black line), and HC_3N (dotted black line) emission within the first contours our of map (e.g., see Fig. 4), and pixels including protostars (hatched gray area).

In terms of line brightness, our IRAM-30 m maps display a wide dynamic range over up to two orders of magnitude in integrated-intensity, and with regions such as OMC-1 showing peak values >200 Jy beam^{-1} km s^{-1} (or >30 K km s^{-1}; e.g., Fig. 4). Moreover, the total luminosity of our maps systematically changes between the regions in our sample (see the comparison between the total number of pixels included in the histograms in Fig. 5), and it is closely related to their physical characteristics (see Sect. 5.1). Regions reaching high column densities (OMC-1/2/3 and LDN 1641N) and/or high temperatures (OMC-1 and Flame Nebula) show the brightest lines and total intensity maps in all transitions. In contrast, the emission of the same lines in more diffuse and colder regions (NGC 2023 and OMC-4 South) is always weaker.

The results of our molecular maps match the expected behavior for the suite of selective tracers included in our survey (see line upper energies and effective densities in Table 3). Formed after the depletion of CO (see Bergin & Tafalla 2007, for a review), N_2H^+ is a density-selective species tracing the cold ($\lesssim 20$ K) and dense ($>10^4$ cm^{-3}) gas (Tafalla et al. 2002), while its ground transition is associated to dense cores (Caselli et al. 2002) and filaments (Hacar et al. 2017b, 2018). Complementary to N_2H^+, the HNC (1–0) line (as well as the HCN (1–0)) is effectively excited at densities of a few 10^3 cm^{-3} (Shirley 2015), and its emission traces the cloud gas from low to intermediate column densities before it is also depleted at high densities (Tafalla et al. 2021, 2023). On the other hand, the excitation of the HC_3N (10–9) transition requires higher temperatures ($E_u = 24.1$ K) and is only effectively excited in lukewarm regions exposed to feedback (Hacar et al., in prep.).

5.1.3. Evolutionary state and dense gas star formation efficiency

Additional insight into the evolutionary state of our targets can be inferred from the comparisons of their populations of P and D stars, respectively. From the comparison of the number of Class II (similar to D) and Class I (similar to P) stars in nearby clouds, and assuming a typical age of 2 Myr for a typical Class II objects, IR surveys of nearby stars have inferred a median evolutionary timescale for a Class I object of ∼0.5 Myr (Evans et al. 2009). Based on these typical ages, the observed P/D ratio can be used as proxy of the relative ages between targets (see Table 1). Young regions in which most of the stars were formed within the past Myr are expected to show high P/D$\gtrsim$1. Clouds that continuously form stars for long periods of time should approach a steady-state ratio in which P/D = $\frac{0.5\ \mathrm{Myr}}{2.0\ \mathrm{Myr}} \sim 0.25$. Finally, older regions with an already declining star formation rate should exhibit ratios of P/D<0.25 and approach P/D→0 when they exhaust their gas reservoir for forming stars. Within our sample, OMC-1, Flame Nebula, and OMC-4 South appear to be older and more evolved (P/D$\lesssim$0.2) than younger regions such as OMC-2/3 and LDN 1641N (P/D>0.3) and particularly NGC 2023 (P/D=0.75).

We quantified the potential of these regions to form new stars from the amount of gas detected in our *Herschel* maps (see Fig. 4a). We obtained three measurements from adding all *Herschel* column density measurements: (a) the total mass of gas in our maps (M_{tot}), (b) the mass of dense gas M_{dense} showing significant N_2H^+ emission ($\gtrsim$15 Jy beam^{-1} km s^{-1} or $\gtrsim$2.5 K km s^{-1} in T_{mb} units; see also Sect. 5.1.4) and (c) the fraction of dense gas (f_{dense}) obtained from the ratio of M_{tot} and M_{dense}. Active regions such as OMC-1, OMC-2, OMC-3, and LDN 1641N exhibit large mass reservoirs ($M_{tot} > 500\ M_\odot$) with significant amount of dense gas ($M_{dense} > 300\ M_\odot$ and $f_{dense} \gtrsim 0.4$) within their central ∼1 pc^2. Although comparable in total mass, OMC-4 South and NGC 2023 are more diffuse ($M_{dense} < 90\ M_\odot$ and $f_{dense} \lesssim 0.2$). Flame Nebula is a particularly interesting case with the lowest available mass ($M_{tot} < 300\ M_\odot$), but with a significant fraction of dense gas ($f_{dense} = 0.33$). In its latest stages of evolution (see above), the dense gas in Flame Nebula (as well as in OMC-1) appears to be more resistant to the disruptive effects of feedback, which agrees with simulations (Dale et al. 2014).

Despite the above exceptions, the values of M_{dense} and f_{dense}, both correlated with the amount of N_2H^+ emission, appear as better proxies of the current star formation than the total gas mass M_{tot} of the region. We illustrate these correlations in Fig. 6 by comparing the number of protostars (P) with the amount of dense gas M_{dense} (upper panel) and the fraction of dense gas f_{dense} (middle panel) across our sample (see values in Table 1). As shown in Fig. 6 (upper panel), the number of protostars in all our targets shows a tight linear dependence (dashed red line) with M_{dense} (with a Pearson coefficient $p = 0.90$) that is stronger than the dependence obtained with M_{tot} ($p = 0.77$; not shown). The correlation with M_{dense} further improves when Flame Nebula is removed ($p = 0.96$; not shown). A similar linear dependence is shown in Fig. 6 (middle panel) when comparing the number of protostars with f_{dense} ($p = 0.91$). Figure 6 reinforces the direct correspondence between the amount of dense gas (traced by N_2H^+) and the ongoing star formation in our targets.

The linear correlation between total number of YSOs (P+D) and the amount of gas at column densities above $A_V \gtrsim 8$ mag, used as proxy of M_{dense} for the dense gas above 10^4 cm^{-3}, has been well documented in the past in large-scale surveys of nearby molecular clouds (e.g., Lada et al. 2010). A similar dependence is found when comparing the surface density of protostars (P) (alone) and high column density gas (Heiderman et al. 2010). This positive correlation indicates how high-density material is needed to efficiently form stars within these clouds (e.g., Lada 1992). Figure 6 indicates that this correlation tightens for the

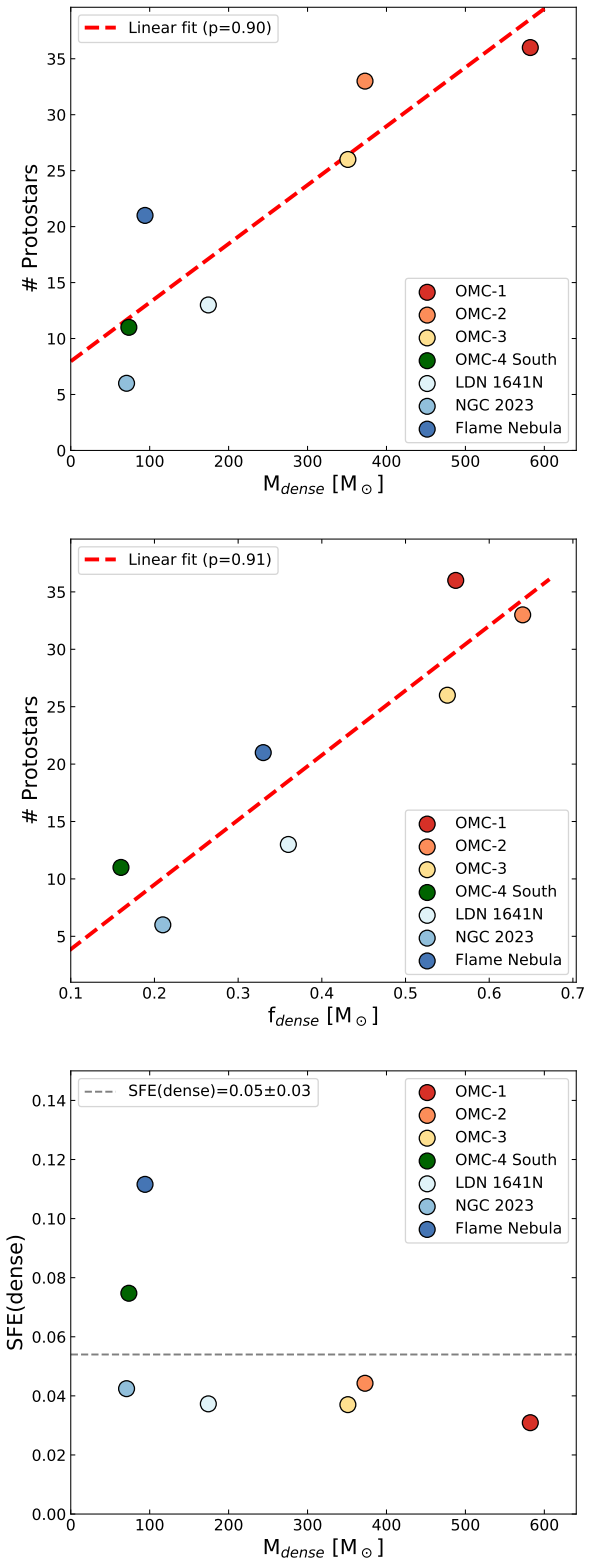

Fig. 6. Star formation activity in the different regions explored in our sample (see legend). *From top to bottom*: (*upper panel*) number of protostars (P) with respect to the total amount of dense gas $M_{\rm dense}$, (*middle panel* fraction of dense gas $f_{\rm dense}$, and (*lower panel*) star formation efficiency in the dense gas (see the values in Table 1). We display the linear fit of all our targets (dashed red line) and their corresponding Pearson p-coefficients (see the values in legend) in the upper and middle panels.

youngest protostars (P) and the highest-density material (traced by N_2H^+) in our targets.

The linear correlation found in Fig. 6 (upper panel) suggests a roughly constant star formation efficiency per unit of dense gas SFE(dense) of a few percent. In Fig. 6 (lower panel), we estimate this efficiency as SFE(dense) = $\frac{M_{\rm proto}}{M_{\rm dense}+M_{\rm proto}}$ (e.g., Megeath et al. 2022), where $M_{\rm proto}$ (= $0.5 \times \#Protostars$) corresponds to the total mass in protostars each assumed with a typical mass of 0.5 $M_\odot$. Our targets exhibit a typical SFE(dense) of about 0.05 ± 0.03 (or ~5%; dashed gray line) for $M_{\rm dense}$ values between ~100 and 600 $M_\odot$. Compared to the ~0.1 Myr free-fall time $\tau_{\rm ff} = \sqrt{3\pi/32G\rho}$ for a typical dense gas density of $n \sim 10^5$ cm^{-3} traced by N_2H^+, the above values for SFE(dense) translate into an approximate efficiency per free-fall time of $\epsilon_{\rm ff} \sim 1\%$, which agrees with previous predictions (Krumholz & Tan 2007) and surveys (Pokhrel et al. 2021). The roughly constant SFE(dense) derived from our data indicates that the total mass of dense gas is the primary factor determining the current star formation activity of our regions, leading into a number of young protostars that is directly proportional to the amount of $M_{\rm dense}$ that is available in each case (see also Lada et al. 2010).

This is expected as the cold ($T_{\rm K} \lesssim 20$ K) and dense ($n(H_2) > 10^5$ cm^{-3}) material traced by N_2H^+ (Sect. 5.1.1) promotes the conditions for gravitational collapse and is therefore prone to forming stars. Active star-forming regions showing more protostars (e.g., OMC-1) are simply the result of their (currently) higher content of this dense and cold material, showing higher star formation rates (SFR(dense)=$\frac{\epsilon_{\rm ff}}{\tau_{\rm ff}} \times M_{\rm dense}$. When this dense gas is generated, it forms stars at a constant SFE(dense), regardless of the star formation regime (low- vs. high-mass) and evolutionary state (young vs. old, or P/D value; see above) of the cloud.

5.1.4. Timescales for the evolution of the dense gas

As demonstrated in previous sections, N_2H^+ appears to be the most direct proxy of the young star-forming gas in all targets of our survey. Following the method introduced by Hacar et al. (2017b), we show in Fig. 7 the cumulative distribution of the N_2H^+ integrated intensity (in K km s^{-1}) at the position of protostars (P; solid colored line) and disk stars (D; dotted colored line) found in the seven regions of our study (see panels). In all cases, the distribution for P stars always runs toward the right side of the one for D star distribution, indicating that protostars are typically associated with brighter N_2H^+ emission in our maps, as expected for this dense-gas tracer.

We compared these observational results with the expected cumulative plots for a similar total number of objects per cloud (see legend) located following a series of simulated distributions: (a) a random distribution (solid gray line) and independent distributions in which the position of these objects is favored at the position of the N_2H^+ emission in our maps using (b) linear (dotted gray line) and (c) quadratic (dashed gray line) weights, respectively. In all our targets, the distribution of D stars closely resembles a (quasi-)random distribution with respect to the dense gas traced in N_2H^+ indicating that the D stars and N_2H^+ emission are not correlated. In contrast, P stars have a much stronger connection with these emission features and typically show a linear or even quadratic correlation with the N_2H^+ intensities in regions such as OMC-2/3 and LDN 1641N.

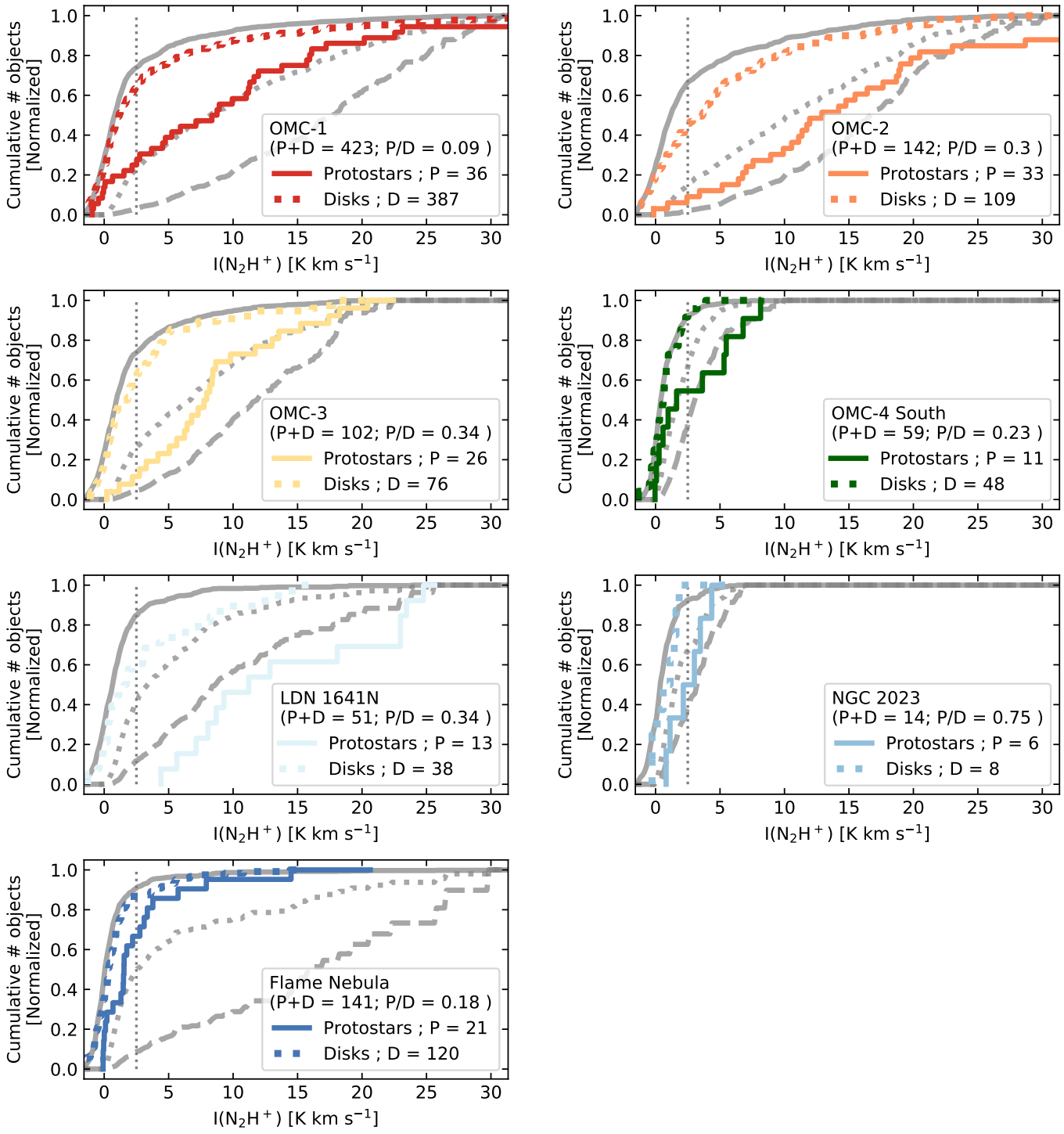

Fig. 7. Cumulative distributions of the N_2H^+ (1–0) integrated emission I(N_2H^+) at the positions of protostars (solid colored line) and disk stars (dashed colored line) within the central 1.5×1.5 pc^2 (~700×700 arcsec^2) in each of the regions that are part of the EMERGE Early Science sample. All panels are displayed within the same ranges of N_2H^+ intensity to facilitate their comparison. In each subpanel, the local number of protostars (P) and disk stars (D), as well as their ratio (P/D), is indicated in the legend (see also Table 1). In each of these regions, the average result of simulating the expected cumulative distributions of a P+D number of objects following (a) a random distribution (solid gray line), (b) a linear dependence (dotted gray line), and (c) a quadratic dependence (dashed gray line) are shown in the corresponding panels.

Deviations from the above general behaviur seen in OMC-1 (linear) and, more prominently, in Flame Nebula (sublinear) can be explained by the evolutionary state of these targets (see Sect. 5.1.3) and the direct impact of radiative and mechanical feedback in regions hosting high-mass stars. Only small amounts of dense gas are detected in the immediate vicinity of the Trapezium stars (including θ^1 Ori C) and NGC 2024 IRS-2b, which demonstrates the rapid dispersion and photodissociation of the molecular material after the formation of these high-mass objects. The still visible correlation between the N_2H^+ emission and the location of the young protostars in these targets indicates, however, that the dense gas is not strongly affected when the

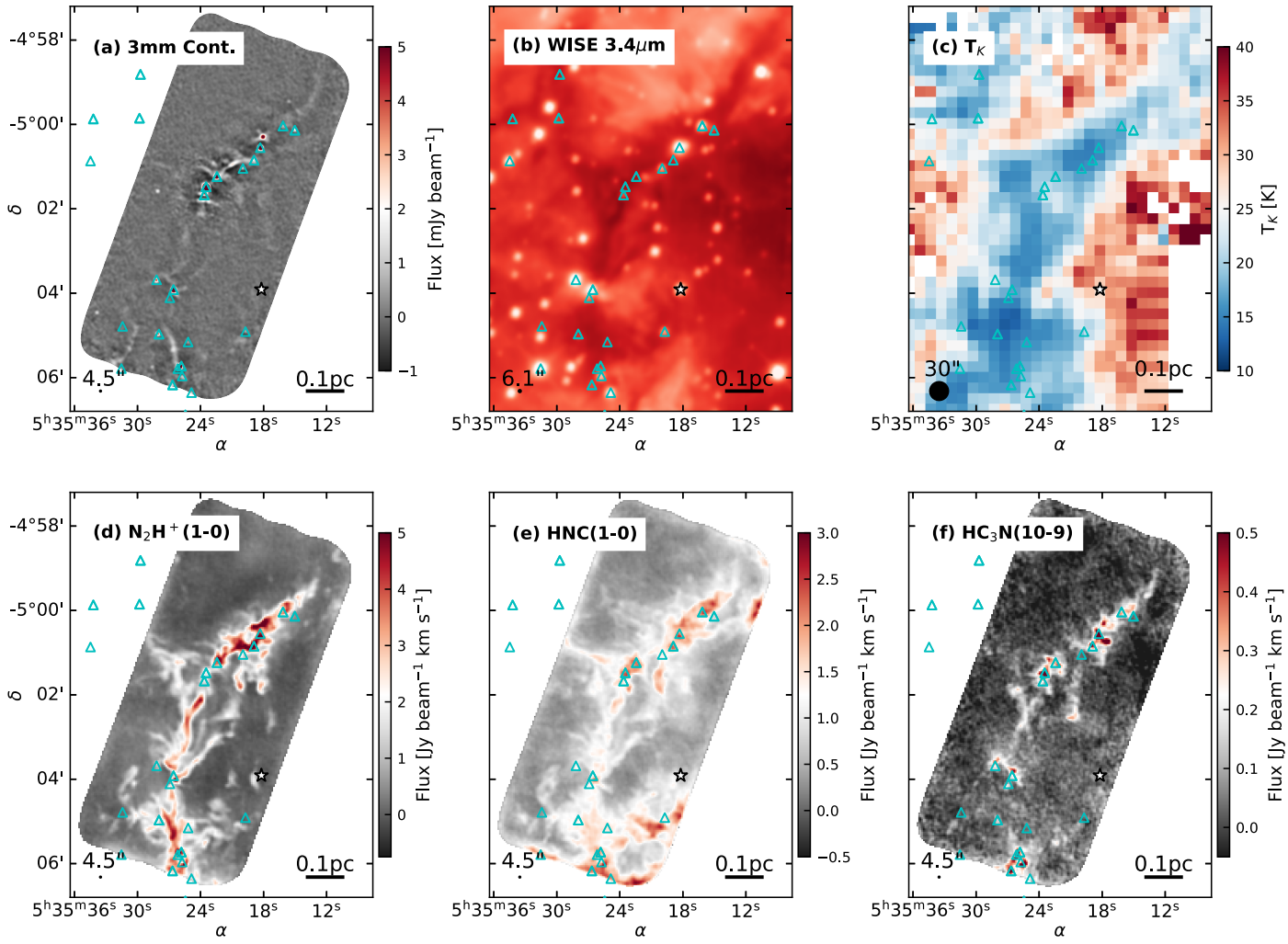

Fig. 8. High-resolution observations in OMC-3. *From left to right and from top to bottom*: (*a*) ALMA interferometric-alone (int-CLEAN) continuum map (4.″5); (*b*) WISE 3.4 μm emission (6.1″); (*c*) gas kinetic temperature, T_K (30″), (*d*) N_2H^+ (1–0), (*e*) HNC (1–0), and (*f*) HC_3N (10–9) integrated-intensity maps obtained in our ALMA+IRAM-30 m MACF observations (4.″5). The symbols are similar to those in Fig. 1. Similar plots for all the other regions can be found in Appendix B. The enhanced resolution of our ALMA+IRAM-30 m maps (panels a, and d–f) can be seen from their comparison with the gas T_K map (panel c) displayed at the single-dish resolution.

cloud can shield this intense radiation field (e.g., O'Dell 2001). In fact, large fractions of dense undisturbed gas are still detected in regions such as OMC-1 Ridge and Orion South at a distance of (∼0.1–0.2 pc in projection) (e.g., Wiseman & Ho 1998).

The results in Fig. 7 suggest a fast evolution of the dense gas in the targets included in our survey. The strong correlation between P stars and N_2H^+ indicates that the dense gas survives at least the typical duration of the protostellar (Class 0/I) phase with ∼0.5 Myr. On the other hand, the random distribution of D stars indicates that this correlation is lost at the typical timescales of ∼2 Myr for this older (Class II) YSO population. Analogous trends were reported in other filamentary clouds such as Taurus (Hartmann et al. 2001, comparing CO maps with the position of YSOs) and Perseus (Hacar et al. 2017b, using a similar analysis based on N_2H^+ observations). The combination of the above results suggests that the typical evolutionary timescales of the dense gas currently observed in N_2H^+ is therefore fast and lasts $\lesssim$1 Myr. A similar fast evolution of the dense star-forming gas traced by N_2H^+ has recently been proposed from simulations (Priestley et al. 2023). Within these timescales, a small fraction of the dense gas would be either accreted onto stars following its small SFE(dense) (Sect. 5.1.3) while most of the remaining dense gas will be recycled by the turbulent motions inside these regions (e.g., Padoan et al. 2016). The diversity of environments and regions explored here rules out exotic ejection mechanisms (e.g., Stutz 2018). Instead, this fast evolution appears to be driven by the continuous assembly of dense gas within these regions and its subsequent local destruction when star formation is ignited in them. In contrast to more traditional quiescent scenarios (e.g., quasi-static fragmentation), these results depict a dynamical evolution for typical star-forming regions such as Orion that agrees with recent simulations (Padoan et al. 2016; Smith et al. 2020; Seifried et al. 2020; Ibáñez-Mejía et al. 2022).

5.2. High-resolution observations: Cloud substructure at 2000 au resolution

Our ALMA observations provide a new perspective of the gas structure down to ∼2000 au resolution (or 4.″5 at the distance of Orion). In analogy to our low-resolution data (Sect. 5.1), we illustrate in Fig. 8 the high-resolution 3 mm continuum (panel a), as well as the integrated emission maps of N_2H^+ (1–0) (panel d),

HNC (1–0) (panel e), and HC_3N (10–9) (panel f) lines observed by ALMA in the OMC-3 region. Similar plots for all the other targets in our sample can be found in Figs. B.7–B.12.

The interferometric-alone (int-CLEAN) 3 mm continuum maps such in Fig. 8a show a combination of two different emission mechanisms in our targets. We find large emission areas with fluxes of >10 mJy beam^{-1} are detected in both OMC-1 (Fig. B.7a) and Flame Nebula (Fig. B.12a) coincident with their bright emission nebulae and contaminated by the free-free (bremssthrahlung) emission in these HII regions up to 100 GHz (see Mason et al. 2020, for a discussion). Outside these areas, the 3 mm emission is dominated by the thermal dust continuum and reaches values of $\lesssim$5 mJy beam^{-1}. In fields such as OMC-3, we identify multiple bright, compact sources coincident with the position of previously identified millimeter compact objects (Takahashi et al. 2013), YSOs (Megeath et al. 2012), and protoplanetary disks (van Terwisga et al. 2019), all of them unresolved in our maps. We detect a more extended cloud emission with fluxes of $\lesssim$ 2 mJy beam^{-1} toward column densities of $N(H_2) \gtrsim 50 \times 10^{21}$ cm^{-2} that show a large number of filamentary structures with sizes of a few 10^4 au. Because of this limited sensitivity, however, regions such as OMC-4 (Fig. B.9a) or NGC 2023 (Fig. B.11a) become mostly undetected in our survey. On the other hand, we identify clear negative side-lobes in regions with extended and bright emission such as OMC-1/2/3. Additional short-spacing information in continuum (via SD data) is thus needed to carry out a full analysis of these data products[9].

The enhanced dynamic range of our molecular ALMA+IRAM-30 m (MACF) maps provides a more complete picture of the gas distribution in our targets. The N_2H^+ (1–0) emission above $\gtrsim$1 Jy beam^{-1} km s^{-1} (>7 K km s^{-1}) shows in a plethora of small-scale, elongated and narrow structures in all our targets (Fig. 8d). The previous characterization of the N_2H^+ emission in OMC-1 and OMC-2 (see Figs. B.7d and B.8d) identified these filamentary features as velocity-coherent, sonic fibers (Hacar et al. 2018). The analogous emission features seen in all our maps suggest a similar fiber organization of the dense gas in all targets in our sample (see Socci et al. 2024, for a full characterization). More striking, this filamentary gas organization continues into the more diffuse material traced by HNC (1–0) at $\gtrsim$1 Jy beam^{-1} km s^{-1} ($\gtrsim$7 K km s^{-1}; Fig. 8e) sometimes running perpendicular to the above dense fibers. Yet, the complexity and emission properties of the gas change across our sample: Dense regions such as OMC-3 (Figs. 8e–d) and LDN 1641N (Figs. B.10e–d) show bright and highly contrasted emission in both N_2H^+ and HNC, while more diffuse clouds such as OMC-4 (Figs. B.9e–d) and NGC 2023 (Figs. B.11e–d) present systematically weaker and shallower emission in these tracers. In contrast, we find the HC_3N (10–9) emission $\gtrsim$0.15 Jy beam^{-1} km s^{-1} (>1 K km s^{-1}) to be typically more clumpy and concentrated toward local regions with dense gas (showing N_2H^+ emission) that are directly exposed to some stellar feedback (either by nebulae or YSOs) as seen in OMC-3 (Fig. 8f) or Flame Nebula (Fig. B.12f), although the lack of ALMA observations of this tracer (as well as HNC) in OMC-1 and OMC-2, both with bright emission in our SD maps (Figs. B.1 and B.2), prevents us from drawing further conclusions about the origin of its emission.

Figure 8 also highlights the ability of our molecular maps to reproduce the internal gas distribution at high spatial resolutions. OMC-3 appears to be a highly extincted MIR-dark region showing multiple narrow structures in the WISE 3.4 μm image in Fig. 8b (see Juvela & Mannfors 2023, for a detailed discussion on the MIR absorption in OMC-3). We find a direct spatial correlation (both in location and size) between these MIR absorption features and the emission distribution of our N_2H^+ maps (Fig. 8d). These similarities extend to the more diffuse gas where we identify several whisker-like features in the WISE image (see northern part of OMC-3) that are also detected in HNC (Fig. 8e). Although less clear, we also observe these similarities in the case of NGC 2023 (Fig. B.11). The nature of these diffuse gas features remains unclear at this point. Yet, this unique correspondence with the MIR absorption again shows the potential of our N_2H^+ and HNC (ALMA+IRAM-30 m) maps to describe the fine gas substructure within our cloud sample with high fidelity.

The inspection of our molecular ALMA+IRAM-30 m (MACF) maps reveals a new and complex cloud substructure at high spatial resolution. This is particularly remarkable in comparison with previous single-dish observations. Regions such as OMC-3 and Flame Nebula have been identified as single elongated (and broad) filaments using low spatial resolution, continuum (Schuller et al. 2021; Könyves et al. 2020) and line (Orkisz et al. 2019; Gaudel et al. 2023) observations. With the higher resolution of ALMA, however, the gas organization in all our fields is clearly more complex, nebulous, and wispy, and sometimes still unresolved by our beam size of ∼2000 au.

Both low- and high-mass star-forming regions exhibit a strong fibrous and gaseous substructure. Far from idealize cylinders, the size, width, radial distributions, and morphology of these slender fibers vary both locally and between regions. These fiber properties appear to vary smoothly in our sample, while show no discontinuity between low- and high-mass star-forming regions. Obtaining high-resolution ALMA observations with a high dynamic range is therefore not only one of the main goals of the EMERGE project, but also crucial for correctly interpreting the ISM structure that leads to the formation of all types of stars. We will characterize the physical properties of these gas substructure in following papers of this series (e.g., Socci et al. 2024) and compare these results with similar ongoing ALMA surveys (e.g., Sanhueza et al. 2019; Anderson et al. 2021; Barnes et al. 2021; Motte et al. 2022; Liu et al. 2024).

6. Conclusions

The EMERGE project describes the origin of high-mass stars and clusters as part of the complex interactions in dense filamentary (fiber-like) networks. This work (Paper I) introduces the EMERGE Early ALMA Survey and a series of novel methods for data reduction and exploration of ALMA observations. Accompanying papers will investigate the effects of filtering and data combination of similar ALMA observations (Bonanomi et al. 2024, Paper II) and will characterize the physical properties of the dense fibers found in these regions (Socci et al. 2024, Paper III). The main results of this paper are summarized below.

1. As part of our EMERGE Early ALMA Survey we systematically investigated seven star-forming regions that are part of the Orion A and B clouds, namely OMC-1, OMC-2, OMC-3, OMC-4 South, LDN 1641N, NGC 2023, and Flame Nebula

[9] We remark that, similar to our spectral observations (see Appendix A for a discussion), the interferometric filtering of large-scale emission can artificially reduce the continuum emission at small scales. The lack of short-spacing information therefore reduces the effective dynamic range and sensitivity of our int-CLEAN maps to recover the true dust continuum emission.

(Sect. 2). We homogeneously surveyed this sample combining large-scale single-dish (IRAM-30 m; 30") observations together with dedicated interferometric mosaics (ALMA-12m array, Cycles 3+7; 4.″5) maps in both (3 mm-)continuum and density-selective molecular lines (N_2H^+ (1–0), HNC (1–0), and HC_3N (10–9)) in Band 3. We complemented these data with additional archival FIR observations and IR surveys (Sect. 3).

2. We developed an optimized framework for the massive and automatic data reduction of our ALMA observations (Sect. 4). This includes the paralellization of the calibration, imaging, and combination process of single-dish and interferometric (ALMA+IRAM-30 m) observations carried out on high-performance computers.
3. The analysis of the large-scale properties of our targets demonstrates the wide variety of physical conditions sampled by our survey including low- (OMC-4 South and NGC 2023), intermediate- (OMC-2, OMC-3, and LDN 1641N) and high-mass (OMC-1 and Flame Nebula) star-forming regions that cover a wide range of surface density of star formation, gas column densities, fractions of dense gas, temperatures, and evolutionary stages (Sect. 5.1).
4. Our suite of selective molecular tracers sample distinct gas regimes in our sample. N_2H^+ (1–0) highlights the densest and coldest gas in our targets at column densities $\gtrsim 20 \times 10^{21}$ cm^{-2}. HNC (1–0) efficiently traces the cloud material at low and intermediate densities down to $\sim 5 \times 10^{21}$ cm^{-2}. On the other hand, HC_3N (10–9) is connected to lukewarm temperatures in regions exposed to feedback (Sect. 5.1.2). Of the three tracers, N_2H^+ **(1–0)** appears the best descriptor of the star-forming gas that leads to the formation of protostars at constant efficiency (Sect. 5.1.3) and within timescales of ∼1 Myr (Sect. 5.1.4).
5. When they are observed in the (low-resolution) *Herschel* column density maps (36″), our targets appear to be clumpy and filamentary. Similar structures are recognized in the N_2H^+ and HNC (1–0), IRAM-30 m maps (∼30" or ∼12 000 au) showing total intensities that positively correlate with the total gas column density and kinetic temperatures of these regions (Sect. 5.1.2).
6. In contrast, our (high-resolution) ALMA+IRAM-30 m observations (4.″5 or ∼2000 au) provide a new perspective of the gas structure within these star-forming regions (Sect. 5.2). At the enhanced interferometric resolution the gas organization in all our targets exhibits an increasing level of complexity down to the beam size of ALMA. Dozens of small-scale, dense fibers can be recognized in the dense gas traced by their N_2H^+ (1–0) emission similar to those previously identified in OMC-1 and OMC-2. Additional filamentary features are seen in more diffuse material that is traced by the HNC emission.
7. The gas organization observed in our high-resolution ALMA maps suggests the common presence of dense fibers in all targets in our sample of our sample, regardless of their gas content (mass of diffuse versus dense gas), stellar density (isolated versus clustered), star formation activity (low- versus high-mass), and evolutionary state (young versus evolved). The properties of these fibers will be fully characterized in other papers of this series (e.g., see Paper III). These results highlight the hierarchical nature of the ISM, which shows filaments down to a resolution of ∼2000 au.

Acknowledgements. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 851435). M.T. acknowledges partial support from project PID2019-108765GB-I00 funded by MCIN/AEI/10.13039/501100011033. D.H. is supported by Center for Informatics and Computation in Astronomy (CICA) grant and grant number 110J0353I9 from the Ministry of Education of Taiwan. D.H. acknowledges support from the National Technology and Science Council of Taiwan through grant number 111B3005191. J.R.G. thanks the Spanish MCIU for funding support under grant PID2019-106110GB-I00. This paper makes use of the following ALMA data: ADS/JAO.ALMA#2019.1.00641.S., ADS/JAO.ALMA#2013.1.00662.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. This work is based on IRAM-30 m telescope observations carried out under project numbers 032-13, 120-20, 060-22, and 133-22. IRAM is supported by INSU/CNRS (France), MPG (Germany), and IGN (Spain). This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. This research has made use of NASA's Astrophysics Data System. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. The authors acknowledge assistance from Allegro, the European ALMA Regional Center node in the Netherlands.

References

Ahmadi, A., & Hacar, A. 2023, Astrophysics Source Code Library [record ascl:2306.025]
Alves, J., & Bouy, H. 2012, A&A, 547, A97
Anderson, M., Peretto, N., Ragan, S. E., et al. 2021, MNRAS, 508, 2964
André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, in Protostars and Planets VI, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning (University of Arizona Press), 27
Arzoumanian, D., André, P., Peretto, N., & Könyves, V. 2013, A&A, 553, A119
Bally, J. 2008, in Handbook of Star Forming Regions, Volume I, ed. B. Reipurth, 4, 459
Bally, J., Langer, W. D., Stark, A. A., & Wilson, R. W. 1987, ApJ, 312, L45
Barnes, A. T., Henshaw, J. D., Fontani, F., et al. 2021, MNRAS, 503, 4601
Behrend, R., & Maeder, A. 2001, A&A, 373, 190
Bergin, E. A., & Tafalla, M. 2007, ARA&A, 45, 339
Bernasconi, P. A., & Maeder, A. 1996, A&A, 307, 829
Berné, O., Marcelino, N., & Cernicharo, J. 2010, Nature, 466, 947
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Bonanomi, F., Hacar, A., Socci, A., et al. 2024, A&A, in press, https://doi.org/10.1051/0004-6361/202348920
Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 2001, MNRAS, 323, 785
Brinkmann, N., Wyrowski, F., Kauffmann, J., et al. 2020, A&A, 636, A39
Brown, A. G. A., de Geus, E. J., & de Zeeuw, P. T. 1994, A&A, 289, 101
Cao, Y., Qiu, K., Zhang, Q., & Li, G.-X. 2022, ApJ, 927, 106
CASA Team (Bean, B., et al.) 2022, PASP, 134, 114501
Caselli, P., Benson, P. J., Myers, P. C., & Tafalla, M. 2002, ApJ, 572, 238
Chahine, L., López-Sepulcre, A., Podio, L., et al. 2022, A&A, 667, A6
Chen, H.-R. V., Zhang, Q., Wright, M. C. H., et al. 2019, ApJ, 875, 24
Chown, R., Sidhu, A., Peeters, E., et al. 2024, A&A, 685, A75
Clark, B. G. 1980, A&A, 89, 377
Clarke, S. D., Whitworth, A. P., Duarte-Cabral, A., & Hubber, D. A. 2017, MNRAS, 468, 2489
Clarke, S. D., Whitworth, A. P., Spowage, R. L., et al. 2018, MNRAS, 479, 1722
Cotton, W. D. 2017, PASP, 129, 094501
Dale, J. E., Ngoumou, J., Ercolano, B., & Bonnell, I. A. 2014, MNRAS, 442, 694
Emonts, B., Raba, R., Rau, U., et al. 2022, in ASP Conf. Ser., 532, eds. J. E. Ruiz, F. Pierfedereci, & P. Teuben, 389
Evans, Neal J., I., Dunham, M. M., Jørgensen, J. K., et al. 2009, ApJS, 181, 321
Fehér, O., Tóth, L. V., Ward-Thompson, D., et al. 2016, A&A, 590, A75
Fernández-López, M., Arce, H. G., Looney, L., et al. 2014, ApJ, 790, L19
Furlan, E., Fischer, W. J., Ali, B., et al. 2016, ApJS, 224, 5
Gaudel, M., Orkisz, J. H., Gerin, M., et al. 2023, A&A, 670, A59
Gildas Team 2013, Astrophysics Source Code Library [record ascl:1305.010]
Goicoechea, J. R., Pabst, C. H. M., Kabanovic, S., et al. 2020, A&A, 639, A1
Greene, T. P., Wilking, B. A., Andre, P., Young, E. T., & Lada, C. J. 1994, ApJ, 434, 614
Großschedl, J. E., Alves, J., Meingast, S., et al. 2018, A&A, 619, A106
Großschedl, J. E., Alves, J., Teixeira, P. S., et al. 2019, A&A, 622, A149
Hacar, A., & Tafalla, M. 2011, A&A, 533, A34
Hacar, A., Tafalla, M., Kauffmann, J., & Kovács, A. 2013, A&A, 554, A55

Hacar, A., Alves, J., Burkert, A., & Goldsmith, P. 2016, A&A, 591, A104
Hacar, A., Alves, J., Tafalla, M., & Goicoechea, J. R. 2017a, A&A, 602, L2
Hacar, A., Tafalla, M., & Alves, J. 2017b, A&A, 606, A123
Hacar, A., Tafalla, M., Forbrich, J., et al. 2018, A&A, 610, A77
Hacar, A., Bosman, A. D., & van Dishoeck, E. F. 2020, A&A, 635, A4
Hacar, A., Clark, S. E., Heitsch, F., et al. 2023, ASP Conf. Ser., 534, 153
Hartmann, L., Ballesteros-Paredes, J., & Bergin, E. A. 2001, ApJ, 562, 852
Heiderman, A., Evans, Neal J., I., Allen, L. E., Huard, T., & Heyer, M. 2010, ApJ, 723, 1019
Henshaw, J. D., Caselli, P., Fontani, F., Jiménez-Serra, I., & Tan, J. C. 2014, MNRAS, 440, 2860
Hoemann, E., Heigl, S., & Burkert, A. 2021, MNRAS, 507, 3486
Högbom, J. A. 1974, A&AS, 15, 417
Ibáñez-Mejía, J. C., Mac Low, M.-M., & Klessen, R. S. 2022, ApJ, 925, 196
Inutsuka, S.-i., & Miyama, S. M. 1997, ApJ, 480, 681
Johnstone, D., & Bally, J. 1999, ApJ, 510, L49
Juvela, M., & Mannfors, E. 2023, A&A, 671, A111
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
Kauffmann, J., Goldsmith, P. F., Melnick, G., et al. 2017, A&A, 605, L5
Kirk, H., Johnstone, D., Di Francesco, J., et al. 2016, ApJ, 821, 98
Kong, S., Arce, H. G., Feddersen, J. R., et al. 2018, ApJS, 236, 25
Könyves, V., André, P., Arzoumanian, D., et al. 2020, A&A, 635, A34
Kounkel, M., Hartmann, L., Loinard, L., et al. 2017, ApJ, 834, 142
Krumholz, M. R., & McKee, C. F. 2008, Nature, 451, 1082
Krumholz, M. R., & Tan, J. C. 2007, ApJ, 654, 304
Kumar, M. S. N., Palmeirim, P., Arzoumanian, D., & Inutsuka, S. I. 2020, A&A, 642, A87
Kumar, M. S. N., Arzoumanian, D., Men'shchikov, A., et al. 2022, A&A, 658, A114
Lada, E. A. 1992, ApJ, 393, L25
Lada, E. A., Bally, J., & Stark, A. A. 1991, ApJ, 368, 432
Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687
Leroy, A. K., Hughes, A., Liu, D., et al. 2021, ApJS, 255, 19
Li, P. S., & Klein, R. I. 2019, MNRAS, 485, 4509
Liu, X., Liu, T., Zhu, L., et al. 2024, Res. Astron. Astrophys., 24, 025009
Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, A&A, 566, A45
Mairs, S., Johnstone, D., Kirk, H., et al. 2016, MNRAS, 461, 4022
Mason, B., Dicker, S., Sadavoy, S., et al. 2020, ApJ, 893, 13
McKee, C. F., & Tan, J. C. 2003, ApJ, 585, 850
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
Megeath, S. T., Gutermuth, R. A., & Kounkel, M. A. 2022, PASP, 134, 042001
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Moeckel, N., & Burkert, A. 2015, ApJ, 807, 67
Molinari, S., Swinyard, B., Bally, J., et al. 2010, A&A, 518, A100
Monsch, K., Pineda, J. E., Liu, H. B., et al. 2018, ApJ, 861, 77
Motte, F., Bontemps, S., & Louvet, F. 2018, ARA&A, 56, 41
Motte, F., Bontemps, S., Csengeri, T., et al. 2022, A&A, 662, A8
Myers, P. C. 2009, ApJ, 700, 1609
Nishimura, A., Tokuda, K., Kimura, K., et al. 2015, ApJS, 216, 18
O'Dell, C. R. 2001, ARA&A, 39, 99
Orkisz, J. H., Peretto, N., Pety, J., et al. 2019, A&A, 624, A113
Pabst, C. H. M., Goicoechea, J. R., Teyssier, D., et al. 2017, A&A, 606, A29
Pabst, C., Higgins, R., Goicoechea, J. R., et al. 2019, Nature, 565, 618
Pabst, C. H. M., Goicoechea, J. R., Teyssier, D., et al. 2020, A&A, 639, A2
Padoan, P., Pan, L., Haugbølle, T., & Nordlund, Å. 2016, ApJ, 822, 11
Palmeirim, P., André, P., Kirk, J., et al. 2013, A&A, 550, A38
Peretto, N., Fuller, G. A., Duarte-Cabral, A., et al. 2013, A&A, 555, A112
Peterson, D. E., & Megeath, S. T. 2008, in Handbook of Star Forming Regions, Volume I, ed. B. Reipurth, 4, 590
Pety, J. 2005, in SF2A-2005: Semaine de l'Astrophysique Française, eds. F. Casoli, T. Contini, J. M. Hameury, & L. Pagani, 721
Pety, J., Schinnerer, E., Leroy, A. K., et al. 2013, ApJ, 779, 43
Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98
Pineda, J. E., Arzoumanian, D., Andre, P., et al. 2023, in ASP Conf. Ser., 534, Protostars and Planets VII, eds. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 233
Plunkett, A., Hacar, A., Moser-Fischer, L., et al. 2023, PASP, 135, 034501
Pokhrel, R., Gutermuth, R. A., Krumholz, M. R., et al. 2021, ApJ, 912, L19
Priestley, F. D., Clark, P. C., Glover, S. C. O., et al. 2023, MNRAS, 526, 4952
Remijan, A., Biggs, A., Cortes, P. A., et al. 2019, ALMA Technical Handbook, ALMA Doc. 7.3, ver. 1.1
Rieke, G. H., & Lebofsky, M. J. 1985, ApJ, 288, 618
Rodriguez-Franco, A., Martin-Pintado, J., Gomez-Gonzalez, J., & Planesas, P. 1992, A&A, 264, 592
Sanhueza, P., Contreras, Y., Wu, B., et al. 2019, ApJ, 886, 102
Santa-Maria, M. G., Goicoechea, J. R., Pety, J., et al. 2023, A&A, 679, A4
Sato, A., Takahashi, S., Ishii, S., et al. 2023, ApJ, 944, 92
Schneider, S., & Elmegreen, B. G. 1979, ApJS, 41, 87
Schöier, F. L., van der Tak, F. F. S., van Dishoeck, E. F., & Black, J. H. 2005, A&A, 432, 369
Schuller, F., André, P., Shimajiri, Y., et al. 2021, A&A, 651, A36
Seifried, D., Walch, S., Weis, M., et al. 2020, MNRAS, 497, 4196
Shimajiri, Y., Kitamura, Y., Saito, M., et al. 2014, A&A, 564, A68
Shimajiri, Y., André, P., Palmeirim, P., et al. 2019, A&A, 623, A16
Shimajiri, Y., André, P., Peretto, N., et al. 2023, A&A, 672, A133
Shirley, Y. L. 2015, PASP, 127, 299
Shu, F. H., Adams, F. C., & Lizano, S. 1987, ARA&A, 25, 23
Smith, R. J., Glover, S. C. O., Clark, P. C., Klessen, R. S., & Springel, V. 2014, MNRAS, 441, 1628
Smith, R. J., Treß, R. G., Sormani, M. C., et al. 2020, MNRAS, 492, 1594
Socci, A., Hacar, A., Bonanomi, F., et al. 2024, A&A, submitted
Sokolov, V., Wang, K., Pineda, J. E., et al. 2019, ApJ, 872, 30
Stanke, T., & Williams, J. P. 2007, AJ, 133, 1307
Stanke, T., Arce, H. G., Bally, J., et al. 2022, A&A, 658, A178
Stutz, A. M. 2018, MNRAS, 473, 4890
Stutz, A. M., & Kainulainen, J. 2015, A&A, 577, A6
Stutz, A. M., Tobin, J. J., Stanke, T., et al. 2013, ApJ, 767, 36
Suri, S., Sánchez-Monge, Á., Schilke, P., et al. 2019, A&A, 623, A142
Tafalla, M., & Hacar, A. 2015, A&A, 574, A104
Tafalla, M., Myers, P. C., Caselli, P., Walmsley, C. M., & Comito, C. 2002, ApJ, 569, 815
Tafalla, M., Usero, A., & Hacar, A. 2021, A&A, 646, A97
Tafalla, M., Usero, A., & Hacar, A. 2023, A&A, 679, A112
Takahashi, S., Ho, P. T. P., Teixeira, P. S., Zapata, L. A., & Su, Y.-N. 2013, ApJ, 763, 57
Tatematsu, K., Kandori, R., Umemoto, T., & Sekimoto, Y. 2008, PASJ, 60, 407
Treviño-Morales, S. P., Fuente, A., Sánchez-Monge, Á., et al. 2019, A&A, 629, A81
van der Tak, F. F. S., Lique, F., Faure, A., Black, J. H., & van Dishoeck, E. F. 2020, Atoms, 8, 15
van Terwisga, S. E., Hacar, A., & van Dishoeck, E. F. 2019, A&A, 628, A85
van Terwisga, S. E., Hacar, A., van Dishoeck, E. F., Oonk, R., & Portegies Zwart, S. 2022, A&A, 661, A53
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Wilner, D. J., & Welch, W. J. 1994, ApJ, 427, 898
Wiseman, J. J., & Ho, P. T. P. 1998, ApJ, 502, 676
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Zamora-Avilés, M., Ballesteros-Paredes, J., & Hartmann, L. W. 2017, MNRAS, 472, 647
Zucker, C., Smith, R., & Goodman, A. 2019, ApJ, 887, 186

Appendix A: Assessing data quality in interferometric observations

A.1. Effects of the short-spacing information

As recognized in multiple studies in the literature (see Leroy et al. 2021; Plunkett et al. 2023, and references therein), the lack of short-spacing information fundamentally impacts the recovery of extended emission in interferometric observations. We illustrate the severity of these issues in real observations in Fig. A.1 by comparing different high spectral resolution N_2H^+ (1–0), HNC (1–0), and HC_3N (10–9) maps in OMC-3 obtained using the int-CLEAN, Feather, and MACF data combination methods (see descriptions in Sect. 4.1). Similar comparisons for all the other targets observed in these tracers can be found in Figs. B.13–B.16. An inspection of these integrated-intensity maps by eye reveals thatinterferometric-alone (int-CLEAN; left panels) reductions miss most of the line emission of all our tracers. These filtering effects are particularly prominent for molecules with extended and bright emission, such as N_2H^+ and HNC (top and bottom panels), as is easily recognized by the prominent negative sidelobes surrounding large emission features. Less obviously, filtering also affects more compact emission features such as those seen in HC_3N (bottom panels) down to the ALMA resolution. Similar issues are seen in all targets of our sample.

The addition of the short-spacing information significantly changes the amount of recovered emission in all our spectral maps. Both Feather (central planels) and MACF (right panels) reductions reveal significant levels of extended emission in bright tracers such as N_2H^+ (top panels). Yet, data combination improves the flux recovery of both compact and extended emission features seen. The most dramatic improvements are seen in our HNC maps (middle panels), however, which is the most diffuse tracer in our sample (see Sect. 5.1.2). There, Feather and MACF recover a large amount of extended emission that is systematically filtered our by the interferometer. Little or no sign of remaining negative sidelobes is observed in these combined maps.

The above improvement in the recovery of extended emission after data combination can be understood from the uv-coverage in our ALMA+IRAM-30 m observations. We show in Fig. A.2 the uv-sampling of our ALMA-12m data (blue) in OMC-3 within the inner 50 meters of the uv-plane. The large number of baselines in the ALMA-12m array produces a densely sampled coverage of uv-plane above (projected) distances of $\geq$ 12 meters. On the other hand, the inner gap in the uv-plane is uniformly sampled by our IRAM-30 m data (gray area) up to 30 meters and with a sufficient overlap with the ALMA-12m array in C43-1 configuration (hatched area). While ALMA provides unprecedented sensitivity at long baselines (~12-300 meters in our case), the inclusion of the IRAM-30 m data as short-spacing information becomes essential to recover the extended emission in these targets. As demonstrated in Paper II (Bonanomi et al. 2024), the homogeneous uv-sampling provided by a large SD such as IRAM-30 m significantly improves on the image quality during the study of the filamentary ISM compared to the more sparse coverage obtained using equivalent ALMA Atacama Compact Array (ACA; sampling intermediate uv-distances) plus total power (TP; sampling the innermost baselines) observations.

A.2. Statistical results

Beyond the visual comparisons in maps and spectra (see above), we performed a quantitative analysis of the QA of our molecular datasets implementing the assessment metrics presented in Plunkett et al. (2023). In particular, we focused our assessments on the analysis of the accuracy parameter (A-par):

$$A_\nu(x,y) = \frac{I_\nu(x,y) - R_\nu(x,y)}{|R_\nu(x,y)|}. \quad \text{(A.1)}$$

Intuitively, A-par measures the relative error (in percentage) between a reference ($R_\nu(x,y)$) and a target ($I_\nu(x,y)$) image while its sign indicates whether the flux is over- ($A_\nu(x,y) > 0$) or underestimated ($A_\nu(x,y) < 0$). A-par can be evaluated in space (x,y) and velocity (ν) and applied to maps and cubes alike within a dedicated assessment region (see Plunkett et al for a full discussion).

In the case of our molecular observations, we adopted our IRAM-30 m datasets as the reference and evaluated the performance of the target int-CLEAN, Feather, and MACF reductions. To do this, we convolved and regridded each combined map (θ =4.″5 resolution) into the single-dish frame (θ =30") and applied an assessment threshold of approximately five times the noise level in our ALMA maps. From the statistical analysis of $A_\nu(x,y)$, we then quantitatively measure the amount of flux recovered by our data combinations in comparison with those found in the single-dish data, which are assumed to represent the true sky flux distribution at least at the single-dish resolution. Reduced with the same deconvolution parameters, our QA allowed us to systematically compare the goodness of each data reduction method applied in our study (Sect. 4.1).

We summarize the results of our QA in Table A.1 in three representative regions in our sample selected by presenting different line intensities and source structure: OMC-3 (bright and extended), Flame Nebula (compact and bright), and NGC 2023 (weak and extended). We display the values of A-par statistics (mean and standard deviation) obtained from the analysis of all voxels within the assessment area as well as the fraction of total flux recovered (FR) in the final combined data cubes of our three lines of interest (N_2H^+, HNC, and HC_3N). We display the results for both high (top three rows; narrow) and low (bottom two rows; broad) spectral resolution.

Our QAs confirm the trends suggested during the visual inspection of our maps (Fig. A.1). All interferometric-alone reductions (int-CLEAN) are subject to widespread filtering missing >70% of the flux both per pixel (A<-0.70) and in absolute (FR$\lesssim$40%) terms. Conversely, data combination techniques (Feather and MACF) exhibit a significant improved performance by combining an enhanced local accuracy (|A|$\lesssim$0.10) and flux recovery (FR$\gtrsim$85%) in most targets and lines. Data combination is particularly relevant in regions with bright extended emission (e.g., OMC-3 and Flame Nebula). Nonetheless, data combination shows some limitations in targets with either sharp features (e.g. HC_3N in OMC-3) or weak extended emission (e.g., HC_3N in NGC 2023), probably connected to the quality of our single-dish data. Regardless of these minor issues, our QAs statistically quantify the great impact of advanced data combination techniques during the reduction of ALMA observations (see also Plunkett et al. 2023).

The differences between Feather and MACF are more subtle. The results in Table A.1 indicate some systematic differences between the Feather and MACF results. Overall, MACF produces the most stable results per pixel with A-par values close to |A|$\lesssim$0.05 and a better total flux recovery (FR> 92%). In some

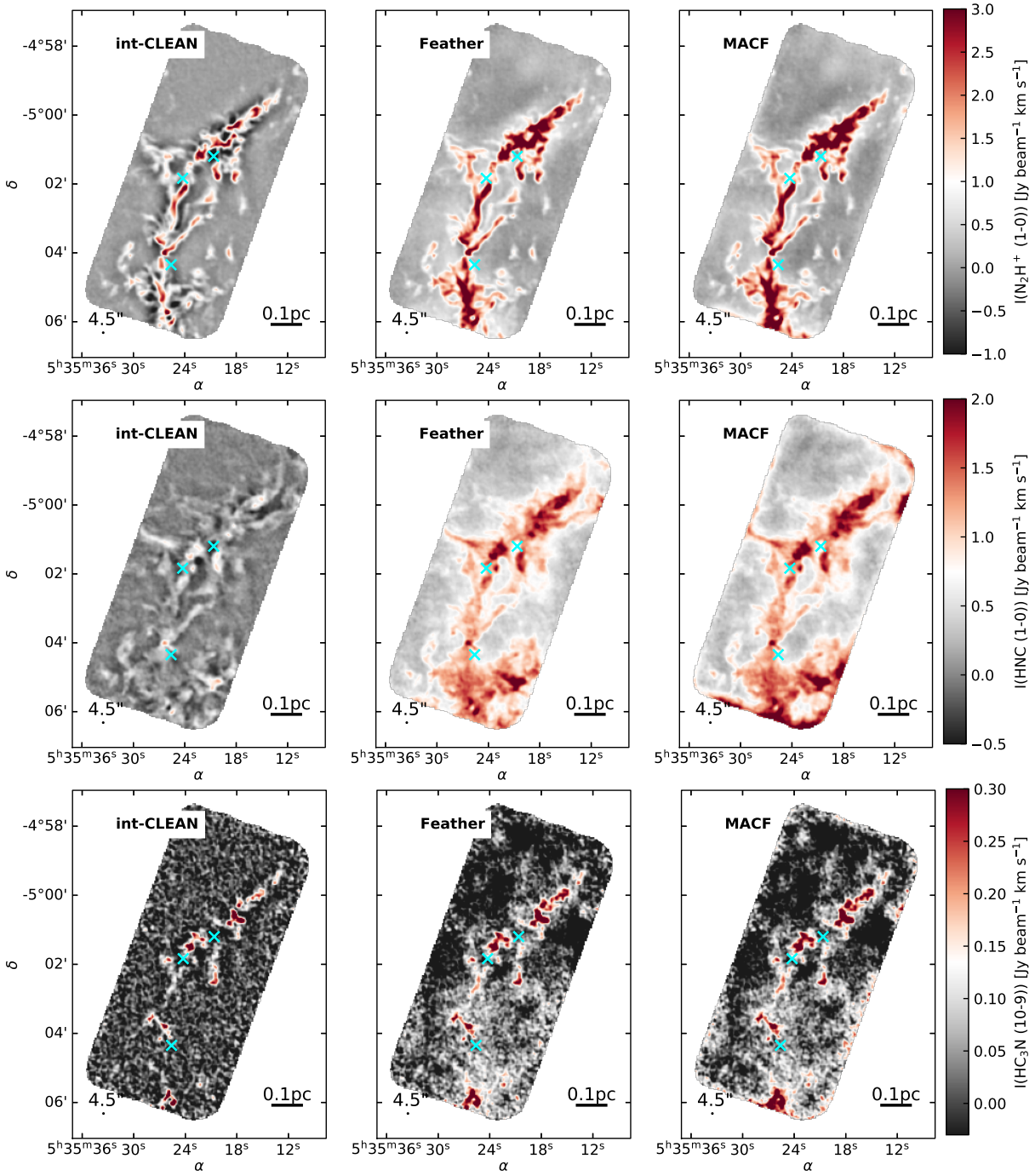

Fig. A.1. Effects of the different data combination methods in OMC-3: **(Top)** N_2H^+ (1–0), **(middle)** HNC (1–0), and **(bottom)** HC_3N (10–9) integrated intensity maps. The different columns show the results for the **(left)** interferometric-alone (int-CLEAN), **(centre)** feather (Feather), and **(right)** Model Assisted CLEAN plus Feather (MACF) data combination methods. To facilitate their comparison all integrated-intensity maps of the same molecule are displayed with the same intensity scale (see color bars). The location of the spectra shown in Fig. A.3 is indicated by cyan crosses in all maps. Compared to our standard PB cut above 0.5 (Sect. 4.1), these maps are shown down to PB values of 0.2 in order to identify potential issues at their edges (e.g., HNC MACF map). We remark here that these edges issues do not affect our scientific results as they are removed when our stricter PB cut is applied. Similar features can be identified in all the ALMA fields observed in these three tracers shown in Figs. B.13-B.16.

cases, MACF appears to produce noisier edges, although this effect is easily mitigated by a sensible PB cut after combination (Sect. 4.1). Based on these results, and for consistency with the analysis in OMC-1 and OMC-2 (Hacar et al. 2018), we adopted the MACF method as our standard reduction. Still, we remark that in specific cases, Feather may produce similar or even slightly better results (by a few percent).

While both Feather and MACF produce satisfactory results, we find systematic improvements in their data products when we compare high- (narrow) and low- (broad) spectral resolution

Table A.1. Quality assessment results

Transition	Method	OMC-3		Flame Nebula		NGC 2023	
		A-par(⋆)	FR(⋆⋆)	A-par	FR	A-par	FR
N_2H^+ (1–0) (narrow)	int-CLEAN	-0.79±0.04	18.8%	-0.72±0.06	35.3%	-0.90±0.02	4.7%
	Feather	-0.02±0.05	90.6%	-0.13±0.05	84.9%	-0.07±0.01	91.4%
	MACF	-0.00±0.03	99.5%	-0.08±0.03	91.7%	-0.02±0.01	97.0%
HNC (1–0) (narrow)	int-CLEAN	-0.94±0.01	6.3%	-0.94±0.01	8.6%	-0.97±0.02	4.5%
	Feather	-0.01±0.03	94.6%	-0.03±0.03	95.7%	-0.05±0.04	93.6%
	MACF	-0.03±0.01	99.3%	0.00±0.02	99.1%	-0.01±0.03	98.1%
HC_3N (10–9) (narrow)	int-CLEAN	-0.69±0.03	32.9%	-0.59±0.01	43.7%	-0.72±0.04	32.7%
	Feather	-0.16±0.01	83.5%	-0.04±0.03	91.6%	-0.16±0.03	80.6%
	MACF	-0.07±0.01	92.9%	0.00±0.02	98.1%	-0.09±0.02	87.1%
HNC (1–0) (broad)	int-CLEAN	-0.94±0.01	6.3%	-0.94±0.01	8.6%	-0.97±0.01	4.4%
	Feather	-0.02±0.07	93.0%	-0.02±0.03	94.3%	-0.03±0.02	93.6%
	MACF	-0.07±0.06	84.9%	0.02±0.02	89.9%	0.01±0.02	88.5%
HC_3N (10–9) (broad)	int-CLEAN	-0.79±0.03	24.0%	-0.59±0.05	43.6%	-0.67±0.04	38.1%
	Feather	-0.27±0.04	72.2%	-0.09±0.03	86.4%	-0.18±0.02	81.0%
	MACF	-0.20±0.02	75.6%	-0.04±0.04	92.4%	-0.11±0.02	90.5%

Notes. (⋆) Mean ± standard deviation A-par values. (⋆⋆) Percentage of the total flux recovered (FR) respect to the one detected in our single-dish data. We note that most of the best results are usually produced by the MACF method.

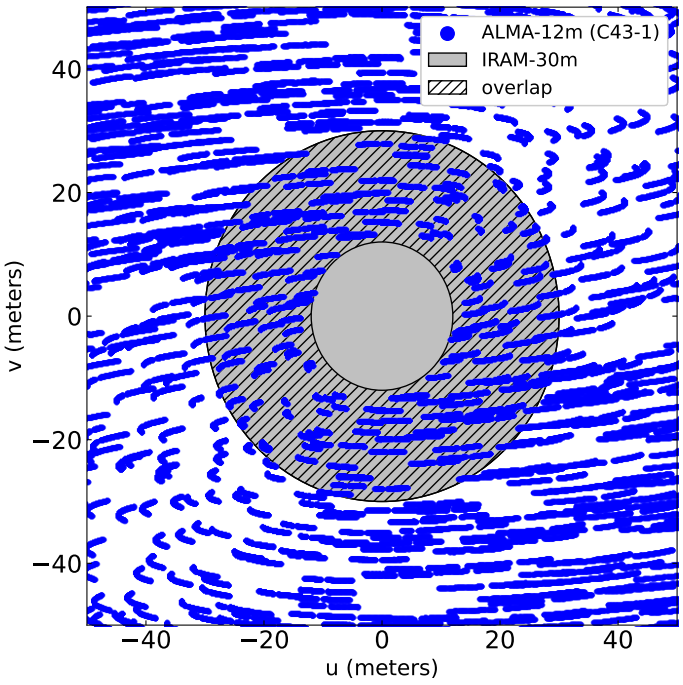

Fig. A.2. ALMA-12m array uv-coverage (blue), with effective baselines between ∼12 and 300 meters (C43-1), obtained in our OMC-3 observations. We highlight the uv-area sampled by the IRAM-30 m data (≤ 30 m; gray) as well as its overlap with the ALMA baselines (hatched area). Note that we display only those baselines at u,v-distances ≤ 50 meters.

reductions (see the results of HNC and HC_3N in Table A.1). Reductions including single-dish data at high spectral resolutions consistently increase the accuracy (A-par) and flux recovery (FR) of the combined datasets. Resolving the line structure (and its possible filtering in different channels) seems to help in recovering high-fidelity images. Further analyses are needed to confirm this trend.

A.3. Velocity-dependent filtering effects in molecular line observations

Plunkett et al. (2023) discussed that the lack of short-spacing information can have a critical impact on the recovered emission in space and velocity (i.e., spectral cubes). We illustrate these effects in Fig. A.3 by comparing the int-CLEAN (blue), Feather (dashed black), and MACF (red) spectra in three representative positions in our OMC-3 region. As seen in the hyperfine structure of our N_2H^+ spectra (top panels), filtering can act selectively in velocity, changing the emission profile of all our lines, regardless of their brightness.

Filtering can critically alter all line properties (e.g., line flux and peak, centroid and FWHM, and hyperfine distribution) and distort any physical property derived from them. This is clear in many int-CLEAN spectra exhibiting physically unrealistic hyperfine ratios, artificially narrow FWHM, and/or negative artifacts. To explore the effects of filtering on the line parameters, we automatically fit one Gaussian component to all HNC (1–0) spectra in OMC-3 obtained by our different data combination methods. In Fig. A.4 we show a pixel-by-pixel comparison of the peak temperature (T_{peak}; left panel), full width at half maximum (FWHM; central panel), and line velocity (V_{LSR}; right panel) between the int-CLEAN and Feather results with respect to MACF, the latter assumed as reference. As expected from the results in Sect. A.2, we find an excellent agreement between the line properties derived in the Feather and MACF methods, with most variations within the spectral resolution (e.g., see the FWHM). The results of the int-CLEAN reduction are dramatic, however. Interferometric filtering not only produces significant changes in the line peak (similar to the flux), but also completely corrupts the recovered FWHM and V_{LSR} parameters. As shown from the comparison in velocity (right panel), pure filtering artifacts can alter the entire velocity structure of the cloud and produce artificially narrow profiles and systematic drifts in the field.

Selective filtering effects also depend on the tracer. The different brightness, velocity distribution, and emission structure of the lines and continuum produce differential effects in the recovered emission at the same position. This is shown in Fig. A.3, where we compare N_2H^+ (top panels) and HNC (middle panels)

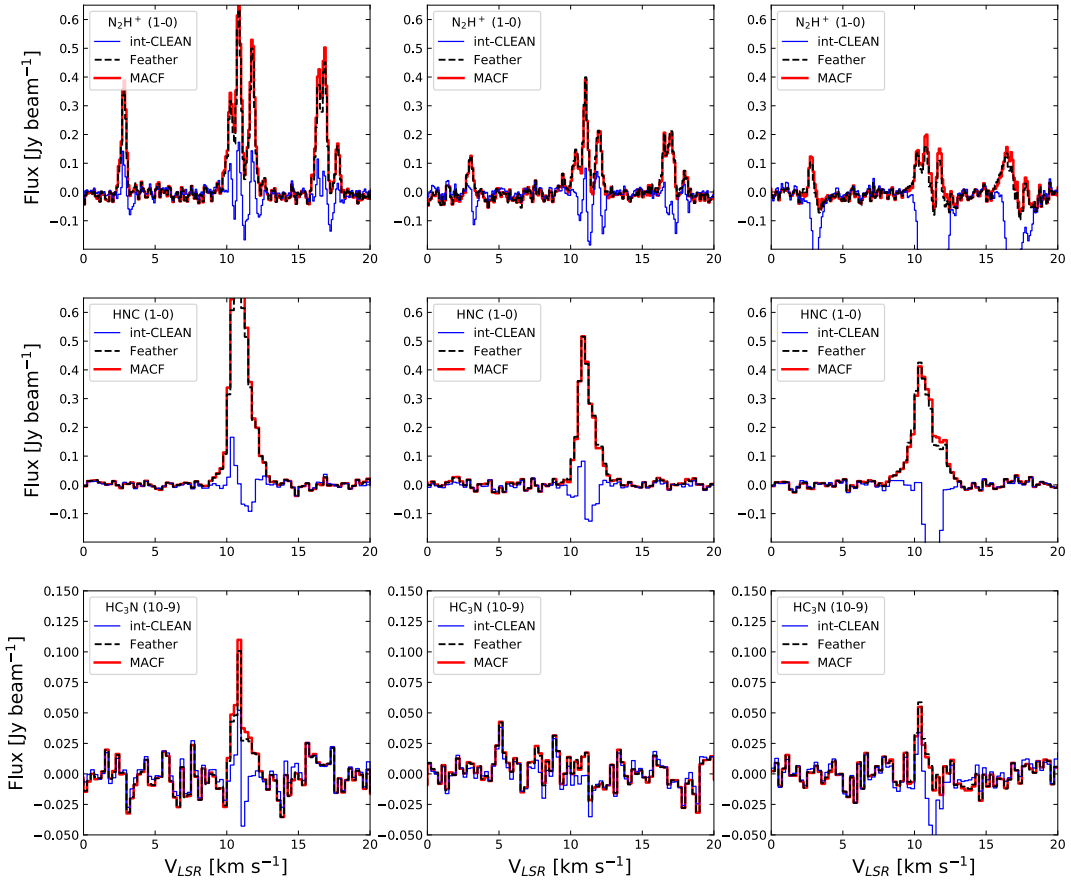

Fig. A.3. ALMA spectra obtained using int-CLEAN (black), Feather (dashed blue), and MACF (red) data combination methods. We display the (narrow) N_2H^+ (1–0) **(Top)**, HNC (1–0) **(middle)**, and HC_3N (10–9) **(bottom)** spectra (rows) in three representative positions (columns) within the OMC-3 region (see Fig. A.1): **(Left panels)** (α, δ)=(05:35:19.6,-05:00:29.5), **(central panels)** (α, δ)=(05:35:25.5,-05:04:19.0), and **(right panels)** (α, δ)=(05:35:20.7,-05:01:11.9). Note how the filtering effects experienced in the int-CLEAN maps are unpredictable and change depending on position, velocity, and tracer consider in each case.

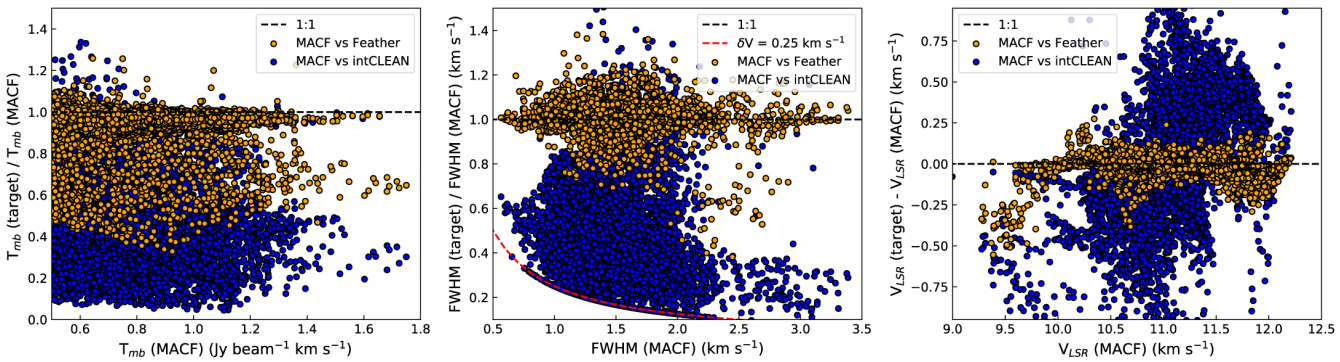

Fig. A.4. HNC (1–0) line properties derived from the automatic fit of one Gaussian component in OMC-3 obtained from our int-CLEAN (blue points) and Feather (yellow points) reductions (targets) in comparison with MACF. For simplicity, only fits with I(HNC, MACF)≥0.5 Jy beam^{-1} and S/N≥5 are shown. From left to right: **(Left panel)** Line peak ratio; **(central panel)** full width at half maximum ratio; **(right panel)** Velocity difference. The lower envelope of the central panel is determined by the spectral resolution of our (narrow) reductions with δV=0.25 km s^{-1}.

spectra at similar positions. The bright and extended HNC distribution throughout the entire OMC-3 cloud completely suppresses the emission of this molecule in different parts of the cloud more efficiently than in the case of N_2H^+ (e.g., left middle panel). Other regions cannot compensate for the negative sidelobes coming from nearby bright spots and produce more prominent negative features (e.g., right middle panel). We remark here that these effects are also seen in the weaker HC_3N spectra, even though this molecule shows a more compact distribution (see bottom panels).

We also note that the large-scale filtering produced by sources with an extended and complex emission structure can be highly deceptive. Negative sidelobes can be compensated for by bright emission features at similar velocities and produce almost

flat spectra (e.g. Fig. A.3, left middle panel). An analysis of these spectra could lead to measurements with artificially low noise values that give the incorrect impression of high-quality data reductions. Conversely, Feather and MACF reductions show bright emission features in many of these positions. These effects are apparent in the case of our continuum int-CLEAN maps. Most of the continuum emission is clearly filtered out, while our maps show a surprisingly low noise level.

The above maps (Sect. A.1) and spectra (this section) illustrate the perils of using interferometric-alone reductions for the analysis of spatially resolved sources. Our results discourage the use of interferometer-alone reductions (int-CLEAN) for characterizing the emission properties of (resolved) physical structures larger than few times the interferometric beamsize. Projects exploring similar targets are strongly encouraged to carefully evaluate the impact of the missing short-spacing information in both maps and spectra. In an accompanying paper (Bonanomi et al. 2024), we quantify how the lack of the single-dish information severely corrupts the characterization of the physical properties (masses, sizes, widths, etc.) of cores and filaments in our ALMA observations. Data combination becomes therefore essential for our line emission and kinematic studies in our ALMA sample.

Appendix B: Data products

In this appendix, we present all the data products of our EMERGE Early ALMA survey in OMC-1, OMC-2, OMC-4 South, LDN 1641N, NGC 2023, and Flame Nebula (see also Figs. 4 and 8 for those on OMC-3). We include all the integrated-intensity and continuum maps of our IRAM-30 m and ALMA observations together with the additional *Herschel* and WISE data. We display the individual molecular maps of each of these regions using low spatial resolution (single-dish) observations in Figs. B.1-B.6 (see Sect. 5.1). Similar maps for the high spatial resolution (ALMA+IRAM-30 m) datasets can be found in Figs. B.7-B.12 (see Sect. 5.2). Complementing Fig. A.1, we display the comparisons between the different data combination methods for all our ALMA fields observed in N_2H^+ (top panels), HNC (middle panels), and HC_3N (lower panels) in Figs. B.13-B.16. To facilitate the comparison between targets, we display all regions within the same plotting ranges in each of the panels.

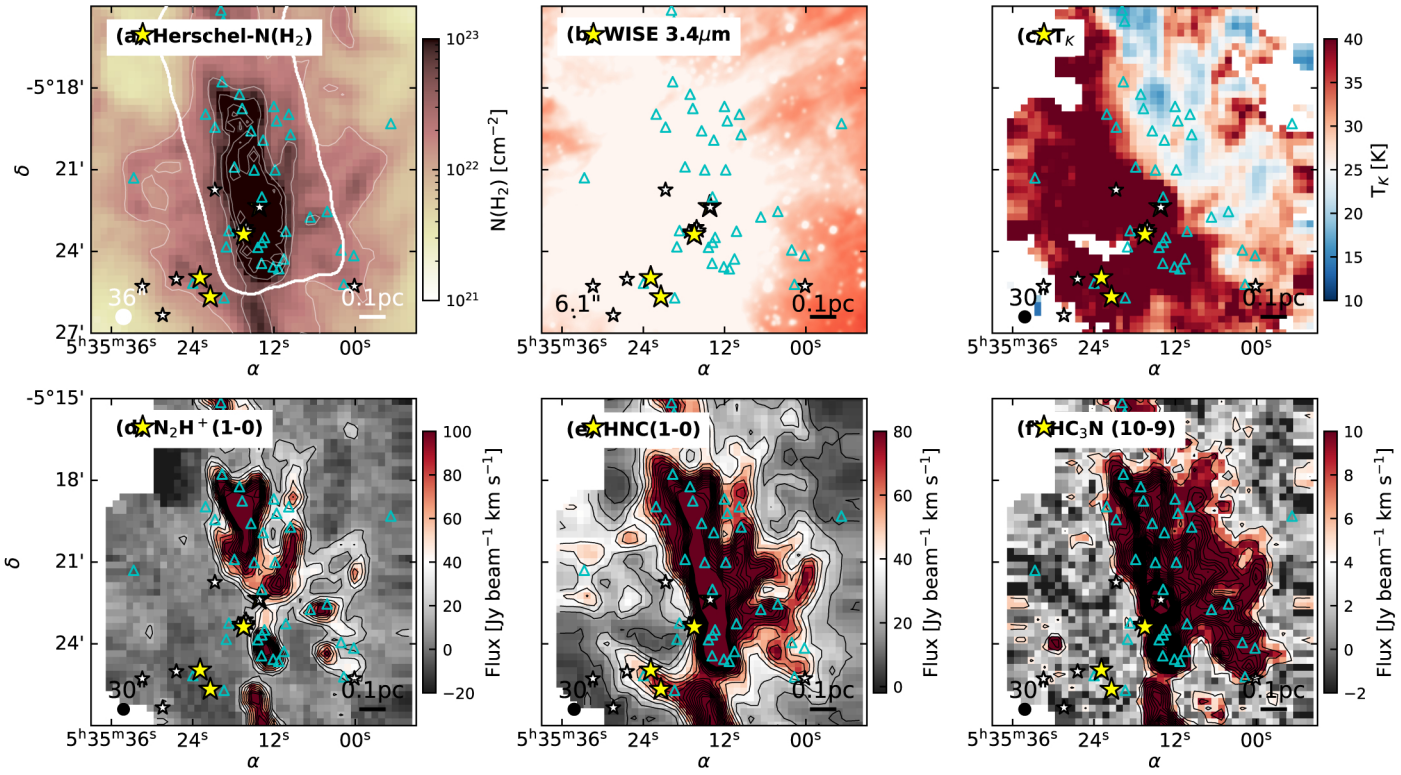

Fig. B.1. Low-resolution observations in OMC-1 similar to Fig. 4

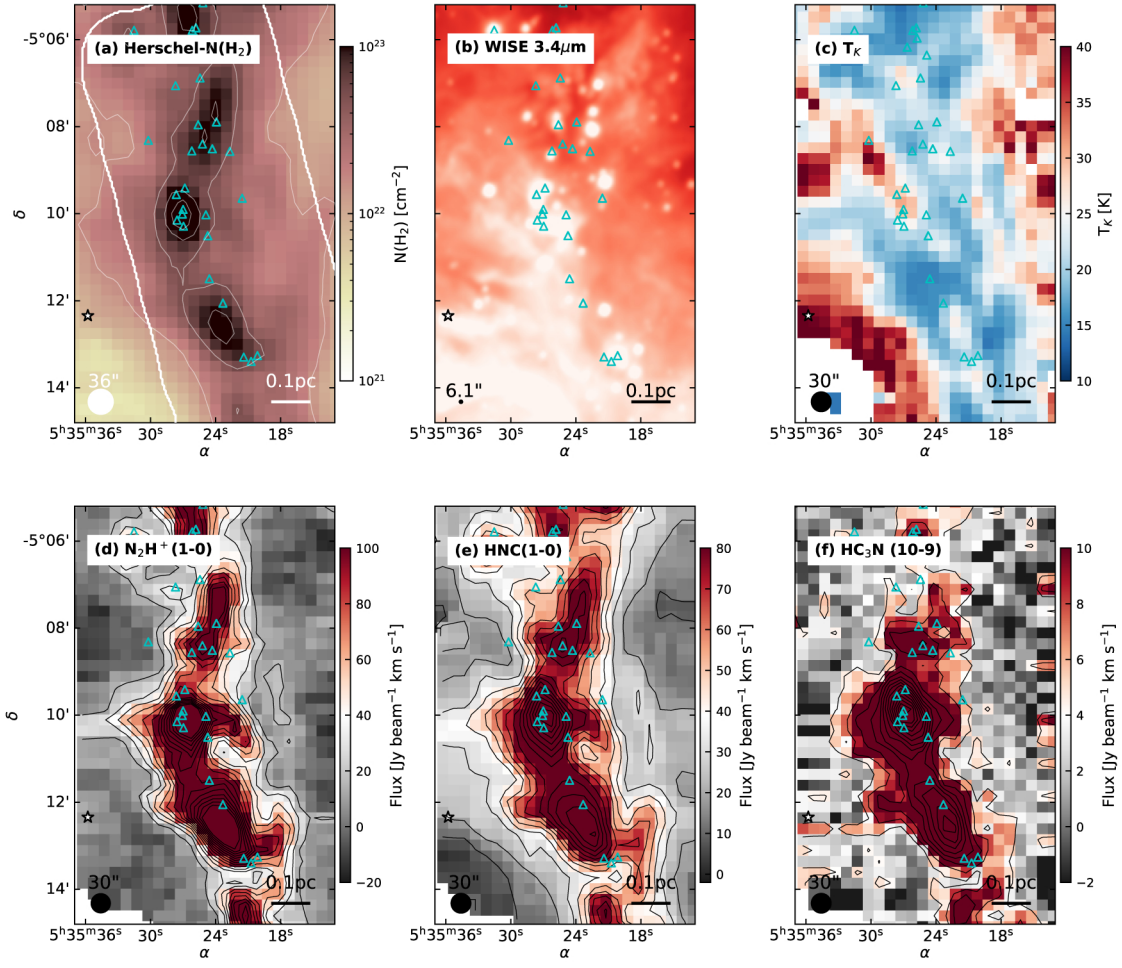

Fig. B.2. Low-resolution observations in OMC-2 similar to Fig. 4

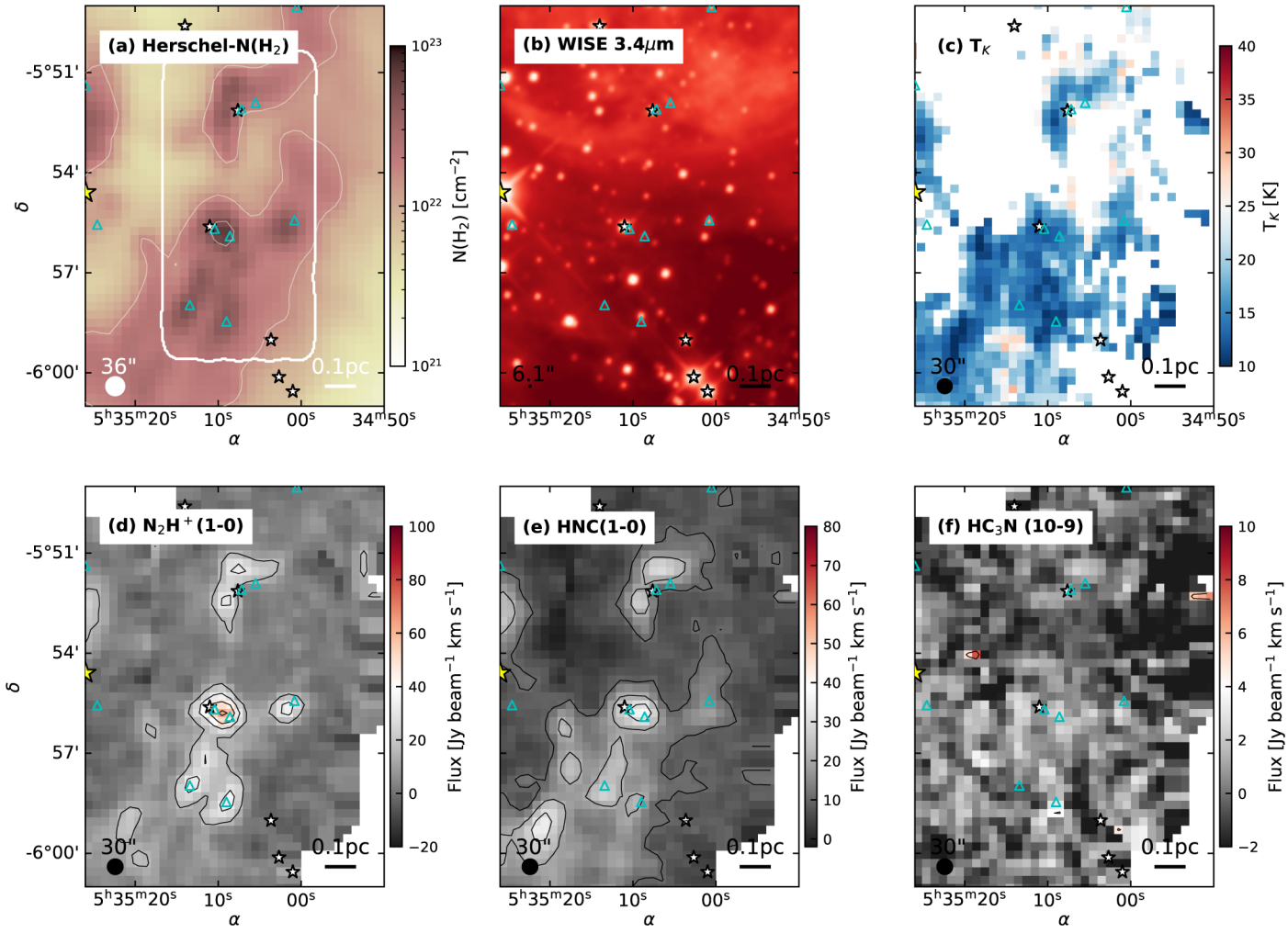

Fig. B.3. Low-resolution observations in OMC-4 South similar to Fig. 4

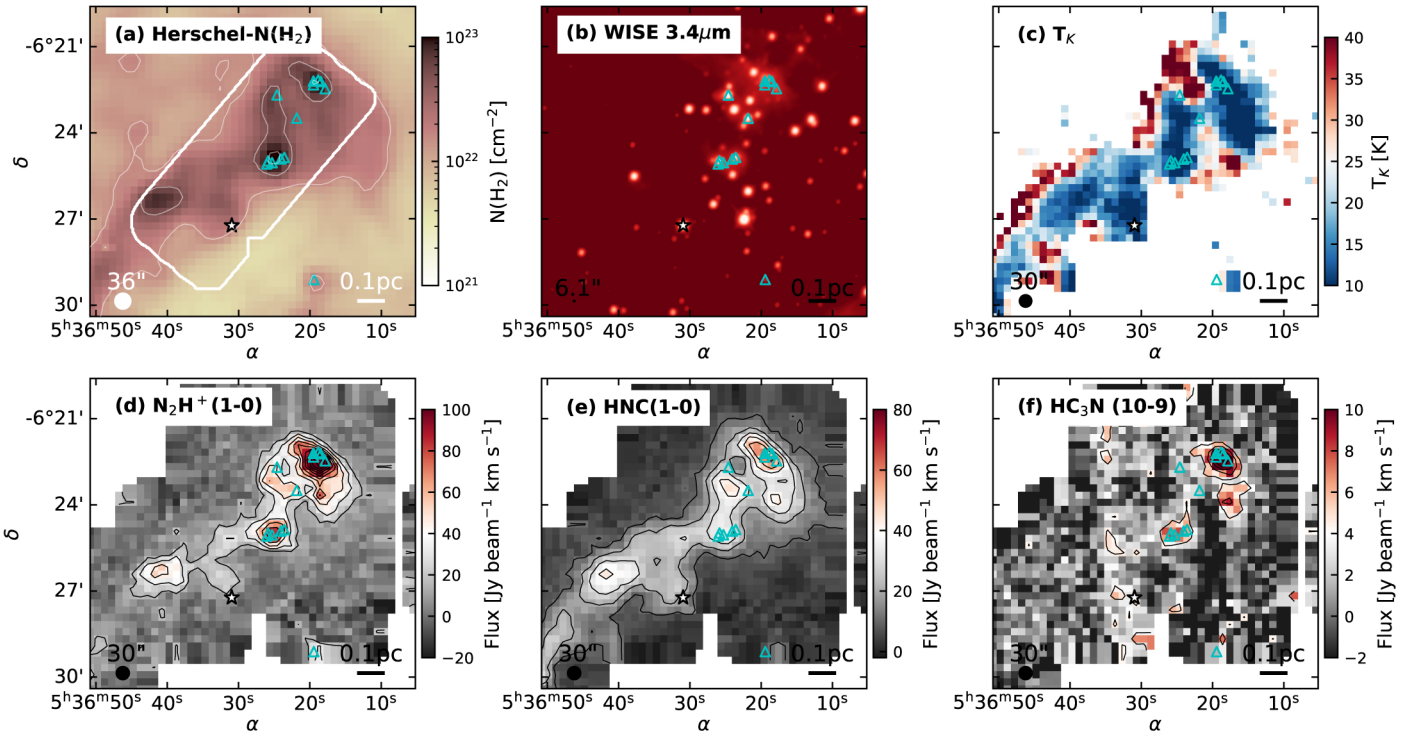

Fig. B.4. Low-resolution observations in LDN 1641N similar to Fig. 4

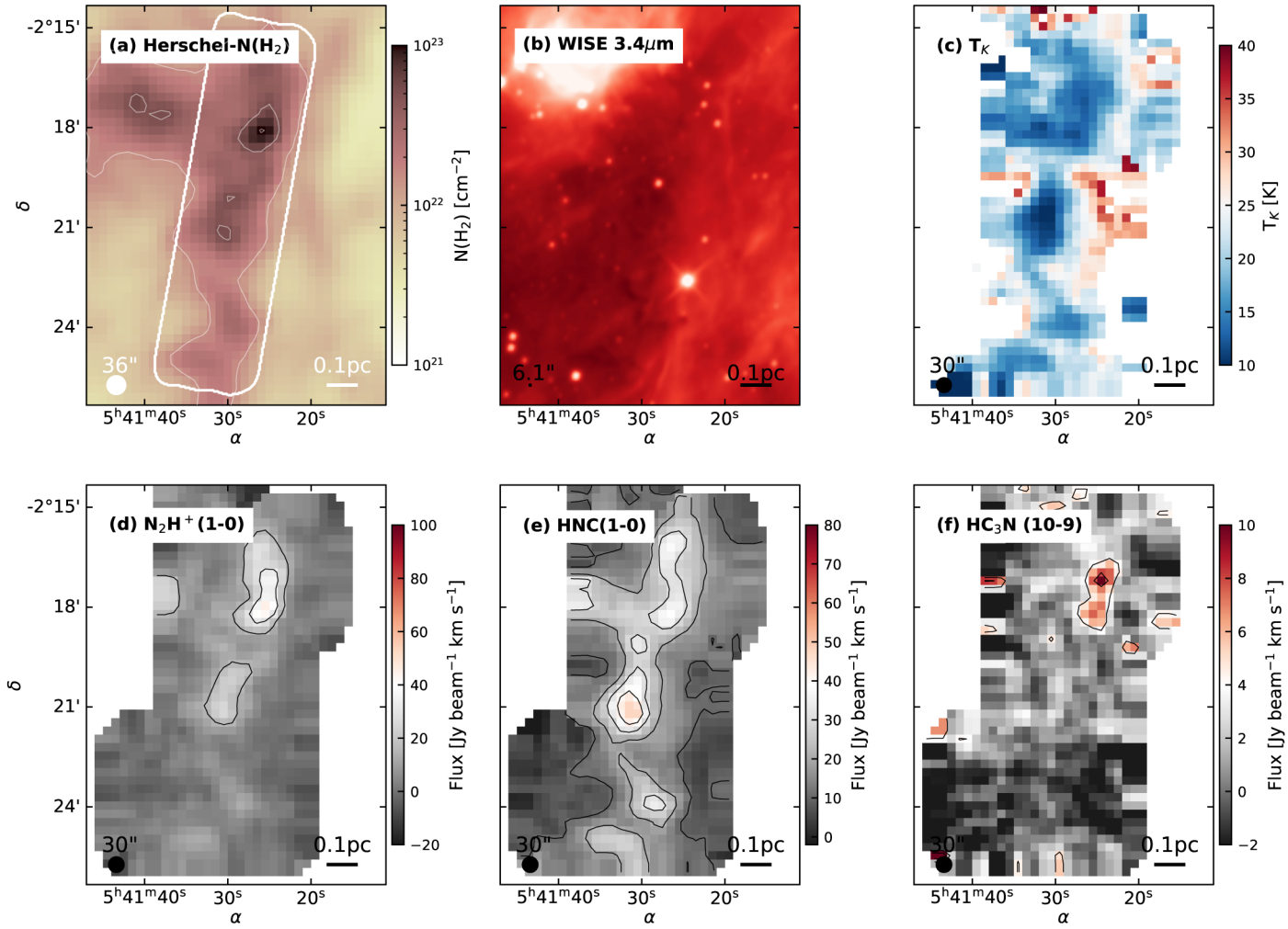

Fig. B.5. Low-resolution observations in NGC 2023 similar to Fig. 4

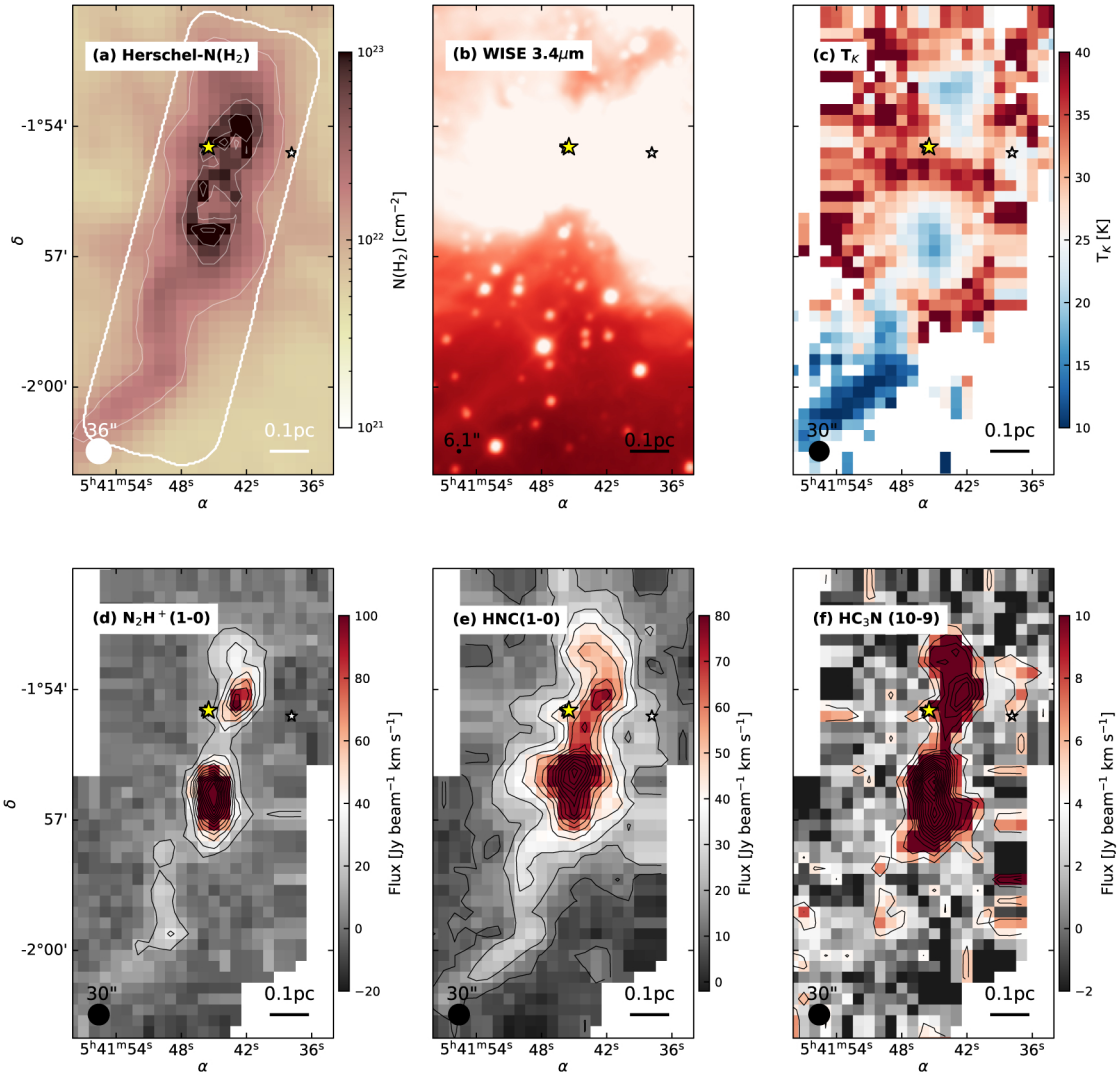

Fig. B.6. Low-resolution observations in Flame Nebula similar to Fig. 4

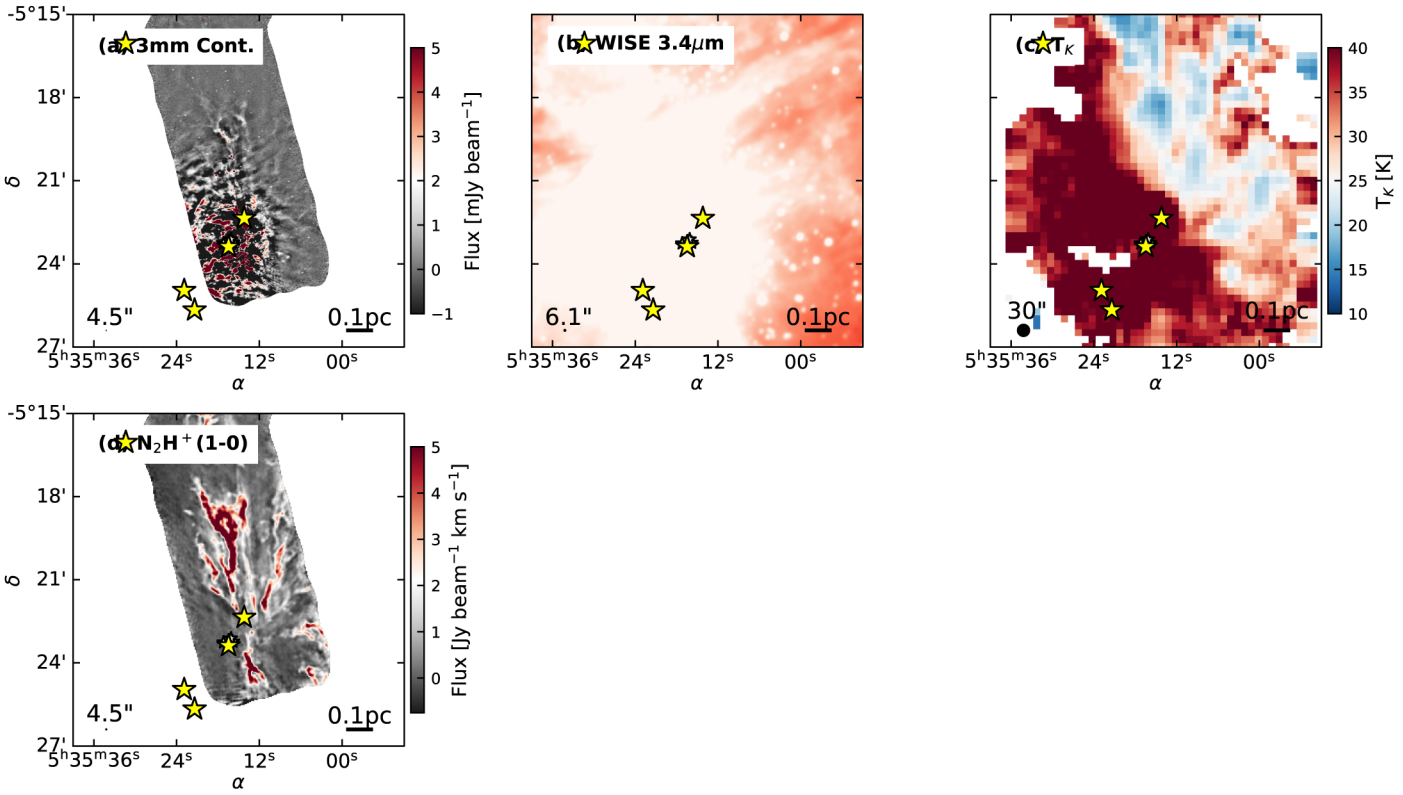

Fig. B.7. High-resolution observations in OMC-1 similar to Fig. 8. No ALMA maps are available for the HNC (1–0) or HC_3N (10–9) lines (see Sect. 3.1).

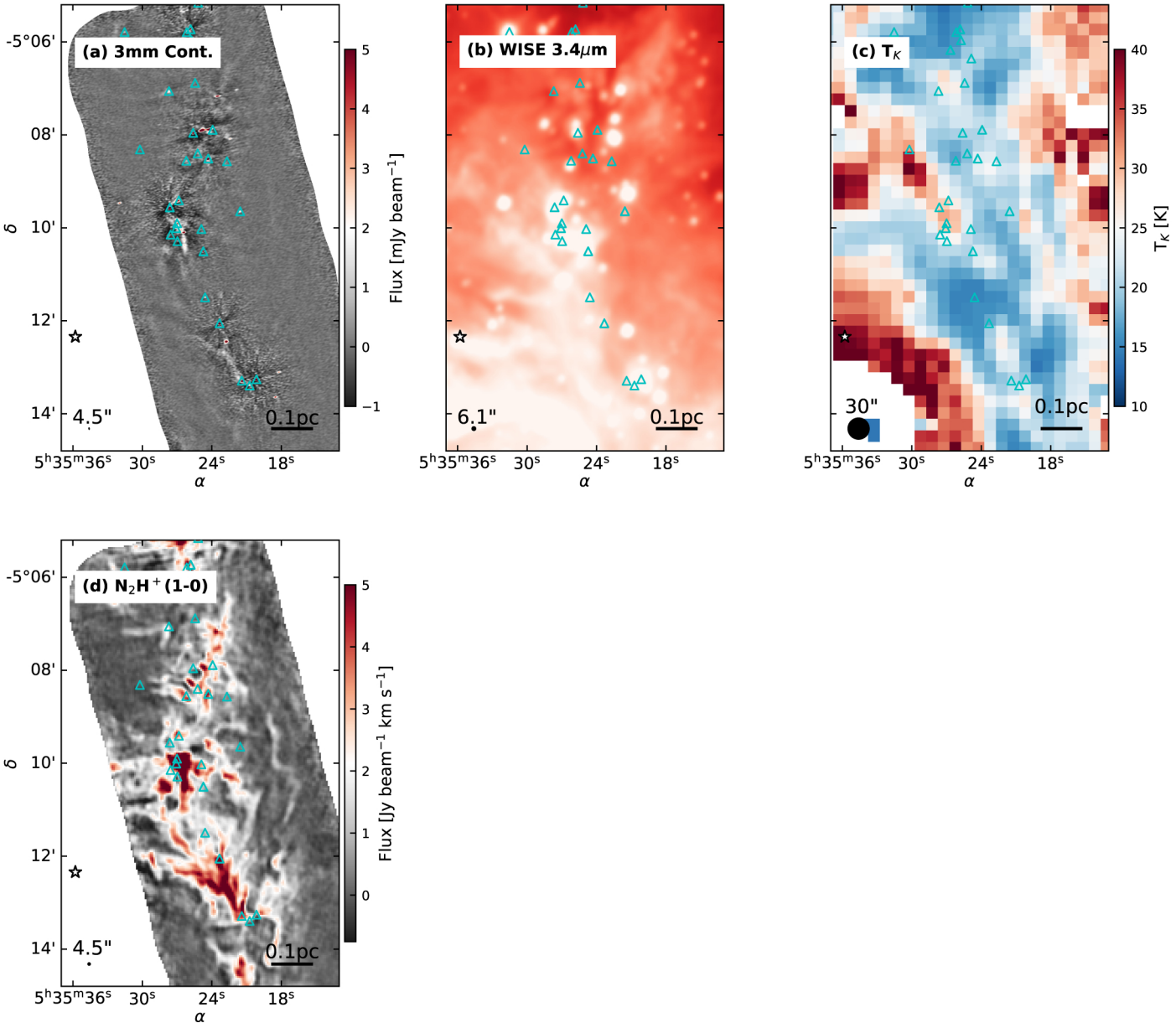

Fig. B.8. High-resolution observations in OMC-2 similar to Fig. 8. No ALMA maps are available for the HNC (1–0) or HC_3N (10–9) lines (see Sect. 3.1).

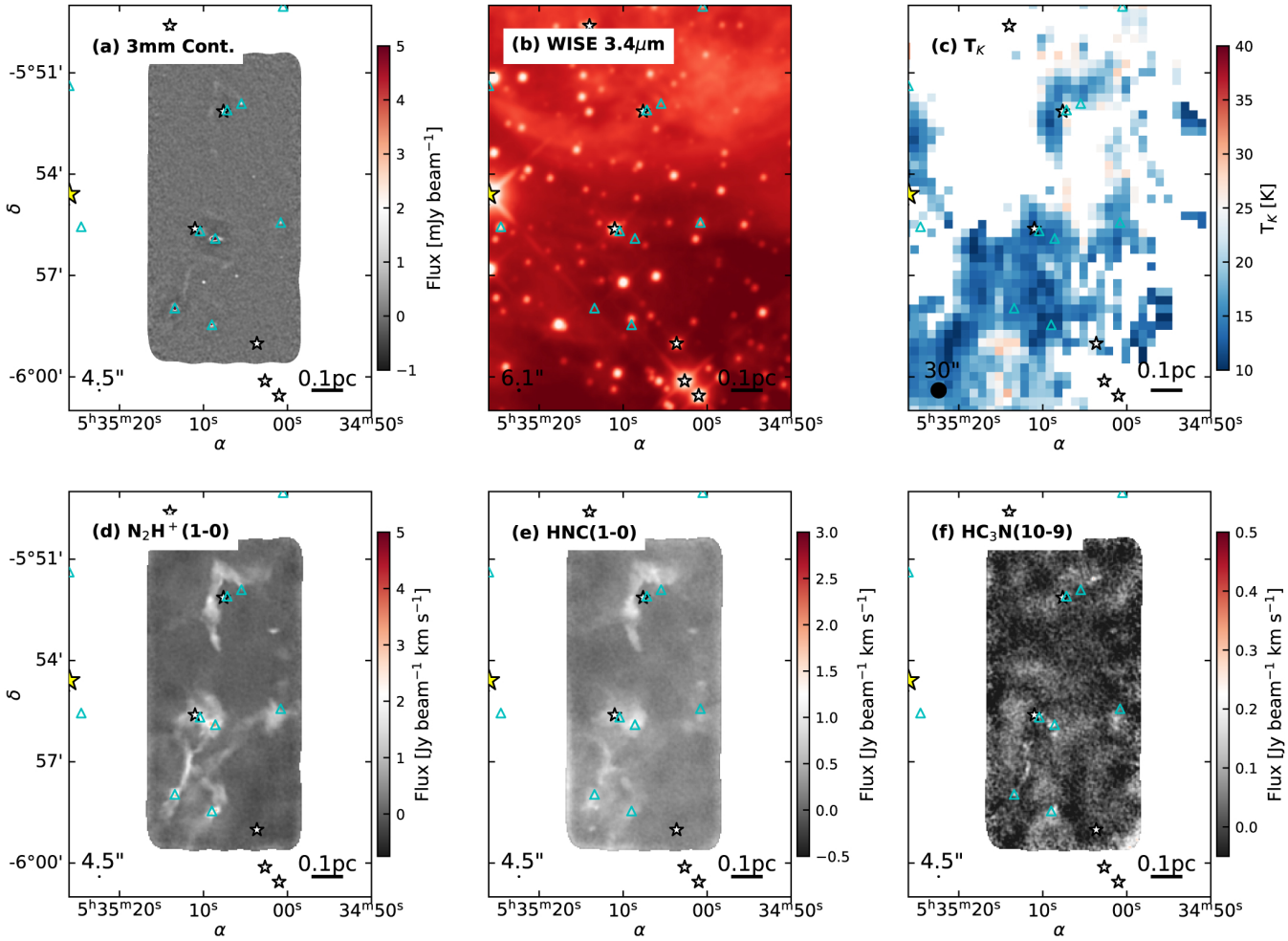

Fig. B.9. High-resolution observations in OMC-4 South similar to Fig. 8.

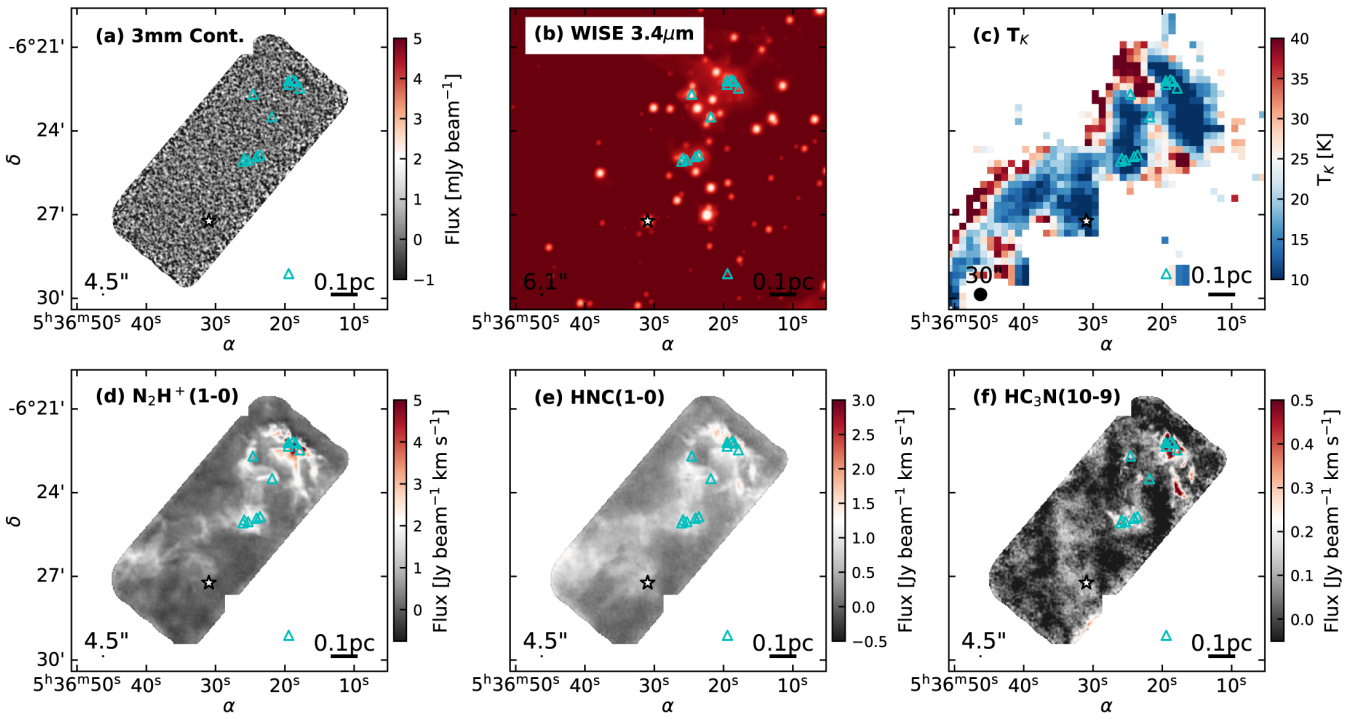

Fig. B.10. High-resolution observations in LDN 1641N similar to Fig. 8.

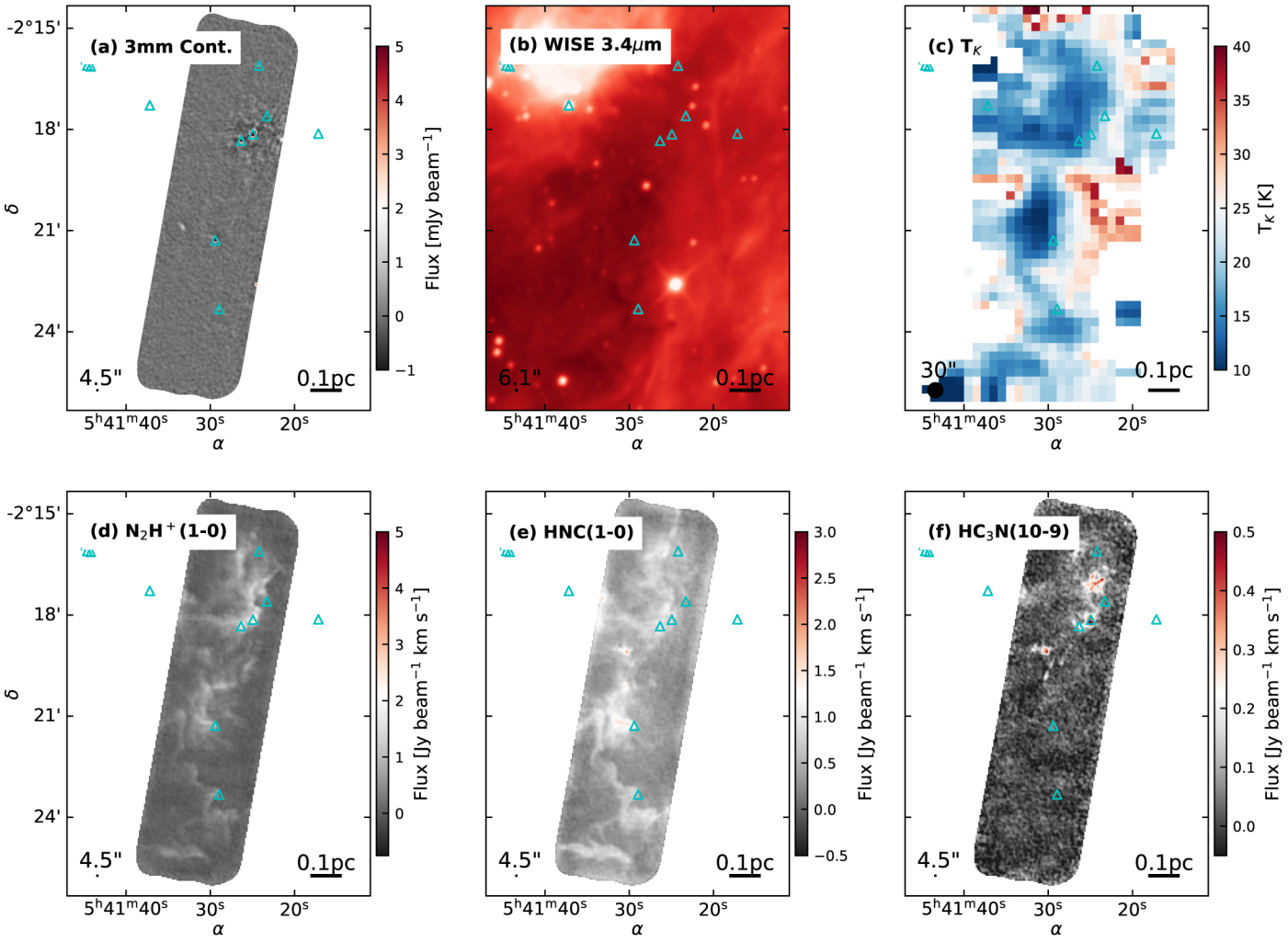

Fig. B.11. High-resolution observations in NGC 2023 similar to Fig. 8.

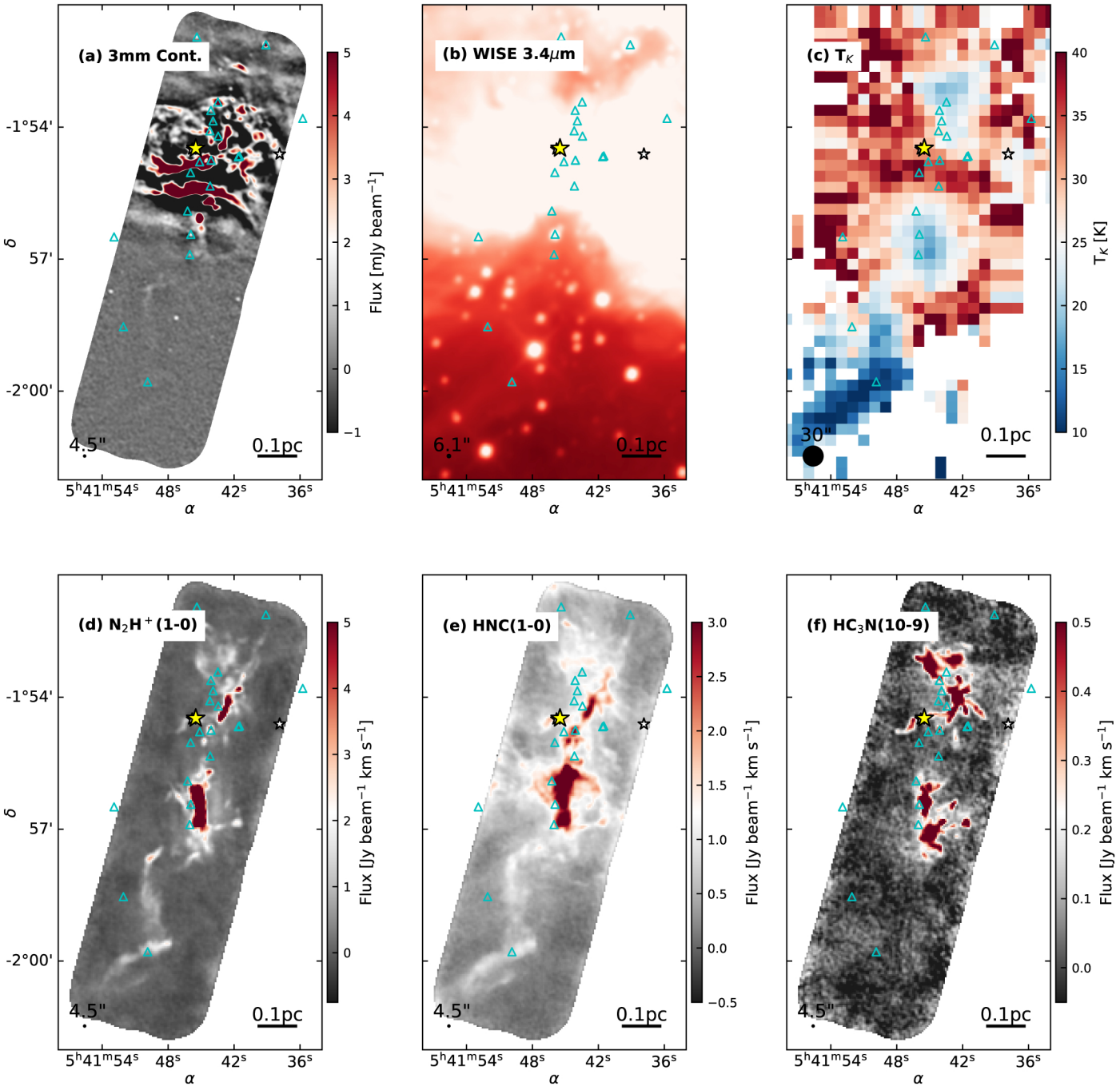

Fig. B.12. High-resolution observations in Flame Nebula similar to Fig. 8.

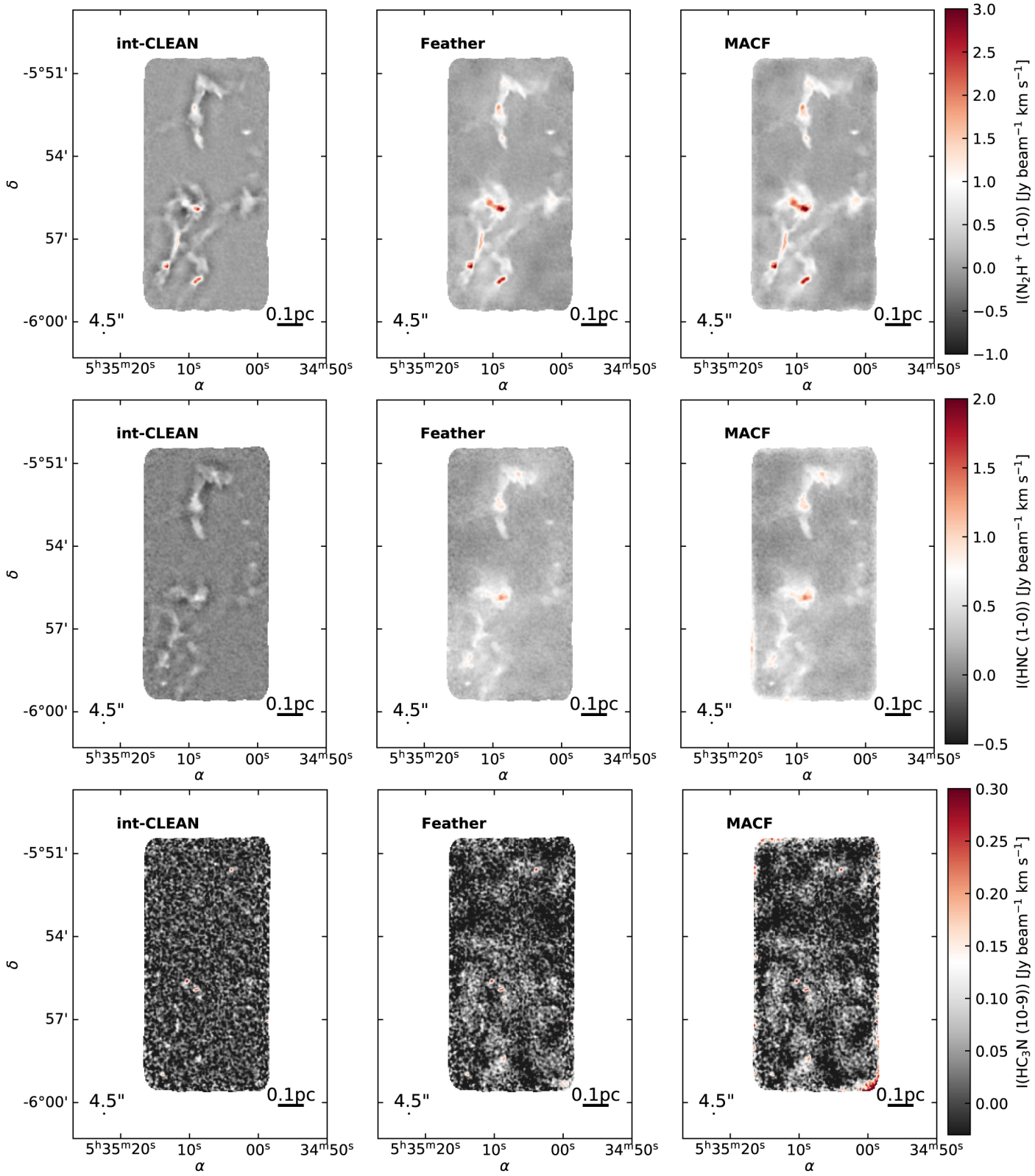

Fig. B.13. Comparison of the different data combination methods in OMC-4 South, similar to Fig. A.1

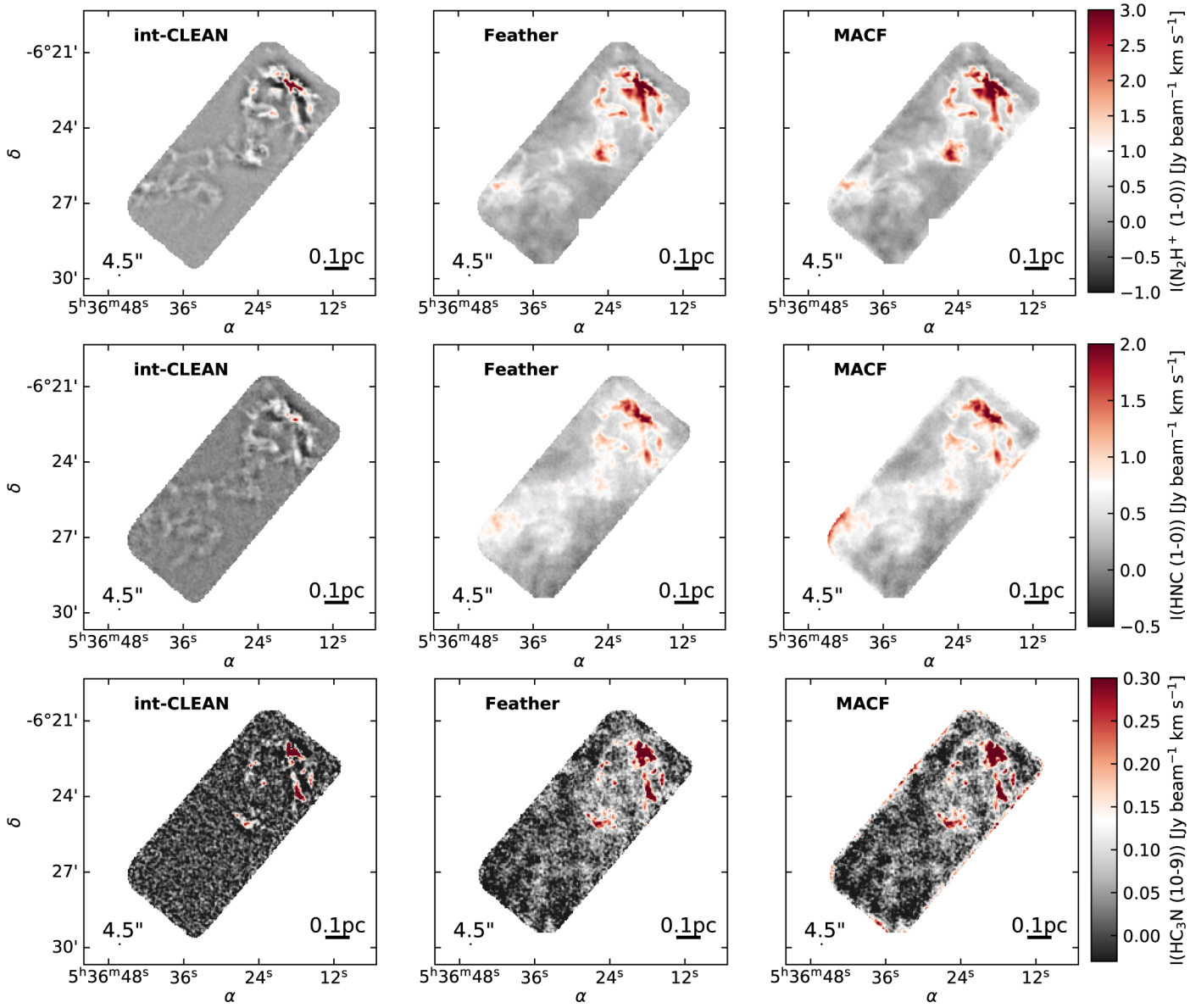

Fig. B.14. Comparison of the different data combination methods in LDN 1641N, similar to Fig. A.1

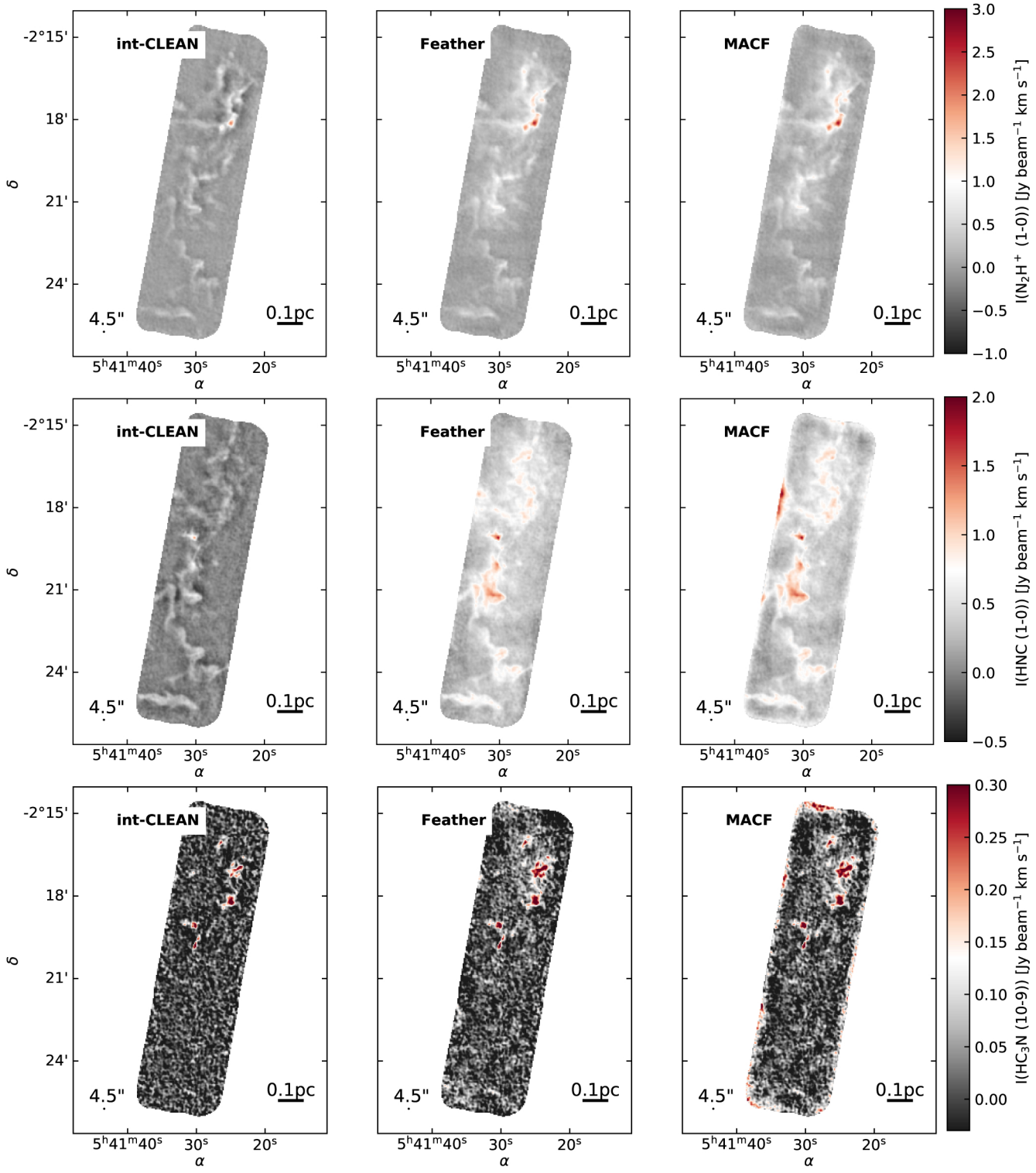

Fig. B.15. Comparison of the different data combination methods in NGC 2023, similar to Fig. A.1

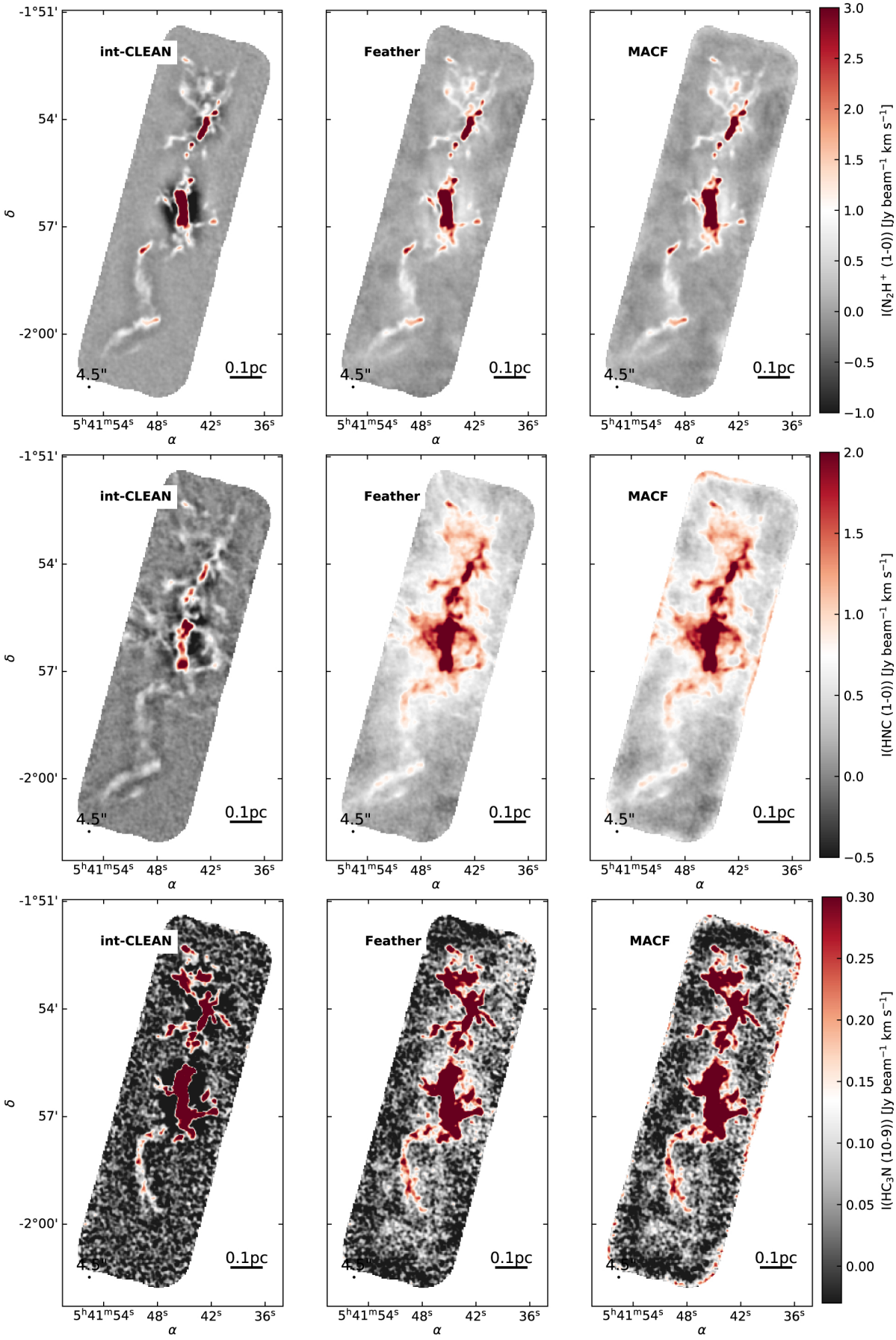

Fig. B.16. Comparison of the different data combination methods in Flame Nabula, similar to Fig. A.1