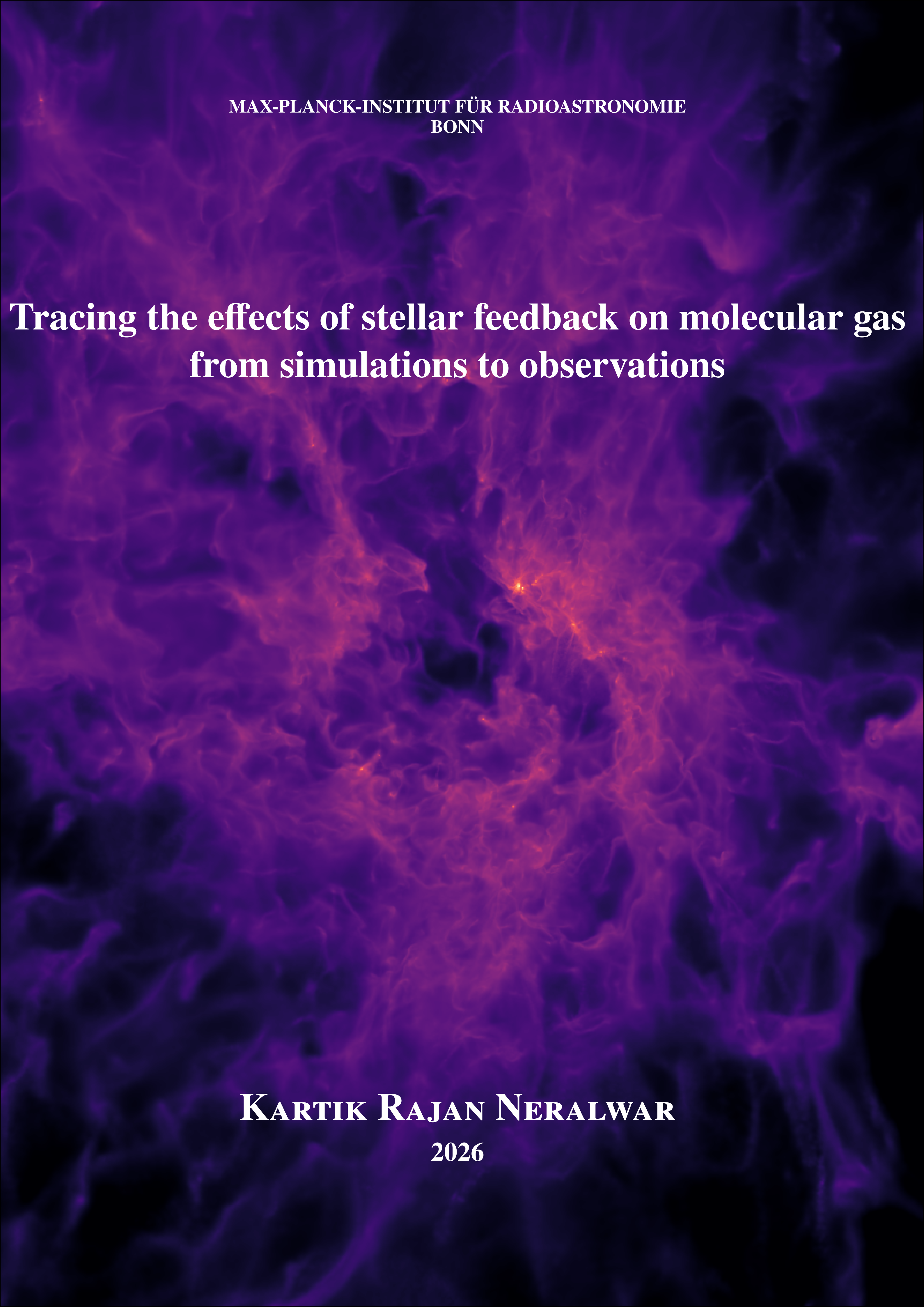


MAX-PLANCK-INSTITUT FÜR RADIOASTRONOMIE
BONN

Tracing the effects of stellar feedback on molecular gas from simulations to observations

Kartik Rajan Neralwar

2026

Tracing the effects of stellar feedback on molecular gas from simulations to observations

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Kartik Rajan Neralwar
aus
Chandrapur, Maharashtra, India

Bonn, 02.02.2026

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer:	Privatdozent Dr. Daniel Seifried
Gutachter:	Prof. Dr. Pavel Kroupa
Tag der Promotion:	29.04.2026
Erscheinungsjahr:	2026

RHEINISCHE FRIEDRICH–WILHELMS–UNIVERSITÄT BONN

Abstract

by Kartik Rajan Neralwar

for the degree of

Doctor rerum naturalium

The interstellar medium (ISM) is a turbulent, multi-phase medium with a hierarchical structure consisting of molecular clouds (MCs), clumps, and cores. Stars form in dense cores and, over their evolution, inject mass, momentum, and energy back into the ISM through stellar feedback processes. This thesis investigates the effects of various stellar feedback mechanisms on molecular gas structures across different spatial scales and evolutionary stages. Our analysis is based on the STARFORGE simulations, which model the evolution of a giant molecular cloud (GMC) at high resolution while self-consistently incorporating all relevant star-formation processes. The simulations explicitly track individual stellar feedback processes within a GMC across multiple timesteps (snapshots), enabling the isolation of individual feedback signatures. Using hierarchical clustering, we identify molecular gas structures within a 20 000 $M_\odot$ GMC and analyse their properties across multiple projects.

We first study the effects of protostellar outflows, stellar winds, and supernovae on molecular gas cores identified in the H_2 density maps of a GMC. Stellar feedback increases the velocity dispersion and virial parameter of the cores by injecting momentum into them. In addition, cores affected by different feedback mechanisms show structural differences, e.g. cores affected by outflows and winds appear smaller and cores affected by supernovae appear larger compared to pristine cores. These results highlight the role of individual feedback mechanisms in shaping the cores.

To enable a direct comparison between the MCs in simulations and observations, we use the RADMC-3D radiative transfer algorithm to post-process the STARFORGE simulation in the likeliness of ^{13}CO(2-1) observations from the SEDIGISM survey. From these synthetic datacubes, we identify MCs using hierarchical clustering and analyse their observable properties in multiple snapshots. The distributions of the synthetic MCs occupy different average positions in the Larson's and Heyer's relations plots. This suggests an evolutionary sequence in which MCs form as small diffuse cloudlets, grow into massive, filamentary, star-forming complexes, and are subsequently transformed into bubbles by stellar feedback before being dispersed.

To study the evolution of clumps embedded in synthetic MCs, we have created an algorithm to follow their evolution by iteratively matching clumps with their corresponding counterparts in subsequent snapshots. Analysing the clump properties as a function of their lifetime, we find that long-lived clumps are typically larger and more massive. A fraction of these survive long enough to form protostars and remain relatively unaffected by strong radiative feedback. In contrast, short-lived clumps are small, less massive structures that are rapidly dispersed by turbulence and radiation.

We further use these synthetic observations to train a convolutional neural network CASI-3D to identify stellar feedback signatures in an observational dataset. The neural network successfully recovers the entire shell of a large bubble, but also identifies wind-affected gas in surrounding noisy regions. This preliminary analysis shows the ability of the neural network to detect stellar feedback signatures in previously unseen observational datasets.

Throughout this thesis, we find that stellar feedback operates across a wide range of spatial scales in molecular gas. On cloud scales, feedback mainly drives fragmentation, creates cavities, and governs the morphological evolution of molecular clouds. On clump scales, feedback reduces the sizes and masses of structures by eroding them and injects momentum to enhance their internal gas motions. On core scales, feedback typically increases the velocity dispersion and the virial parameter of structures. Overall, stellar feedback is a key driver of molecular cloud evolution, shaping gas structures throughout galaxies.

To my family, who mean the world to me.

List of Publications

Publications related to this thesis

The following list indicates the relevant first-author publications that are discussed and presented in this thesis. These are the two peer-reviewed and published research articles and a manuscript that is under preparation.

1. *Effects of stellar feedback on cores in STARFORGE*
   K. R. Neralwar, D. Colombo, S. Offner, F. Wyrowski, K. M. Menten, A. Karska, M. Y. Grudić, S. Neupane (2024), Astronomy & Astrophysics, Volume 690, id.A345, 14 pp. DOI: 10.1051/0004-6361/202451156

2. *A multi-scale evolutionary study of molecular gas in STARFORGE: I. Synthetic observations of SEDIGISM-like molecular clouds*
   K. R. Neralwar, D. Colombo, S. Offner, F. Wyrowski, A. Karska, M.Figuera, G. Wyrowski, S. Neupane (2025)
   Astronomy & Astrophysics, Volume 704, id.A38, 17 pp. DOI: 10.1051/0004-6361/202557095

3. *A multi-scale evolutionary study of molecular gas in STARFORGE: II. Evolution of individual clumps and cores*
   K. R. Neralwar et al. in prep

Other publications during Ph.D.

1. *The SEDIGISM survey: The influence of spiral arms on the molecular gas distribution of the inner Milky Way*
   D. Colombo, A. Duarte-Cabral, A. R. Pettitt ..., **K.R. Neralwar** et al. (2022)
   Astronomy & Astrophysics, Volume 658, id.A54, 38 pp. DOI: 10.1051/0004-6361/202141287

2. *The SEDIGISM survey: Molecular cloud morphology. I. Classification and star formation*
   K. R. Neralwar, D. Colombo, A. Duarte-Cabral et al. (2022)
   Astronomy & Astrophysics, Volume 663, id.A56, 39 pp. DOI: 10.1051/0004-6361/202142428

3. *The SEDIGISM survey: Molecular cloud morphology. II. Integrated source properties*
   K. R. Neralwar, D. Colombo, A. Duarte-Cabral, et al. (2022)
   Astronomy & Astrophysics, Volume 664, id.A84, 27 pp. DOI: 10.1051/0004-6361/202142513

4. *Large-scale velocity-coherent filaments in the SEDIGISM survey: Association with spiral arms and the fraction of dense gas*
   Y. Ge, K, Wang, A. Duarte-Cabral, ..., **K. R. Neralwar** et al. (2023)
   Astronomy & Astrophysics, Volume 675, id.A119, 28 pp. DOI: 10.1051/0004-6361/202245784

5. *OGHReS: star formation in the outer Galaxy II ($l = 180 - 280^{\circ}$)*
   J. S. Urquhart, C. König, D. Colombo, ..., **K. R. Neralwar** et al. (2025)
   Monthly Notices of the Royal Astronomical Society, Volume 539, Issue 4, pp. 3105-3121, 17 pp. DOI: 10.1093/mnras/staf665

6. *The Timescales of Embedded Star Formation as Observed in STARFORGE*
   T. M. Wainer, J. J. Dalcanton, M. Y. Grudic, ..., **K. R. Neralwar**, et al. (2025)
   eprint arXiv:2509.18322 DOI: 10.48550/arXiv.2509.18322

Contents

CHAPTER 1

Introduction

Star formation is one of the most widely studied fields in modern astrophysics, as it affects the formation and evolution of structures from galaxies to planetary systems. This is also reflected in the three big questions[1] used by NASA to guide their science divisions: *How does the universe work? How did we get here? Are we alone?* A large focus of the second question is on understanding how galaxies, stars, and planets form and evolve across cosmic time. Stars form within cold, dense molecular gas found in galaxies, making this gas the fundamental building block of stellar systems. Molecular gas is usually analysed by classifying gaseous structures across spatial scales into giant molecular clouds (GMCs), clumps, and cores. These structures evolve under the combined influence of gravity, turbulent motions within the gas, and ejections from stars. Understanding how these molecular gas structures form, change, and eventually disperse is therefore important in revealing the history and future of star formation.

Over the past two decades, our understanding of star formation and molecular clouds (MCs) has progressed significantly, driven by major advancements in both observations and simulations (modelling). Observational facilities such as ALMA[2] and JWST[3] have produced a large number of high-resolution and high-sensitivity images of star forming regions across a wide range of environments. In parallel, numerical simulations have become more realistic by incorporating better models of stars and large-scale galactic environments. Recent developments in radiative transfer algorithms enable the generation of synthetic observations from simulations in the likeliness of real observations, allowing for detailed comparisons between the two. Moreover, applying observational analysis techniques directly to synthetic datasets provides better interpretations of observed structures. One way to achieve this is by linking specific observational signatures (e.g. bubbles) to the gas conditions and physical processes that produce them (e.g. stellar winds). These comparisons also test the reliability of the simulations, allowing them to be further improved. Together, these advancements have led to improved constraints on the interactions between stars and clouds, while providing better estimates of the cloud lifetime.

One of the main scientific objectives of this thesis is to explore how stellar feedback (ejections

---

[1] `https://science.nasa.gov/astrophysics/science-questions/`

[2] Atacama Large Millimeter/submillimeter Array is a ground based telescope based in Chile. `https://www.almaobservatory.org/en/home/`

[3] James Webb Space Telescope is a space based observatory equipped with high-resolution and high-sensitivity instruments to conduct detailed observations of the universe `https://science.nasa.gov/mission/webb/`

from stars) affects the properties and evolution of molecular gas structures at different scales. This is achieved using the STARFORGE simulations (Grudić, Guszejnov, Hopkins et al., 2021) that model the lifetime of a GMC at high resolution and incorporate a comprehensive treatment of various stellar feedback mechanisms. The analysis begins in the simulation domain by studying the properties of the molecular gas directly. We then move towards the observational domain through the use of synthetic observations, and ultimately connect these to real observations using machine learning (ML) techniques. This work presented in this thesis is organised into three major projects, each focused on structures at different spatial scales, followed by a preliminary ML-based study.

Project I: *Tracing the effects of individual stellar feedback mechanisms on high-resolution gas cores*
This project involves understanding the differences in the population of cores affected by different types of feedback, such as protostellar outflows, stellar winds, and supernovae. By dividing the cores into bins based on the degree to which they are affected by feedback, we aim to answer the following questions.

- Does stellar feedback influence sub-parsec-scale molecular gas structures?
- How do feedback mechanisms affect the physical properties of cores?
- Is it possible to study the effects of individual stellar feedback mechanisms on molecular gas?

Project II: *An evolutionary study of synthetic molecular clouds*
This project aims to create realistic synthetic observations of simulated MCs to understand their evolution. After comparing the synthetic clouds with an observational benchmark, we study the changes in their properties as a function of time to address the following questions:

- To what extent do current simulations reproduce observed MCs?
- How do the observable properties of MCs vary as they evolve?
- Can scaling relations provide insights into the evolutionary stages of observed MCs?
- What are the visible effects of stellar feedback on the evolution of MCs?

Project III: *Following the evolution of individual synthetic clumps*
The objective of this project is to extend the study of the properties and evolution of gas structures to smaller scales. We have developed an algorithm to follow the evolution of individual clumps embedded in MCs to understand how their properties and lifetimes are affected by individual feedback mechanisms. In doing so, our aim is to address the following:

- How does the natal cloud environment influence the evolutionary pathways and lifetime of clumps?
- How do the physical properties of individual clumps evolve over time as they undergo fragmentation and merge with other structures?
- Are the effects of individual feedback mechanisms visible on the observable properties of clumps over time?

Project IV: *Machine learning approach to identifying stellar feedback*
This project explores the use of a neural network algorithm to automatically identify feedback-affected structures in observational datasets. By training simulations in which stellar feedback signatures can be isolated, the aim is to replicate this capability in observations. Ultimately, this will enable a pixel-level characterisation of feedback-driven structures, allowing the science question mentioned above to be addressed directly in observational data. The basic concepts and techniques necessary to understand the algorithm are described in Sect. 4.2.

The thesis is structured as follows. Chapter 2 presents an overview of the literature relevant to this work. Chapter 3 describes the STARFORGE simulations used for the analysis and the SEDIGISM survey, which is the observational benchmark. Chapter 4 describes the methods used to create the synthetic observations and the machine learning techniques used in the thesis. The synthetic observations described in Sect. 4.1 represent the main dataset used in the thesis. Section 4.2 introduces the machine learning concepts necessary to understand the neural network algorithm used in the final project. The four projects are described in Chapters 5, 6, 7, and 8, respectively. Finally, Chapter 9 summarises the main findings of this work, together with an outlook on future projects.

CHAPTER 2

Astrophysical background

Matter in galaxies goes through a continuous cycle between gas and stars, referred to as the Baryon cycle (Fig. 2.1). Galaxies accrete gas from their surrounding medium into the turbulent, multi-phase interstellar medium (ISM), where diffuse atomic and molecular phases coexist. The gas in these diffuse phases cools, becomes self-shielded, and settles in the form of molecular clouds. Within these clouds, hierarchical structures such as filaments, clumps, and cores form, which ultimately give rise to stars and stellar systems. Throughout their lifetime, stars eject matter and energy back into the ISM through stellar feedback mechanisms, which shape the surrounding ISM by driving turbulence and creating low density bubbles. Low- and intermediate mass stars undergo high mass loss during the asymptotic giant branch (AGB) phase, forming planetary nebulae. Massive stars, in contrast, end their lives as supernovae, ejecting large amounts of energy and metals into the ISM. These feedback processes enrich the ISM with heavy metals produced in the stellar interiors and contribute to the redistribution of gas across different phases and scales. A fraction of the feedback gas is expelled from the galactic disk through galactic fountains, replenishing the gas reservoir responsible for future star formation. Feedback also disperses gas, transforming it from molecular to atomic and ionised

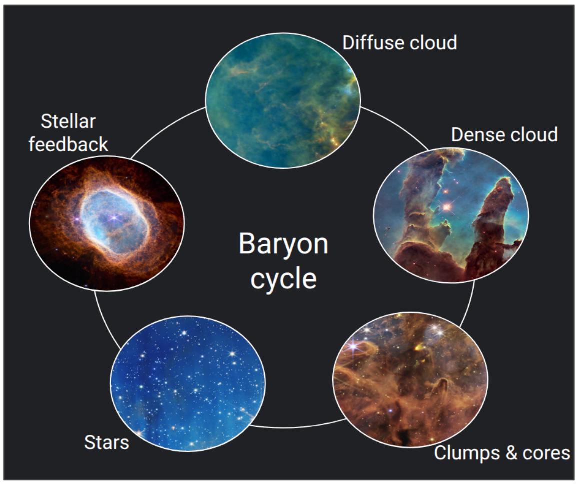


Figure 2.1: Baryon life cycle in the ISM. Image credits: NASA, ESA, CSA, STScI, Herschel/PACS/SPIRE/V. Roccatagliata

phases, continuing the Baryon cycle. In this thesis, we discuss multiple stages of this cycle, focussing on molecular gas, stars, and stellar feedback, as described in the following sections.

2.1 The interstellar medium

The interstellar medium (ISM) refers to the material that fills the space between stars in galaxies (Cox, 2005; Draine, 2011b; Klessen and S. C. O. Glover, 2016; Kennicutt and Evans, 2012; Saintonge, 2025). It is inhomogeneous, dynamic, and hosts multiple physical and chemical processes. It encompasses various energy sources that drive the flow of matter within and around galaxies. It also provides the raw material for the formation of stars. The ISM thus serves as an interface for energy and matter transfer between the large galactic scales and the smaller stellar scales.

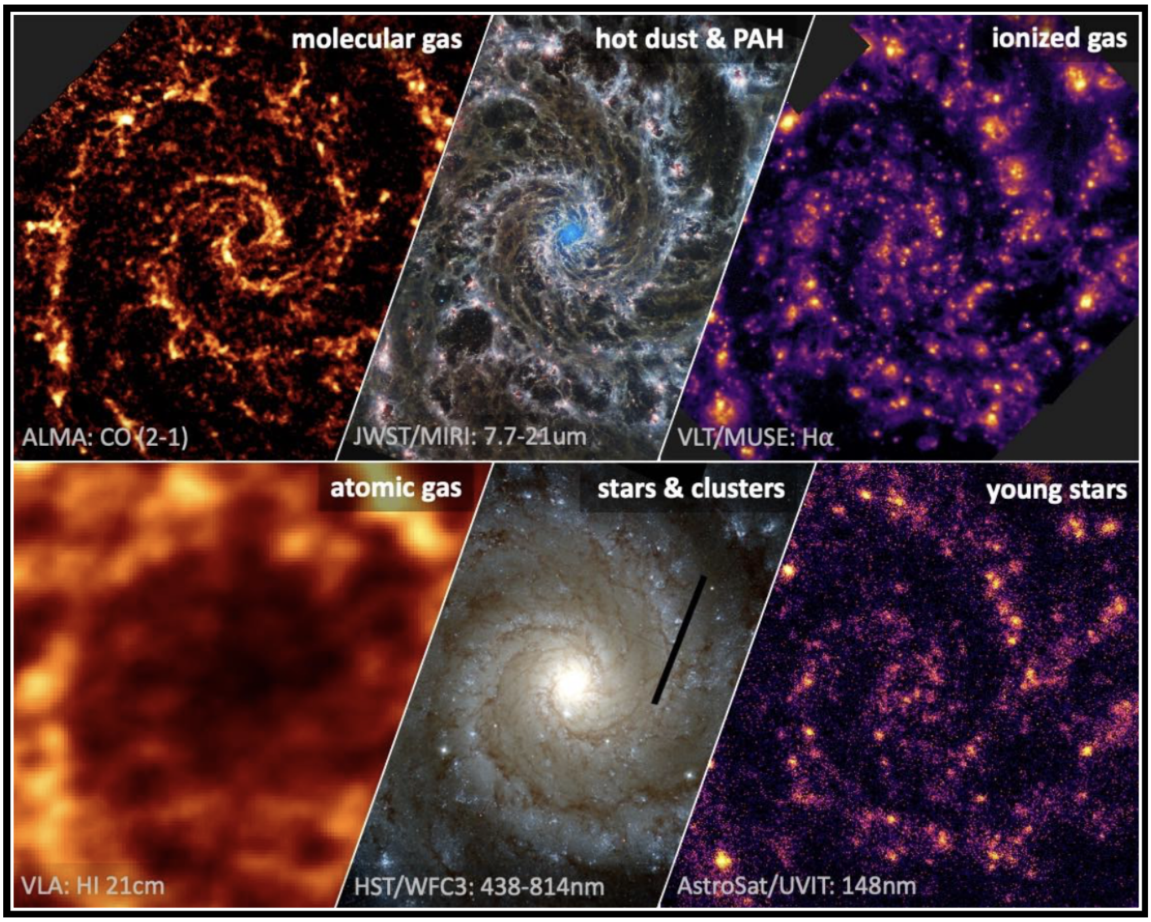


Figure 2.2: Multi-phase ISM and stellar population in the galaxy NGC628. Iage credit: PHANGS Team / J. Sun (Princeton) / J. Schmidt (ASCL) / NASA / ESA/ CSA. Adapted from Saintonge, 2025.

2.1.1 Components of the ISM

Although it appears empty to the naked eye, the space between stars is filled with gas, dust, electromagnetic radiation, cosmic rays, magnetic fields, and the gravitational field (Draine, 2011b; Klessen and S. C. O. Glover, 2016). Gas is the dominant component of the ISM, with the ionised gas occupying most of its volume and the neutral gas containing the majority of the mass (Fig. 2.3). Hydrogen, the most abundant element, accounts for roughly 70% of the ISM gas mass, followed by helium at about 28%. The high abundance of these elements is mostly attributed to their formation during the Big Bang. The remaining 2% consists of heavier elements that are termed metals. These are produced during thermonuclear reactions in stars and supernova explosions. Despite their low abundance, metals play a significant role in regulating the thermal behaviour of the ISM and setting the environment for star formation (Maciel and Costa, 2010).

The interstellar gas is also mixed with dust (Fig. 2.3), with a typical dust-to-gas ratio of ~ 1 % in the solar neighbourhood (Spitzer, 1998; Draine, 2003; Calura, 2025). The dust grains range from 0.3 nm to approximately 1 μm in size, with their composition mainly based on silicates and carbonaceous material (Witt, 2000; Sneppen and D. Watson, 2025). Catalytic reactions on the surface of dust grains provide a dominant pathway for the formation of H_2 in the ISM (Perets and Biham, 2006; Wakelam et al., 2017). By absorbing and scattering radiation, dust causes the extinction of starlight, reddening emission from distant sources. The absorbed UV light is reemitted as infrared emission, with smaller grains emitting near-infrared (NIR) and mid-infrared (MIR) and large grains emitting far-infrared (FIR) radiation (Witt, 2000; Draine, 2003; Paradis et al., 2023).

The ISM is permeated by high energy (10^{21} eV), relativistic protons, electrons, and alpha particles called cosmic rays (Draine, 2011b; Klessen and S. C. O. Glover, 2016). These rays can penetrate deep into the ISM, ionising molecular and atomic gas to initiate chemical networks within clouds (Jacob, Neufeld et al., 2022). Cosmic ray ionisation of H, H_2 and Ar triggers ion-neutral reaction chains that drive the formation of oxygen- and argon-bearing hydride ions such as OH^+, H_2O^+ and ArH^+ (Neufeld, Goicoechea et al., 2010; Neufeld and M. G. Wolfire, 2016). The ISM also contains magnetic fields that are believed to originate from weak primordial seed fields. Seed fields are amplified by turbulence, differential rotation, and dynamo processes driven by galactic-scale flows and stellar feedback (Seta and McClure-Griffiths, 2025). Magnetic fields play an important role in star formation by directing large-scale gas flows in molecular clouds to accrete material onto star-forming filaments and hubs (Pattle, Fissel et al., 2023). They also provide significant support against gravitational collapse in molecular clouds, influence cloud fragmentation, and increase their Jeans mass (Crutcher, 2012; Klos, Bonnell and R. J. Smith, 2025).

The ISM is filled by electromagnetic radiation from multiple sources, such as starlight, thermal emission from dust, emission from ions, atoms, and molecules, free-free emission from plasma, syncrotron radiation from relativistic electrons, gamma rays, and the cosmic microwave background (CMB). These radiations span over a wide range of wavelengths (10^{-7} cm – 10 cm), and are collectively referred to as the interstellar radiation field (ISRF). The interstellar gas and dust heavily interact with the ISRF to set the physical, chemical, and thermal state of the gas in the ISM. These interactions can be better studied by dividing the radiation into several emission bands, of which the ones relevant to this study are the following.

- The extreme ultraviolet (EUV; $\lambda < 912$ Å) regime includes ionising photons emitted by hot, massive early-type stars. EUV photons are absorbed by diffuse neutral hydrogen in the

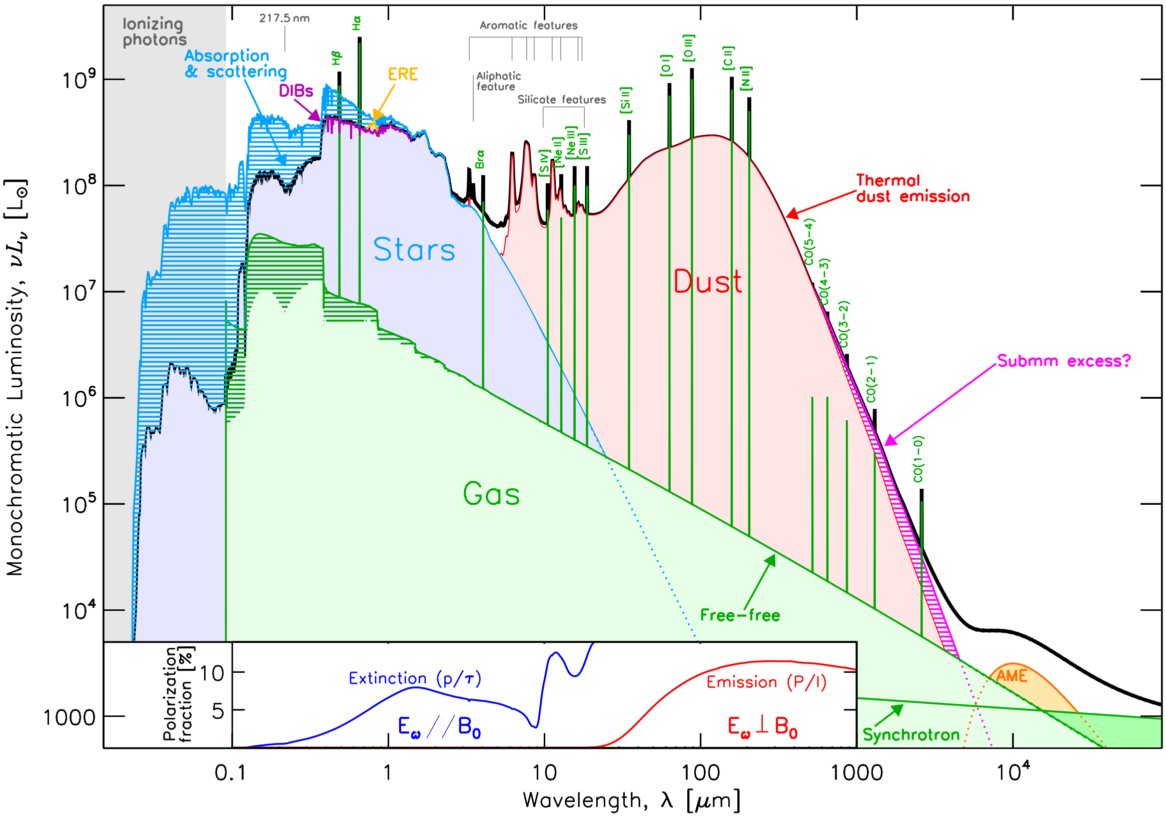


Figure 2.3: Spectral energy distribution of a typical late-type galaxy for multiple components. The blue hatched region shows the power absorbed by dust. The image shows the spectra for Diffuse Interstellar Band (DIB), Extended Red Emission (ERE) and Anomalous Mircrowave Emission (AME), with various gas lines. Taken from Galliano, Galametz and A. P. Jones (2018).

photoionised (H II) regions and are responsible for their dynamics. Photoionisation is often theorised to be the most important feedback effect from massive stars on MC scales (Dale, Ercolano and Bonnell, 2012; Grudić, Hopkins et al., 2019; Olivier et al., 2021).

- The Far-ultraviolet (FUV; 912 Å $< \lambda <$ 1550 Å) band includes the majority of the ISRF responsible for heating and photodissociation in the neutral ISM. It is dominated by stellar emission from young massive OB stars and controls processes such as photoelectric effects on dust and dissociation of molecules such as H_2 (Korpela, Bowyer and Edelstein, 1998; E. C. Ostriker, McKee and A. K. Leroy, 2010; Kulkarni, 2022).

- The Near-UV (NUV; 1550 Å $< \lambda <$ 3600 Å) contains most of the photon energy and momentum from young stars. It contributes to dust heating, has indirect effects on ISM chemistry, and is the major contributor to radiation pressure from stars (Fall, Krumholz and Matzner, 2010; J.-G. Kim, W.-T. Kim and E. C. Ostriker, 2018; Akshaya et al., 2019).

- Visible / near infrared (NIR; 3600 Å $< \lambda <$ 3 μm) emission originates from old stellar populations and Polycyclic Aromatic Hydrocarbons (PAHs). It sets the energy budget for the extinction of background sources and carries a significant fraction of the photon momentum (Bianchi, 2024).

- The Mid- /far-IR (FIR; 3 µ $< \lambda <$ 1000 µm) emission is dominated by thermal emission from interstellar dust, which absorbs short-wavelength radiation and re-emits it at longer wavelengths. It is the primary cooling channel of the ISM and represents a large fraction of its irradiative energy budget (Krumholz, Klein, McKee et al., 2009; Tsang and Milosavljević, 2018; Ramos Padilla et al., 2023).

2.1.2 Phases of the ISM

In a typical star-forming galaxy, the interstellar medium spans a wide range of densities and temperatures, indicating that it is regulated primarily by an approximate pressure equilibrium rather than by a global thermal equilibrium. Based on these features, ISM is often described using a three-phase model: hot ionised medium (HIM), warm medium (WIM and WNM), and cold neutral medium (CNM) (McKee and J. P. Ostriker, 1977; Cox, 2005; Saintonge, 2025).

The hot ionised medium (HIM) is the hottest ($\sim 10^6$ K) and least dense (~ 0.004 cm^{-3}) component of the ISM. It accounts for a small fraction of the total ISM mass, but almost half of its volume. It is believed to be produced mainly by supernovae driven shocks, as it appears in the form of bubbles in the galactic disks, which can be traced through diffuse X-ray emission and metal absorption lines at UV wavelengths.

The warm interstellar medium comprises an ionised phase (the warm ionised medium; WIM) and a neutral phase (the warm neutral medium; WNM). The WIM is a mostly diffused component with typical densities of ~ 0.2 cm^{-3} and temperatures ~ 8000 K. It is dominated by ionised hydrogen and contains 90% of the total ionised gas in the ISM, often traced by Balmer lines (Haffner et al., 2009). The ionised state of the WIM is mostly a result of the photonisation from stars; however, it is not the same as ionised regions such as H II regions and planetary nebulae that appear as localised and dense structures. The WNM has temperatures ~ 6000 K and densities ~ 0.4 cm^{-3} and is traced using the 21 cm emission line of atomic hydrogen. The same emission line also traces the CNM, and therefore these two phases are often considered a two phase medium in pressure equilibrium (Bialy and Sternberg, 2019). However, WNM has a higher temperature and is more diffuse than the CNM, tracing the diffuse gas between the molecular clouds. In addition, it is more spatially extended than the molecular and stellar components of the galactic disks, and is considered a reservoir of material available for future star formation.

The cold neutral medium (CNM) is the coldest phase of the ISM and dominates the total ISM mass budget. Observations and simulations of star forming galaxies reveal that the CNM exhibits filamentary structures and is confined to a very thin galactic disk (Fig. 2.2). The CNM spans a wide range of temperatures and densities with typical values of T~ 80 K and n~ 40 cm^{-3}. The densest structures in CNM are commonly identified as molecular clouds (MCs), which are treated as discrete physical entities and are typically delineated using thresholds in molecular line emission or boundaries in dust column density.

2.2 Molecular Clouds

Molecular clouds are defined as semi-virial structures in ISM that are separated from their surroundings by the rapid change of some property, such as pressure, density, or chemical state (Kennicutt and Evans, 2012). They are some of the densest ($n_H \sim 100$ cm^{-3}) and coldest ($T \sim 10$ K) regions in the interstellar medium and have a hierarchical structure (Dobbs, Krumholz et al., 2014a; E. W. Rosolowsky, Pineda, Kauffmann et al., 2008a). MCs are magnetically super-critical (Crutcher, 2012), self-gravitating turbulent structures (Ballesteros-Paredes, L. W. Hartmann et al., 2011a) that have masses of $\sim 10^2 - 10^7$ M$_\odot$, with the mass distribution function typically described by a power law form (Colombo, Hughes et al., 2014; Miville-Deschênes, Murray and E. J. Lee, 2017). Their volume densities are predicted to follow a log-normal distribution with a power-law tail (Ballesteros-Paredes, L. W. Hartmann et al., 2011a; Burkhart, Stalpes and Collins, 2017; Jaupart and Chabrier, 2020; Khullar et al., 2021). MCs were originally presented as structures with constant surface densities (Larson, 1981a), however, the analysis of a larger sample later disproved this (M. Heyer, Krawczyk et al., 2009). The typical surface densities of MCs are on the order of 100 M$_\odot$ pc^{-2} (Rebolledo et al., 2012) and show variations with environment (Colombo, Schinnerer et al., 2013; E. Rosolowsky, Hughes et al., 2021). MCs also exhibit a range of morphologies, with the majority being filamentary. (Jaffa et al., 2018; Yuan et al., 2021; Neralwar, Colombo, Duarte-Cabral et al., 2022). Although MCs are often defined as regions where hydrogen is mainly in molecular form, it should be noted that the presence of H_2 is a necessary but not a sufficient condition for a gas agglomeration to be called a cloud.

The most massive molecular clouds are referred to as giant molecular clouds (GMC, Chevance, Kruijssen et al., 2022). They span several parsecs to tens of parsecs with masses $\sim 10^4 - 10^7$ M$_\odot$, densities $\sim 1 - 3 \times 10^2$ M$_\odot$ cm^{-3}, and temperatures $\sim 10 - 20$ K. GMCs show a hierarchical structure with regions of higher density called clumps, which are considered progenitors of star clusters (Rigby, Moore, Eden et al., 2019). These clumps, in turn, contain denser structures called cores (Fig. 2.4), which are the birthplaces of individual and binary stars (Schisano, Molinari, Coletta et al., 2025). In this thesis, the terms GMC and MC have been used interchangeably.

Figure 2.4 shows the molecular gas structures across a wide range of spatial scales and morphologies. In the following, we first discuss the formation, evolution, and destruction of clouds, followed by an overview of the different morphological structures, i.e., filaments and bubbles. We then move to smaller scales, focussing on the structures within molcular clouds, namely clumps and cores.

2.2.1 The lives of molecular clouds

The formation of MCs can be explained as a consequence of multiple non distinct processes, including gravitationally-induced compression of Jeans unstable regions, converging filamentary flows driven by ISM turbulence, shock wave-induced compression at edges of supernoave driven shells, agglomeration by cloud-cloud collisions, compression due to shearing motions and spiral density waves in galaxies, and compression in large scale galaxy mergers (Dobbs, Krumholz et al., 2014a). The various theories generally involve converging flows of gas, but the scales on which they act are different, resulting in a wide range of MC properties. Interstellar turbulence, shocks, and stellar feedback (Sect. 2.4) can cause streams of warm HI gas to collide, decreasing their temperature, and eventually causing a change of phase from atomic to molecular. A free-fall collapse of this molecular gas then gives rise to molecular clouds (L. Hartmann, Ballesteros-Paredes and Bergin, 2001; Vázquez-Semadeni, Banerjee

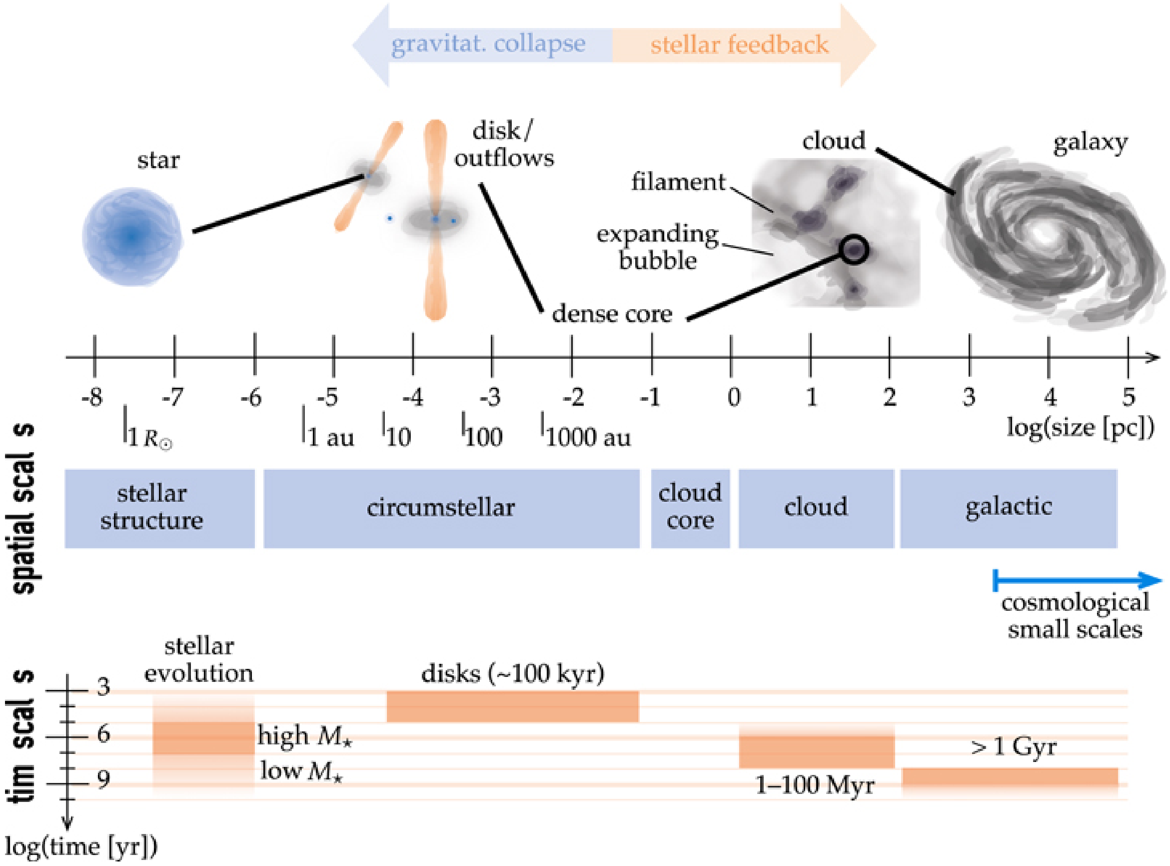


Figure 2.4: Structures in the ISM at different spatial scales. Taken from Geen, Agrawal et al. (2023).

et al., 2011a; Inoue and S.-i. Inutsuka, 2009) similar to those observed in the solar neighbourhood ($M < 10^4$ $M_\odot$), e.g., Taurus (Soler, Zucker et al., 2023). Magnetic fields in the Galactic plane can create low density regions through Parker instability, which are replenished by gas (Mouschovias, Kunz and Christie, 2009). The gas can then undergo a phase shift as a result of thermal instability and form molecular clouds of mass $\sim 10^5 M_\odot$ (Kosiński and Hanasz, 2007). The formation of molecular clouds with $M \sim 10^6 M_\odot$ can be attributed to the merging of smaller clouds (Dobbs, Pringle and Burkert, 2012; Dobbs and Pringle, 2013a; Dobbs, 2008) as well as the magneto-Jeans instability in spiral arms and spurs of galaxies (W.-T. Kim and E. C. Ostriker, 2006). In these crowded regions, frequent interactions and cloud collisions further promote this growth (Casoli and Combes, 1982; Kwan and Valdes, 1983). Massive molecular clouds ($M \sim 10^7$ $M_\odot$) are theorised to form due to gravitational instabilities in the Galactic disk (Y. Li, Mac Low and Klessen, 2005; Tasker and Tan, 2009). These mechanisms act in concert to create various scenarios of cloud formation, causing MCs to have different lifetimes. However, it is difficult to attribute the formation of an individual MC to a particular formation mechanism, as turbulence and gravity balance each other in the ISM without either being dominant (C.-G. Kim and E. C. Ostriker, 2017).

The internal structure of MCs is shaped by a combination of gravity, turbulence, shocks, magnetic fields, heating, cooling, chemistry, and stellar feedback (Chevance, Krumholz et al., 2023). The correlations between cloud properties (e.g. velocity dispersion, surface density, and the virial parameter) have resulted in multiple interpretations of their structure and mechanisms responsible for their evolution. Ballesteros-Paredes, L. W. Hartmann et al. (2011b) argues that clouds are in free-fall collapse; Field, Blackman and E. R. Keto (2011) presents clouds as marginally self-gravitating pressure-confined objects, whereas Larson (1981b) concludes that clouds are self-gravitating objects in virial equilibrium. Other works claim that gravity acting on clouds is counterbalanced by internal forces, giving an average virial parameter around unity (Fukui, Kawamura et al., 2008; Solomon, Rivolo et al., 1987a; Roman-Duval et al., 2010).

Two MC evolution models have been the most predominant over the last decades. These are the gravitational hierarchical collapse (GHC, Vázquez-Semadeni, Palau, Ballesteros-Paredes et al., 2019) and the turbulent support (TS, also known as inertial inflow, Padoan, L. Pan, Juvela et al., 2020) models. The GHC model presents MCs as dynamic entities that are part of continuous gravitationally driven flows, with structures at different spatial scales accreting from their parent structure to grow. This causes a scenario where the small dense substructures collapse later and have shorter lifetimes compared to their parent structures. The turbulent support model is based on the idea of turbulent fragmentation of MCs. It considers MCs as static structures in virial equilibrium, and only a small fraction of the MC mass, typically locked within dense cores, can collapse and form stars. Although the two models are conceptually different, differentiating between them on the basis of observations is extremely difficult (Vázquez-Semadeni, Palau, Gómez et al., 2024).

Observations of GMCs constrain their lifetimes to $\sim 10 - 30$ Myr, which are significantly shorter than their depletion times (Meidt, Hughes et al., 2015; Chevance, Krumholz et al., 2023). Moreover, the low star formation efficiencies of MCs have been a topic of active research for decades. These results suggest that MCs do not end their lives due to the gradual and complete transformation of their gas into stars, but are rather destroyed by other mechanisms. Star formation requires MCs to be gravitationally bound, and the destruction of these star-forming MCs is dominated by stellar feedback (Sect. 2.4). In contrast, marginally bound and unbound clouds are more susceptible to external stresses such as galactic-scale dynamics and turbulent interstellar flows. On galactic scales, differential rotation and shear can stretch and fragment clouds into small elongated structures such as spurs, feathers, and branches (Dobbs and Bonnell, 2006; Dobbs and Pringle, 2013b; Duarte-Cabral and Dobbs, 2017; Schinnerer et al., 2017).

2.2.2 Filaments

Filaments are ubiquitous in the ISM ranging from sub-parsec scales to tens of parsecs (Hacar, Tafalla, Kauffmann and Kovacs, 2013; Hacar, S. E. Clark et al., 2023; Mattern et al., 2018). The filamentary nature of the ISM is evident in both observations (André, Di Francesco, Ward-Thompson, S. .-. Inutsuka et al., 2014; Mattern et al., 2018) and simulations (Padoan, Juvela et al., 2001; Vázquez-Semadeni, Banerjee et al., 2011b; Nakamura and Z.-Y. Li, 2008; Hennebelle, Banerjee et al., 2008; Klessen and Burkert, 2000; Heitsch et al., 2008). Herschel images have revolutionised the study of filaments in the Galaxy, exhibiting their abundance within molecular clouds and providing constraints on the formation and evolution of filaments (Molinari, Swinyard et al., 2010a; Arzoumanian, André, Didelon et al., 2011; Palmeirim, André et al., 2013; Pezzuto et al., 2021). These observations have motivated multiple algorithms to identify filaments from cloud maps, e.g. `DISPERSE` (Sousbie, 2011), `getsf` (Men'shchikov, 2013; Men'shchikov, 2021), `FILFINDER` (Koch and E. W. Rosolowsky, 2015), J-plots (Jaffa et al., 2018; Clarke, Jaffa and Whitworth, 2022), `FIVE` (Hacar, Tafalla, Kauffmann and Kovács, 2013; Hacar and van Dischoeck, 2018), `fil-3D` (D. A. Kim et al., 2023).

Filaments can be grouped together into observational "families", namely giant filaments, Galactic filaments and IRDCs, nearby filaments, hubs, fibres, and striations (Fig. 2.5; Hacar, Tafalla, Kauffmann and Kovacs, 2013). Giant filaments are the largest and most massive structures in the Milky Way present on scales of hundreds of parsecs (Contreras et al., 2013a; Zucker, Battersby and A. Goodman, 2018; Colombo, König et al., 2021). They are extremely elongated and massive (up to 10^6 $M_\odot$) and are associated with large-scale structures in the Galaxy such as the spiral arms (Zucker, Battersby and A. Goodman, 2018). A large number of filaments and infrared dark clouds (IRDCs) have been discovered in Galactic plane surveys such as ATLASGAL (G.-X. Li et al., 2016), HiGal (Molinari, Swinyard et al., 2010b; Schisano, Molinari, Elia et al., 2020) and SEDIGISM (Schuller, Csengeri et al., 2017; Neralwar, Colombo, Duarte-Cabral et al., 2022). These are collectively referred to as Galactic filaments and IRDCs. High resolution observations of diffuse H I emission have also revealed slender and linear filaments, which are termed H I filaments or H I fibres (S. E. Clark, Peek and Putman, 2014).

The filaments in nearby molecular clouds (nearby filaments) are often visible as networks of filamentary structures that, although covering a small fraction of the area, dominate the mass budget of MCs (Arzoumanian, André, Könyves et al., 2019; Chung et al., 2021). Filaments can form complex associations that extend up to several parsecs from a central clump. Such structures are visible both in line and continuum observations and are termed "hubs" or "hub-filament structures (HFS)" (Schneider, Csengeri et al., 2010; Peretto, Fuller et al., 2013; Kumar et al., 2020). On sub-pc scales, dense velocity-coherent filaments identified in PPV space are termed fibres (Hacar, Tafalla, Kauffmann and Kovács, 2013; Eswaraiah et al., 2021), and diffuse elongated structures are termed striations (Chapman et al., 2011; Panopoulou, Psaradaki and Tassis, 2016). They have been observed in line emission maps (e.g. CO, $C^{18}O$, N_2H^+), have transsonic velocity dispersions ($\sigma_{\rm tot} \leq 2 \times c_s$) and low mass ($M \sim 5 - 10\ M_\odot$).

The formation and evolution of filaments are attributed to various galactic phenomena such as supersonic turbulence (Padoan, Å. Nordlund et al., 2007), converging flows, and gravitational collapse and instabilities (Schisano, Rygl et al. 2014; G.-X. Li et al. 2016 & references therein). The classical picture of filament formation has been presented as the formation of sheets due to gravitational instability, which further form filaments (Miyama, Narita and Hayashi, 1987; Van Loo, E. Keto and Q. Zhang, 2014). However, under realistic ISM conditions, several other methods have been proposed to explain filament formation. The turbulent motions of gas in the ISM can create shocks that accumulate gas and compress it to form networks of filaments (Padoan and Å. Nordlund, 1999; Pudritz and Kevlahan, 2013; Federrath et al., 2021). Magnetic fields support filament formation by directing the gas along a flow and providing magnetic tension to stretch the gas, forming long continuous structures (Hennebelle, 2013; Inoue and Fukui, 2013; S. Xu, Ji and Lazarian, 2019; Fukui, Inoue et al., 2021). Stellar feedback mechanisms create filamentary structure into ISM by sweeping materials into shells, stretching gas into streamers, and eroding the edges of structures (Peretto, André et al., 2012; Suri et al., 2019; Pabst et al., 2020a). On large scales, spiral arms and the differential rotation of galaxies stretch MCs into large filamentary structures (R. J. Smith, S. C. O. Glover et al., 2014; Duarte-Cabral and Dobbs, 2017; Körtgen et al., 2018).

Filaments are transient structures that undergo a highly dynamical evolution (C.-Y. Chen, Mundy et al., 2020; Suin et al., 2025). They are coherent structures surrounded by molecular gas from which they accrete. Over time, filaments fragment into smaller structures. Observations of filaments present a complex picture of fragmentation, with the spacing between fragments being hierarchical (Meidt, E. Rosolowsky et al., 2023). The end point of the fragmentation process is the formation of dense

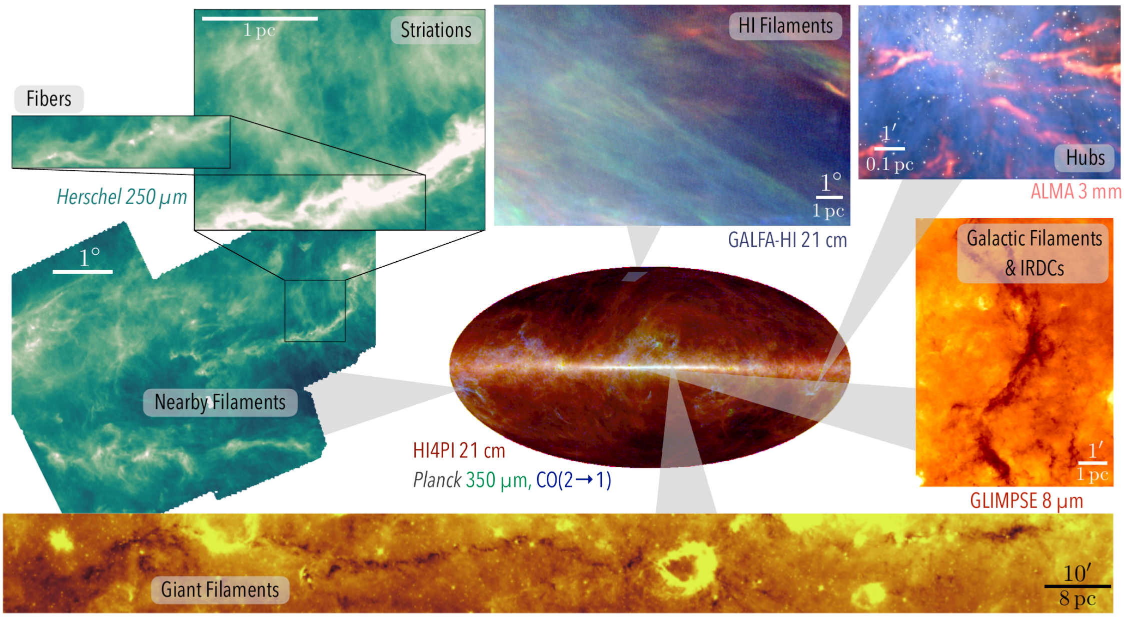


Figure 2.5: Different families of filaments. Clockwise from top: (1) HI filaments in GALFA-HI maps with red, green and blue representing HI brightness temperature in sequential velocity channels with $\Delta v \sim 3\ \mathrm{km s^{-1}}$. (2) Hub-filament structures in OMC-1 observed with ALMA. (3) IRDC from GLIMPSE. (4) The giant filament "Nessie" (5) Herschel 250 µm maps of filaments in Taurus. Taken from Hacar, S. E. Clark et al., 2023.

cores that may form stars.

2.2.3 Bubbles

The molecular gas in clouds is often dispersed due to stellar radiation and winds, resulting in the generation of ring-like objects, called bubbles or shells (Quillen et al., 2005; Nakamura, Miura et al., 2012; Schneider, Simon et al., 2020). Infrared bubbles are often associated with triggered star formation and usually encompass an H II region (Deharveng, Schuller et al., 2010a). Whether the ring-like appearance of bubbles is a consequence of their 3D structure (Pabst et al., 2020b; Churchwell, Povich et al., 2006) or a projection effect is debatable. The expansion of H II regions due to massive stars causes the entrapment of neutral material between the ionised and shock fronts giving rise to dense rings of molecular gas. In addition, stellar winds from massive stars create X-ray-dominated regions, which contributes to bubble expansion and can further destabilise the swept-up material (Zavagno, Deharveng et al., 2006; Deharveng, Zavagno et al., 2009a). These compressed shells are prone to gravitational instabilities. Small-scale instabilities, such as the Jeans instability, favour the formation of low-mass stars, while larger-scale instabilities can eventually give rise to high-mass star formation (Elmegreen and Lada, 1977; Habing, Israel and de Jong, 1972; Deharveng, Schuller et al., 2010b). As a result, the shells host young stellar objects (YSOs) and are good candidates for triggered star formation (Zavagno, Anderson et al., 2010).

The first detection of bubbles came from infrared images by the Infrared Space Observatory and the Midcourse Space Experiment, which culminated in high-resolution surveys of the Milky Way such as

GLIMPSE (Churchwell, Povich et al., 2006), WISE (Anderson, Bania et al., 2014) and Hi-GAL (Carey, 2016; Palmeirim, Zavagno et al., 2017). More recently, JWST has extended such studies to nearby galaxies, enabling high resolution observations of infrared bubbles and feedback-driven structures (Watkins et al., 2023; Schultz et al., 2025). These have been used to understand the mechanisms behind bubble formation, associating them with objects like supernovae, planetary nebulae, open clusters, Wolf-Rayet stars, and OB stars. Churchwell, D. F. Watson et al. (2007) postulates that most bubbles are produced by stars with strong winds, O stars, and B stars. The H ii regions are associated with young OB stars, whereas late-B stars produce small bubbles by exciting PAH bands without forming H ii regions.

Deharveng, Schuller et al. (2010a) classified the bubbles into different types based on their morphology. The "complete bubbles" are young and have a small size compatible with scales at which the ISM is homogeneous. This, along with the dense environment, minimises turbulence that could disrupt their circular structure. "Elongated bubbles" are formed due to differences in densities along the shell. They are well-defined towards the higher density region and begin to open at the lower density region. The last type bubble is "large-sized bubbles" with a partial shell. This is considered a consequence of evolution in low density environments and long age.

One of the most famous Galactic H ii regions is the RCW 120 bubble (Zavagno, Russeil et al., 2010; Anderson, Deharveng et al., 2015; Torii et al., 2015; Walch, Whitworth et al., 2015; Kirsanova, Pavlyuchenkov, Wiebe et al., 2019; Figueira, Zavagno, Deharveng et al., 2017; Figueira, Bronfman et al., 2018). RCW 120 is an ovoid region of 3.8 pc diameter, located 0.5° above the Galactic plane at a distance of 1.34 ± 0.6 kpc (Rodgers, Campbell and Whiteoak, 1960; Russeil, 2003). The region is a hotbed for studying various feedback mechanisms and has been observed in multiple tracers such as 24 µm emission (Zavagno, Pomarès et al., 2007; Deharveng, Zavagno et al., 2009b), dust continuum at 70 – 870 µm (Siringo et al., 2009; Deharveng, Zavagno et al., 2009b; Figueira, Zavagno, Deharveng et al., 2017), 1.3 mm emission Zavagno, Pomarès et al., 2007 and in CO molecular lines (Anderson, Deharveng et al., 2015; Figueira, Zavagno, Bronfman et al., 2020). RCW 120 is surrounded by a dense shell of gas and dust; however, the geometry of the region cannot be described as a simple 2D ring or a 3D bubble (Anderson, Deharveng et al., 2015; Kirsanova, Pavlyuchenkov, A. O. H. Olofsson et al., 2023).

2.2.4 Clumps

Clumps are parsec scale overdensities within molecular clouds and contain a significant fraction of dense star-forming gas (Urquhart, König, Giannetti et al., 2018; Rigby, Moore, Eden et al., 2019; Pazukhin, Zinchenko and Trofimova, 2025). Their inferred densities ($\sim 10^2 - 10^5$ cm^{-3}) and masses ($\sim 0.1 - 1000$ M$_\odot$) vary widely and depend on the choice of the observational tracer (Vázquez-Semadeni, Gómez et al., 2007; Rigby, Moore, Eden et al., 2019; Shetty, Collins et al., 2010; Olmi, Brand and Elia, 2023). They are considered progenitors of star clusters, with their mass function closely resembling the stellar initial mass function (IMF; P. C. Clark, Klessen and Bonnell, 2007).

Molecular line observations of clumps are carried out using various gas tracers such as ^{12}CO, ^{13}CO, CH_3OH, HCO^+, CS, NH_3 and N_2H^+ (discussed in Sect. 4.1.2; Rigby, Moore, Eden et al., 2019; Eden, Moore, Plume, Rigby et al., 2021; Lin, Wyrowski, H. B. Liu, Y. Gong, Sipilä, A. Izquierdo et al., 2024; Olmi, Brand and Elia, 2023; Pazukhin, Zinchenko and Trofimova, 2025). In these studies, clumps are defined as compact density enhancements with size $\sim 0.1 - 1$ pc and densities $\sim 10^4 - 10^5$ cm^{-3} (Olmi, Brand and Elia, 2023). Dust continuum observations reveal clumps as

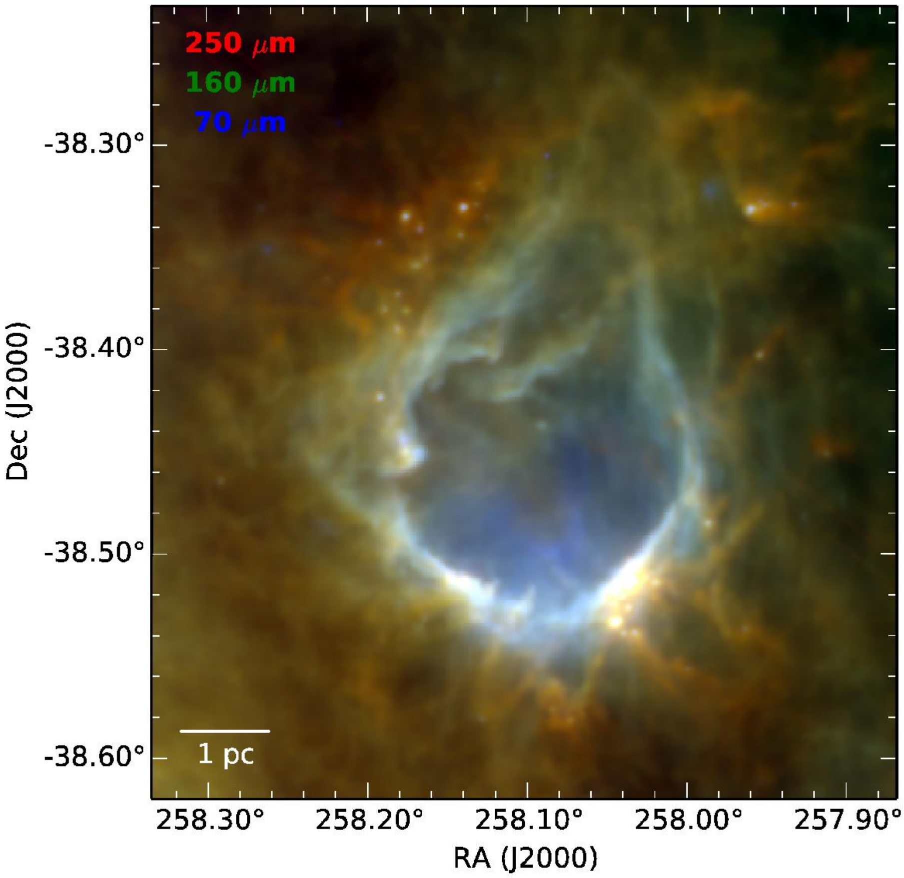


Figure 2.6: RCW 120: Herschel-PACS 70 μm (blue), 160 μm (green) and Herschel-SPIRE 250 μm (red). The field size is 21.8′ × 24.5′, with north as up, and east as left. Taken from Figueira, Zavagno, Deharveng et al., 2017.

localised enhancements in column density, with temperature and density gradients in and around these structures (Contreras et al., 2013b; Urquhart, Csengeri et al., 2014; Traficante, Duarte-Cabral et al., 2018; Zetterlund, Glenn and E. Rosolowsky, 2018). These observations present clumps as structures with typical size ~ 0.5 pc, mass ~ 650 $M_\odot$, and surface density 70 $M_\odot pc^{-2}$ (Table 5 in Urquhart, Wells et al., 2022). Although the sizes of the clumps are fairly robust, their masses, densities, velocity dispersions, and virial parameters vary significantly depending on the tracer's critical density, depletion effects, and optical depth and the galactic environment (Zetterlund, Glenn and E. Rosolowsky, 2018). For example, low-J rotational transitions of ^{13}CO trace a more extended lower-density gas surrounding a clump, while dust continuum traces a more compact, dense, and gravitationally bound clump. The choice of the clump identification algorithm (e.g. `GAUSSCLUMP` Stutzki, 2014; Contreras et al., 2013b, dendrograms E. Rosolowsky and A. Leroy, 2006; Pazukhin, Zinchenko and Trofimova, 2025) also creates disparities in the observed properties. These studies underscore the importance of careful cross-comparison between various observations to define a consistent clump population across star-forming environments.

Recent studies of clumps show that they dynamically decouple from their parent MCs and exhibit different density and velocity dispersion profiles, suggesting that they undergo an evolution independent of the global evolution of the MC (Peretto, Rigby et al., 2023). Clumps can be classified into four categories based on their morphologies and the presence of different emissions (König et al., 2017; Urquhart, Figura et al., 2021). 'Quiescent' clumps are cold (10-15 K) structures with simple velocity structures and absence of embedded star-formation. 'Protostellar' clumps are at the earliest stage of star-formation as denoted by the presence of a warm emission from protostars. The subsequent 'YSO'-dominated stage is characterised by a growing population of infrared-bright sources embedded within the clump. The final 'H II region' stage is the warmest and most clearly observable. The massive stars formed within the clump produce ionising radiation that alters its structure and results in a bubble like morphology. Although clumps at these stages represent different star-formation activities, their properties, e.g. size, mass, and density, do not vary significantly (Traficante, Duarte-Cabral et al., 2018; Urquhart, Figura et al., 2021; Pazukhin, Zinchenko and Trofimova, 2025). This suggests that the physical structures of the clumps do not change significantly during their evolution.

2.2.5 Cores

The most compact and densest regions within molecular clouds are known as cores (Fig. 2.4; Federrath, 2020; Nony et al., 2023; Ishihara et al., 2024; Caselli, Spezzano et al., 2025; S. S. R. Offner, J. Taylor and Grudic, 2025). In Galactic molecular clouds, typical dense cores have sizes ≤ 0.1 pc and number densities $\geq 10^4$ cm^{-3} (Pineda, Arzoumanian et al., 2023). Dense cores are precursors of individual stars and binary star systems and provide important insights about the initial condition of the star-forming gas. Multiple studies have investigated the relationship between dense cores and stars by comparing the core mass function with the IMF (Padoan, B. J. T. Jones and Å. P. Nordlund, 1997; R. J. Smith, P. C. Clark and Bonnell, 2009; S. S. R. Offner, P. C. Clark et al., 2014; Smullen et al., 2020; Pelkonen et al., 2021; Louvet et al., 2024).

Dust continuum surveys have revealed a large number of dense cores within filamentary molecular clouds, with most cores located in supercritical filaments (Ward-Thompson et al., 2016; Bresnahan et al., 2018; Könyves et al., 2020). Molecular line observations provide information on the kinematics and chemical properties of the cores, and help to access their stability using virial analysis (Pineda, A. A. Goodman et al., 2010; Schisano, Molinari, Coletta et al., 2025; Valeille-Manet et al., 2025). Cores are often approximated as Bonnor-Ebert spheres (Ebert, 1955; Bonnor, 1956), with their internal and gravitational potential energy compared using the virial parameter (Bertoldi and McKee, 1992). However, observations indicate that a significant fraction of cores are gravitationally unbound and are instead confined by external pressure (Maruta et al., 2010; Pattle, Ward-Thompson et al., 2015; H. Kirk et al., 2017; H. H.-H. Chen et al., 2019). Such pressure-confined structures may represent an early turbulent phase of cores that collapse over time (S. S. R. Offner, J. Taylor, Markey et al., 2022). Several methods have been developed to identify cores in emission maps, including the curvature threshold extractor (CuTEx; Molinari, Schisano et al., 2011; Coletta et al., 2025), hybrid photometry and extraction routine (Hyper; Traficante, Fuller et al., 2015), cardiff sourcefinding algorithm (CSAR; J. M. Kirk, Ward-Thompson, Palmeirim et al., 2013), `getsf` (Men'shchikov et al., 2010), astrodendro (Kauffmann, Pillai et al., 2010), the minimimum spanning tree method (MST; Schisano, Molinari, Coletta et al., 2025), and the line emission detection and background subtraction (Valeille-Manet et al., 2025). The use of different tracers and identification techniques has led to the use of the term 'core' for structures with a wide range of physical properties.

The formation of cores is mostly associated with the fragmentation of filaments (André, Di

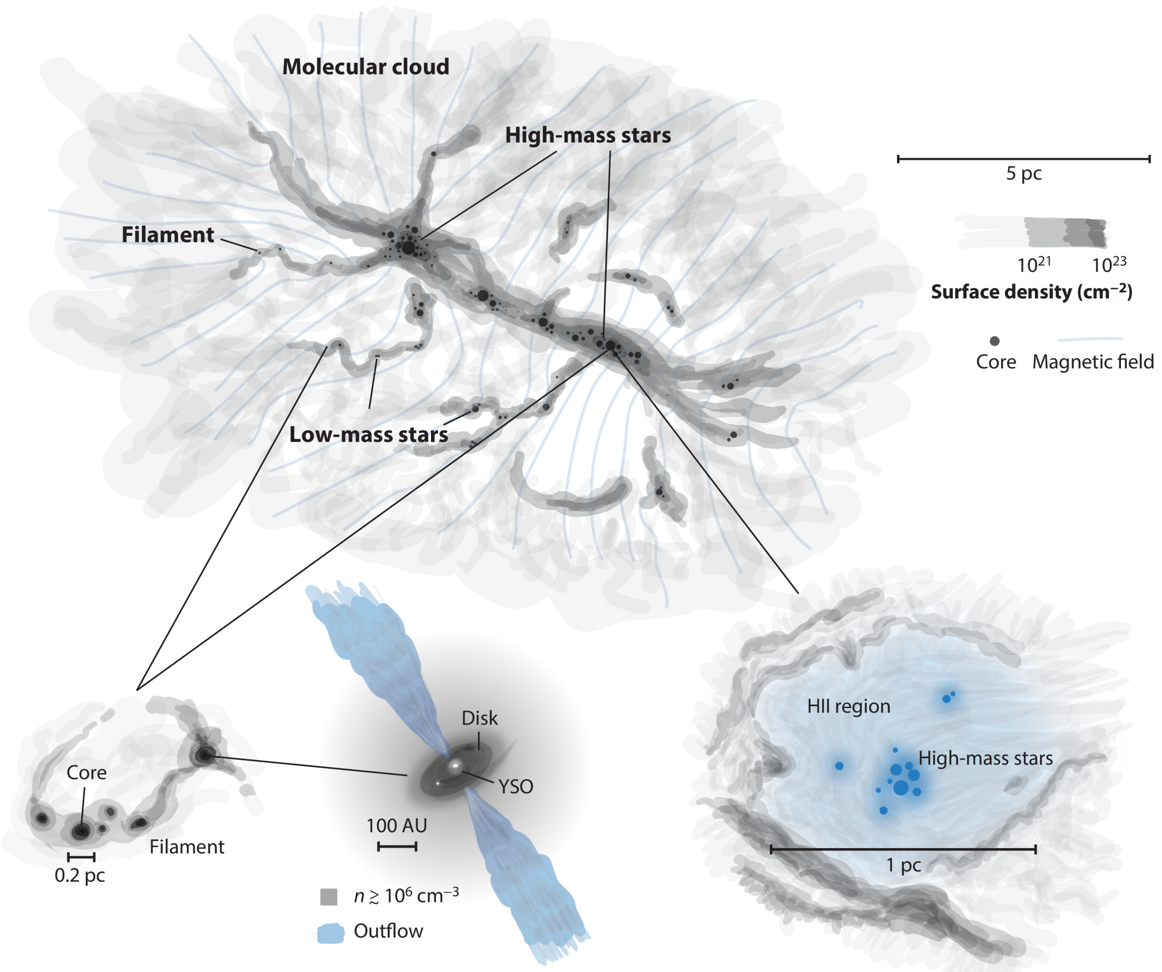
Francesco, Ward-Thompson, S.-I. Inutsuka et al., 2014; Pineda, Arzoumanian et al., 2023). Filament fragmentation can be described in the context of both gravity-dominated and turbulence-dominated models of MCs (Vázquez-Semadeni, Palau, Gómez et al., 2024). In the GHC model, the filaments undergo large-scale collapse and dense cores form within these filaments while continuously accreting mass from their surroundings (Vázquez-Semadeni, González-Samaniego and Colín, 2017; Vázquez-Semadeni, Palau, Ballesteros-Paredes et al., 2019). This is supported by numerical simulations that show the simultaneous formation of filaments and cores (C.-Y. Chen and E. C. Ostriker, 2014; Van Loo, E. Keto and Q. Zhang, 2014; M. Gong and E. C. Ostriker, 2015). Additionally, internal perturbations in filaments can also cause turbulent fragmentation. The observed rotation of cores is consistent with the angular momentum imparted by large-scale cascading turbulent motions. (Padoan, B. J. T. Jones and Å. P. Nordlund, 1997; Padoan, L. Pan, Juvela et al., 2020).

The evolution of cores is highly stochastic, reflecting the nature of turbulent flows in ISM. This stochasticity is further enhanced by the absence of well-defined boundaries between the cores and their surroundings (Smullen et al., 2020; S. S. R. Offner, J. Taylor, Markey et al., 2022; S. S. R. Offner, J. Taylor and Grudic, 2025). Cores are commonly classified as starless, pre-stellar, or protostellar based on their evolutionary stage and the presence of an embedded protostar (André, Men'shchikov et al., 2010; Tobin and Sheehan, 2024). Starless cores do not contain a protostar, whereas protostellar cores host one or more central protostars. Pre-stellar cores form a subset of starless cores that are centrally concentrated, gravitationally unstable, and are expected to collapse to form a protostar. They host an active organic chemistry and are predicted to form as substructures in low density clouds. However, the sensitivity and resolution of most CO surveys prevent them from isolating these low-mass self-gravitating structures at large distances (Chevance, Krumholz et al., 2023; Caselli, Spezzano et al., 2025; Valeille-Manet et al., 2025). As collapse progresses, protostellar cores become more centrally condensed than prestellar cores, reflecting the growth of a central mass concentration (J. M. Kirk, Ward-Thompson and André, 2005). This central object formed during the collapse is a protostar and it evolves over time to form a star.

2.3 Star formation

Star formation has been actively studied since the late 1960s and continues to be one of the widely studied fields in astronomy today (McKee and J. P. Ostriker, 1977; Bergin and Tafalla, 2007; André, Di Francesco, Ward-Thompson, S.-I. Inutsuka et al., 2014; S. S. R. Offner, P. C. Clark et al., 2014; Megeath, Gutermuth and Kounkel, 2022; Beuther, Kuiper and Tafalla, 2025). Significant progress has been made in understanding the physical and chemical processes involved in the formation of stars across a wide range of stellar masses ranging from 0.08 $M_\odot$ to ~ 100 $M_\odot$. The overview of star formation presented here follows the framework of Beuther, Kuiper and Tafalla, 2025. Stars are often classified into two types based on their mass, low-mass stars ($M < 8\ M_\odot$) and high-mass stars ($M > 8\ M_\odot$). The classification relates the Chandrasekhar limit (~ 1.4 $M_\odot$; Chandrasekhar 1931), which sets the maximum stable mass of a white dwarf core. Low-mass stars end their lives as a planetary nebula and leave behind a white dwarf. High mass stars explode as supernovae and result in formation of a neutron star or a black hole. In addition, high mass stars spend a very short time in the pre-main-sequence phase and emit high-energy ionising photons during their lifetime, creating H II regions (Palla and Stahler, 1993; Spitzer, 1998).

Low mass stars form in isolated low mass cores or in clusters distributed in star-forming regions,



Beuther H, et al. 2025
Annu. Rev. Astron. Astrophys. 63:1–44

Figure 2.7: A sketch of star-formation complex encompassing both low and high-mass star formation. The abbreviation YSO refers to young stellar object. The figure is inspired by observational and numerical works by Soler, Beuther et al., 2019; Grudić, Guszejnov, S. S. R. Offner et al., 2022; Traficante, B. M. Jones et al., 2023; Oliva and Kuiper, 2023, and a Hubbe Image of Orion (credit: NASA, C.R. O'Dell and S.K. Wong). The figure was created by André Olivia and is taken from Beuther, Kuiper and Tafalla, 2025.

whereas high mass stars typically form in clusters alongside both low- and high mass stars. The differences between the two regimes (low-mass and high-mass star forming regions) appear as quantitative variations between their physical and chemical properties. The similar accretion timescales of the two regimes (Evans et al., 2009) suggest higher accretion rates in high-mass star-forming regions. This is further evidenced by the increased outflow and infall rates in high-mass clumps and cores compared to their low mass counterparts (Di Francesco et al., 2001; Maud et al., 2015; Wyrowski et al., 2016). The volume and column density threshold for star formation in high-mass regions is almost an order of magnitude higher than that of the low-mass regions (Kauffmann and

Pillai, 2010; Duarte-Cabral, Colombo et al., 2021). In addition, the distance between dense cores in high mass regions (Beuther, Ragan et al., 2015) is smaller than in low mass regions (Tafalla and Hacar, 2015). These results indicate that the formation of high mass stars requires an environment where gas is compressed to high densities, resulting in smaller thermal Jeans lengths (Motte, Bontemps and Louvet, 2018).

High-energy radiation can photoionise the ISM and form the H II regions (Sect. 2.4.3). The high-accretion rates in high mass regions can result in bloating of protostars to delay the formation of H II regions, allowing longer star-formation timescales uninterrupted by ionising radiations (Hosokawa, Yorke and Omukai, 2010; Haemmerlé and Peters, 2016). The ionising radiation emitted by massive protostars and the associated feedback mechanisms is therefore considered a qualitative difference between high-mass and low mass regions. Higher feedback also leads to more fragmentation and high multiplicity around high-mass stars (Mazumdar et al., 2021).

The star-formation process for both the low mass and high mass regions begins with the collapse of molecular clouds under self-gravity, to form the hydrostatic Larson cores (Larson, 1981a; Ahmad et al., 2023). Therefore, the two regimes share similar processes that do not depend on the stellar mass, e.g. disk formation, protostellar jets, molecular outlows are observed at all mass regimes (Maud et al., 2015; Ahmadi et al., 2023). This is further highlighted by the universality of the IMF (Kroupa, 2002; S. S. R. Offner, P. C. Clark et al., 2014; Kroupa et al., 2026). These similarities motivate the need for a unified model of star formation, especially with the latest high-resolution, high-sensitivity telescopes (e.g. JWST) that enable simultaneous observations of low- and high-mass regimes, thus linking star formation across hierarchical gas structures over a wide range of spatial scales.

2.4 Stellar feedback

Stellar feedback refers to the injection of mass, momentum, and energy into the interstellar medium (ISM) by stars (Krumholz, 2014; Gatto et al., 2017; Peters et al., 2017; Kruijssen et al., 2019; Rosen, S. S. R. Offner et al., 2020; Guszejnov, Grudić, S. S. R. Offner et al., 2022; Chevance, Krumholz et al., 2023; Geen, Agrawal et al., 2023). It is one of the key processes that regulates star formation in galaxies by dispersing gas to inhibit star formation and compressing gas to trigger star formation. Stellar feedback mechanisms include protostellar outflows, stellar winds, radiation pressure, photoionisation, and supernova explosions (Krumholz, Bate et al., 2014; Girichidis et al., 2020; Chevance, Krumholz et al., 2023; Geen, Agrawal et al., 2023). The various stellar feedback mechanisms operate on different scales, are linked to different evolutionary stages of stars, and create distinct features in the ISM such as bubbles (Arce and A. A. Goodman, 2001; Arce, Borkin, A. A. Goodman, Pineda and Beaumont, 2011; Hopkins, Quataert and Murray, 2012; Geen, Rosdahl et al., 2015; Grudić, Guszejnov, S. S. R. Offner et al., 2022; Geen, Agrawal et al., 2023). We describe these individual feedback mechanisms below.

2.4.1 Protostellar outflows

The first observed signature of the birth of a new star is often a bipolar outflow of molecular gas (Bally, 2016). Protostars accrete gas from the surrounding ISM and eject a small fraction of it in the form of bipolar outflows to conserve angular momentum (Frank et al., 2014; Nisini et al., 2018). The ejections from protostars appear as (i) highly collimated jets and (ii) wide angle molecular outflows (Arce, Shepherd et al., 2007; Bally, Reipurth and C. J. Davis, 2007; T. Ray et al., 2007). Jets contain both atomic and molecular gas and exhibit velocities on the order of 100 km/s (Tafalla, Santiago et al., 2004; Kristensen et al., 2011; Tychoniec et al., 2019). The origin of jets is attributed to the dynamical interaction of accreted matter with the magnetic field of the star and stellar disk (Frank et al., 2014). These dense jets escape from protostellar envelopes and appear in the ISM as parsec-scale outflows (Karska et al., 2025). Molecular outflows typically appear as slow, wide-angle ejections that carved out expanding cavities in the ISM ($v \sim 1 - 30$ kms^{-1}, Frank et al., 2014). However, they may have an extremely high velocity (EHV) component, which is often associated with jets ($v \geq 30$ kms^{-1}, Hirano et al., 2010; Tychoniec et al., 2019). Most of the measured outflow mass is entrained envelope gas, put in motion by jets and winds (Arce, Shepherd et al., 2007).

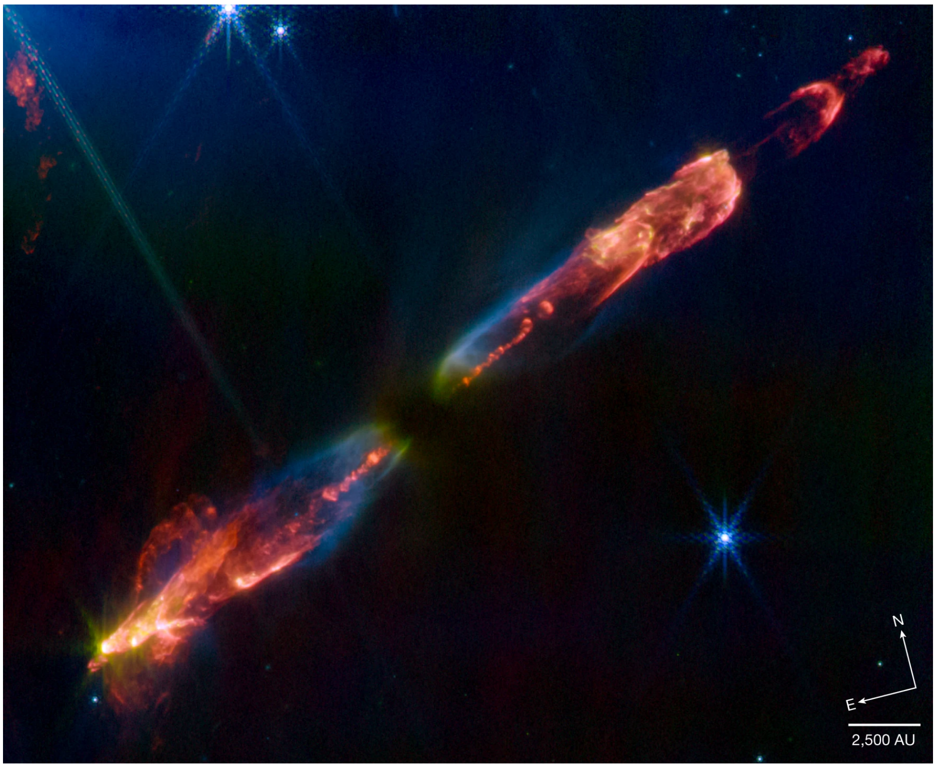


Figure 2.8: A three-colour NIRCam composite of the bipolar HH 211 outflow consisting of images made in the F335M (blue), F460M (green) and F470N (red) filters. Taken from T. P. Ray et al., 2023.

Protostellar jets were first detected in the 1950s with the identification of Herbig-Haro (HH) objects (Herbig, 1951; Haro, 1952). These objects are bright nebulae in dark clouds, and form as a result of high-speed stellar jets interacting with the ISM. In 1970, the first observation of CO revealed gas flowing away from stars, indicating a supersonic molecular gas outflow (Wilson, Jefferts and Penzias, 1970; Kwan and Scoville, 1976). These detections showed that young stellar objects shape the surrounding ISM through energetic, high-velocity mass-loss events. Subsequent observations showed that outflows are common in star forming regions, establishing the role of these bipolar structures as indicators of ongoing star formation activity (Bally, 2016).

Protostars are surrounded by turbulent accretion disks of sizes ranging between 1 and 1000 AU. The origin of the outflows is often attributed to the magnetocentrifugal acceleration of the material in these disks (Fendt and Čemeljić, 2002). There are two prominent models used to describe the launching mechanisms of outflows, namely the D-wind model (Ferreira, 1997; Pudritz, Ouyed et al., 2007) and the X-wind model (Camenzind, 1990; Fendt, 2009; Čemeljić, Shang and Chiang, 2013). In the D-wind model, the collapse, rotation, and accretion of a protostellar core distort the ambient magnetic field, resulting in an hourglass-shaped magnetic field. Gas from the disk surface is centrifugally expelled along the open field lines, with higher-velocity components originating near the protostar and progressively slower components being launched further out. The X-wind model proposes that strong magnetic field from the protostar truncates the accretion disk, creating open field lines along which material is accelerated resulting in fast collimated outflows. Determining which model is better is a subject of active research.

Outflows are an integral component of star formation and significantly impact the physical and chemical properties of the ISM (Frank et al., 2014; Bally, 2016). They inject momentum into collapsing clumps and unbind gas, influencing the stellar IMF, which determines the distribution of stars in galaxies (Hennebelle and Grudić, 2024). The ejected momentum and energy help maintain interstellar turbulence, disperse the surrounding gas, and regulate star formation (Matzner and McKee, 2000; Nakamura and Z.-Y. Li, 2007; Maret et al., 2009; Appel et al., 2025; López-Vázquez et al., 2025). The propagation of outflows through the dense ISM produces shocks that heat and compress the gas, driving chemical processes enriching the ISM (Flower and Pineau Des Forêts, 2010; Burkhardt et al., 2019; Yang, Qiu and X. Pan, 2024). The chemical composition of the ISM is also altered through the molecules expelled from the disk, as well as new molecules that are produced through collisions of outflows with clouds and between multiple outflows. (Bachiller, 1996; Zapata et al., 2018; Toledano-Juárez et al., 2023).

2.4.2 Stellar winds

Stellar winds refer to the ejection of matter from the surface of a star (Martínez-Núñez et al., 2017; Vink, 2024). Observational studies provide strong evidence that stellar winds are present in all stars (N. Smith, 2014; Höfner and H. Olofsson, 2018). Mass loss transports chemically processed material to the stellar surface, leading to different evolutionary scenarios (Langer et al., 1994; Vink, 2011). In massive stars, winds determines the stellar mass before the collapse and, therefore, the type of compact remnant left behind (e.g., black hole). Therefore, stellar winds are a important phenomenon in understanding the evolution of stars.

In low mass stars, the mass loss is driven thermally and magnetically, whereas, in massive stars, radiation pressure drives the winds (Romanova and Owocki, 2015). The mechanisms driving winds are often used to classify them as pressure driven winds, line-driven winds, cool star winds, and continuum driven super-eddington winds (Puls, Herrero and Allende Prieto, 2024).

The study of stellar winds began with observations of the Sun. Parker (1958) theorised that the hot solar corona could not exist for a long time without accelerating some of the gas into interplanetary space, introducing the concept of solar winds. This theory was extended to rotation- and magnetic-field-driven stellar winds by Weber and L. Davis J. (1967) and to radiation pressure driven stellar winds by Lucy and Solomon (1970). In cool main sequence stars, the external forces other than gravity are insignificant, and winds are initiated by high-temperature corona, similar to solar winds. These pressure driven winds result in mass-loss rates $\sim 10^{-14} \mathrm{M}_\odot\ \mathrm{yr}^{-1}$ and terminal velocities $\sim 500\ \mathrm{km\ s}^{-1}$.

UV based observations of massive stars revealed that OB stars possess strong winds, due to their strong P Cygni profiles. In these stars, the mass loss is driven by radiative forces of multiple ionic spectral line transitions. These line-driven winds were first quantitatively described as stationary,

Figure 2.9: Stellar wind bubbles in NGC 346. The pink gas represents energized hydrogen ($\geq 10^4$ K) and the orange gas represents dense, molecular hydrogen (100 K) and associated dust. Image credit: NASA, ESA, CSA, STScI, A. Pagan (STScI)

.

homogeneous, and spherically symmetric outflows (Castor, Abbott and Klein, 1975). Subsequent work showed that winds are subject to strong instability (line-driven instability Owocki, Castor and Rybicki, 1988; Feldmeier, Kudritzki et al., 1997; Feldmeier, Puls and Pauldrach, 1997; Dessart and Owocki, 2005). The high luminosity from massive stars results in line-driven winds with high velocities ($\sim 100 - 3000$ km s^{-1}) and significant mass loss rates ($10^{-8} - 10^{-5}$ M$_\odot$ yr^{-1}) (Puls, Vink and Najarro, 2008; Vink, 2022).

Cool stars such as red supergiants and asymptotic giant branch (AGB) stars have large radii, giant convective cells, and low gravity. They can transport stellar plasma to cool regions ($T \leq 1500$ K) to form dust and expel mass by radiative dust driving (Freytag and Höfner, 2023). Winds are further accelerated by turbulent pressure in red supergiants and strong shocks in AGB stars (Kee et al., 2021). These cool star winds have low terminal velocities (a few tens of km s^{-1}) and high mass loss rates $10^{-7} - 10^{-5}$ M$_\odot$ yr^{-1}. Hot massive stars near or above the Eddington limit can form winds due to Thomson-scattering (Owocki, Gayley and Shaviv, 2004; N. Smith and Owocki, 2006; van Marle, Owocki and Shaviv, 2008). These continuum driven super-eddington winds can achieve mass loss $\geq 10^{-3}$ M$_\odot$ yr^{-1} and velocities $\sim 100 - 1000$ km s^{-1}.

Massive stars with high velocity winds can shock and heat the surrounding gas and create parsec-scale circular cavities or bubbles (Sect. 2.2.3). Winds have a significant impact on parent clouds by injecting significant momentum and driving strong flows into the ISM (Arce, Shepherd et al., 2007; Zinnecker and Yorke, 2007; Arce, Borkin, A. A. Goodman, Pineda and Halle, 2010; Crowther et al., 2016; Pabst et al., 2020a). The energy deposited by stellar winds in bubbles accumulates over time and becomes significant (Weaver et al., 1977; Geen, Agrawal et al., 2023). However, they do not significantly alter the general picture of star formation in the clouds (Guszejnov, Grudić, S. S. R. Offner et al., 2022).

2.4.3 Photoionisation

Photons carry energy and momentum, which are conserved during their interaction with matter. When a photon exceeds the ionisation threshold of an atom (13.6 eV for neutral hydrogen in the ground state), the absorption of the photon can ionise the atom. Ionising radiation from massive stars can propagate and deposit energy on different spatial scales, ranging from the surrounding of the star to the intergalactic medium (Sales et al., 2014). Individual massive stars emit a large amount of Lyman continuum photons, with typical luminosities $Q_H \sim 10^{48} - 10^{50}\ \mathrm{s}^{-1}$, corresponding to the total ionisation energy $\sim 10^{51}$ erg during their lifetimes (N. Smith, 2006). The absorption of radiation heats the surrounding gas to temperature T$\sim$ 7000 – 15000 K, creating hot ionised regions in the cold ISM ($\sim$ 100 K). The resulting pressure contrast between the ionised and neutral gas drives the expansion and acceleration of the surrounding material, representing the process of photoionisation.

Figure 2.10: The H II region Rosette Nebula in narrowband sulfur-hydrogen alpha-oxygen modified Hubble palette with a 384 mm telescope by astrophotographer Stephan Hamel.

UV photons from massive stars ionise the surrounding neutral hydrogen to create hot bubbles called ‘Strömgren spheres’ or ‘H II regions’ (Strömgren, 1939; Khan et al., 2026). Strömgren spheres were first modelled as fully ionised spherical regions of uniform density and consisting purely of hydrogen (Chapter 15 in Draine, 2011b). The radius of this sphere R_{S0} (Eqn. 2.1) corresponds to the distance from the central star at which the ionisation of neutral hydrogen and the recombination of ionised hydrogen are balanced. The Strömgren radius increases with the temperature and luminosity of the ionising source and decreases with the hydrogen density.

$$R_{S0} = \left(\frac{3\,Q_0}{4\pi n_H^2 \alpha_B}\right)^{1/3} , \tag{2.1}$$

where Q_0 is the emission rate of hydrogen ionising photons ($h_v > 13.6$ eV), n_H is the hydrogen density, α_B is the case-B recombination coefficient. The transition from ionised to neutral gas occurs at a smaller distance compared to R_{S0}. Therefore, the idealised discontinuous transition boundary between the two (ionisation front) is a good approximation (Draine, 2011b). Subsequent extensions of the Strömgren model incorporated Helium, introducing nested ionisation zones whose relative sizes depend on the rate of emission of photons.

Photoionisation is mainly driven by the emission from the earliest (hottest) stars (Ramachandran et al., 2018). The overall effect of photoionisation on the ISM strongly depends on the temperature of the parent star and the efficiency with which radiation couples with the gas. In addition, accreting compact objects such as white dwarfs and X-ray binaries contribute to the ionisation budget of the ISM (Schaerer, Fragos and Izotov, 2019; Senchyna et al., 2020; Souropanis et al., 2022). Recent studies suggest that cluster winds and superbubbles could also be potential sources of ionising radiation (Oskinova and Schaerer, 2022).

In Milky Way like galaxies, photoionisation is considered to be the most important feedback mechanism (Chevance, Kruijssen et al., 2022; Chevance, Krumholz et al., 2023). H II regions significantly affect their surroundings through shocks and radiation, and their expansion compresses the surrounding molecular gas, which might trigger the next generation of star formation (Deharveng, Schuller et al., 2010b; Elmegreen, 2011). Photoionisation plays an significant role in dispersing a large fraction of GMC before the onset of SNe and constraining the star formation efficiency (Grudić, Guszejnov, Hopkins et al., 2021). It is also responsible for the phase and thermodynamics of the molecular gas that forms massive stars, which explode as SNe.

2.4.4 Radiation pressure

The interaction of photons with the ISM creates a pressure on the surrounding matter due to the conservation of momentum, which is termed ‘radiation pressure’ (Krumholz, Klein and McKee, 2005; Draine, 2011a). An optically thick gas shell around an ionising source with luminosity L experiences an increase in momentum P as follows (Murray, Quataert and Thompson, 2005).

$$\frac{dP}{dt} = \frac{GM(r)M_g(r)}{r^2} + \frac{L}{c} , \tag{2.2}$$

where M and M_g are the total enclosed mass and the gas mass of the shell, respectively. c is the speed of light. This creates a radially outward pressure on the gas and dust that evolves the shell (J.-G. Kim, W.-T. Kim and E. C. Ostriker, 2016; Rahner et al., 2017).

The radiation pressure from UV and optical radiation from stars is termed direct radiation pressure (Sect. 5.2.2 in Chevance, Krumholz et al., 2023). MCs with surface densities $\sim 10 - 100\ \mathrm{M_\odot\ pc^{-2}}$

are optically thick to this emission and could be dispersed by it. IR emission from dust can couple momentum with clouds that have surface densities $\geq 1000\,\mathrm{M_\odot\ pc^{-2}}$. This process is termed reprocessed radiation pressure (Menon, Federrath and Krumholz, 2023). Regions affected by extreme external pressure such as ultra-luminous infrared galaxies and dwarf starbursts, IR photons can undergo multiple cycles of absorption and emission, continuously excerting radiation pressure (Murray, Quataert and Thompson, 2010).

In dusty galaxies, radiation pressure is an important mechanism in regulating star formation (Fall, Krumholz and Matzner, 2010; Murray, Quataert and Thompson, 2010; Thompson and Krumholz, 2016). It can drive winds with velocities of $\sim 20\ \mathrm{km\ s^{-1}}$ over $1-5$ Myr at typical temperatures and densities of a MC (Murray, Quataert and Thompson, 2005; Wise et al., 2012). By driving the expansion of H II regions, radiation pressure can diffuse out compressed gas, limit the SFE of GMCs, and contribute to shortening their lifetimes (J.-G. Kim, W.-T. Kim and E. C. Ostriker, 2016; Chevance, Kruijssen et al., 2022). However, it does not play a major role in the dispersal of GMCs and cannot prevent the formation of massive stars by accretion (Krumholz, Klein and McKee, 2005; Reissl et al., 2018; Kruijssen et al., 2019; Reissl et al., 2018).

2.4.5 Supernovae

The end of lives of massive stars is in the form of an explosion that expels the stellar interior into their surroundings. These powerful explosions release energies of $\sim 10^{51}$ ergs and mass of $\sim 7-100\ \mathrm{M_\odot}$, and are termed supernova explosions (SNe) (Chevalier, 1977; Soker, 2024; Kundu, 2025). The ejecta from SNe travel at high velocities ($\sim 5\%$ the speed of light) and carve beautiful structures in the ISM (Fig. 2.11). These shells are called supernova remnants (SNR) and might be spherical depending on the homogeneity of the ISM (Dokara et al., 2023). SNR are often classified according to their time evolution into different stages, such as ejecta-dominated, Sedov-Taylor, snowplough, and dispersion (Woltjer, 1972; Yasuda, S.-H. Lee and Maeda, 2021).

The majority of supernovae ($> 80\%$) occur due to the collapse of the core of a high-mass star (Tammann, Loeffler and Schroeder, 1994). These core collapse supernovae occur when a star exhausts its nuclear fuel and the electron degeneracy pressure can no longer oppose its gravitational collapse (Jerkstrand, Milisavljevic and Müller, 2026). Depending on the mass of the collapsing stars, core collapse supernovae are classified into electron capture, iron core gravitational collapse, pair instability, and photodisintegration.

The other 20% of supernovae are the result of thermal runaway of a white dwarf and are termed Type Ia supernova (Turatto, 2003; Parrent, Friesen and Parthasarathy, 2014). Once a degenerate star accumulates mass from a companion star and reaches the Chandrashekhar limit, it explodes with a consistent brightness. These supernovae are therefore referred to as standard candles and are used to measure cosmic distances. The presence and absence of hydrogen lines classify them as Type I and Type II, respectively.

Supernovae are a crucial stellar feedback mechanism that plays a significant role in maintaining the energy balance of the interstellar medium (E. C. Ostriker and Shetty, 2011; Walch and Naab, 2015) The turbulent support theory for MC assumes that supersonic turbulence from SNe is the main driver of MC evolution and fragmentation (Padoan, L. Pan, Juvela et al., 2020). SNe contribute significantly to the cosmic metal budget by injecting metals into the ISM and at the same time destroying interstellar dust (Walch, Girichidis et al., 2015; Collacchioni et al., 2018; Ibrahim and Kobayashi, 2025; Vasiliev, 2025). They also produce cosmic rays through diffusive shock acceleration (Berezhko and Völk,

Figure 2.11: The supernova remnant Cassiopeia A, located at ~ 3 kpc from Earth. The top and right side of the image (orange; warm dust emission) marks where ejected material from the exploded star is ramming into surrounding gas and dust. Inside are bright pink filaments with clumps and knots. Image credits: NASA, ESA, CSA, Danny Milisavljevic (Purdue University), Tea Temim (Princeton University), Ilse De Looze (UGhent); Image Processing: Joseph DePasquale (STScI).

2007). However, SNe occur at a late stage of the GMC lifecycle and a large amount of the energy from SNe energy escapes through low-density channels. Therefore, they do not significantly affect the star-formation history of the natal GMC, but rather influence the large-scale ISM (M. C. Smith et al., 2021; Guszejnov, Grudić, S. S. R. Offner et al., 2022).

2.5 The Milky Way Galaxy

The Milky Way is a barred spiral galaxy with loosely-bound spiral arms, resembling the Sbc type in the Hubble classification (Fig. 2.12; Hubble, 1926; Kennicutt and Evans, 2012; Rezaei Kh. et al., 2024). The spiral arms are rich in molecular gas and stars, but their exact structure and number are still a topic of debate (Gaia Collaboration et al., 2018; Colombo, Duarte-Cabral et al., 2022). The centre of the Galaxy contains a boxy/peanut-shaped bulge that holds 30 - 40% of the Galactic stellar mass (Portail, 2016; Shen and Zheng, 2020). It also hosts a supermassive black hole of mass 4×10^6 $M_\odot$ and Schwarzschild radius of 10^7 km (Samuel, 2020). The Galactic disk contains two vertical components separated by their chemistry and radial structure and are referred to as thin and thick disks (H. .-. Guo et al., 2020). The components of the Milky Way beyond these disks are the stellar halo containing the old metal-poor stars and globular clusters, as well as the dark matter halo, which is used to account for the total mass of the Galaxy.

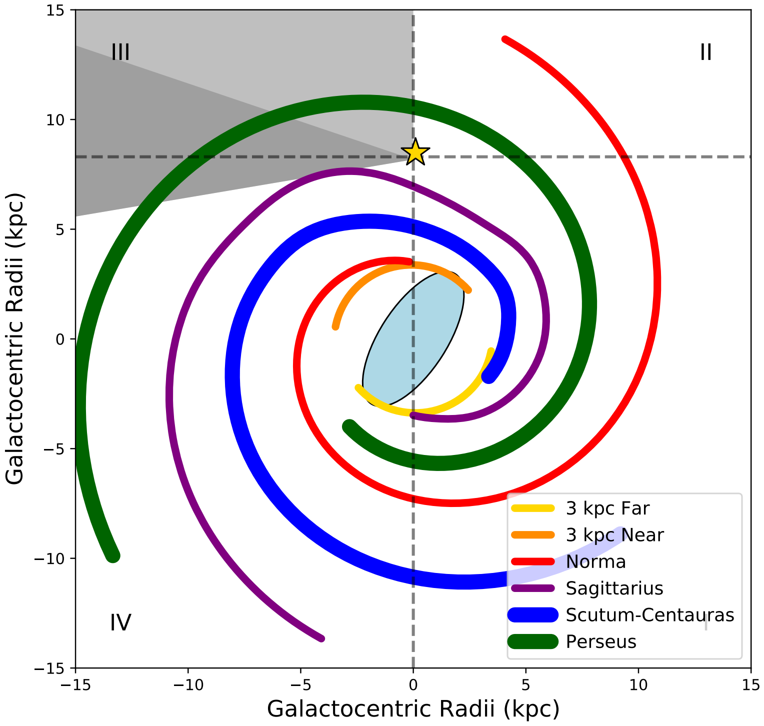


Figure 2.12: Schematic showing the position of spiral arms following model by J. H. Taylor and Cordes (1993) and Cordes (2004). The star shows the position of the Sun and the Roman numerals show the Galactic quadrants. The light blue ellipse at the centre represents the Galactic bar. The grey shaded region shows the coverage of the Outer Galaxy High Resolution Survey (OGHReS). Taken from Urquhart, König, Colombo, Karska, Wyrowski et al. (2024).

The different regions of the Galaxy exhibit distinct environments that are comparable to different types of galaxies. The Galactic Centre with galactocentric distance $R_{gc} < 2$ kpc hosts extreme conditions, including high uv-radiation, cosmic rays, turbulence, and density, making it similar to starburst galaxies (Battersby et al., 2025). The inner disc (2 kpc $< R_{gc} <$ 8 kpc) traces conditions similar to those of typical nearby spiral galaxies (Schuller, Urquhart et al., 2021). The outer Galaxy ($R_{gc} >$ 8 kpc) is characterised by low densities and metallicity, resembling environments of irregular and dwarf galaxies (Urquhart, König, Colombo, Karska, Wyrowski et al., 2024; Urquhart, König, Colombo, Karska, Giannetti et al., 2025). These diverse environments have been observed through multiple Galctic plane surveys in both continuum and line emission. Continuum surveys span infrared (Churchwell, Babler et al., 2009a; Churchwell, Babler et al., 2009b; Molinari, Swinyard et al., 2010b), submillimeter (Schuller, Menten et al., 2009; Eden, Moore, Plume, Urquhart et al., 2017) and radio (Purcell et al., 2013; Beuther, Bihr et al., 2016; Medina et al., 2019) wavelengths, while line-emission surveys observe gas tracers CO, ^{13}CO, C^{18}O (Dame, D. Hartmann and Thaddeus, 2001; Jackson et al., 2006; Rigby, Moore, Plume et al., 2016; Schuller, Csengeri et al., 2017; Umemoto et al., 2017; Benedettini et al., 2020; Schuller, Urquhart et al., 2021; Urquhart, König, Colombo, Karska, Wyrowski et al., 2024).

Our Sun is situated at a distance of around 8 kpc from the Galactic centre in the Orion spur and has a circular velocity of 238 km/s (Sofue, 2020). Like most stars in the Milky Way, the Sun is present in the Galactic disk. This leads to the Milky Way being projected as a thin band in the sky and causes problems in detailed studies of its structure (Fig. 2.13). Apart from stars, there exist other objects orbiting the Galaxy such as dwarf galaxies, which are often associated with the Galactic warp. The two main satellite dwarf galaxies around the Milky Way are the Large Magellanic Cloud and the Small Magellanic Cloud (LMC & SMC; McConnachie, 2012).

Figure 2.13: A wide-field ground-based image of the Milky Way plane towards the Cygnus constellation. Image by ESA/Hubble (A.Fujii).

CHAPTER 3

Data

3.1 The SEDIGISM survey

The Structure, Excitation, and Dynamics of the Inner Galactic InterStellar Medium (SEDIGISM ; see Schuller, Csengeri et al. 2017; Schuller, Urquhart et al. 2021 for an overview) survey spans a range of 84 deg^2 covering $-60^\circ \leq l \leq +18^\circ$ and latitude $|b| \leq 0.5^\circ$ (with some regional variations). It covers the inner Galaxy in multiple molecular tracers, specifically targeting the $J = 2\text{-}1$ transitions of ^{13}CO and C^{18}O. These observations are well suited to trace the moderately dense molecular gas ($\sim 10^3$ cm^{-3}) in the Galaxy. The survey was observed from 2013 to 2017 using the 12-m Atacama Pathfinder EXperiment (APEX) telescope (Güsten et al., 2006). The data products are provided as a contiguous data set divided into 77 data cubes, each covering approximately $2^\circ \times 1^\circ$, with a velocity coverage of -200 to 200 km s^{-1} and a pixel size of $9.5''$. The first data release (DR1) contains ^{13}CO observations with a full width at half maximum beam size of $28''$ and a typical 1-σ sensitivity of 0.8–1.0 K per 0.25 km s^{-1}, providing a detailed view of the ISM on scales ranging from GMCs to star-forming clumps.

SEDIGISM observations were used to build a catalogue of 10,663 MCs with their physical properties Duarte-Cabral, Colombo et al. 2021. This includes the deconvolved equivalent radius (radius_dec_pc[1]), cloud mass (Mass), velocity dispersion (sigv_kms), virial parameter (alpha_vir), and surface density (Surf_density_dec_Mpc2), which were used in this study. MCs were identified using the Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES) algorithm (v.0.3.2, Colombo, E. Rosolowsky, Ginsburg et al. 2015; Colombo, E. Rosolowsky, Duarte-Cabral et al. 2019). Urquhart, Figura et al. (2021) updated the cloud catalogue to include star formation efficiencies and dense gas fractions . Colombo, Duarte-Cabral et al. (2022) studied the positions of these clouds with respect to the spiral arms of the Galaxy and Neralwar, Colombo, Duarte-Cabral et al. (2022) obtained the morphologies of the clouds. Individual catalogues and the final catalogue with updates from different works are publicly available on the SEDIGISM website[2].

---

[1] Column name in the catalogue of Duarte-Cabral, Colombo et al. (2021).

[2] https://sedigism.mpifr-bonn.mpg.de/cgi-bin-seg/SEDIGISM_DATABASE.cgi

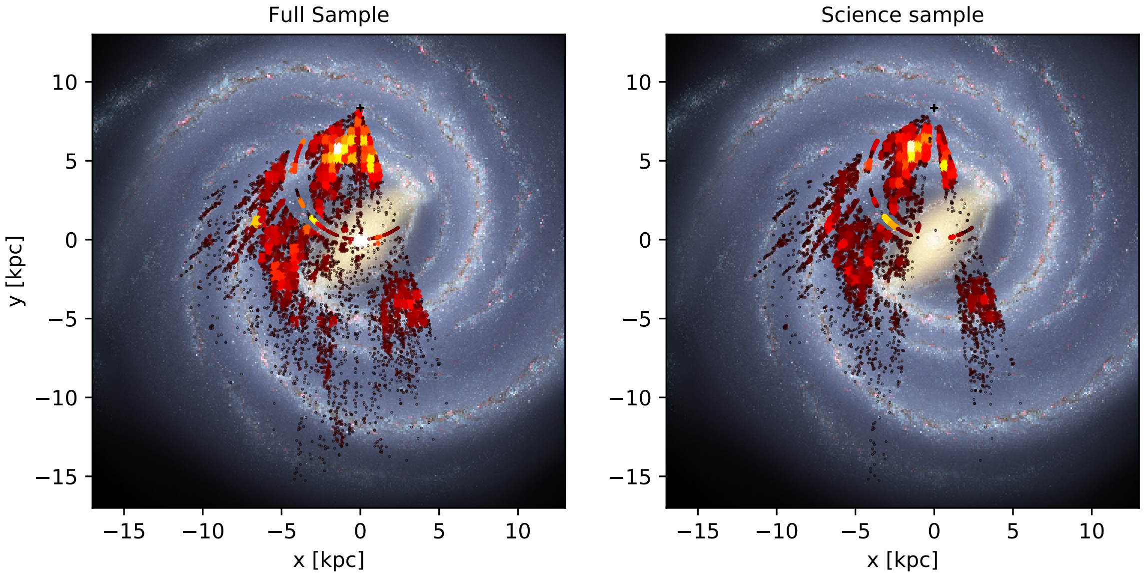


Figure 3.1: Top down view of the Galaxy, with deprojected position of SEDIGISM clouds overplotted on an artistic impression of the Milky Way. The position of the Sun is marked with a ‘+’ in both panels. The left panel shows the density plot of the entire catalogue and the right panel shows the well-resolved clouds with kinematic distances, labelled as the science sample. The color scale and the size of the symbols reflects the local density of clouds, with more crowded areas shown in white and with larger symbols. Adapted from Fig. 10 in Duarte-Cabral, Colombo et al. (2021).

3.2 STARFORGE simulations

Star Formation in Gaseous Environments[3] (STARFORGE) are three-dimensional radiation magnetohydrodynamic (MHD) simulations that follow the formation, evolution, and destruction of GMCs (Grudić, Guszejnov, Hopkins et al., 2021). They combine the physics of MHD, gas self-gravity, stellar dynamics, thermodynamics, and all major protostellar and stellar feedback mechanisms. The thermodynamic treatment includes photoelectric heating, cosmic ray heating, hydrogen photoionisation, dust heating and cooling, line cooling, and collisional excitation of both hydrogen and helium. The simulation framework is built on the GIZMO code (Hopkins, 2015), which uses the meshless finite mass (MFM) method to solve MHD equations (Hopkins and Raives, 2016). MFM is a Lagrangian mesh-free Godunov method that discretises fluid into gas cells of finite mass (Springel, 2010). The interactions between neighbouring cells are defined through effective face areas and the fluxes of conserved quantities (mass, momentum, energy, etc.) are computed by solving Riemann problems using the fluid states reconstructed at the interfaces. A comprehensive explanation of the numerical methods and validation tests is provided in Grudić, Guszejnov, Hopkins et al. (2021).

The simulations scale up to massive GMCs with box sizes of up to $(100\ \mathrm{pc})^3$, while achieving spatial resolutions of a few 10 AU. They also model the formation, evolution, and dynamics of individual stars within a GMC, resolving the stellar initial mass function down to $\sim 0.1\ \mathrm{M}_{\odot}$. As the simulation progresses, the GMC collapses under self-gravity to form protostars with outflows. The

---

[3] `https://www.starforge.space`

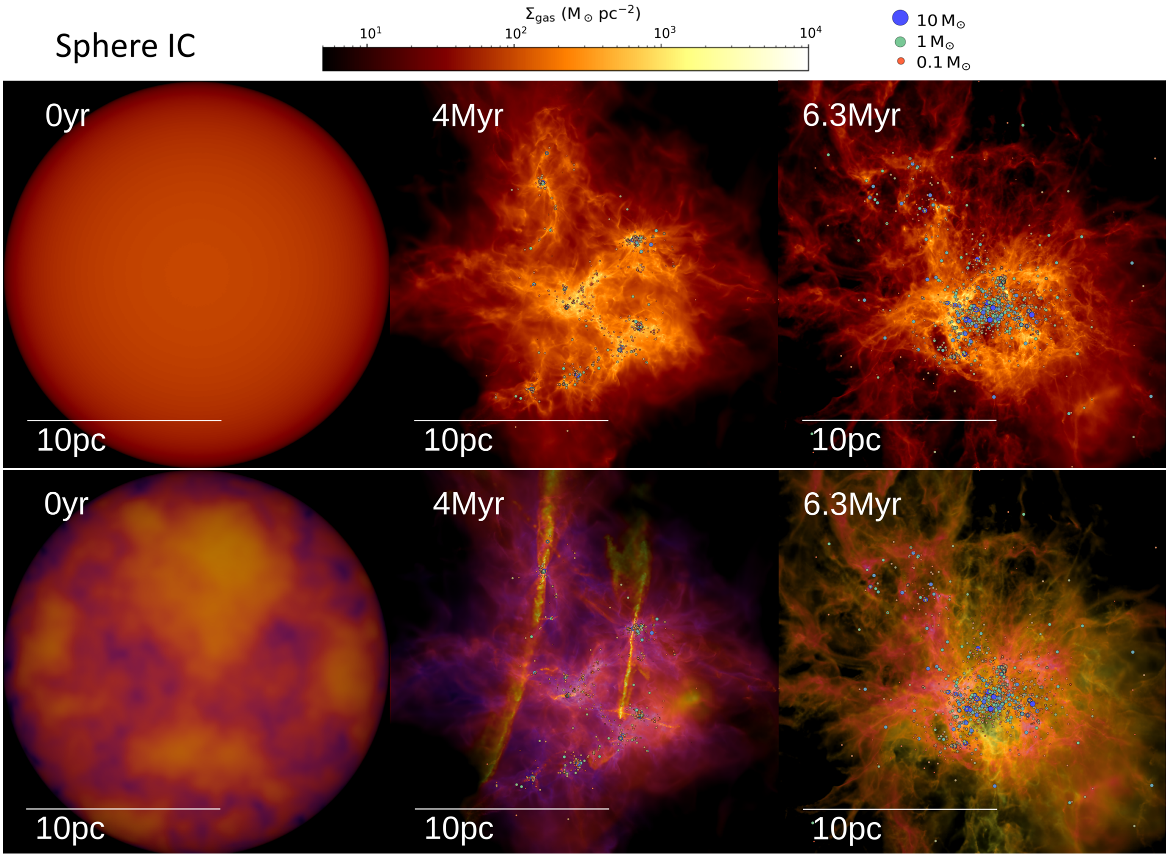


Figure 3.2: Evolution of a GMC in the STARFORGE M2e4a2 simulation. The color scale is logarithmic and the circles represent sink particles (stars) that form in high-density regions. The size and colors of the circles represent the stellar mass. Top: Surface density maps at different times. Bottom: 1D line-of-sight velocity dispersion maps, with surface density encoded as transparency. Taken from Guszejnov, Grudić, Hopkins et al. (2021).

stars evolve to main sequence and eject photoionising radiation and stellar winds. Most of the GMC is dispersed by these feedback mechanisms before the first supernova occurs. This evolution follows a global hierarchical collapse scenario with various stellar feedback mechanisms that can contribute to the injection of local turbulence and provide support against local collapse (Grudić, Guszejnov, S. S. R. Offner et al., 2022; Guszejnov, Grudić, S. S. R. Offner et al., 2022).

3.2.1 Sink particles

Stars are modelled as sink particles in the simulation. A sink particle is formed once it satisfies a list of criteria designed to identify the centres of runaway collapse (see Sect. 2.5.1 in Grudić, Guszejnov, Hopkins et al., 2021). To form a sink particle, the corresponding gas cells should be above the density criteria and be the densest of the neighbouring cells, with the density increasing over time. The cells should be gravitationally unstable and the tidal tensor should be fully compressive. Additionally, the local free-fall time should be shorter than the timescales for approaching a sink particle and the orbital time around it. The evolution of stars follows the one-zone subgrid model that uses the protostellar mass and the accretion rate to evolve the protostellar radius (S. S. R. Offner, Klein et al., 2009). The protostar is updated at each timestep based on its current mass, accretion rate, and evolutionary phase. There are five distinct evolutionary phases in the simulation: pre-collapse, no burning, deuterium burning at fixed core temperature, core burning at variable core temperature, shell deuterium burning, and main sequence. All sink particles are treated as a blackbody to calculate the emitted flux in five wavelength bands. STARFORGE solves the time-dependent frequency-integrated radiation hydrodynamics equations using `GIZMO`'s mesh-free finite-volume solver (Hopkins and Grudić, 2019; Hopkins, Grudić et al., 2020). This enables STARFORGE to follow the emission, transport, and absorption of photons in the different wavelength bands. These are hydrogen ionising ($\lambda < 912\,\text{Å}$), far-UV/photoelectric ($912\,\text{Å} < \lambda < 1550\,\text{Å}$), near-UV ($1550\,\text{Å} < \lambda < 3600\,\text{Å}$), optical/near-IR ($3600\,\text{Å} < \lambda < 3\,\mu\text{m}$), mid- and far-IR ($\lambda > 3\mu\text{m}$) (Sect. 4.5 in Grudić, Guszejnov, Hopkins et al., 2021).

3.2.2 Treatment of feedback mechanisms

Jets, stellar winds, and supernovae are introduced as mass injection events in the simulations (detailed description in Sect. 4 of Grudić, Guszejnov, Hopkins et al., 2021). The sink particles inject mass into the simulation domain through two processes: local injection and cell spawning. Local injection is used for photon injection and to propagate stellar winds when the free-expansion radius is unresolved. In this method, the fluxes of mass, momentum and energy from stars are distributed in a weighted manner to neighbouring cells. Cell-spawning refers to the process of creating new finite-mass gas cells at a certain rate, directly resolving the flow of feedback around the star. This method is used for protostellar jets and for winds when the free-expansion radius can be resolved at the given mass resolution.

The protostellar outflows are modelled in STARFORGE based on the prescription described in Cunningham et al. (2011). This does not favour a particular jet launching mechanism, i.e. X-wind or D-wind model, since these are a subject of active research. The model uses three parameters: (i) the fraction of mass accreted by the disk that is diverted to the jet ($f_w = 30\%$), (ii) the fraction of Keplerian velocity (v_{jet}) at the protostellar radius $R_\star$ (Eqn. 3.1), and (iii) the collimation angle θ_0 (Eqn. 3.2), which sets the angular distribution of the injected jet and the outflow momentum.

$$v_{\text{jet}} = f_{\text{K}} \sqrt{\frac{GM_\star}{R_\star}} \tag{3.1}$$

$$\xi\left(\theta, \theta_0\right) = \left(\ln\left(\frac{2}{\theta_0}\right) \sin^2\theta + \theta_0^2\right)^{-1}, \tag{3.2}$$

where θ is the angle with respect to the angular momentum axis of the sink. The default values are $f_k = f_w = 0.03$ and $\theta_0 = 0.01$ following Cunningham et al. (2011), however the choice of θ_0 does not significantly influence the simulation (Sect. 4.2.1 in Grudić, Guszejnov, Hopkins et al., 2021).

All main-sequence stars with mass $M > 2\,M_\odot$ generate stellar winds. The mass loss rates $\dot{M}$ of stars are modelled on the basis of their main-sequence luminosity L_{MS} and metallicity $Z_\star$ (Eqn. 3.3).

$$\frac{\dot{M}_{\text{wind}}}{M_\odot \text{yr}^{-1}} = \min\left(10^{-15} L_{\text{MS}}^{1.5}, 10^{-22.2} L_{\text{MS}}^{2.9}\right) Z_\star^{0.7} \tag{3.3}$$

The velocity of the winds is calculated following Lamers, Snow and Lindholm (1995) as

$$v_{\text{wind}} = \sqrt{\frac{2GM_\star}{R_\star}} \times \begin{cases} 0.7 & T_{\text{eff}} < 12500\text{ K} \\ 1.3 & 12500\text{ K} < \text{T}_{\text{eff}} < 21000\text{ K} \\ 2.6 & T_{\text{eff}} \geq 21000\text{ K} \end{cases} \tag{3.4}$$

All stars with mass M > 8 $\text{M}_\odot$ are modelled to end as supernovae at the end of their lifetimes. These stars stop all other feedbacks and expel mass with the velocity

$$v_{\text{SN}} = \sqrt{\frac{2E_{\text{SN}}}{M_{\text{ejecta}}}} = 3200\text{ km s}^{-1} \left(\frac{E_{\text{SN}}}{10^{51}\text{erg}}\right)^{1/2} \left(\frac{M_{\text{ejecta}}}{10M_\odot}\right)^{-1/2}, \tag{3.5}$$

where $\text{E}_{\text{SN}} = 10^{51}$ erg. The evolution of these massive stars is modelled beyond the five evolutionary phases, with their lifetime being mass dependent and varying between ~ 10 Gyr for solar-type stars, ~ 40 Myr for 8 $\text{M}_\odot$ stars, and ~ 3 Myr for the most massive stars (> 100 $\text{M}_\odot$).

STARFORGE tracks gas originating from jets (protostellar outflows), stellar winds, and supernovae using tracer fields. Each cell has an array of mass fractions associated with it that correspond to the three feedback mechanisms. These *feedback fractions* represent the mass of the feedback material in a cell relative to the total gas mass of the cell. The different feedback fractions are not directly comparable to each other because they trace only the mass of the gas originating from the feedback. Since they do not include any information about the gas velocity, they fail to completely capture the relative importance of the feedback.

Most of the analysis in this thesis is based on the STARFORGE M2e4a2 simulations (Table 1 of Guszejnov, Grudić, S. S. R. Offner et al. 2022). This particular suite of simulations follows the evolution of a 20 000 $\text{M}_\odot$ GMC to ~ 11 Myr and saves all properties every 24.7 kyr. The GMC is initialised as a uniform surface density sphere with $R = 10$ pc and T = 10 K surrounded by a diffuse medium (density contrast of 1000)[4]. The initial uniform magnetic field is in the $\hat{z}$ direction and corresponds to a mass-to-magnetic flux ratio (μ) of 4.2. The initial turbulence follows a Gaussian random field with k^{-2} power spectrum, resulting in a virial parameter of 2, i.e., a marginally bound cloud.

---

[4] The equations to calculate these properties are described in Guszejnov, Grudić, Hopkins et al. (2021).

CHAPTER 4

Methodology

One of the main goals of this thesis is to study the impact of stellar feedback on molecular gas by connecting simulations and observations. Our methodology is built on two main pillars: radiative transfer and machine learning. First, we describe the theory of radiative transfer and the tool used to generate synthetic data from numerical simulations, i.e. RADMC-3D. These synthetic observations form the primary dataset analysed in Chapters 6 and 7, and allow us to study the evolution of molecular gas structures that closely resemble observations. They also allow us to study the effects of individual stellar feedback processes on molecular clouds. However, our ultimate aim is to characterise these feedback-affected gas structures in observational datasets. We do this in Chapter 8 by training a neural network to identify stellar feedback signatures in molecular gas. The training data set consists of the full synthetic datacubes as inputs and the isolated feedback signatures as outputs (ground truth). The trained model is then used to automatically detect similar structures in the observational data. Since this ML algorithm has a complex structure, we introduce it by first presenting the basic concepts on which it is built. Section 4.2 begins with the introduction of these machine learning concepts that build up to the U-Nets and ResNets architectures that are used to create our algorithm.

4.1 Connecting simulations to synthetic observations

The synthetic observations in this thesis are generated using the RADMC-3D radiative transfer algorithm (version 2.0, Dullemond et al., 2012). The inputs to RADMC-3D are obtained from the meshless STARFORGE snapshot data using a Python-based pipeline[1]. Once synthetic observations are generated, they are postprocessed to mimic the observational characteristics of the SEDIGISM survey. This includes convolution with a telescope beam and addition of noise. Subsequently, molecular clouds are identified using the astrodendro algorithm (E. Rosolowsky and A. Leroy, 2006). The complete data processing and analysis procedure is described in detail in Chapter 7 and Appendix B. In this section, we instead describe the basic working principles of radiative transfer and the RADMC-3D algorithm.

---

[1] `https://github.com/Kartik-Neralwar/gizmo_carver`

4.1.1 Theory of radiative transfer

Most of our knowledge of the Universe is based on radiation that interacts with matter and is detected by telescopes. These interactions cause the absorption, emission, and scattering of radiation, which are studied under the theory of radiative transfer. The strength of the radiation field is expressed using multiple equivalent quantities – specific intensity (I_υ), photon occupation number (n_γ), brightness temperature (T_B), antenna temperature (T_A), and specific energy density (u_υ) – of which the brightness temperature is widely used in submillimeter observations. In this section, we describe the theory of radiative transfer in astrophysics following the descriptions in Rybicki and Lightman, 1979; Draine, 2011b.

Radiative processes arise from the transitions between quantised energy states of atoms, ions, and molecules. In a simplified description of the quantum mechanical theory of multielectron atoms, electrons are treated as occupying single-electron orbitals characterised by discrete quantum numbers. Transitions between different energy states (levels) occur through the absorption and emission of photons, with energies set by the difference between the two levels. A single transition of an electron can be modelled using a two level system, which only allows transitions between the upper (u) and lower(l) energy levels. The populations of two levels can change as a result of the following processes (Mitofsky, 2018):

1. *Spontaneous emission* is the process by which an electron in a higher excitation state transitions to a lower energy state without an external radiation field. This causes the release of a photon with energy equal to the energy difference between the two states. The rate of change of the number density in the two levels is calculated as

$$\frac{dn_l}{dt} = -\frac{dn_u}{dt} = A_{ul}\, n_u \,,$$

where n_u and n_l represent the number densities of the upper and lower levels, respectively. A_{ul} is the Einstein A coefficient that represents the probability of the transition $u \rightarrow l$.

2. *Absorption* refers to the conversion of optical energy to the internal energy of electrons, atoms, and molecules. The absorption of a photon with energy $E > h\upsilon$ causes the electron to move to a higher excitation state, changing its internal momentum and quantum numbers. The rate of change of the number density in the two levels is given by

$$\frac{dn_u}{dt} = -\frac{dn_l}{dt} = B_{lu}\, n_l\, u_{ul}$$

where u_{ul} is the radiation energy density per unit frequency and B_{lu} is the Einstein B coefficient that represents the probability of the transition $l \rightarrow u$.

3. *Stimulated emission* is the process in which an excited electron interacts with a photon and decays to a lower energy level. The energy of the excited electron is converted to the energy of the photon, and the stimulated photon has the same frequency, phase, direction, and polarisation as the incoming photon. The probability of the transition is given by the Einstein coefficient B_{ul}.

$$\frac{dn_l}{dt} = -\frac{dn_u}{dt} = B_{ul}\, n_u\, u_{ul}$$

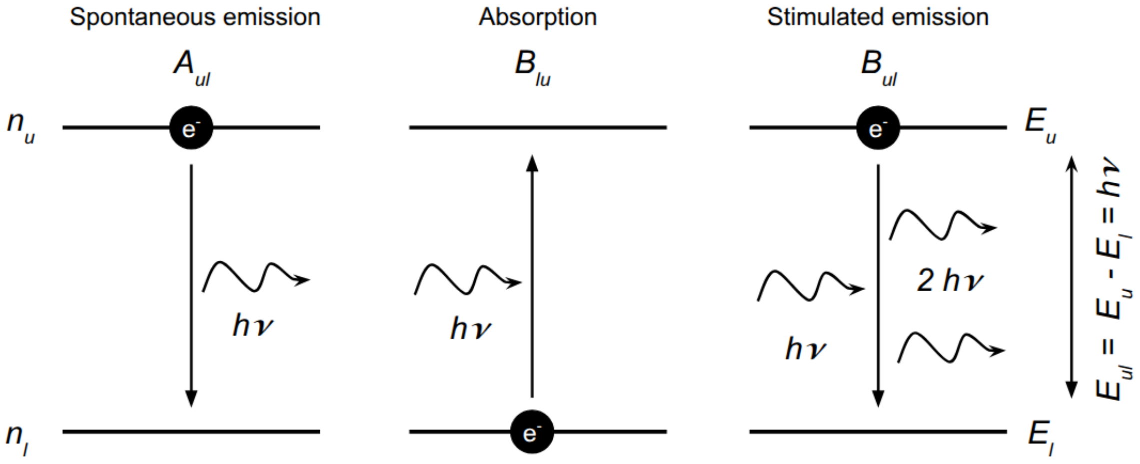


Figure 4.1: Absorption and emission of photons due to transition of electrons between energy levels. The black circles represent the electrons and the curved arrows represent the photons.

The three Einstein coefficients are interdependent, and knowing one of them is sufficient to calculate the other coefficients.

$$g_l\, B_{lu} = g_u\, B_{ul} \qquad ; \qquad B_{ul} = \frac{c^3}{8\pi h \nu_{ul}^3} A_{ul}\ , \tag{4.1}$$

where g_u and g_l are the statistical weights of the upper and lower levels, respectively. c is the speed of light, h is the Planck constant, and υ_{ul} is the transition frequency. The relative populations of the two levels and the statistical weights can be used to calculate the excitation temperature (T_{ex}) as

$$T_{ex} = \frac{E_{ul}/k}{ln(\frac{n_l/g_l}{n_u/g_u})},$$

where $E_{ul} = E_u - E_l$ is the energy difference between the upper and lower levels. Excitation temperature characterises the relative importance of absorption and emission and is frequently used to infer the physical conditions of the gas in ISM.

The combined effects of absorption, stimulated emission, and spontaneous emission can be studied by passing a beam of light through a slab of thickness ds. Neglecting scattering, the propagation of radiation through a medium is governed by the radiative transfer equation.

$$dI_\upsilon = -I_\upsilon\, \kappa_\upsilon\, ds + j_\nu\, ds \tag{4.2}$$

The first term is the net change in intensity (I_υ) due to absorption and stimulated emission, while the second term is the change in intensity due to spontaneous emission. The attenuation coefficient (κ_υ) characterises how easily a beam of light can penetrate a material, and the emissivity (j_ν) is the power radiated per unit frequency per unit solid angle per unit volume. Equation 4.2 can be integrated to obtain the radiation intensity that escapes the slab as

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + \int_0^{\tau_\nu} e^{-(\tau_\nu-\tau')}\frac{j_\nu}{\kappa_\nu}\,d\tau' \tag{4.3}$$

, where τ_ν is the optical depth[2] and $I_\nu(0)$ represents the intensity value at $\tau_\nu = 0$. The entire derivation is described in Sect. 7.2 – 7.4 in Draine, 2011b.

Observations of molecular clouds include spectral line emission and dust continuum emission (Sect. 2.2). Modelling these two types of observations requires different approximations of the radiative transfer equation. Line radiative transfer involves calculating the emerging specific intensity of a emission line of a particular molecule using Eqn. 4.3 by setting the incident intensity $I_\nu(0)$ to zero. The absorption and emission coefficients are calculated from the Einstein coefficients, the density and level population of the given molecule. The Einstein A coefficients, collisional rates and statistical weights for multiple species have been computed and stored in the atomic and molecular line databases such as the LAMBDA[3] and the CDMS[4]. These can be used to solve the radiative transfer equation under the Local Thermodynamic Equilibrium (LTE) approximation, which assumes that energy levels of molecules and atoms are populated according to a thermal distribution. However, the proper calculations require consideration of collisions and scattering, and are solved under non-LTE conditions, discussed in detail in Sect. 4.1.3. The photons from stars heat dust grains that absorb radiation and reemit in the far-infrared range. The continuum emission is therefore dominated by the absorption and thermal emission of radiation by dust grains, producing a smooth spectrum that traces the column density and temperature of clouds. Therefore, the emissivity and the attenuation coefficient do not depend on the Einstein coefficients but on the temperature and opacity of the dust grains.

4.1.2 Observations of molecular gas

The interstellar gas is mainly composed of hydrogen and helium with trace amounts of heavier elements (metals) formed via primordial and stellar nucleosynthesis. Molecules in the ISM form due to chemical transitions in photodissociation regions (PDRs), with H_2 molecules forming on the surface of dust grains (Hollenbach and Salpeter, 1971) and aiding in the formation of CO. This gives rise to a layered structure, where the outer parts consist of atomic ISM (HI and CI) and the inner parts contain molecular gas (H_2 and CO) (M.-Y. Lee et al., 2012). Far ultraviolet radiation dissociates the molecules into their atomic counterparts, which is prevented in clouds because of self-shielding by the outer layers. The shielding is further aided by interstellar dust (Solomon and Wickramasinghe, 1969) which aids in star formation, leading to a correlation between the H_2 fraction and star formation in galaxies (Krumholz, A. K. Leroy and McKee, 2011; S. C. O. Glover and P. C. Clark, 2012a).

Molecular hydrogen (H_2) exhibits electronic, rotational, and vibrational transitions; however, direct observations of H_2 are difficult (Habart et al., 2005). The electronic transitions occur in UV, which are not observable from the ground. The vibrational and ro-vibrational transitions are faint due to their quadrupolar origin, and can only be observed with high-sensitivity telescopes such as JWST. The rotational transitions have the lowest energies; however, their observations are difficult because of large spacing in their rotational energy levels and the lack of a permanent dipole moment. The lowest

---

[2] Optical depth is a measure of the absorption of radiation as it travels through a medium.

[3] Leiden Atomic and Molecular Database `https://home.strw.leidenuniv.nl/~moldata/`

[4] The Cologne Database for Molecular Spectroscopy `https://cdms.astro.uni-koeln.de/cdms/portal/`

possible rotational transition of H_2, J= 2 → 0, has an energy difference of 510 K, which cannot trace the cold molecular gas ($\leq$ 100 K) associated with star formation (Draine, 2011b). In addition, the transition has a relatively low Einstein A-value of 3×10^{-11} s^{-1}, and produces weak emission lines. Therefore, molecular observations are often performed using the second most abundant molecule, i.e. carbon monoxide (Bolatto, M. Wolfire and A. K. Leroy, 2013a; Groves and Schinnerer, 2015). CO has a weak permanent dipole moment ($\mu \sim 0.11$ D), is easily excited (J=1-0 at T ~ 5.5 K) and has strong emission in radio wavelengths. It is therefore a common tracer of molecular gas in galaxies (M. Heyer and Dame, 2015; Colombo, E. Rosolowsky, Duarte-Cabral et al., 2019; Duarte-Cabral, Colombo et al., 2021). The isotopes with low energy difference between the excitation levels, ^{12}CO, ^{13}CO, and $C^{18}O$, are the most common amongst these. These are ideal for observations of the cold molecular gas at millimetre and submillimeter wavelengths. However, in low density and low metallicity environments, CO often cannot trace the entire H_2 reservoir. This CO-dark gas is studied using other tracers such as C I and C II (Papadopoulos et al., 2008; S. C. O. Glover and P. C. Clark, 2012b; Madden et al., 2020).

The translation of CO intensity to the total H_2 gas mass is critical to understanding star formation. The conversion is carried out using the CO-to-H conversion factor H_2, which often appears as X_{CO} or α_{CO}. X_{CO} correlates the observed CO integrated intensity W(CO) and the column density of molecular gas $N(H_2)$ as

$$N(H_2) = X_{CO}\ W(CO),$$

while α_{CO} relates the CO luminosity (L_{CO}) to the molecular gas mass (M_{mol}) as

$$M_{mol} = \alpha_{CO}\ L_{CO}.$$

The typical value of X_{CO} in Milky Way is assumed to be 2×10^{20} cm^{-2} (K km s^{-1}), which corresponds to $\alpha_{CO} = 4.3\ M_{\odot}\ (K\ km\ s^{-1}\ pc^{-2})^{-1}$ (Bolatto, M. Wolfire and A. K. Leroy, 2013b). It is mostly consistent on cloud scales (Madden et al., 2020; Seifried et al., 2020; Hu et al., 2022), but might vary for individual environments (Gorski and Murchikova, 2025). In addition, isotopic abundance ratios such as $^{12}CO/^{13}CO$ vary across Galactic environments due to differences in chemical enrichment, fractionation, and selective photoionisation (Jacob, Menten et al., 2020), with a typical value of 60 (Schuller, Csengeri et al., 2017).

4.1.3 RADMC-3D

RADMC-3D is a versatile radiative transfer algorithm that can perform spectral line transfer and dust continuum radiative transfer to provide images and spectra of a simulated environment. The algorithm does not have any physical data implemented and is just a computational engine completely dependent on the input files provided to it. The main task of RADMC-3D is to predict how an image or spectra of a cloud appears for a given dust, gas, and stellar distribution in a 3D space. The user specifies the properties of the simulated environment, including the density, temperature, and composition of the gas, as well as the source of radiation. Additionally, the size and resolution of the grid, the wavelength of radiation, and the level of radiation intensity can be set to control the simulation parameters. The code then generates a three-dimensional grid of points that are used to sample the environment and calculate the radiative transfer. It can model various radiative processes such as dust thermal emission and absorption, dust scattering, and atomic and molecular line transitions.

RADMC-3D is capable of modelling radiative transfer in molecular and atomic lines by obtaining the level population in each grid cell. A detailed description of the line radiative transfer is provided in the algorithm's manual[5]. RADMC-3D performs line radiative transfer by solving Eqn. 4.2. It does so by assuming a co-moving line profile:

$$\tilde{\varphi}_{ij}(\nu) = \frac{c}{a_{\mathrm{tot}}\nu_{ij}\sqrt{\pi}} \exp\left(-\frac{c^2(\nu-\nu_{ij})^2}{a_{\mathrm{tot}}^2\nu_{ij}^2}\right), \tag{4.4}$$

where ν_{ij} is the central frequency of the line and a_{tot} is the linewidth in units of cm/s. The default setting in RADMC-3D is to assume the line width as the thermal broadening of the line

$$a_{\mathrm{tot}} = a_{\mathrm{therm}} = \sqrt{\frac{2kT_{\mathrm{gas}}}{m_{\mathrm{mol}}}},$$

where T_{gas} is the temperature of the gas, μ is the molecular weight and k is the Boltzmann constant. However, there is an option to specify the local turbulent width using the `microturbulence.inp` file to calculate the linewidth as

$$a_{\mathrm{tot}} = \sqrt{a_{\mathrm{turb}}^2 + a_{\mathrm{therm}}^2} = \sqrt{a_{\mathrm{turb}}^2 + \frac{2kT_{\mathrm{gas}}}{m_{\mathrm{mol}}}}. \tag{4.5}$$

This allows the code to include unresolved non-thermal broadening self-consistently, improving the accuracy of line opacities, excitation, and emergent line profiles without explicitly resolving small-scale turbulence. We calculated the microturbulence as the product of the cell size and its velocity gradient. The velocity gradients for the cells are obtained using the numerical differentiation functionality of the `MESHOID`[6] script.

The co-moving linewidth (Eqn. 4.5), being smaller than the typical velocity field (Doppler shift), can introduce numerical artefacts in the PPV cube. If some cells along the line of sight are red shifted while the neighbouring cells are blue shifted, the doppler shift between the two should be smooth. However, if the co-moving linewidths are small, neither of these cells contributes to the emission. To prevent these *Doppler jumps*, RADMC-3D includes the *Doppler catching*[7] method (detailed discussion in Pontoppidan et al., 2009). This adaptively divides ray segments whenever a large Doppler jump is detected, ensuring that the line profile is sampled finely enough to properly capture the line through the wavelengths of interest.

The default line transfer model in RADMC-3D assumes LTE, which is valid in dense regions such as planetary atmospheres, but not for MCs that span a wide range of densities. The level populations for MCs must be computed to be consistent with the local density, temperature, and radiation field. Solving the level populations self-consistently is known as non-LTE radiative transfer, but in most cases it is either too computationally demanding or impractical, and a simplified approximation of non-LTE effects suffices. One of the approximations implemented in RADMC-3D for non-LTE radiative transfer is the Large Velocity Gradient (LVG) method or the Sobelov approximation (Ossenkopf, 1997; Shetty, S. C. Glover et al., 2011). This method assumes the existence of large velocity gradients

---

[5] https://github.com/dullemond/radmc3d-2.0/blob/96d52c22c60325c0f867d107334e63f40d46672d/manual/sphinx/lineradtrans.rst

[6] https://github.com/mikegrudic/meshoid

[7] https://github.com/dullemond/radmc3d-2.0/blob/96d52c22c60325c0f867d107334e63f40d46672d/manual/sphinx/lineradtrans.rst#preventing-doppler-jumps-the-doppler-catching-method

across a region, and the radiation interaction length is set by these gradients rather than the physical size of the cloud. It is widely used to mimic turbulent motions with moderate velocity gradients. RADMC-3D combines the LVG method with the escape probability (Esc) method. This method estimates the probability that a photon emitted in a given transition escapes the local region without being absorbed or scattered. The probability depends on the optical depth and geometry of the region. All RADMC-3D calculations in this work are performed using the non-LTE approximation together with the Large Velocity Gradient mode and the Escape Probability method. It allows the source function for each emission line to be computed self-consistently under non-LTE conditions without explicitly solving the full, non-local radiative transfer problem.

The global configuration of radiative transfer calculations is controlled using the `radmc3d.inp` file, which contains the keywords relevant to molecular line imaging. We used `nphot` = 1000 photons for the Monte Carlo simulation and disabled dust scattering (`scattering_mode_max` = 0) to focus on the ^{13}CO emission. `lines_mode` = 3 enables the non-LTE LVG+Esc mode and `catch_doppler_resolution` = 0.05 ensures that the line is integrated with small enough steps so that the Doppler shift is at most 5% the local line width. Enabling the Doppler shift also enables the second order integration, providing high-resolution images with minimal pixelation.

The `line.inp` file is the master file that specifies which atoms and molecules are to be modelled and the collisional partners. We modelled the ^{13}CO molecule using the molecular data from the LAMBDA database, with H_2 as the collision partner. The molecular data file (`molecule_13CO.inp`) contains energy levels, statistical weights, Einstein A coefficients, and collision rate coefficients for ^{13}CO. The spatial distribution of H_2 and ^{13}CO is defined by `numberdens_h2.inp` and `numberdens_13co.inp`, which present the local number density of the two species.

The `dustopac.inp` file specifies general information about the various species of dust in the model. We used a single population of thermal silicate grains as dust and specified the opacities in `dustkappa_silicate.inp`. Dust density and dust temperature are calculated from the simulation using `gizmo_carver` to allow for dust absorption and thermal emission consistently. The ^{13}CO(2-1) transition emits at 220.398 GHz or 1360.227 μm. We used this as the central wavelength to obtain PPV cubes with 65 channels separated by 0.25 kms^{-1}, leading to a velocity coverage of ± 8 kms^{-1}. This is specified in the `camera_wavelength` file.

The synthetic observations were generated using the command

```
radmc3d image loadlambda incl 0 phi 0 npix 448 sizepc 60 dpc 3000 useapert.
```

This creates PPV cubes of size 65 × 448 × 448 placing them at a distance of 3 kpc, replicating the SEDIGISM pixel size of 9.5″. The ^{13}CO(2-1) cubes are constructed with a line of sight pointed toward the x, y, and z directions, using the phi={90, 0, 0} and incl={90, 90, 0} parameters for projections along the x, y and z axes, respectively. The `useapert` option is used to set the aperture size to the APEX beam size (28″), ensuring consistency with SEDIGISM.

4.2 Machine learning

Machine learning (ML) is the science of programming computers so that they can learn from data (Goodfellow, Bengio and Courville, 2016; Geron, 2019). Data refers to anything that can be quantified, be it simple numbers or complex things such as images, which can be converted to tensors. Traditional computer algorithms follow a set of logical structures and predefined rules to convert the input to the output. ML algorithms, on the other hand, analyse data to learn patterns, structures, and the mapping between input and output. Their goal is to discover meaningful information from the data and the decision on what is meaningful lies with the user. Therefore, they excel in predictive analysis, from simple tasks, such as filtering spam emails, to complex ones, such as large language models.

Astronomical datasets are experiencing rapid growth in both size and complexity. Due to this accelerated expansion, machine learning algorithms have become a tool of choice for the astronomical community, being used across a wide range of spatial scales and wavelengths. ML algorithms have been used for the analysis of gravitational lenses (Hezaveh, Perreault Levasseur and Marshall, 2017), estimation of Galaxy cluster mass accretion rates (Ntampaka et al., 2019; Soltis et al., 2025), inference of the dense-matter equation of state of neutron stars from their observable properties (Fujimoto, Fukushima and Murase, 2021), identification of feedback driven bubbles (Beaumont, Williams and A. A. Goodman, 2011; Beaumont, A. A. Goodman et al., 2014; D. Xu, S. S. R. Offner et al., 2020b). Additionally, various frameworks have been designed for testing and benchmarking machine learning methods on astronomical data, e.g. the hello universe project[8].

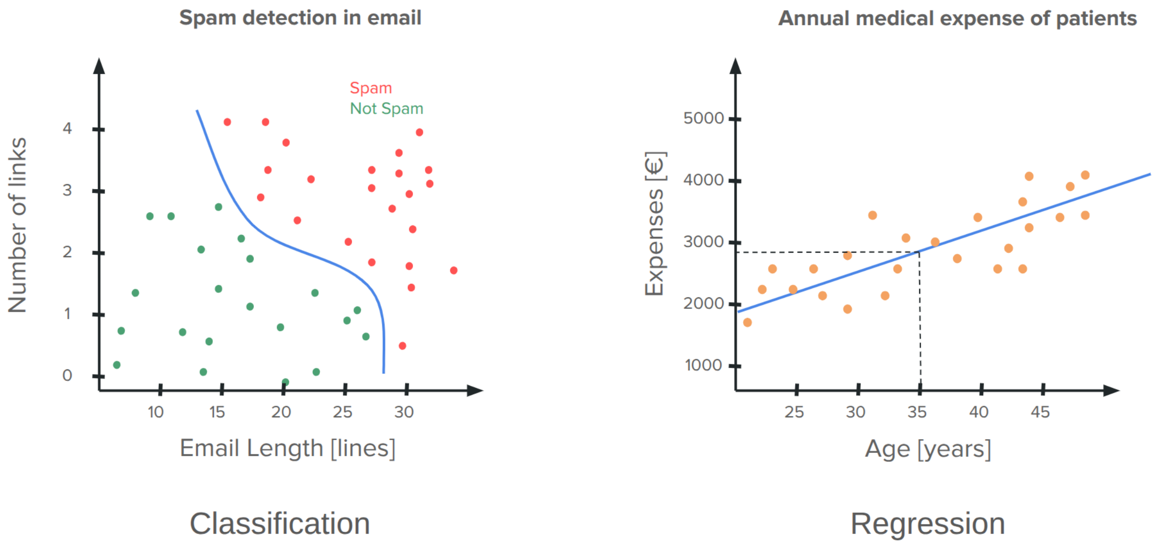


Figure 4.2: Applications of supervised learning. The input data and the output (labels) are known. Left (classification): The algorithm separates an email into two classes: spam or not spam. Right (regression): The algorithm estimates an expense based on the age of the patient.

ML algorithms are often classified according to how they are trained, into *unsupervised* learning and *supervised* learning (Geron, 2019; Neupert et al., 2021). Supervised learning is an approach defined by the use of labelled data. The machine is provided with pairs of values, i.e. an item drawn from a dataset along with the label assigned to it. It then iteratively predicts[9] the label for the training data and is corrected by the corresponding true labels. The two main tasks solved under supervised learning are classification and regression (Fig. 4.2), which differ by the target labels being discrete and continuous values, respectively.

---

[8] `https://archive.stsci.edu/hello-universe`

[9] Prediction is a commonly used terminology in machine learning, which refers to the output of the algorithm given a particular input and a model.

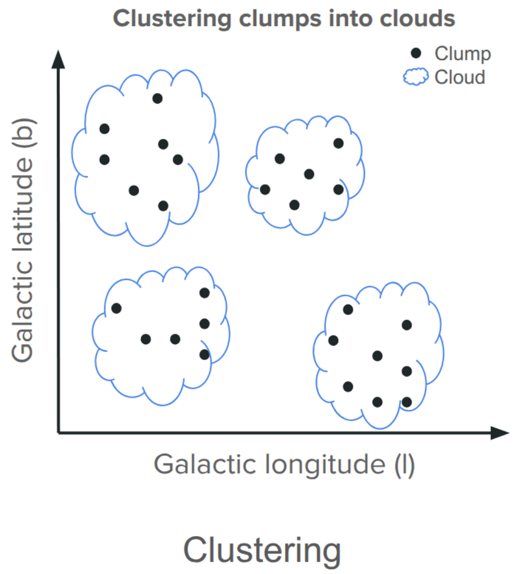


Figure 4.3: Applications of unsupervised learning. The algorithm detects patterns based on the unlabelled input data and clusters clumps that have similar properties, and likely belong to the same cloud.

Unsupervised learning consists of data without labels. These algorithms learn the relationships between the input elements rather than between the input and a label. They can extract new knowledge from existing datasets and are therefore important for new scientific discoveries. Unsupervised algorithms are frequently used to cluster data based on their similarity (Fig. 4.3). They are also used to reduce the dimensionality of data by identifying low-dimensional structures that capture dominant patterns.

4.2.1 Deep learning

Deep learning refers to ML algorithms that are inspired by the structure and function of neural networks in the brain (Goodfellow, Bengio and Courville, 2016). The models used in deep learning are referred to as artificial neural networks (ANNs). ANNs are computing systems designed to mimic the way human brains process information. These networks rely on a set of interconnected elements known as artificial neurons. The connections between neurons transmit signals, which are processed by the receiving neuron and forwarded to the subsequent neuron. ANNs consist of neurons organised into layers, namely the input layer, the hidden layers, and the output layer. The different layers perform different transformations on their input, with the computation handled in the hidden layers. The number of hidden layers determines the classification of an algorithm into shallow learning (single hidden layer) and deep learning (multiple hidden layers).

Neural networks are often presented in the popular press as a synonym for an electronic brain capable of consciousness, general intelligence, and emotions. However, neural networks are mathematical models that can be understood in the simplest sense as solving a problem of parametric regression. This interpretation can be understood using the concept of perceptron, which is a simplified mathematical model of an artificial neuron (Rosenblatt et al., 1962). A perceptron computes the weighted sum of its inputs, adds a bias, and applies a nonlinear function to produce an output. In case of a single input,

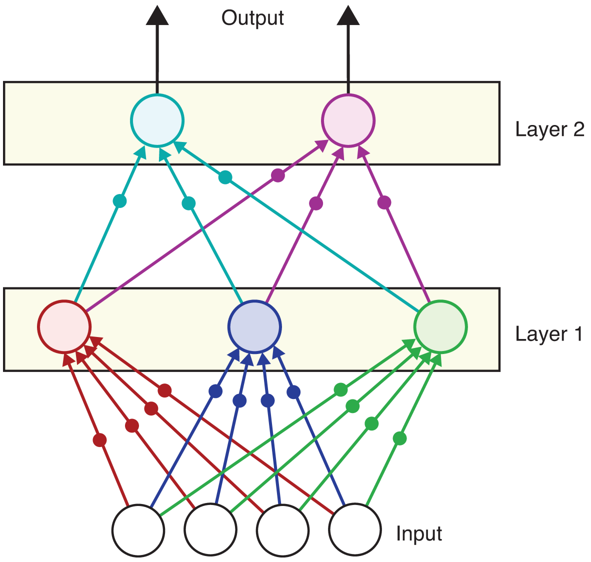


Figure 4.4: A simple deep neural network. Taken from Glassner, 2021

.

the perceptron reduces to a linear model (Fig. 4.5). The weight determines the slope of a line fitted to a set of data points and the bias corresponds to the intercept.

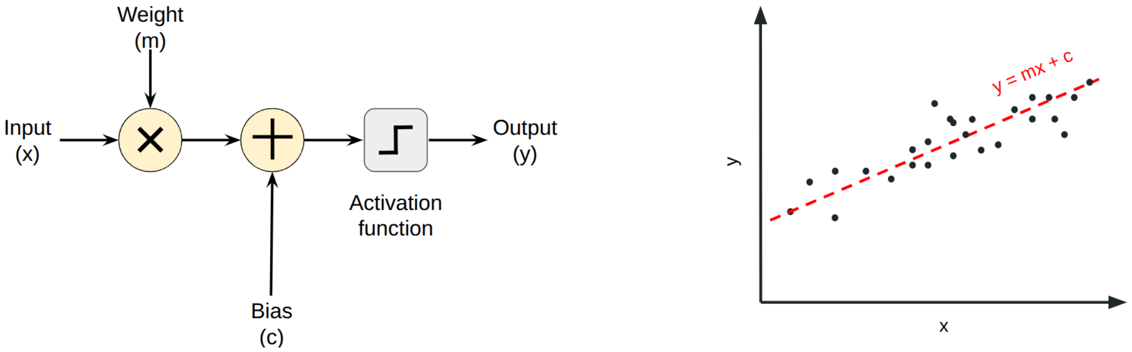


Figure 4.5: Left: A single input perceptron. Right: The representation of the perceptron as a linear model.

Obtaining the best-fit model requires reducing the distance of each point (true output) with its corresponding y-intercept on the linear model (predicted output). The sum of the differences between the predicted and the true output for each point is referred to as “loss”, and is quantified by a “loss function”. A neural network “trains” by iteratively adjusting the weights and biases to minimise this loss. These parameters are updated following the gradient of the loss function with respect to each parameter, such that subsequent updates move the model towards a better fit. Optimisation is performed using gradient-based methods, with gradients computed using the backpropagation process. These concepts are described in detail in Goodfellow, Bengio and Courville, 2016; Glassner, 2021.

A single linear model is insufficient to represent complex data distributions. To capture non-linear relationships, multiple linear models can be combined and connected through non-linear activation

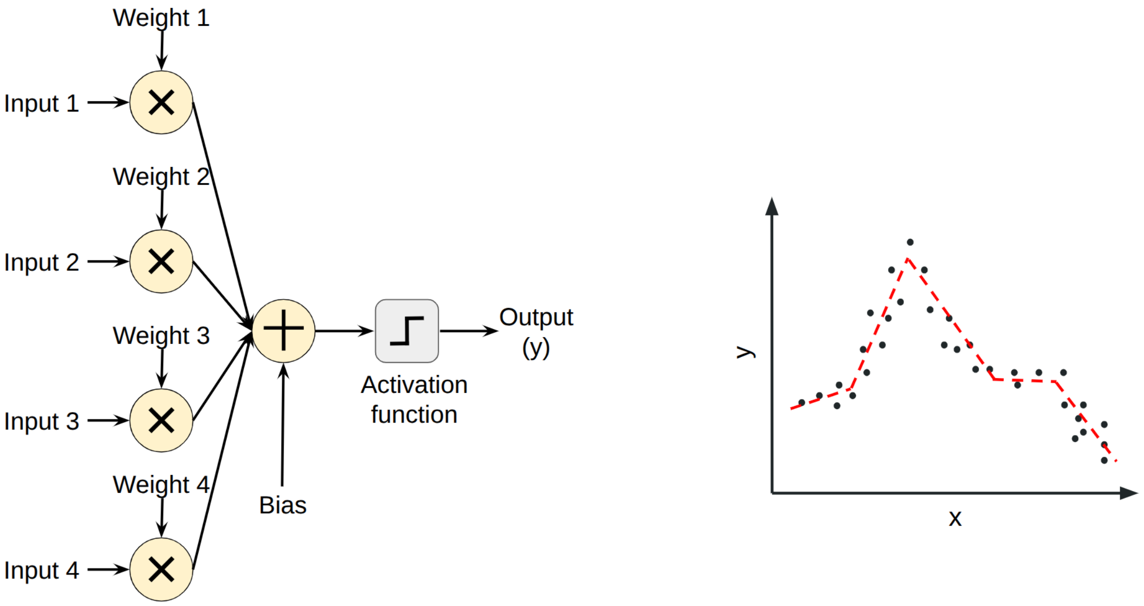


Figure 4.6: Left: A perceptron with bias. Right: The model representing the perceptron.

functions. An artificial neural network can therefore be viewed as a composition of linear models combined with nonlinear transformations, resulting in a nonlinear function capable of approximating complex distributions (Fig. 4.6). A neural network with a single hidden layer consisting of n neurons can represent a nonlinear function of $n+1$ linear segments. A shallow neural network can approximate nonlinar functions, but its capacity is limited to simpler distributions. The increase in the complexity of the model requires efficient stacking of multiple layers of neurons. A neural network with every neuron in a given layer being connected to every other neural in the neighbouring layers is referred to as a dense neural network or a multilayer pereptron (MLP). An MLP with L layers, each containing n neurons, can represent a function composed of up to n^L linear segments, allowing for the modelling of complex datasets. Therefore, real-world problems with complex data distributions usually require deep learning models with several layers.

4.2.2 Convolutional neural networks

Although MLPs are capable of modelling complex data, they are highly inefficient in analysing images. This is because such architectures require the image to be flattened and each pixel of the image to be connected to a separate neuron. This requires a large number of neurons and also discards the spatial structure of the image, therefore preventing the network from detecting features in the images. Convolutional neural networks (CNNs) address these limitations by operating directly on images and preserving spatial information through the learning process (LeCun et al., 1989; Krizhevsky, Sutskever and Hinton, 2012; Girshick et al., 2014). The core principle of CNNs is convolution, which refers to the matrix multiplication of input with a filter or kernel (Fig. 4.7). The filters slide across the image computing weighted sums of local neigbourhoods.

CNNs adopt the classical approach of feature extraction using fixed mathematical bases (e.g. Gaussian derivatives to detect edges) and go one step further by making the feature extractors learn

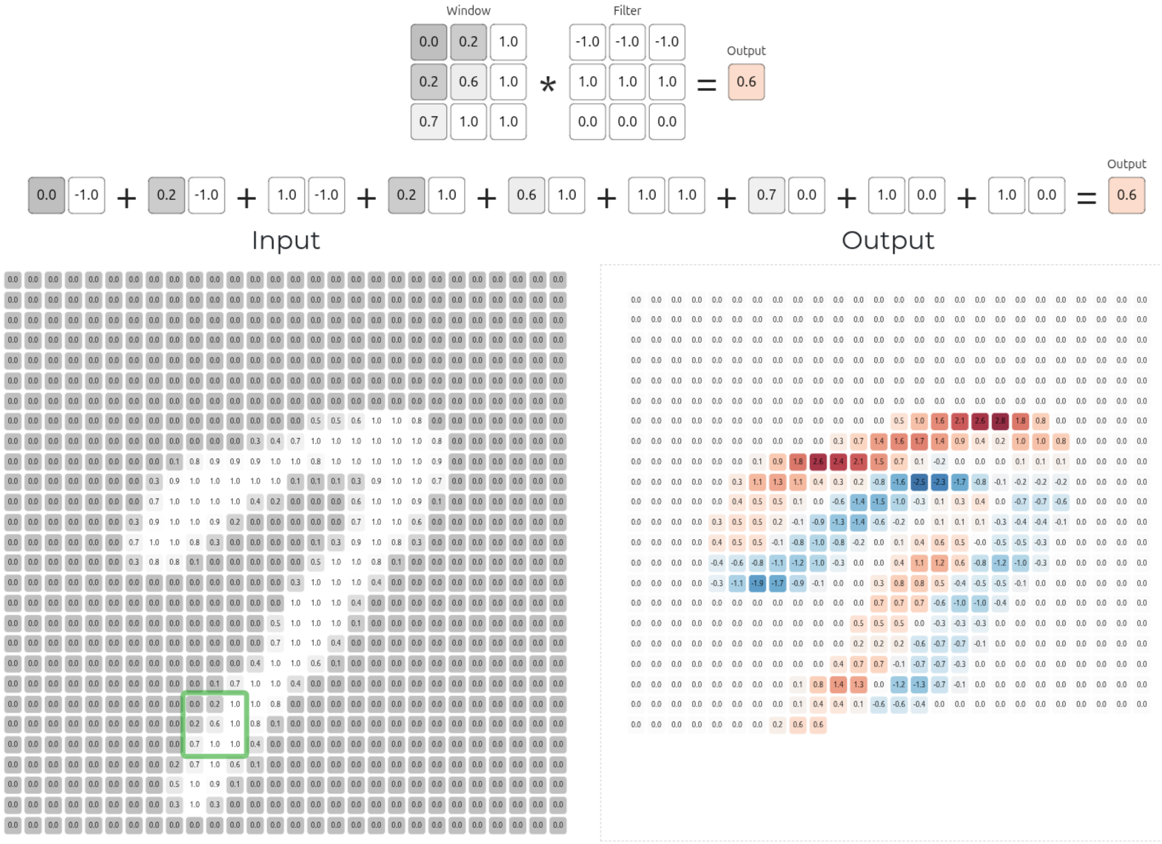


Figure 4.7: Convolution of the image of the digit “7” with a filter to detect the top edge. The values in the window represent the green box on the input data. The corresponding matrix multiplication operations are shown on the top of the input and output images. In the output image, the colors represent the value of the output with red representing positive values and blue representing negative values. Since this is a top-edge filter, the negative values represent the opposite feature, which is a bottom edge. The image is generated using the convolution demo tool on deeplizard.com.

from the data. They are built through hierarchical feature learning, with subsequent layers transforming representation from previous layers to progress from local to global features. For example, the earlier layers encode simpler features such as edges, the intermediate layers detect structures such as large filaments and spiral arms, and the deeper layers capture the morphology of an entire galaxy. The images are also downsampled along the CNN to increase the region covered by a single filter, thus capturing complex structures.

The downsampling of images in a CNN is performed by the pooling layers (Gholamalinezhad and Khosravi, 2020). These layers reduce the spatial dimensions of the input data while retaining the important features. Pooling involves dividing the input image into smaller regions and getting the maximum, minimum, average, or sum of each region, and combining these values to obtain a smaller output image. The most common types of pooling are max pooling and average pooling. Max pooling involves sliding a filter over an input image and selecting the maximum value of the pixels covered by the filter. Thus, performing max pooling on an input image of size 27×27 with a filter of size 3×3

and stride[10] 3 results in an output image of 9 × 9, with each pixel representing the maximum value of a 3 × 3 sub-region of the input image. Average pooling takes the average values of the pixels in the sub-regions. It is often used when working on astronomical datasets, since averaging helps suppress random noise fluctuations and improves the signal-to-noise ratio.

4.2.3 U-Nets

Convolutional neural networks are typically used for classification tasks. However, many image identification problems require that an image be segmented by assigning a label to each pixel. U-Net is a fully-convolutional neural network architecture that specialises in semantic segmentation of images. It is able to create masks of the important features by assigning individual pixels to particular objects. It was initially developed as a single class segmentation model to segment medical images (Ronneberger, Fischer and Brox, 2015). Further advancements have resulted in its use for multi-class segmentation, where pixels can be assigned to multiple classes, enabling the simultaneous identification of multiple objects in an image (Tommasino et al., 2023; Z. Xu et al., 2024). The ability to learn features robustly and efficiently has made U-Nets the go-to architecture in the field of pixel-level computer vision research. They are used for many tasks such as image segmentation (Bolya et al., 2019; M.-H. Guo et al., 2022), image restoration (L. Chen et al., 2022), and image matting (Yu et al., 2021).

U-Nets contain an encoder, a decoder, and skip connections between them, with the encoder and decoder containing various layers (Fig. 4.2.3). The encoder part tries to understand the “what” of an image, i.e. it extracts the important features from an image. At each layer, the input signal is convolved and the features are saved as feature maps. The input is then downsampled (pooling layers; Sect. 4.2.2), and the number of channels is increased to improve the receptive field of the filters and identify complex features. The decoder in the U-Net tries to understand the “where” of the image, i.e. it reconstructs the selected features of the image at their original location. At each level, the condensed information from the previous layer is upsampled by using transposed convolutions (deconvolution), recovering the original image size. Skip connections preserve the position of features in the image by concatenating the feature maps from each level of the encoder to the transposed convolutions in the corresponding layers of the decoder.

Modern neural networks like U-Nets can solve various image segmentation problems by training on different datasets. Modified U-Nets like UNet++ and attention-based mechanisms such as TransUNet and Vision Transformers show the potential to segment complex astrophysical structures (D. Xu and Zhu, 2024). These methods effectively capture non-local dependencies and complex spatial relationships, vital for accurately identifying complex structures such as feedback-driven shells. Furthermore, these models offer enhanced interpretability, allowing for clearer insights into why specific feedback features are identified in certain layers, which may also help link particular feedback features to specific physical phenomena.

4.2.4 ResNets

A common phenomenon observed in very deep neural networks is the degradation problem, which refers to the decrease in performance with increasing depth. The degradation arises because, as signals propagate through layers, they are repeatedly transformed by non-linear operations, making optimisation difficult and leading to distorted representations. One contributing factor is that, in

---

[10] The step size by which a filter moves across the input data

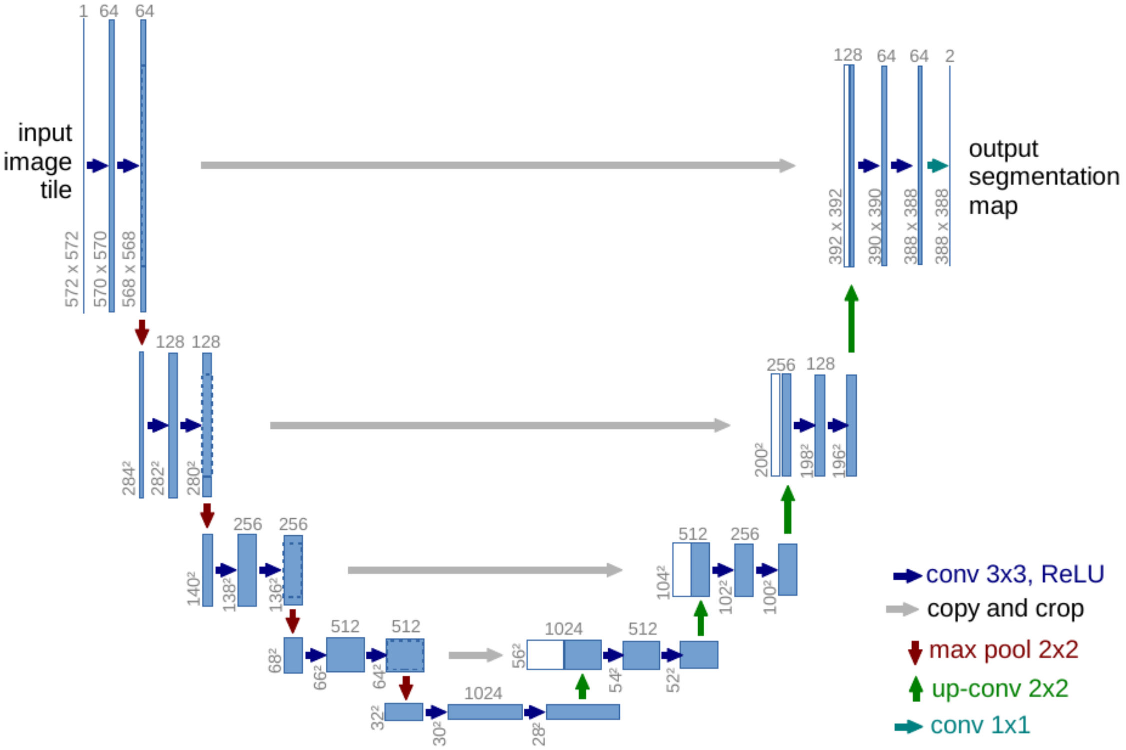


Figure 4.8: U-net architecture example. The left side of the "U-shape" is the encoder part and the right side is the decoder part. Each blue box corresponds to a multi-channel feature map and a white box represent copied feature map. The number of channels is denoted on the top and the size (in pixels) is denoted at the lower left edge of a box. The arrows represent various operations between the layers, such as convolution and skip connections. Taken from Ronneberger, Fischer and Brox, 2015.

traditional deep networks, each layer is required to simultaneously (i) preserve the input and (ii) learn a transformation that maps the local input to a more useful representation. As depth increases, the dual burden makes learning identity mapping extremely difficult.

The introduction of residual networks (ResNet) marked a turning point in the development of deep neural networks (He et al., 2016). By incorporating skip (or residual) connections, ResNets enable layers to learn a residual mapping, i.e., a difference between the desired output and the input (X. Liu and Goh, 2025). This allows the network to function without learning the entire input signal. Each residual block learns only the transformation $\mathcal{F}(\mathbf{x})$ such that the output becomes $\mathbf{x} + \mathcal{F}(\mathbf{x})$ (Fig. 4.9). This reformation allows the layers to focus on learning incremental refinements to the input. Residual connections preserve the identity mapping and maintain signal fidelity throughout the depth. As a result, information and gradients can be propagated more effectively through the network, preventing degradation and enabling the creation of deep architectures.

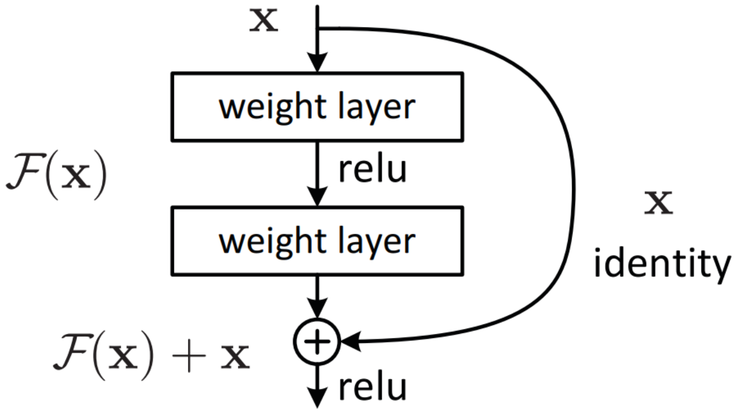


Figure 4.9: Residual block: a building block of the residual neural network. The input **x** is passed through the convolutional (weight) layers and the activation function (relu). The skip connections from the output to the input enable the neurons to analyse residual maps. Taken from He et al., 2016.

CHAPTER 5

Effects of stellar feedback on molecular gas cores

This chapter is a summary of the article published in Astronomy and Astrophysics (see Appendix A) titled "**Effects of stellar feedback on cores in STARFORGE**" by:

K. R. Neralwar, D. Colombo, S. Offner, F. Wyrowski, K. M. Menten, A. Karska, M. Y. Grudić and S. Neupane

5.1 Context

The interstellar medium (ISM) in galaxies exhibits a hierarchical structure, with molecular gas organised into clouds on large scales, which fragment into clumps and, on smaller scales, into dense cores (Dobbs, Krumholz et al., 2014b; Duarte-Cabral, Colombo et al., 2021; Chevance, Krumholz et al., 2023). Stars form in these cores embedded in clouds and over time inject momentum and energy back into the ISM through stellar feedback processes such as protostellar outflows, stellar winds, radiation, and supernovae. These feedback mechanisms, in turn, influence the dynamics and kinematics of the molecular gas structures.

Although the regulation of star formation in galaxies by stellar feedback is widely established (Krumholz, Bate et al., 2014; Agertz and Kravtsov, 2015; Peters et al., 2017; Guszejnov, Grudić, S. S. R. Offner et al., 2022), the impact of feedback on cores is still being investigated. Observational surveys such as ALMAGAL (Molinari, Schilke et al., 2025) have led to detailed analyses of the properties of cores; however, they are unable to separate the effects of individual feedback mechanisms on cores. Numerical simulations offer a means to isolate the effects of individual feedback processes within controlled environments. The STARFORGE simulations provide an ideal framework for this purpose, as they model the evolution of GMCs at high resolution while including all the physics relevant for star formation. The simulations also track the impact of individual feedback mechanisms on each gas cell in the simulation domain, while resolving the formation of individual stars.

We analyse a large sample of ~ 400 000 cores to perform a comprehensive statistical study of the impact of feedback on molecular gas. To take advantage of the high spatial resolution of STARFORGE , we define the cores as the smallest resolvable dendrogram structures (leaves). This allows us to probe the effects of feedback on sub-pc scales, which is challenging in observations because of limited telescope resolution. The large and well-defined sample of cores allows us to quantify how individual feedback mechanisms affect the molecular gas, achieving one of the main goals of the thesis. In addition, this study ensures that simulated structures show meaningful feedback-driven trends, serving as a foundation for the subsequent analysis of molecular gas and stellar feedback using synthetic observations.

5.2 Aims

We aim to quantify how various stellar feedback mechanisms influence the physical properties of molecular gas cores. Specifically, we examine how the radius, mass, velocity dispersion, and virial parameter change with the influence of protostellar outflows, stellar winds, and supernovae. We further investigate how these individual feedback mechanisms affect the correlations between core properties, using the Larson's (Larson, 1981a), Heyer's (M. Heyer, Krawczyk et al., 2009), and size-mass (Duarte-Cabral, Colombo et al., 2021) scaling relations.

5.3 Methods

We analysed the STARFORGE M2e4 fiducial simulations that follow the evolution of a 20000 $M_\odot$ GMC initialised as a 10 pc sphere with virial parameter of 2. Using a dendrogram analysis on the simulations, we extracted leaves and termed them cores. We further calculated several physical properties of cores, namely radius, mass, velocity dispersion, and virial parameters using the output of dendrograms. Because the simulations store the fraction of gas affected by individual stellar feedback mechanisms, we were able to obtain the corresponding feedback fractions for our cores. We used these feedback fractions to categorise the cores into *no*, *low*, *moderate*, and *high* feedback bins. These bins were created for outflows, winds, supernovae and their combined effects, i.e. global feedback.

5.4 Results

We find that stellar feedback affects the kinematics and the boundedness of the cores, visible as a monotonous increase in the velocity dispersion and the virial parameter with feedback fractions (Fig. 5.1). Cores experiencing strong outflows, winds, and supernovae have significantly higher velocity dispersion and virial parameter (Fig. 5.2).

The structural properties of the cores show more subtle changes (Fig. 5.2). The cores affected by feedback are smaller and denser than the pristine cores. This is most likely the result of feedback compressing and dispersing the molecular gas. However, these feedback mechanisms do not significantly affect the core mass.

The size-linewidth relation retains the traditional form of $\sigma_v \propto \mathrm{R}^{0.5}$ for pristine cores but shows deviations for cores with stronger feedback. Similarly, the Heyer relation shifts upward (higher virial parameter) for feedback affected cores, emphasising the role of feedback in injecting momentum in the ISM. These relations further suggest that outflows mainly disperse molecular gas, stellar winds compress the cores and disperse the gas, whereas supernova explosions push gas outward, leading to large cores.

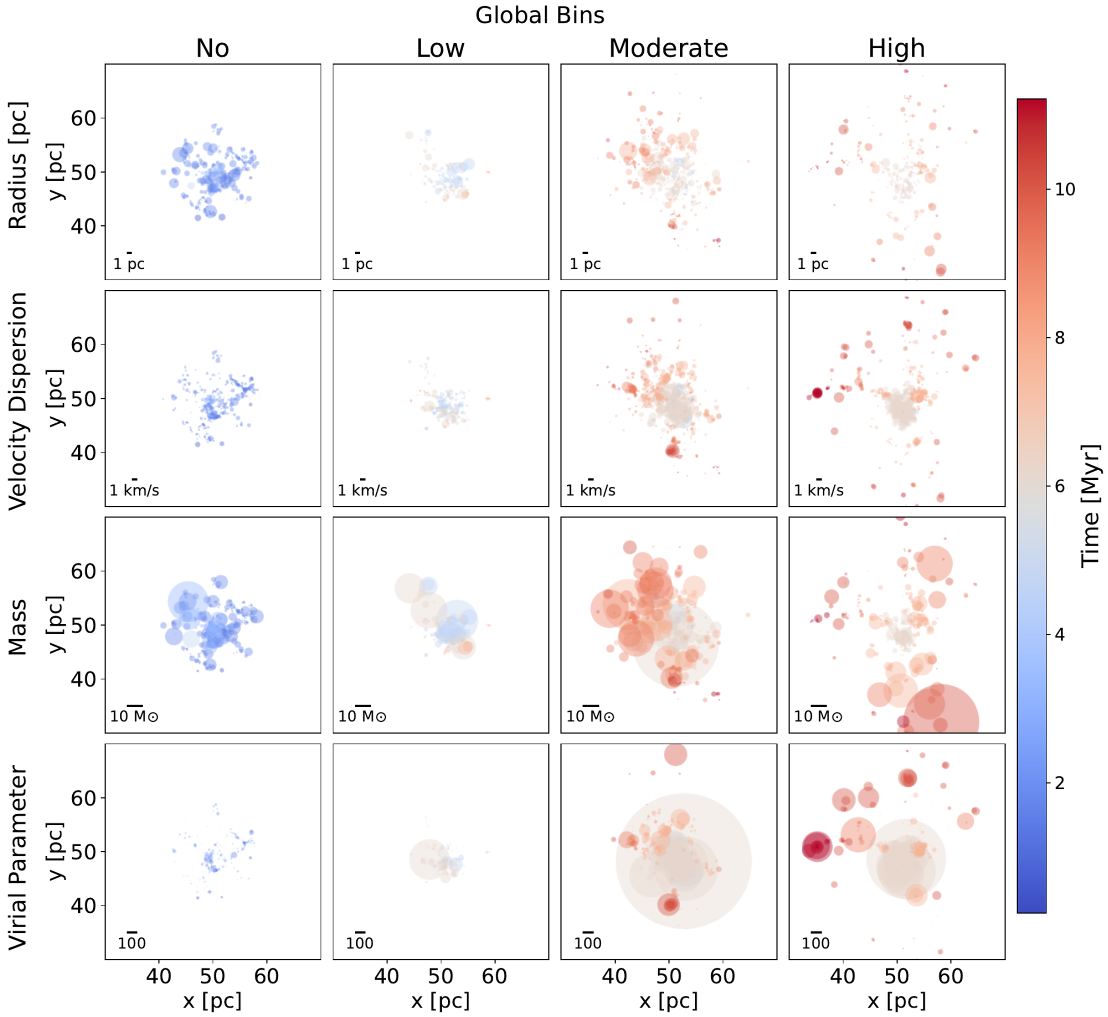


Figure 5.1: Distribution of cores in different snapshots along with their properties in various global feedback bins. The positions of the cores represent their locations in the simulation box. The cores are are assumed to be circular, and the core properties (Sect. 5.3) are scaled with the size of these circles. First row: Radius of the core. Second and third rows: Velocity dispersion and mass of the cores, respectively. Bottom row: Virial parameter. The colours represent the time elapsed (in Myr) since the start of the simulation.

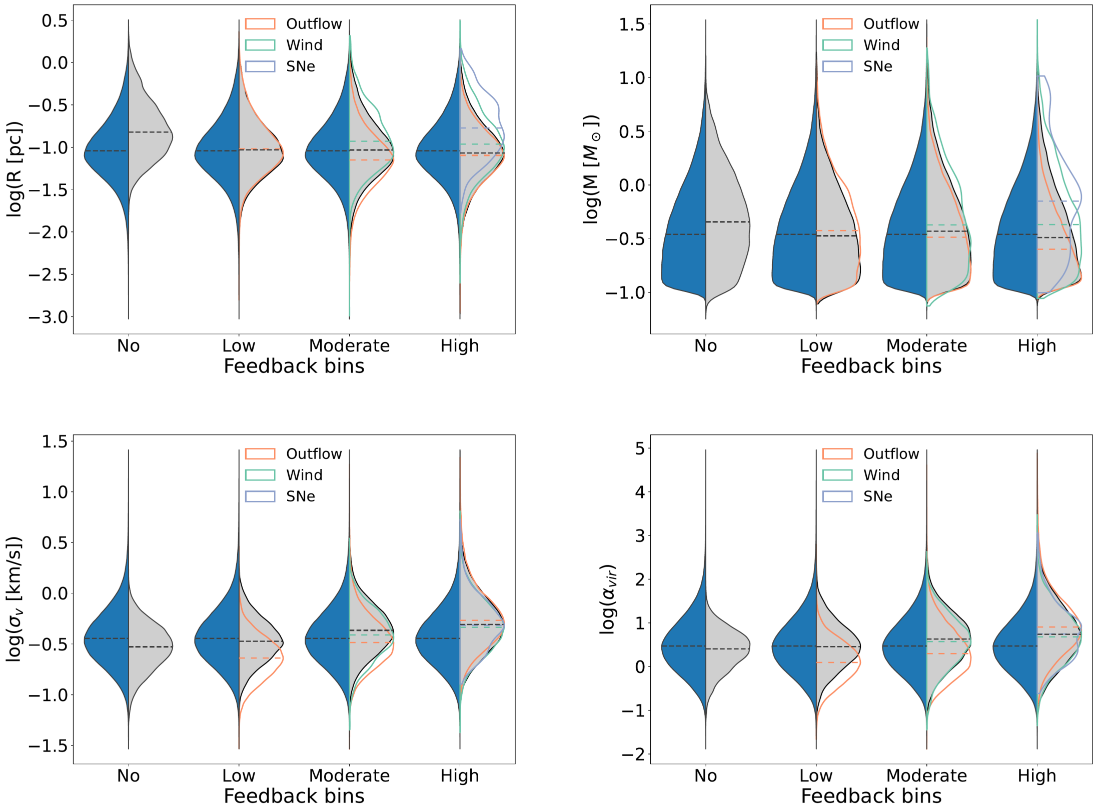


Figure 5.2: Distributions of the core radius (top left), mass (top right), velocity dispersion (bottom left), and virial parameter (bottom right) in the different feedback bins. The split violin plot presents the total distribution (blue) and the distribution of cores in the global feedback bins (grey) for each axis. The colours representing the outflow, wind, and supernova bins are marked in the legends. The dashed lines represent the medians of the different distributions. Violin plots present the density of the data at different values, which were smoothed using the kernel density estimator.

5.5 Conclusions

Our analysis shows that stellar feedback shows measurable impacts on molecular gas cores. By injecting momentum into the ISM, feedback increases the velocity dispersion and the virial parameter of the cores. Although individual feedback mechanisms do not alter the mass of cores, they affect the core sizes. Protostellar outflows and winds compress and disperse gas, leading to smaller cores, whereas supernovae expand the cores to increase their sizes. Among the different mechanisms, protostellar outflows have the strongest impact on cores, suggesting that they are the most dominant feedback mechanisms on sub-pc scales.

Although this study analyses the trends in core properties, the absolute values are not directly comparable to observations. This is because the analysis is performed in simulation space and the properties definitions differ from observational estimates, e.g. virial parameters are derived as ratio of kinetic to potential energy rather than the ratio of virial and luminous mass. To enable a direct comparison with observations, the following chapters present synthetic molecular-line observations of gas structures with properties consistent with observations.

CHAPTER 6

An evolutionary study of synthetic molecular clouds in STARFORGE

This chapter is a summary of the article published in Astronomy and Astrophysics (see Appendix B) titled "**A multi-scale evolutionary study of molecular gas in STARFORGE: I. Synthetic observations of SEDIGISM-like molecular clouds**" by:

K. R. Neralwar, D. Colombo, S. Offner, A. Karska, M. Figuera, F. Wyrowski, S. Neupane, J.S. Urquhart and A. Duarte-Cabral

6.1 Context

Molecular clouds (MCs) are active sites of star formation in galaxies and are often referred to as cradles of star formation. They are large ($\sim$ 1–200 pc; Ballesteros-Paredes, André et al., 2020; Duarte-Cabral, Colombo et al., 2021), massive ($\sim 10^2 - 10^7 M_\odot$; Rebolledo et al., 2012) and dense ($\Sigma \sim 1 - 1000 M_\odot$ pc^{-2} Barnes et al., 2018) structures with virial parameter around unity ($\alpha_{vir} \sim 1$ Fukui, Kawamura et al., 2008; Roman-Duval et al., 2010). MCs host a complex interplay between gravity, turbulence, magnetic fields, and various stellar feedback mechanisms, such as protostellar outflows, stellar winds, radiation, and supernovae. They are fractal structures containing denser substructures, called clumps and cores, which are precursors of stellar clusters and individual stars (Ballesteros-Paredes, André et al., 2020; Lin, Wyrowski, H. B. Liu, Y. Gong, Sipilä, A. F. Izquierdo et al., 2024).

Recent large scale high-resolution surveys such as SEDIGISM (Schuller, Urquhart et al., 2021) have produced high-resolution maps of the Galaxy, providing statistical catalogues of MCs in molecular line emission. This has further enabled the derivation of integrated properties of MCs, such as radii, masses, velocity dispersion, virial parameter, and surface densities. However, interpreting the correlation between these properties, i.e. scaling relations, and the scatter therein is often challenging because MCs are often approximated as a single population of quasi-static entities in near equilibrium. This does not account for the different evolutionary stages of MCs or the fact that stellar feedback might cause a deviation from the typically observed scaling relations.

Numerical simulations can follow the evolution of MCs, but a comparison with observational data requires mapping the simulation output to the same space as observations. To achieve this, we produce ^{13}CO(2-1) synthetic observations of the STARFORGE simulations corresponding to the observational setup of the SEDIGISM survey. Molecular clouds are then identified using a hierarchical clustering approach, where MCs correspond to the largest structures, i.e., dendrogram trunks. This definition allows us to associate the MCs with a snapshot number as a proxy for their evolutionary stage. Since the synthetic MCs have properties comparable to observed MCs, we directly compare them using scaling relations and investigate how these relations and scatter evolve over the cloud lifetime under the influence of stellar feedback.

6.2 Aim

The first goal of this work is to use state-of-the-art simulations that incorporate all the relevant physics of star formation and produce synthetic observations that closely resemble observational data. As a primary diagnostic, we compare the synthetic observations and the extracted MCs with our observational benchmark, i.e., SEDIGISM (Chap. 3). This helps us evaluate the extent to which current simulations can reproduce observed MCs. The second goal of this work is to study the changes in properties of MCs over time, as they are impacted by different feedback mechanisms, to provide insight into the interpretation of observed MCs as they evolve.

6.3 Methods

We produced the synthetic MCs by first creating synthetic observations mimicking the SEDIGISM ^{13}CO(2-1) emission cubes using RADMC-3D and then applying hierarchical clustering to extract the MCs. The data from STARFORGE snapshots were first regridded to reproduce the pixel size in the SEDIGISM observations using a custom-built pipeline. The gas density, temperature, and velocity were then provided to RADMC-3D along with a wavelength grid matching the spectral resolution of SEDIGISM. The output from RADMC-3D was convolved with a 2D gaussian kernel of 28” reproducing the beam convolution in APEX observations. Lastly, the emission from noisy channels of a SEDIGISM ppv cube was added to the synthetic cubes to replicate the noisy ppv cubes in SEDIGISM.

These ppv cubes were masked using the dilated making technique (E. W. Rosolowsky, Pineda, Kauffmann et al., 2008b; Grishunin et al., 2024) to match the SEDIGISM data closely. From these, we extracted a population of MCs using the hierarchical clustering cloud algorithm – astrodendro (E. W. Rosolowsky, Pineda, Foster et al., 2008) – matching the cloud identification procedure used in SEDIGISM (Duarte-Cabral, Colombo et al., 2021). We further obtained the observable properties of the MCs, i.e. mass, radius, velocity dispersion, virial parameter, and surface density. We also obtained additional properties such as the true molecular mass and number of stars by gridding simulation data fields to the same grid as the one implemented by RADMC-3D. Furthermore, we used the algorithm based on the moment of inertia – RJ plots (Clarke, Jaffa and Whitworth, 2022) to obtain the morphologies of our MCs.

6.4 Results

We find a strong agreement between the flux distributions of the SEDIGISM ^{13}CO(2-1) emission cubes and our synthetic obsevations. This serves as a validation that our synthetic emission maps replicate

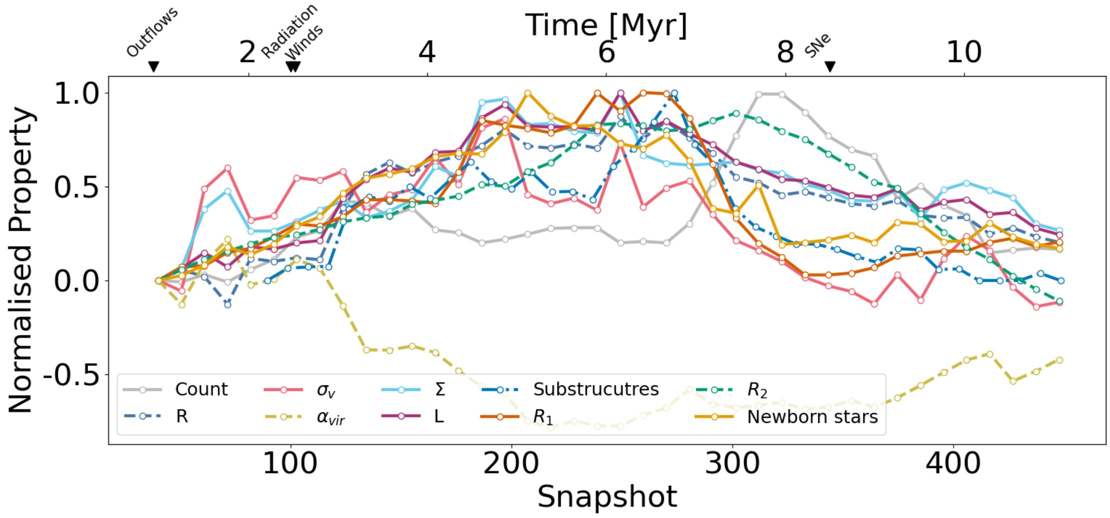


Figure 6.1: Adapted from Fig. 5, 6 and 7 in **neralwar2025arXiv251007393N**. Normalised medians of MC properties (log scales) as a function of time. The properties shown are radius (R), velocity dispersion (σ_v), molecular gas mass (M), virial parameter (α_{vir}), surface mass density (Σ), and luminosity (L), number of substructures (substructures). Count and newborn stars refers to the total number of clouds and stars with ages less than 250 kyr in a time bin, respectively. The R_1 and R_2 values represent the morphologies of the molecular gas complexes. The normalisation process included subtracting the initial value (first bin) of the property from themselves, and this was followed by a min-max standardisation. The symbols on the top represent the times at which outflows, photoionisation radiation, stellar winds, and supernovae begin in the simulation.

the SEDIGISM data to the first order. We further compare the properties of our synthetic MCs with those of the SEDIGISM clouds and find a good overlap between the two samples. This further highlights the reliability of our approach in not only creating emission cubes similar to SEDIGISM, but also extracting MCs similar to those in SEDIGISM.

Tracking the properties, morphology, and substructures of MCs over time reveals their formation, evolution, and destruction mechanisms (Fig. 6.1). The initial turbulence in the simulations creates gas overdensities that collapse under self-gravity and are detected as ^{13}CO(2-1) emission. These reflect the early cloudlets in observations that accrete gas from the larger environment and appear as moderately dense gas structures. Gas flows from large- to small scales shape MCs into elongated filamentary structures with multiple substructures. The fractal substructures in MCs form stars, which eject matter and radiation into the surrounding environment, driving the formation of gas bubbles. These 3D bubble-like MCs are often associated with stellar winds, radiation, and H II regions. Our analysis presents MCs as evolving from small, diffuse structures to dense filamentary MCs followed by 3D gas bubbles, and these evolutionary trends are consistent with simulations initialised differently. This confirms the key hypothesis from our previous observational work that MCs evolve from concentrated to elongated to ring-like structures.

Our synthetic MCs occupy the same parameter space as the SEDIGISM clouds on Larson and Heyer scaling relation plots (Fig. 6.2). However, the average distributions of the MCs at different evolutionary times occupy different regions in these plots. The early MCs have higher velocity

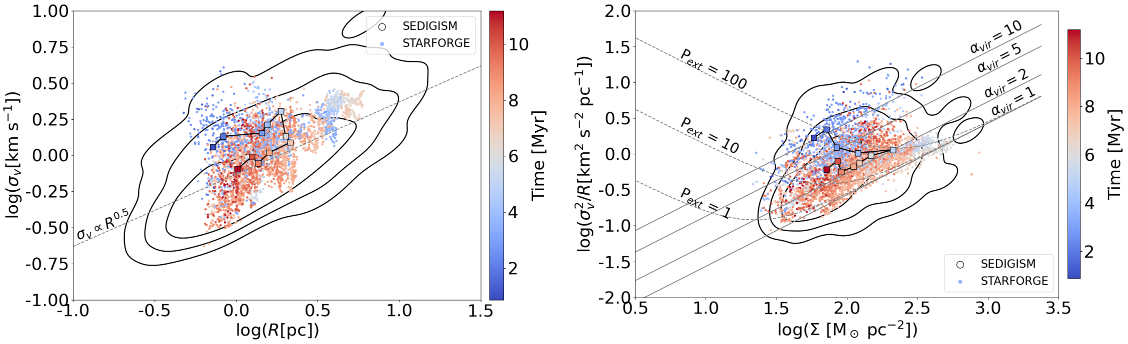


Figure 6.2: Larson's first relation (σ_v versus R; left) and Heyer' relation (σ_v^2/R vs Σ; right) for our MCs (scatter points), colour-coded with respect to the time elapsed (in Myr) since the start of the simulation. The squares represent medians of distributions in $\sim$ 1 Myr bins. The black contours represent the 1σ, 2σ, 3σ levels for the SEDIGISM clouds. The dashed line on the left plot represents Larson's first relation (Larson, 1981a; Solomon, Rivolo et al., 1987b). On the Heyer's relation plot, the solid grey lines represent isocontours of virial parameters. The dashed lines represent α_{vir} = 1 when including an external pressure of P_{ext} = 1, 10, and 100 $\text{M}_\odot$ pc^{-3} km^2 s^{-2}.

dispersion and virial parameters, the mid stage MCs are large and dense, whereas the late stage clouds move towards higher virial parameter and lower surface densities. This confirms the progression from compact clouds to large filamentary structures, which are shaped into bubbles by stellar feedback.

6.5 Conclusion

Our evolutionary analysis demonstrates that MCs evolve from small, diffuse cloudlets to dense filamentary structures to 3D bubbles and are eventually dispersed by stellar feedback. We reproduced the statistical property distributions of the MCs in the inner Milky Way observed in the SEDIGISM survey using STARFORGE simulations and robust data processing. MCs at different evolutionary stages occupy different parameter spaces in scaling relation plots, suggesting that the cloud ages drive the scatter in observed scaling relations. Our findings suggest that the apparent diversities in the MC population largely reflect their evolutionary stage.

While defining MCs as dendrogram trunks allows us to analyse their average evolutionary trends, it limits our ability to study individual structures. In some snapshots, the trunks represent substructures within the GMC; whereas in others, they correspond to the entire GMC. This effectively reduces our analysis to the evolution of a single large GMC and highlights the need to investigate multiple smaller-scale structures within MCs that evolve independently. In the next chapter, we move beyond snapshot-based evolution and follow individual clumps over their lifetime using a self-developed pipeline. This enables a detailed study of the clump evolution under the effects of various stellar feedback processes.

CHAPTER 7

Evolution of individual clumps in STARFORGE

This chapter will be submitted to the journal Astronomy and Astrophysics titled "**A multi-scale evolutionary study of molecular gas in STARFORGE: II. Feedback-driven evolution of individual clumps**" by:

K. R. Neralwar, D. Colombo, S. Offner, A. Karska, M. Figueira, F. Wyrowski, S. Neupane, J.S. Urquhart and A. Duarte-Cabral

7.1 Context

In Chapter 6, we studied the evolution of molecular clouds by generating ^{13}CO(2-1) synthetic observations using a suite of STARFORGE simulations. Hierarchical segmentation of these cubes revealed clumps inside these MCs. We now extend our evolutionary analysis to these smaller structures using the ^{13}CO(2-1) synthetic observations with improved CO chemistry treatment. Recent studies show that clumps dynamically decouple from their parent MCs and undergo evolution on their own timescales (Peretto, Rigby et al., 2023). This raises the question of how clumps evolve in different parental MCs whose properties have been shaped by different galactic processes. With the goal of understanding the evolution of individual clumps, we have created a pipeline that matches a clump in a given snapshot to its counterparts in the following snapshots. We study the changes in the properties of clumps as they evolve, fragment into multiple substructures, and merge with neighbouring gas structures. We also investigate the influence of protostellar outflows, stellar winds, and radiation on molecular gas clumps during their evolution.

7.2 Introduction

The fractal nature of molecular gas structures in ISM and their classification based on size and masses into clouds, clumps, and cores have been studied extensively (Ballesteros-Paredes, André et al., 2020; Chevance, Krumholz et al., 2023). Clumps are parsec scale (1 - 5 pc) molecular gas structures with densities $\sim 10^2 - 10^5$ cm^{-3} and masses $\sim 0.1 - 1000$ M$_\odot$ (Vázquez-Semadeni, Gómez et al., 2007; Rigby, Moore, Eden et al., 2019; Shetty, Collins et al., 2010; Olmi, Brand and Elia, 2023). They typically account for about 5–10% of the mass of their parental clouds and are considered to be progenetors of stellar clusters (Battisti and M. H. Heyer, 2014; Peretto, Rigby et al., 2023). They may contain smaller (≤ 0.1 pc) and denser ($\geq 10^4$ cm^{-3}) structures called cores that host individual stars (S. S. R. Offner, J. Taylor, Markey et al., 2022; Caselli, Spezzano et al., 2025). As stars form in clumps, they inject matter, energy, and momentum back into their surroundings through various stellar feedback mechanisms, thereby regulating the dynamical evolution of clumps.

Stellar feedback is one of the most important processes that maintains the energy balance in galaxies (Krumholz, 2014; Gatto et al., 2017; Peters et al., 2017; Guszejnov, Grudić, S. S. R. Offner et al., 2022). Feedback can both disperse gas, suppressing star formation, and compress gas, thereby triggering new star formation. It therefore regulates the overall star formation in galaxies. Individual feedback mechanisms significantly influence each other's efficiency (Rahner et al., 2017). These mechanisms are broadly classified into five types, namely protostellar outflows, stellar winds, radiation pressure, photoionisation, and supernovae (Girichidis et al., 2020; Geen, Agrawal et al., 2023). In addition, UV photons from stars also dissociate molecules and lead to a phase shift from molecular to atomic and ionic. These feedback mechanisms differ in their origins (i.e. the process by which they arise), the stage of a star's life during which they are dominant, the amount of matter and energy expelled, and the spatial scales on which they act (Hopkins, Quataert and Murray, 2012).

During their accretion phase, protostars launch bipolar outflows (Fendt and Čemeljić, 2002; Bally, 2016). These jets are collimated and directed in opposite directions near the (proto-)star and often lead to bow shocks farther away. Protostellar jets launch outflows with speeds of around a few hundred km s^{-1} (Bally, 2024). The outflows interact with the ISM and can reshape the surrounding gas. Stellar winds refer to the ejection of mass from the surface of a star (Vink, 2024). Winds create small cavities (Ali, Bending and Dobbs, 2022) inject significant kinetic energy into the ISM (Weaver et al., 1977). However, at any given time, winds do not contribute as much momentum as other feedback processes such as outflows (Ali, Bending and Dobbs, 2022).

Stars radiate energy throughout their lifetimes, and stellar emission is often divided into wavelength bands such as the Lyman continuum, far-ultraviolet (FUV), near-ultraviolet (NUV), optical, near-infrared (NIR), and far-infrared (FIR) (Grudić, Guszejnov, Hopkins et al., 2021). Each band traces a specific stellar population and their effects on the ISM (Kennicutt and Evans, 2012). Young and massive OB stars photoionise the surrounding medium, producing H ɪɪ regions. The Hα emission and the inferred lyman continuum photons therefore trace recent star formation ($\sim 10-30$ Myr Dale, Ercolano and Bonnell, 2012; Geen, Soler and Hennebelle, 2017; Grudić, Hopkins et al., 2019; Olivier et al., 2021). The UV continuum traces stars with ages $\sim 10-200$ Myr, with emission from young stars at smaller wavelengths (FUV) and evolved stars at larger wavelengths (NUV) (Murray, Quataert and Thompson, 2010; E. C. Ostriker, McKee and A. K. Leroy, 2010; Raskutti, E. C. Ostriker and Skinner, 2016; J.-G. Kim, W.-T. Kim and E. C. Ostriker, 2018; Hopkins and Grudić, 2019). The optical and UV radiation from stars is absorbed by interstellar dust and is reradiated at infrared wavelengths, which is the primary cooling mechanism in the densest gas. (Krumholz, Klein, McKee et al., 2009; Rosen, Krumholz et al., 2016; Tsang and Milosavljević, 2018).

The influence of stellar feedback mechanisms is visible on different spatial scales, shaping clouds, clumps, and cores; but the effect of individual feedback mechanisms on their evolution is still under investigation (Hopkins, Quataert and Murray, 2012; Bate, Tricco and Price, 2014; Rosen, Krumholz et al., 2016; Guszejnov, Hopkins and Grudić, 2018; Hopkins, Wetzel et al., 2018; Iani et al., 2021; Geen, Agrawal et al., 2023; Neralwar, Colombo, S. Offner, Wyrowski et al., 2024; Neralwar, Colombo, S. Offner, Karska et al., 2025). The goal of this work is to explore how molecular gas clumps evolve as they are affected by various stellar feedback mechanisms, and to explore whether the effects of individual feedback mechanisms can be disentangled. This is achieved using the STARFORGE simulations that can track the gas affected by individual stellar feedback mechanisms in each cell. By constructing feedback tracer maps for the corresponding synthetic observations using the feedback fractions, we investigate the variations in clump properties (e.g. radius, mass, surface density, morphology, and virial parameter) with respect to the degree to which they are affected by different feedback mechanisms. This analysis helps us understand the role of outflows, winds, and radiation in shaping the physical and dynamical evolution of clumps, which is important to get a complete picture of star formation.

The structure of the chapter is as follows. In Section 7.3 we describe the STARFORGE simulations used for our analysis and the SEDIGISM survey, which we used as an observational comparative benchmark. In Section 7.4, we first describe the methodology to obtain CO abundances from the initial C-H-O abundances in the simulation (Sect. 7.4.1). We then describe the creation of the synthetic observations from STARFORGE and structure extraction in Sect. 7.4.2 and our in-house pipeline to follow individual structures over time in Sect. 7.4.3. We analyse the scaling relations to compare the correlation between STARFORGE and SEDIGISM cloud properties and understand the time evolution of clouds on these plots in Sect. 7.5.1. We present the evolution of a typical clump over time in Sect. 7.5.2 and study the evolution of all clumps collectively in Sect. 7.5.3. We conclude by summarising our work in Sect. 7.7.

7.3 Data

7.3.1 STARFORGE

The STARFORGE[1] simulations are 3D radiation magnetohydrodynamic (MHD) simulations that track the evolution of giant molecular clouds (GMCs.), modelling the formation, growth, evolution, and dynamics of individual stars within a GMC. STARFORGE implements protostellar outflows, stellar winds, and supernovae as mass injection mechanisms. Sink particles inject new gas elements into the simulation domain through two mechanisms: local injection and cell spawning. Local injection involves distribution of fluxes (mass, momentum, and energy) among neighbouring cells in a weighted matter, whereas cell spawning refers to generation of new gas cells with finite mass at a specific rate, allowing the feedback around stars to be explicitly resolved. The protostellar outflows are modelled following the framework of Cunningham et al. (2011) and 30% of the mass of the accretion disk is redirected into the jet. Stellar winds originate from all main-sequence stars with masses above 2 $M_\odot$ and follow Eqs. 44–45 in Grudić, Guszejnov, Hopkins et al. (2021). All stars more massive than 8 $M_\odot$ go supernovae at the end of their lifetime (Eqn. 47 in Grudić, Guszejnov, Hopkins et al., 2021). Additionally, the simulations follow the emission of photons in five different wavelength bands (Hopkins, Grudić et al., 2020), namely Lyman-continuum radiation ($\lambda < 912\,\text{Å}$), far-UV/photoelectric ($912\,\text{Å} < \lambda < 1550\,\text{Å}$), near-UV ($1550\,\text{Å} < \lambda < 3600\,\text{Å}$), optical/near-IR ($3600\,\text{Å} < \lambda < 3\,\mu$m), mid/far-IR ($\lambda > 3\mu$m). For each simulation cell, the photon energy in each band is saved as a separate field. A comprehensive explanation of the numerical methods and validation tests is provided in Grudić, Guszejnov, Hopkins et al., 2021.

We used the set M2e4a2 of STARFORGE simulations (Table 1 of Guszejnov, Grudić, S. S. R. Offner et al., 2022), which follows the evolution of a GMC up to ~ 11 Myr while saving all the properties

---

[1] https://www.starforge.space

every 24.7 kyr. The GMC is initiated as a sphere with mass 2×10^4 $M_\odot$ and radius 10 pc surrounded by a warm diffuse gas in a 100 pc × 100 pc box. It evolves, collapses under self-gravity, and leads to formation of protostars (at 0.8 Myr) and main sequence stars with stellar winds (3.6 Myr), radiating photons at different wavelengths, dispersing most of the GMC before the first supernovae (9.8 Myr) (Guszejnov, Grudić, S. S. R. Offner et al., 2022). We restricted our analysis to the inner 60 pc of the simulation box, as it is sufficient to capture the ^{13}CO(2-1) emission in all snapshots.

7.3.2 SEDIGISM

The Structure, Excitation, and Dynamics of the Inner Galactic InterStellar Medium (Schuller, Csengeri et al., 2017; Schuller, Urquhart et al., 2021) is a survey of the inner Milky Way spanning 84 deg^2 ($-60^\circ \leq l \leq +18^\circ$; $|b| \leq 0.5^\circ$). It includes observations using the J = 2-1 transitions of ^{13}CO and C^{18}O. The survey data is provided as 77 PPV cubes, each covering approximately $2^\circ \times 1^\circ$ with a velocity range of −200 to 200 km s^{-1} and a pixel size of 9.5″. The first data release (DR1) comprises ^{13}CO observations with a full width at half maximum beam size of 28″ and 1-σ sensitivity of 0.8–1.0 K per 0.25 km s^{-1}. Duarte-Cabral, Colombo et al. (2021) constructed a catalogue of 10,663 MCs using this dataset. MCs were identified using the Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES) algorithm (v.0.3.2, Colombo, E. Rosolowsky, Ginsburg et al. 2015; Colombo, E. Rosolowsky, Duarte-Cabral et al. 2019). The well-resolved clouds with reliable distance estimates were classified as the science sample. We use MCs from the science sample that are at distances between 2.5 and 3.5 kpc as our observational benchmark in Sect. 7.5.1.

7.4 Methods

7.4.1 Data processing

In this section, we describe the process of generating synthetic observations from the STARFORGE simulations. This includes obtaining the input files for the RADMC-3D algorithm and using it to perform the radiative transfer. The pipelines for the two procedures are described in the following sections.

The `gizmo_carver` Pipeline

We used a python based pipeline `gizmo_carver (pschanges branch)`[2] to interpolate the simulation data onto a uniform grid. This step produces the RADMC-3D input files, including gas temperature, velocity, number densities of H_2 and ^{13}CO, and microturbulence.

Since STARFORGE does not track the abundance of CO, we computed it using the abundances of C, H, and O provided by the simulations. This calculation is carried out within `gizmo_carver` using `pytreegrav`[3] and `UCLCHEM` (Holdship et al., 2017). The first step is the calculation of the H_2 column density using `pytreegrav`. After specifying the snapshot number, `gizmo_carver` extracts the position, mass, volume density, smoothing length, and velocity of gas cells. `pytreegrav` uses these properties to compute the integral of the density field (to infinity) along a grid of rays originating at each particle, outputting the H_2 column density. The pipeline then constructs a chemical grid using

---

[2] `https://github.com/psharda/gizmo_carver/blob/pschanges`

[3] `https://github.com/mikegrudic/pytreegrav`

H_2 column density from `pytreegrav`, together with the volume density, interstellar radiation field, and H_2 abundance. Finally, `UCLCHEM` was used to follow the evolution of C-H-O chemistry on the grid under specific conditions to produce CO abundances.

We obtained the ^{13}CO number density from the CO abundances using a constant isotopic ratio of 42.6 (Jacob, Menten et al., 2020). To model the CO freeze out, we set the ^{13}CO abundance to 50% in regions with $T < 17$ K and $n > 10^5$ cm^{-3} (Caselli, Walmsley et al., 1999; Lippok et al., 2013; Roueff et al., 2021). We also included turbulent line broadening in the radiative transfer by estimating it from the local velocity gradient and cell size, with gradients computed using the `MESHOID`[4] script.

RADMC-3D

We used the line radiative transfer model (LGV + escape probability mode[5]) of RADMC-3D to generate ^{13}CO(2-1) synthetic observations. The resulting PPV cubes consisted of 65 velocity channels with a resolution of 0.25 km s^{-1} centred around 220.398 GHz (^{13}CO(2-1) rest frequency). To match the resolution of SEDIGISM observations at a distance of ~ 3 kpc, we created cubes of size $65 \times 448 \times 448$. For each snapshot, we computed the PPV cubes along three orthogonal lines of sight using the `phi = 90,0,0` and the `incl = 90,90,0` parameters to create a large sample.

7.4.2 Post-processing synthetic observations

We post-processed the output of RADMC-3D to produce more realistic representations of the PPV cubes, mimicking the observational characteristic of SEDIGISM . The cubes were first convolved with a gaussian beam of 28″ to replicate APEX-telescope beam convolution at 220.398 GHz. We added the noise from the first and last fifty channels of the SEDIGISM G305 cube to simulate realistic uncertainties. Following the procedure in Duarte-Cabral, Colombo et al. (2021), we spectrally smoothed the cubes to 0.5 km s^{-1} and subsequently resampled them to 0.5 km s^{-1}. A two-stage dilated masking procedure (Grishunin et al., 2024; Neralwar, Colombo, S. Offner, Karska et al., 2025) was also applied to suppress noise while preserving the real emission. This resulted in a total of 990 data cubes, each corresponding to a simulation snapshot with detectable ^{13}CO(2-1), representing the last 8 Myr of the simulation.

Structure segmentation

Dendrograms capture the hierarchical distribution of isosurfaces in data cubes and have been widely used to decompose molecular gas emission on different scales in both observations and simulations (Colombo, E. Rosolowsky, Duarte-Cabral et al., 2019; Duarte-Cabral, Colombo et al., 2021; S. S. R. Offner, J. Taylor, Markey et al., 2022; S. S. R. Offner, J. Taylor and Grudic, 2025). Therefore, to segment the emission into coherent structures, we constructed dendrograms (`astrodendro` E. W. Rosolowsky, Pineda, Kauffmann et al., 2008b) using the PPV cubes. This approach identifies *leaves, branches, and trunks,* providing a hierarchical catalogue of molecular gas structures. We further used the Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES) clustering algorithm (Colombo, E. Rosolowsky, Ginsburg et al., 2015; Colombo, E. Rosolowsky, Duarte-Cabral

---

[4] `https://github.com/mikegrudic/meshoid`

[5] `https://github.com/dullemond/radmc3d-2.0/blob/96d52c22c60325c0f867d107334e63f40d46672d/manual/sphinx/lineradtrans.rst#the-different-line-modes-the-lines_mode-parameter`

et al., 2019) to identify branches corresponding to physical structures. These SCIMES branches are used in Sect. 7.5.3 to study global trends between properties, feedback and lifetime of clumps.

Property estimation

For each identified dendrogram structure, we extracted the integrated properties, e.g. projected area, effective radius, velocity dispersion, and brightness temperature. The effective radius is defined as $R_{eff} = \sqrt{A/\pi}$, while the deconvolved radius as $R = \sqrt{R_{eff}^2 - R_{beam}^2}$, where $R_{beam} = 0.2$ pc is the physical size of the beam.[6] The cloud luminosity (L) was converted into a mass estimate as M_{lum} [M$\odot$] = αCO L [L$\odot$], adopting $\alpha_{\mathrm{CO}} = 22.43$ M $\odot$ (K km s^{-1} pc^{2})$^{-1}$, consistent with a ^{13}CO(2–1) conversion factor of $X^{13}\mathrm{CO}(2-1) = 1^{+1}_{-0.5} \times 10^{21}$ cm^{-2} (K km s^{-1})$^{-1}$ to match SEDIGISM measurements. From these values, we derived the surface mass density, $\Sigma = M_{lum}/A$ and the virial parameter, $\alpha_{vir} = 5\sigma_v^2 R/GM_{lum}$, assuming uniform spherical clouds.

We also obtained the number of stars, the true molecular mass, and the mass affected by the protostellar outflows and stellar winds by resampling these simulation data fields to the same grid as the one implemented by RADMC-3D. The mass affected by different feedback mechanisms was calculated using the feedback fractions of the cells saved in the simulation[7]. Additionally, we regridded the total photon energy of simulation cells in the five radiation bands (Sect. 7.3). These grids were used to determine the molecular mass, feedback affected masses, and photon energies for each dendrogram structure by summing over the quantities in pixels associated with the corresponding structure. The ratios of the feedback-affected gas mass to the total gas mass provide the outflow and wind feedback fractions for each dendrogram structure. To compare the photon energies of different structures, we divide them by the structure volume and obtain the radiation energy densities (referred to as feedback fractions hereafter). The feedback fractions for outflows, winds, lyman continuum, far-UV, near-UV, far-IR, and near-IR radiations are used to study the variations in clump properties with respect to stellar feedback in Sect. 7.5.3.

7.4.3 CHANDRAST - CHAsiNg DendRogrAm Structures over Time

We have developed a pipeline to track dendrogram-identified structures across consecutive snapshots until the end of their lifetimes. For simplicity, we refer to the structures selected for tracking as *current-clumps*.

For each current-clump in a given snapshot, we identify all the structures in the next snapshot that occupy overlapping voxels. These *next-clump candidates* are retained if the overlap satisfies two criteria (Fig. 7.1): (i) more than 20% of the next-clump candidate overlaps with the current-clump, and (ii) more than 20% of the current-clump overlaps with the next-clump candidate. These criteria remove clumps that marginally overlap with the current-clump to avoid incorrect associations.

Dendrograms often identify nested branches that represent nearly identical physical structures as different branches (Colombo, E. Rosolowsky, Duarte-Cabral et al., 2019). To avoid assigning multiple redundant matches, we select a single structure from the next-clump candidates that overlap with each other. For each overlapping pair in the candidates, we compare each of their similarity to the current-clump (Fig. 7.2). This includes a combination of the (i) intersection-over-union (IoU) of the

---

[6] For reference, the SEDIGISM beam of 28″ FWHM corresponds to 0.4 pc at a distance of 3 kpc.

[7] `https://starforge-tools.readthedocs.io/en/latest/data.html#common-data-fields`

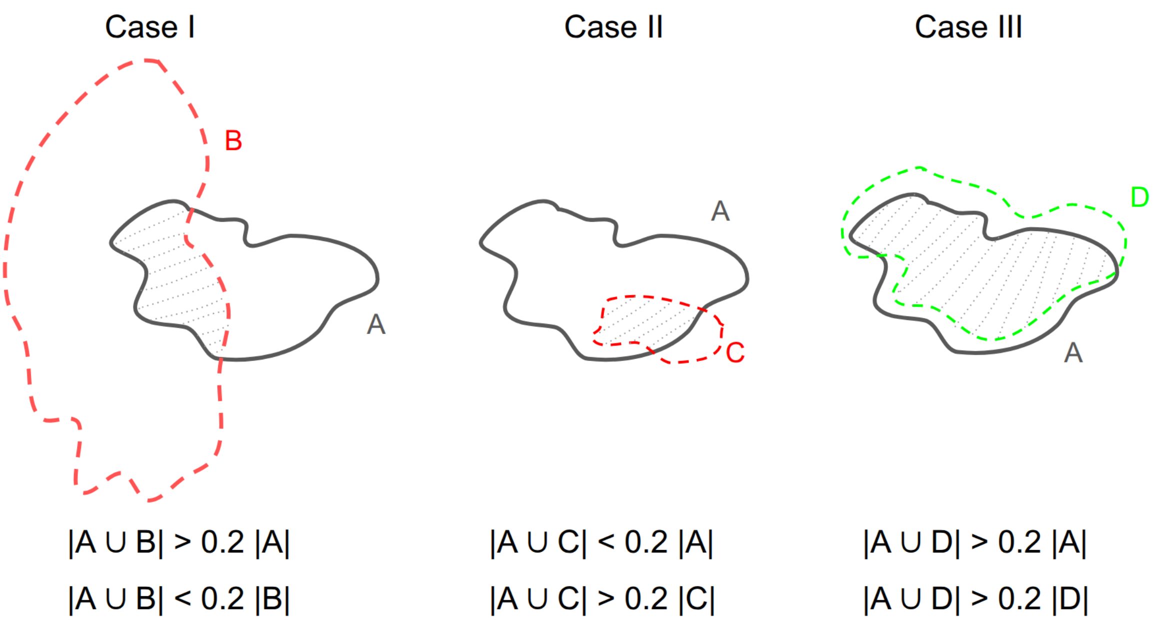


Figure 7.1: A cartoon depicting the selection criteria for the next-clump candidates. Clump A (black solid) represents the current-clump and clumps B, C and D (dashed red and green) represent the next-clump candidates. |A|, |B|, |C| and |D| represent the number of voxels for the respective clump. The grey dotted area represents the overlap between the original clump and a next-clump candidate. We only choose candidate D as the *next-clump* since it satisfies both criteria.

candidate's voxel masks with the current-clump and (ii) the relative difference between candidate's voxel counts and those of the current-clump. The IOU represents overlap between two structures and correlates positively with the similarity between the two. On the other hand, we expect similar structures to have similar sizes; therefore, the size difference correlates negatively with similarity. Contributions from IoU and voxel counts are weighted equally to calculate a similarity score (Eqn 7.1). The structure with the best similarity score is retained as the *next-clump*, resulting in a unique set of next-clumps for each current-clump.

Our approach enables the identification of the most similar clump without relying solely on its overlap or voxel count (Fig. 7.2). For a current-clump 'A' and a next-clump 'B', with the number of pixels represented by |A| and |B| respectively, the similarity score is calculated as

$$\text{Score} = \frac{|A \cap B|}{|A \cup B|} - 0.5 \times \frac{\text{abs}(|A - B|)}{|A|} \tag{7.1}$$

, where the first fraction represents the intersection over union, and the second fraction represents the relative difference in the number of voxels of A and B. The factor 0.5 signifies the relative importance of IOU compared to the difference between the voxel counts of the two structures.

To track a structure over multiple snapshots, we assign a tracer ID to each next-clump. If a single next-clump is found for a current-clump, it inherits the tracer ID of the current-clump. In Figure 7.3 (top), clumps 1 and 4 share the same tracer ID, as clump 4 matches uniquely with clump 1 without

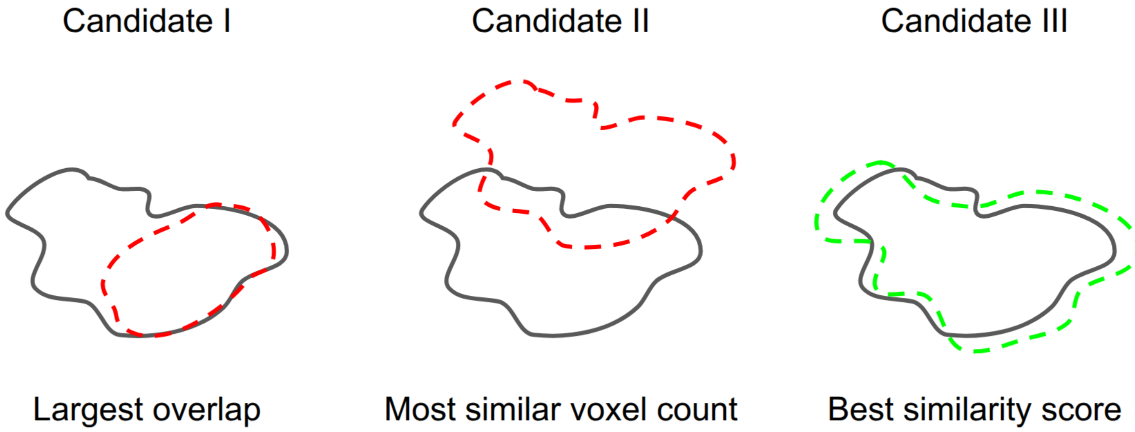


Figure 7.2: Selection of unique next-clump from multiple overlapping candidates. Candidate I, II and III represent the next-clump candidates that partially overlap with each other. Candidate I shows the highest overlap with the current-clump (solid black) while candidate II is the most similar in size. However, visual inspection indicates that candidate III provides the best overall match to current-clump. This is captured in the algorithm by computing a similarity score based on the IOU and the relative difference in voxel counts between the two structures.

overlapping with any other clump. Similarly, clumps 2 and 5 share the same tracer ID; and clumps 3 and 6 share the same tracer ID. If multiple next-clumps are identified for a given current-clump, this is interpreted as fragmentation, and each resulting fragment is assigned a new tracer ID. Figure 7.3 (centre) shows the fragmentation of clump 7 into clumps 8 and 9, each with a different tracer id.

After processing all current-clumps in a snapshot, we check if multiple next-clumps overlap with each other. If multiple current-clumps merge into a single clump, we assign a new tracer ID to the merged next-clumps and follow just the largest of them all. This can be better understood using Figure 7.3 (bottom row). CHANDRAST chooses clump 12 as the next-clump for clump 10 and clump 13 as the next-clump for clump 11. However, since clumps 12 and 13 overlap with each other, we consider this a merger and only follow the evolution of the larger clump, i.e. clump 12. This procedure is applied recursively across multiple consecutive snapshots until no further next-clumps can be identified. The resulting tracer IDs allow us to follow clumps in time, linking dendrogram structures consistently throughout the simulation.

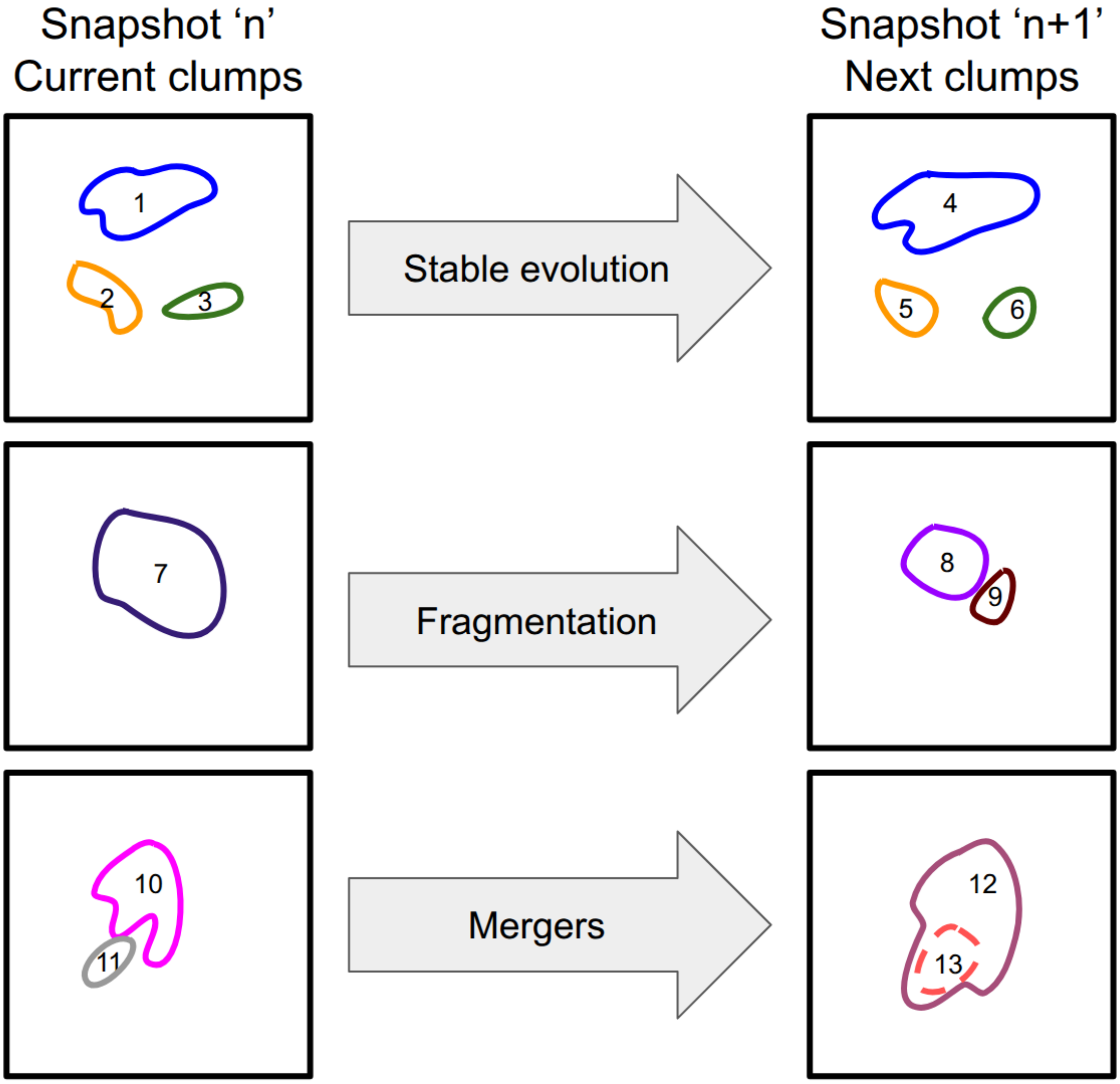


Figure 7.3: Three cases of clump evolution: stable evolution, fragmentation and mergers. The colors of clumps represent their tracer ids.

7.5 Results and Discussions

7.5.1 Scaling relations

Larson's and Heyer's scaling relations show the correlations between the physical properties of clouds. Larson's original analysis of MCs (Larson, 1981a) established a relation between cloud size and velocity dispersion. Later, Solomon, Rivolo et al. (1987b) revised it to form the relation $\sigma_v = 0.74\, L^{0.5}$ (discussed in Colombo, E. Rosolowsky, Duarte-Cabral et al., 2019). The radius and velocity dispersion of MCs following power laws are often interpreted as proof of universal cloud turbulence (Padoan, L. Pan, Haugbølle et al., 2016). It is a simplification of Kolmogorov's law for turbulence, indicating that larger clouds exhibit broader linewidth. However, Larson's laws by themselves do not explain if a cloud is virialised. M. Heyer, Krawczyk et al., 2009 combined these laws to compare the surface density of a MC with its radius and velocity dispersion, formulating Heyer's relation.

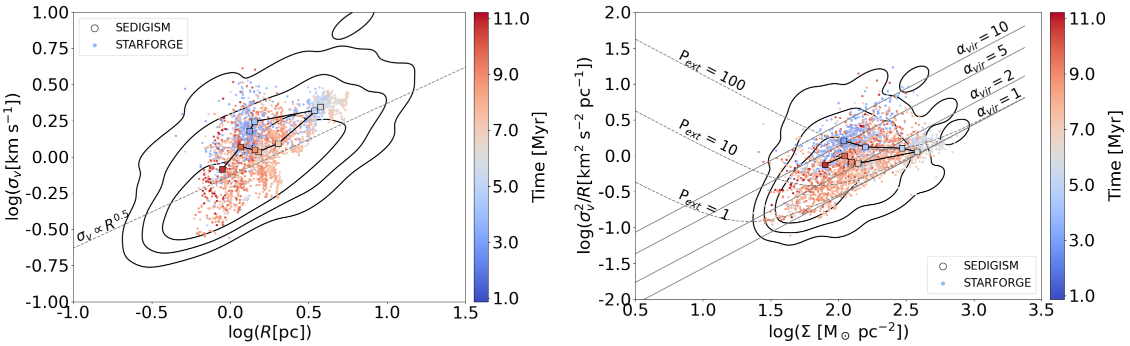


Figure 7.4: Scaling relations for molecular clouds (trunks). Left: The squares represent medians of distributions in ∼ 1 Myr bins. The black contours represent the 1σ, 2σ, 3σ levels for the SEDIGISM clouds. The dashed line represents Larson's first relation (Larson, 1981a; Solomon, Rivolo et al., 1987b). Right: Scaling relation between σ_v^2/R and surface mass density (Σ). The solid grey lines represent isocontours of virial parameters. The dashed lines represent $\alpha_{\mathrm{vir}} = 1$ when including an external pressure of $\mathrm{P_{ext}} = 1$, 10, and 100 $\mathrm{M_\odot\ pc^{-3}\ km^2\ s^{-2}}$.

We present Larson and Heyer scaling relations for dendrogram trunks, branches, and leaes in Fig. 7.4 - 7.5. In doing so, we check if the inclusion of CO chemistry (Sect. 7.4.1) affects the properties of the individual dendrogram identified structures, as well as the average temporal trends in the scaling relations. We also compare the distributions of trunks, branches, and leaves in the scaling relation plots to assess whether they represent similar structures and can therefore be studied collectively.

Figure 7.4 (left) shows an increase in the size and linewidth of structures until ∼ 6 Myr, followed by a sharp decrease in linewidth and a gradual decrease in size. These changes are reflected in the respective Heyer's relation plot (Fig. 7.4, right), with the early structures (< 6 Myr) changing from diffuse and unbound MCs to dense and virialised MCs. The 6 Myr marks a transition towards a decrease in surface density, with stellar feedback fragmenting and dispersing MCs (detailed discussion in Neralwar, Colombo, S. Offner, Karska et al. (2025); hereafter referred to as KN25). These features in the two plots show that despite the differences in individual structures of this work and KN25 (Figs. 8 & 9), the average temporal trends in the properties of both samples agree well with each other.

However, Fig. 7.4 shows fewer trunks compared to KN25, despite using the data-reduction pipeline. In our CO-chemistry included data cubes (Sect. 7.4.3), there is no detectable ^{13}CO emission present in the early snapshots (< 3 Myr). In addition, we do not detect some diffuse structures identified by KN25, which might represent CO-dark clouds and therefore fall below our detection threshold (Sect. 7.4.2). Although the inclusion of chemistry does not significantly alter average temporal trends on large scales (MCs), it leads to different estimates of the properties of individual structures. This suggests that modelling the proper abundances of chemical species is important for obtaining realistic estimates of the properties of individual molecular gas structures. This is especially important for the current study, as we analyse the evolution of individual molecular gas structures.

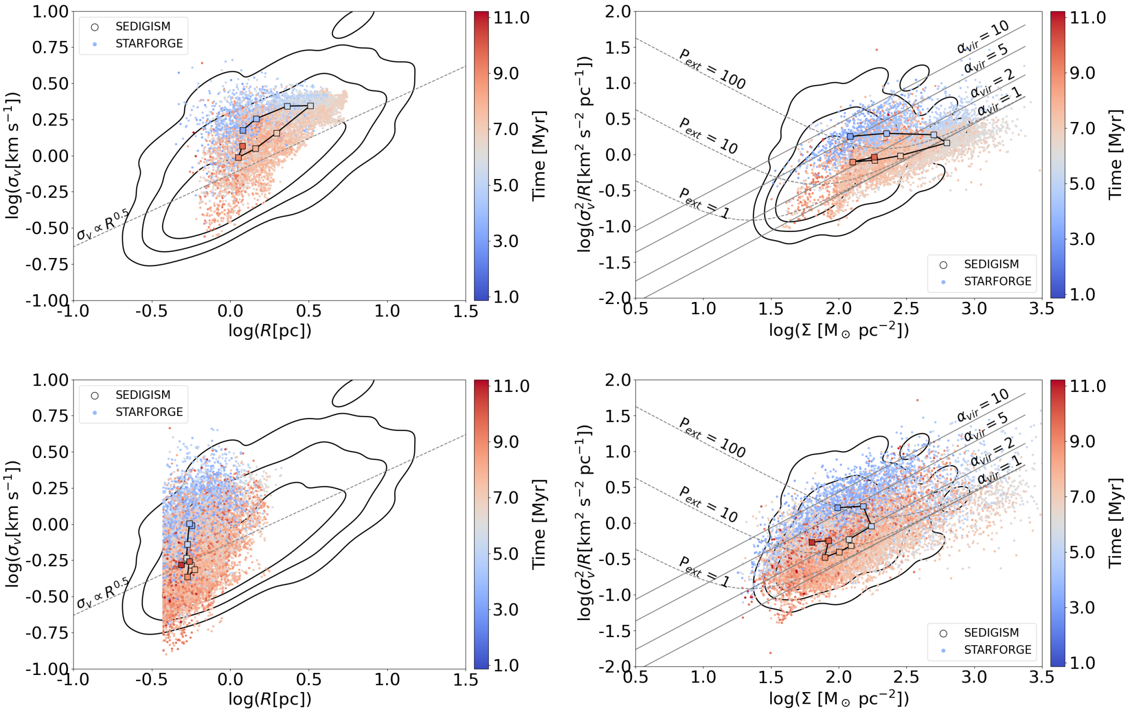


Figure 7.5: Larson (left) and Heyer (right) relations for dendrogram branches (top) and leaves (bottom). The black contours, symbols and notations are same as Fig. 7.4.

The trunks, branches, and leaves of the dendrogram occupy similar parameter spaces in the scaling relation plots and show similar trends over time (Figs. 7.4, 7.5). This is due to the hierarchical nature of dendrograms, which classify closely related emission structures at multiple levels: a structure identified as a trunk can contain substructures with nearly identical properties that are classified as branches or leaves, effectively tracing the same physical entity. As a result, large scatter is observed in the properties of the trunks, branches, and leaves (Figs. 7.4 & 7.5), with a large overlap between their distributions. The scatter is further increased by duplicate identifications of same structure across multiple snapshots and at different hierarchical levels. We therefore collectively refer to these structures as *clumps* in the next sections as we follow them over time.

The clumps embedded in the simulated GMCs evolve on shorter timescales compared to the lifetime of the GMC. Thus, the snapshot numbers do not represent their actual age. Studying the evolution of these clumps requires tracking individual structures over time while accounting for and removing duplicates. This is the main analysis of this work and is described in the following section.

7.5.2 Evolution of individual clumps

In this section, we use the CHANDRAST algorithm to follow the evolution of individual clumps over time. In the following sections, first show visualisations of structures and define them to understand the output of CHANDRAST , and then analyse the properties of a clump as it evolves.

Visualisations and definitions of structures

Figures 7.6 & 7.7 show the evolution of a clump identified at ~ 7 Myr after the start of the simulation. This clump (hereafter called clump 27526) initially evolves as a ring-like structure before fragmenting into two distinct structures at t ~ 7.01 Myr. Such ring-like morphologies and subsequent fragmentation are usually associated with stellar feedback (Deharveng, Zavagno et al., 2009b; Mazumdar et al., 2021; Neralwar, Colombo, S. Offner, Karska et al., 2025). This is consistent as t ~ 7 Myr represents the feedback-affected phase of the GMC (Sect. 7.5.1). After fragmentation, the bottom clump (orange) gradually contracts and drifts away from its original position over time (7.01 Myr < t < 10.18 Myr). The top clump (green) undergoes further fragmentation, producing multiple time substructures that

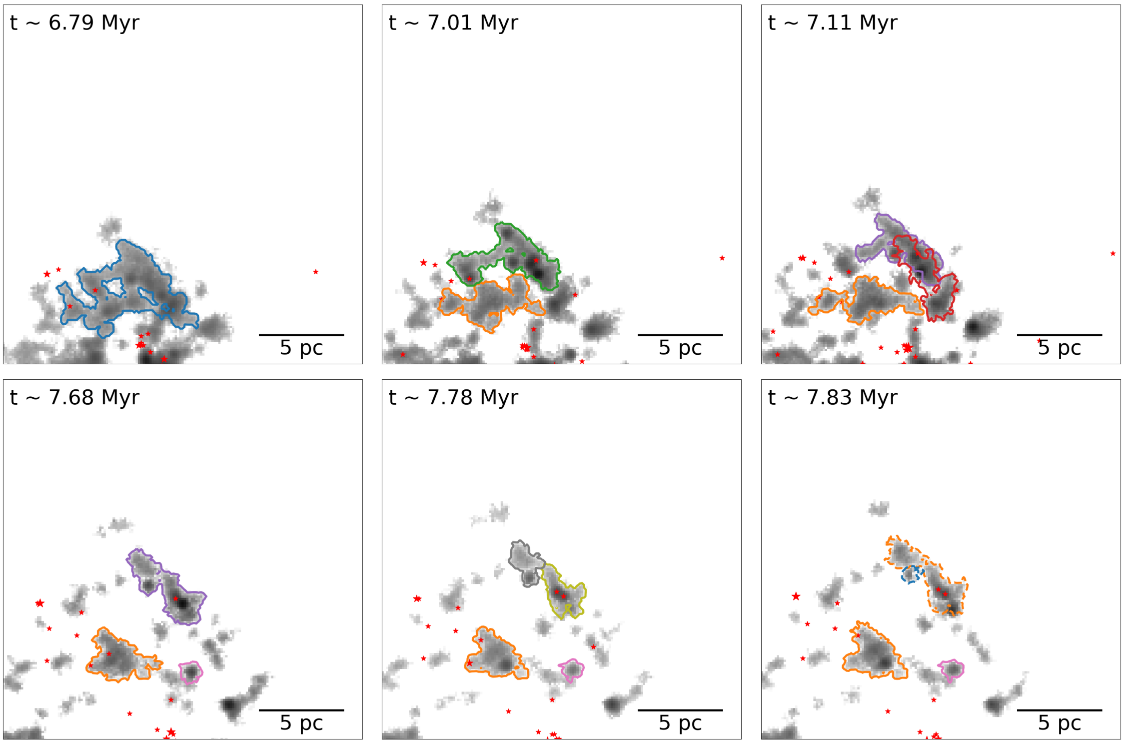


Figure 7.6: Moment 0 maps of the clump 27526 and its time substructures (greyscale). Only a subset of snapshots is shown here; a larger set of equally spaced moment maps is provided in App. C. The contours highlight the time substructures with different tracer ids. The colors of contours follow Fig. 7.8 and the solid and dashed contours are used to represent different structures of same color. The subplots show the same region of the simulation domain with same spatial scales to show the movement of the structure in the region. The red stars shows the positions of stars in the box, with their sizes representing the number on stars on the projected plane.

migrate away from the initial location of the parent clump. Some of these fragments contain sufficiently dense regions that collapse under gravity to form stars (Figs. 7.6 & 7.7). Momentum injected from stars increases local turbulence, while gravity reshapes the distribution of gas, collectively driving the observed motions of the clump and it's time substructures.

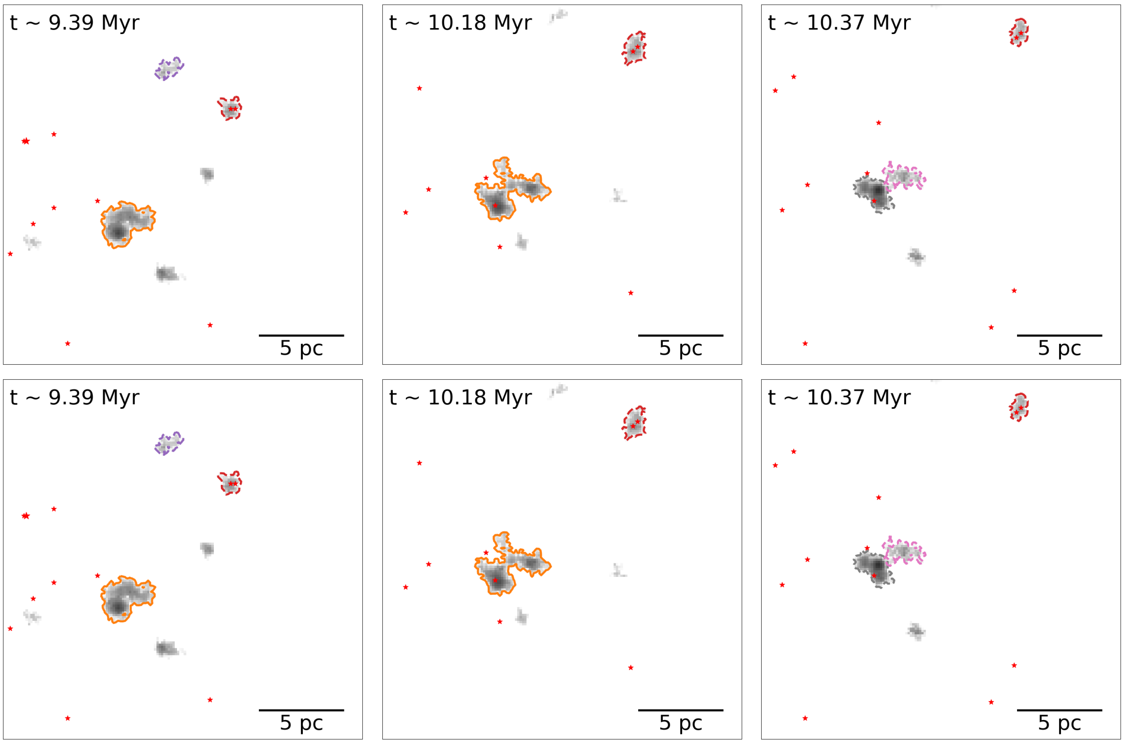


Figure 7.7: Continuation of Fig. 7.6

Following a clump over time allows us to construct a hierarchical map linking it to its counterparts in subsequent snapshots (Fig. 7.8). In the following, we refer to the counterparts of a clump in subsequent snapshots as its *time substructures* (Table 7.1). We will also refer to a clump and all its time structures together as a single *hyperclump*, as it represents a clump in position-position-velocity-time (4-dimensional) space (Table 7.1). The hierarchical representation (Fig. 7.8) provides a systematic way to associate the time substructures with each other. It enables direct tracking of fragmentations and mergers, showing the lifetime of each time substructure.

A significant difference between our tree diagram and the traditional dendrogram tree is the presence of mergers. Fig. 7.8 (bottom right) shows the splitting of structure 8 (grey) into structures 10 (blue) and 11 (orange). At the same time, structure 9 (olive) merges with structure 11 (orange). To be consistent with our definitions, the merged structure will be referred to as a time substructure of both the parent structures; i.e., structure 11 is a time substructure of both structure 8 and structure 9. A more intuitive visualisation of these mergers is provided in the moment maps in Fig. C.2.

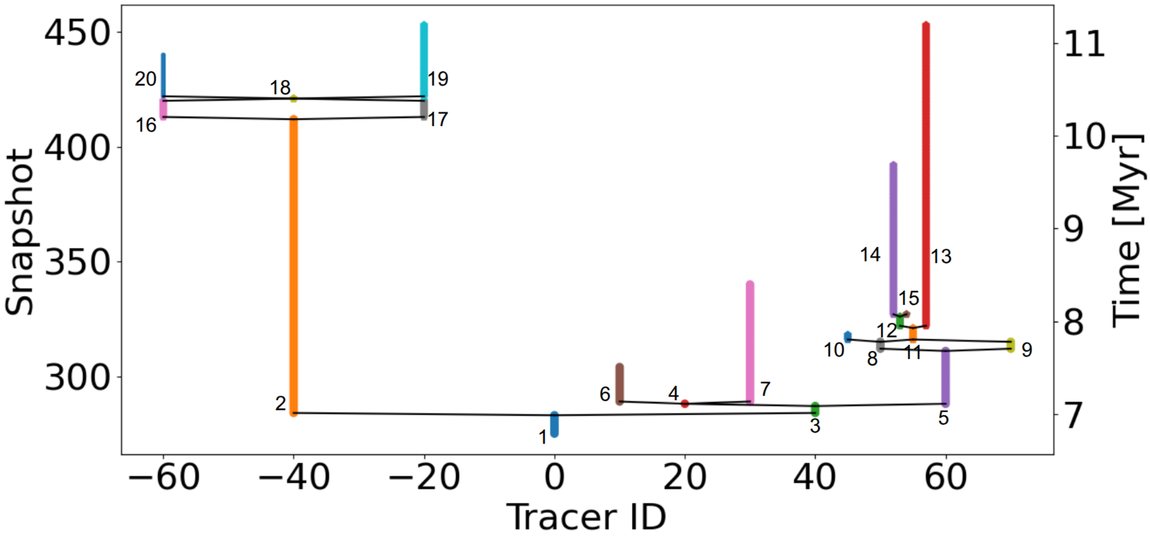


Figure 7.8: Evolution of the clumps in a hyperclump presented in the form of a dendrogram tree. Different colors represent the tracer ids of the clumps, as also noted by the x-axis. The y axis represents the snapshots in which the clumps are identified. The numbers in the figure represent a single trace (time substructures).

Structure	Description	Visualisation in Fig. 7.8
Clump	A structure in PPV space	Each point represents a clump.
Time substructures	Future counterparts of a clump	Individual traces, i.e., different colours, represent distinct time substructures.
Hyperclump	A structure in PPVt space	The entire figure represents a single hyperclump.

Table 7.1: Glossary of terms used in this work.

Evolution of properties

Figure 7.9 presents the evolution of the properties of hyperclump 27526 across multiple snapshots. Among the various properties, the radius and mass distributions have the most coherent and visually traceable evolutionary trends. This is mainly due to two reasons: (i) CHANDRAST following similarly sized clumps, and (ii) radius and mass representing the global structure of a clump.

One of the criteria used by CHANDRAST to match clumps in subsequent snapshots is the similarity in the number of their pixels (Sect. 7.4.3). As a result, the clump radii do not vary significantly between two consecutive snapshots. The mass distribution in Fig. 7.9 shows the true molecular gas mass (Sect. 7.4.2), which is free from observational biases. The close correlation between the mass and radius distributions indicates that the average gas density of these structures remains approximately constant during their evolution.

Radius and mass are likely to trace the global structure of a clump and are affected by a combination of several processes such as gravity, turbulence, and feedback (Colman et al., 2024). Gravitational

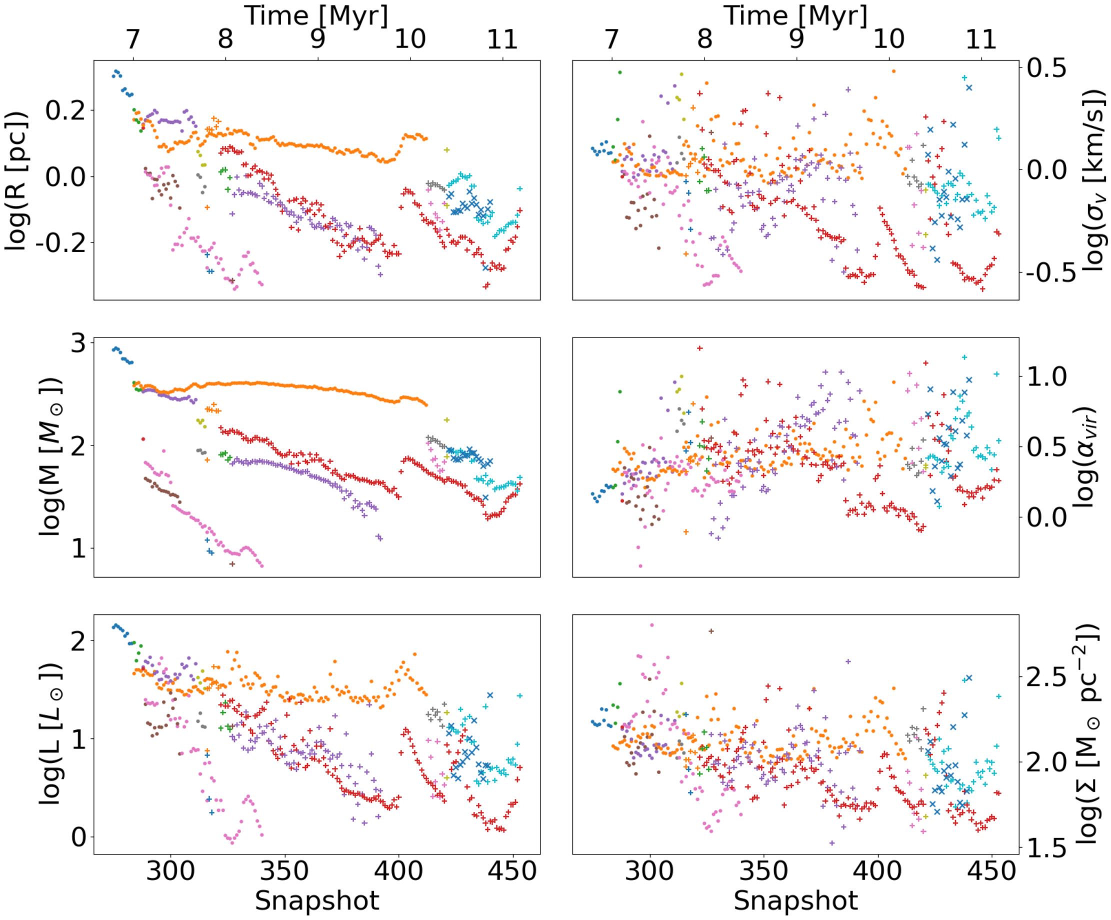


Figure 7.9: Evolution of properties of the clump 27526 and its time-linked structures. The symbols and notation follow Fig. 7.8.

collapse leads to a gradual decrease in the size of the clump as it becomes denser over time (Vázquez-Semadeni, Palau, Ballesteros-Paredes et al., 2019). Stellar feedback from nearby stars can erode the outskirts of a clump, causing a gradual decrease in radius until the clump is fragmented into multiple structures (Colín, Vázquez-Semadeni and Gómez, 2013). If a clump is part of an expanding shell driven by feedback, its effective radius may increase with time (Neralwar, Colombo, S. Offner, Wyrowski et al., 2024). We speculate that the processes typically affect the clump structure monotonically, resulting in smooth distributions of their radii and mass.

Observational studies using emission lines infer clump masses from the line luminosities. Figure 7.9 shows the luminosity distribution of the clumps, which corresponds to the ^{13}CO(2–1) emission. It is influenced by various observational biases, e.g., the presence of CO-dark gas can alter the correlation between gas density and ^{13}CO(2–1) luminosity (Madden et al., 2020). Thus, the luminosity distribution shows more temporal fluctuations compared to the radius and mass distributions. The surface density of the clumps is calculated using the luminosity mass and therefore the fluctuations in luminosity propogate to the surface density (Sect. 7.4.2). These fluctuations also contribute to the large scatter in the scaling relation plots (Fig. 7.5).

The velocity dispersion and the virial parameter can be considered as proxies of the internal structure of the clumps (M. H. Heyer and Brunt, 2004; X. Zhang et al., 2024; Asogwa et al., 2026). The velocity dispersion measures the inner dynamics of the clumps, while the virial parameter closely follows its evolution, as it scales with the square of the velocity dispersion. These properties show stochastic variations and larger scatters for individual time substructures, making their temporal evolution difficult to visually follow (Fig. 7.9). The constant fluctuations represent the constant thermal and turbulent gas motions within the clumps. The momentum from stellar feedback processes also contributes to the rapid changes in the internal structures of these clumps.

So far in this section, we have discussed the evolution of the properties of clump 27526 (and its time substructures). This clump represents a typical long-lived clump in the sample and provides a good representation of trends in clump properties (App. C.2). The trends in the properties discussed for the clump 27526 such as the stability of radius and mass, and the fluctuations in velocity dispersion, are observed for the majority of clumps.

7.5.3 Evolution of global properties

In this section, we analyse the properties of the structures using the time-averaged properties of the hyperclumps. Individual points in Figs. 7.10 & 7.11 represent the ranges and median values of the hyperclump properties measured over their lifetime.

Properties and lifetime

Figure 7.10 shows the distribution of the average properties of the hyperclumps as a function of their lifetime. These average properties are computed as the median values of the corresponding properties of the individual clumps that constitute each hyperclump. Long-lived hyperclumps are on average larger, more massive, luminous and have higher velocity dispersion compared to short-lived ones (Fig. 7.10). This trend is not a direct correlation, but rather describes an underlying causal relationship. Smaller structures evolve on shorter timescales than larger ones and can be rapidly dispersed by stellar feedback (Ballesteros-Paredes, André et al., 2020; Chevance, Krumholz et al., 2023). In contrast, large and massive clumps require stronger feedback and longer timescales to be significantly affected. The higher column density of larger clumps also allows them to self-shield more effectively, increasing their resistance to dispersal. Interstellar turbulence similarly influences smaller clumps more rapidly than larger clumps. As a result, large and massive clumps live longer.

The average values of the median surface densities do not change significantly over time (Fig. 7.10; bottom left). Although most of the hyperclumps have low surface densities, dense hyperclumps are present at all lifetimes. This suggests that the median surface density alone may not be a representative measure of the physical state of the clumps. Some of the most massive hyperclumps with high velocity dispersions have short lifetimes ($\leq$ 10 snapshots). These hyperclumps are also supervirial. This suggests that the large kinetic energy in these clumps rapidly tears them apart (visible in Fig. 7.7). The kinetic energy can be the result of the interstellar turbulence or momentum from stellar feedback. In the next section, we discuss the effect of individual feedback mechanisms on the average properties of hyperclumps.

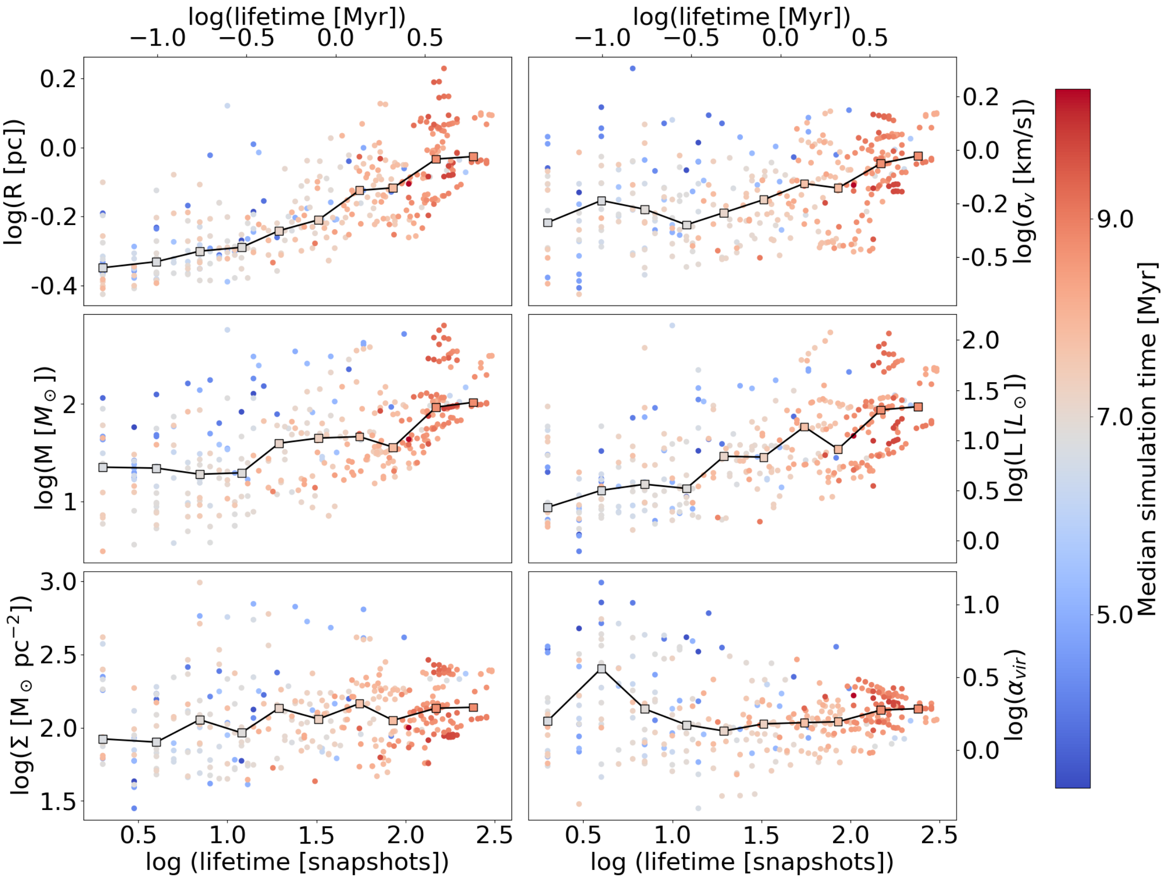


Figure 7.10: Median values of properties of hyperclumps (log scale) as a function of their lifetimes (log scale). The properties shown are radius (R), velocity dispersion (σ_v), molecular gas mass (M), luminosity (L), surface mass density (Σ) and virial parameter (α_{vir}). The hyperclumps are colored to represent the median value of all the snapshots in which it is identified, with the colors being consistent with Fig. 7.4. The squares and the solid black line represent the average distribution of the properties in lifetime bins.

Effects of feedback

Clumps are affected by various feedback mechanisms throughout their lifetime. In this section, we study the influence of protostellar outflows and stellar winds on the properties of hyperclumps. We also study the correlation between hyperclump properties and interstellar radiation in separate bands (Fig. 7.11).

A large number of hyperclumps that are significantly affected by protostellar outflows are long-lived structures (Fig. 7.11). Since strong outflow feedback is expected to be primarily local rather than from external sources, these hyperclumps likely have embedded protostars. This behaviour can be understood using different scenarios of clump evolution. Small, low-mass clumps are rapidly dispersed by stellar feedback and turbulence before they can form protostars and stars. Large and massive clumps evolve over longer timescales, allowing the formation of protostars and subsequent protostellar outflows. We further discuss this in App. C.3, by analysing the evolution of the properties of individual clumps in hyperclumps marginally and significantly affected by outflows.

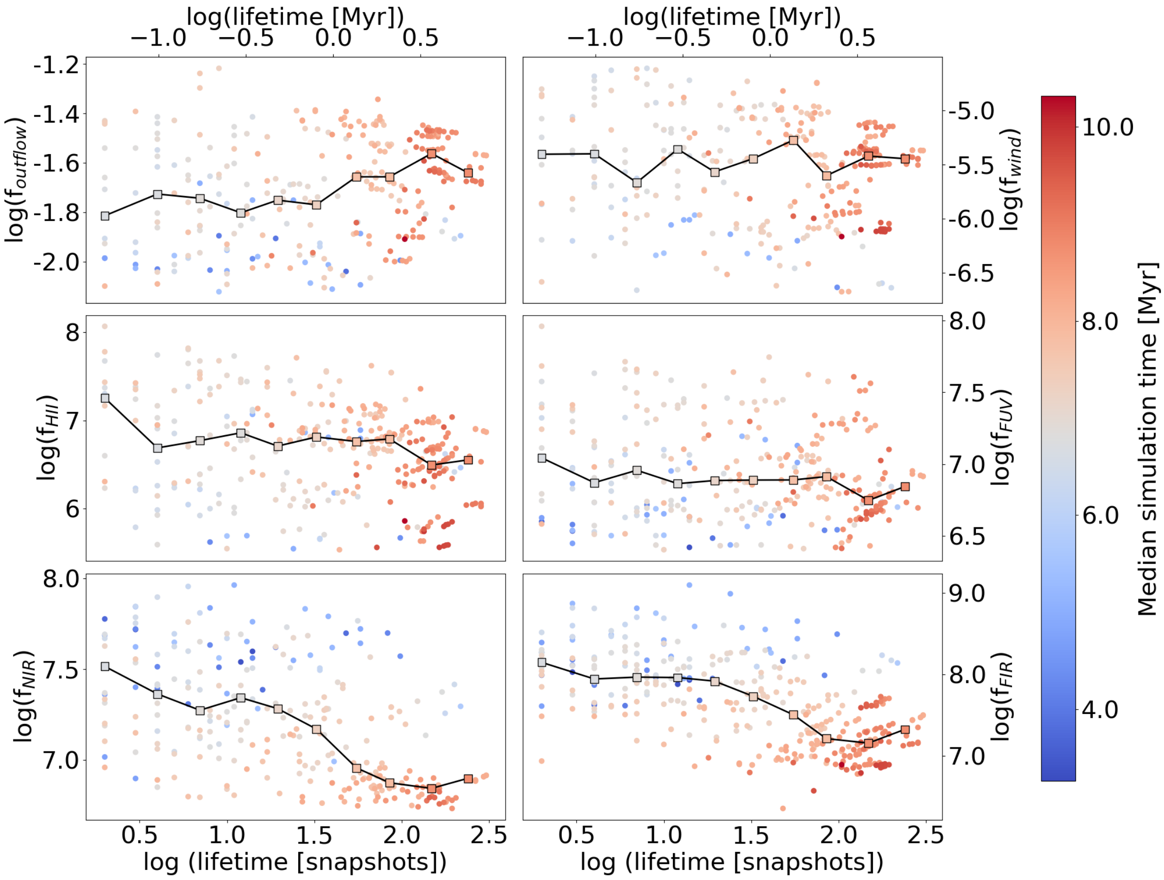


Figure 7.11: Median values of feedback fractions of hyperclumps as a function of their lifetimes. The hyperclumps are colored to represent the median value of all the snapshots (simulation time [Myr]) in which it is identified, with the colors being consistent with Fig. 7.4. The squares represent the average value of the feedback fractions in lifetime bins. As the NUV and FUV distributions are similar, only the FUV is shown for visual consistency.

Stellar winds, photoionisation, and FUV radiations do not show any significant trends with hyperclump lifetimes (Fig. 7.11). In contrast, hyperclumps associated with strong far-infrared (FIR) emission are typically short lived. FIR emission traces reprocessed UV radiation absorbed by dust and therefore indicates regions that have been significantly impacted by radiation. Radiative feedback efficiently disperses gas, resulting in small and low-mass FIR-bright clumps, which likely represent residual gas in radiation-dominated regions (App. C.3). These structures are rapidly dispersed by interstellar processes. Conversely, hyperclumps with marginal FIR emission likely evolve in regions that are not significantly affected by radiation feedback. Thus, they are able to survive longer in these relatively quiescent environments.

7.6 Caveats

To include the proper CO abundances in our synthetic observations, we used the `gizmo_carver` `(pschanges branch)` pipeline (Sect. 7.4.1). Although the code documentation recommends usage of 100 random rays to reduce projection noise, we used 10 rays due to computational costs. This choice reproduces the global column-density structure but might not fully resolve the smallest structures in the ISM. It is also possible that we miss some real MCs, as the H_2 column density in `pytreegrav` is computed using just 10 rays, which may leave some pixels not sufficiently sampled. This is indeed a caveat; however, our data reduction includes convolving the synthetic observations with the telescope beam, which smoothes the data and filters out some of the smaller structures. The comparison of trunks in this work and KN25 shows that although using different number of rays in `pytreegrav` causes individual clouds properties to differ, the reduced ray sampling does not affect our overall analysis significantly (Sect. 7.5.1). Therefore, we do not expect significant changes in our results because we use a lower number of rays.

CHANDRAST follows the evolution of clumps by matching the most similar dendrogram structures in the neighbouring snapshots (Sect. 7.4.3). The similarity metric combines the maximum overlap between two structures and the difference in their sizes (Sect. 7.1). Even in case of fragmentations and mergers, all next structure candidates are compared to the current clump (definitions in Sect. 7.4.3). Since the selection of next structures relies partially on the size similarity, the algorithm may occasionally select an incorrect fragment or merger. This occurs when the code selects the next structure candidates whose sizes are closest to those of the current clump, even though physically the fragments are expected to be smaller (Fig. 7.2). In future work, we plan to improve the selection criterion by incorporating multiple properties such as size, mass, velocity dispersion, into the similarity metric. However, since our analysis focusses on the evolution of time-averaged properties of hyperclumps, we do not expect the improvements in algorithm to cause significant changes in our results.

Minor differences in distribution of flux can cause the dendrograms to identify a specific emission feature as a single clump in one snapshot and as multipe clumps in another. This is evident in Fig. 7.8 (corresponding moment 0 maps in Fig. C.7): structures 16 (pink) and 17 (grey) are identified as a single structure 18 (olive) in a snapshot ($t \sim 10.4$ Myr) and are subsequently identified as two distinct structures 19 (cyan) and 20 (blue). We term this a pseudo merger. In the future, this would be minimised by checking if a merged structure is being identified as a single structure over multiple subsequent snapshots before deciding if the two parent structures have truly merged. In addition, since CHANDRAST is used at multiple snapshots, some clumps may be traced back to earlier snapshots, resulting in multiple clumps sharing the same time substructures. This can introduce bias in the calculations of average properties and property ranges of the clumps in Sect. 7.5.3. Future releases of the code will identify and remove such duplicates. However, these improvements are beyond the scope of this work.

7.7 Summary

In this work, we have developed the algorithm – CHANDRAST – to follow the evolution of individual molecular gas structures in synthetic molecular-line observations. We used the RADMC-3D radiative transfer code to generate synthetic observations of a 20 000 $M_\odot$ GMC from the STARFORGE simulations. From these, we identified molecular gas clumps using a dendrogram analysis and used CHANDRAST to match the clumps to their future counterparts creating hyperclumps. We further generated hierarchical maps of hyperclumps along the time axis and analysed the evolution of properties of individual hyperclumps. We also studied the trends in hyperclump properties as a function of their lifetimes and examined the effects of individual stellar feedback mechanisms on clump evolution.

We present the Larson's and Heyer's relation for synthetic trunks, branches, and leaves to understand the impact of including CO chemistry on their properties and evolution. Although the inclusion of chemistry does not affect the overall temporal trends in cloud properties, it affects the properties and detectability of individual structures. Trunks, branches, and leaves occupy overlapping regions of parameter space in the scaling relation plots and show similar evolutionary trends, suggesting the identification of similar structures by dendrograms at different hierarchies. Thus, we collectively refer to all structures identified by dendrograms as clumps.

We follow the evolution of properties of a typical long-lived hyperclump (27526) separating the individual traces of clumps and their time-substructures. Radius and mass show the most coherent and smooth evolutionary trends, which is partly due to the matching criteria of CHANDRAST based on size. These properties trace the global structure of the clumps, which is monotonically affected over time due to a combination of gravity, turbulence, and stellar feedback, unless in the event of fragmentation and mergers. The distributions of luminosity and surface density of clumps show more stochastic variations compared to radius, since these properties are strongly affected by observational biases such as the presence of CO-dark gas. The velocity dispersion and the virial parameter show the most fluctuating evolutionary behaviour. These properties trace the internal dynamics of the clumps, which changes continuously as a result of the turbulent gas motions and momentum from stellar feedback.

Most of the average properties of hyperclumps show trends as a function of their lifetimes. Long-lived clumps are larger, more massive, and have velocity dispersions higher than their short-lived counterparts. This is expected as smaller and less massive structures evolve and disperse rapidly due to various interstellar processes, whereas larger and massive clumps self-shield due to higher densities and are more resistant to dispersal. The average surface densities of hyperclumps do not show significant trends as a function of lifetime, suggesting that this property alone does not fully capture the physical state of clumps. We also find a sample of massive hypeclumps with high velocity dispersion, virial parameters, and shorter lifetimes, suggesting their rapid dispersal by intense turbulence or feedback. Of the various stellar feedback mechanisms, protostellar outflows and infrared radiation appear to be strongly correlated to clump lifetimes. Long-lived clumps being associated with higher outflow feedback fractions is consistent with the formation of protostars and stars forming in large, massive clumps over time. In contrast, strong infrared emission traces short-lived clumps, as radiation dominated structures are rapidly dispersed compared to those in quiescent environments.

In conclusion, we have generated ^{13}CO(2-1) synthetic observations from the STARFORGE simulations, identified molecular gas structures in them using the dendrogram analysis and followed their evolution using the newly developed CHANDRAST algorithm. We find that on average, long-lived clumps are large, massive structures that are often associated with the presence of protostellar outflows. On the other hand, short-lived structures are small and less massive structures, often present in radiation-dominated regions and dispersed rapidly under the influence of gravity, turbulence, and stellar feedback.

CHAPTER 8

Machine learning approach to identify signatures of feedback in observations

8.1 Context

Stellar feedback mechanisms create distinct features in the ISM: protostellar outflows drive bipolar jets, stellar winds and radiation create parsec-scale cavities, and supernovae produce large structures known as supernova remnants (Sect. 2.4). The bubbles created by stellar winds are ubiquitous in galaxies and have been studied extensively (Sect. 2.2.3). Several citizen science projects have identified bubbles and bow shocks using observational datasets (e.g. using Spitzer surveys Simpson et al., 2012; Jayasinghe et al., 2019). However, visual identification of structures is extremely time consuming. It is also extremely challenging to associate a given bubble with a particular physical process, e.g. wind-driven shells and bubbles produced due to interstellar turbulence may appear the same. Moreover, the presence of irregular bubbles such as broken shells due to anisotropic stellar winds introduces human biases and inaccuracies in bubble identification. These issues result in unreliable property estimates for bubbles (D. Xu, S. S. R. Offner et al., 2020a). In this chapter, we present an artificial intelligence based method to systematically identify and analyse wind-driven shells. We train the neural network algorithm CASI-3D on the STARFORGE simulations to identify the signatures of stellar-wind in SEDIGISM data. Once trained, the algorithm presents a fast and unbiased method for detecting wind-driven bubbles in large observational datasets. This also helps to achieve the goal of the thesis of connecting simulations and observations of molecular clouds to improve our understanding of the interactions between stellar feedback and molecular gas. The analysis in this chapter is preliminary work that will be followed up in the future.

8.2 The algorithm CASI-3D

Convolutional neural networks (CNN) are an excellent tool for identifying features in images (Sect. 4.2.2). Van Oort et al., 2019 developed a CNN based architecture “Convolutional Approach to Shell Identification” (CASI) to identify stellar wind bubbles in 2D CO emission and gas density slices. The algorithm uses a U-net (Sect. 4.2.3) and Res-net (Sect. 4.2.4) based architecture to segment images. D. Xu, S. S. R. Offner et al., 2020b modified the CASI architecture to operate directly on PPV cubes by replacing 2D convolutions with 3D convolutions. This enables training of the neural network on both spatial and spectral features simultaneously. CASI-3D consists of two parts, an encoder that extracts important features, and a decoder that reconstructs these to obtain a feature map (Sect. 4.2.3). The input data cube undergoes three down-sampling in the encoder part, with the filters in the first layer (filter size = $7 \times 7 \times 7$) extracting the large scale features, and the filters in the subsequent layers (filter size = $3 \times 3 \times 3$) capturing the finer details of the bubble morphology in PPV space. The first layer consists of 32 filters, which are doubled after each average pooling layer, following the general architecture of U-nets (Sect. 4.2.2).

8.3 Model training and preliminary analysis

8.3.1 Model ME1

D. Xu, S. S. R. Offner et al. (2020b) developed two strategies to identify stellar wind bubbles. The first strategy "ME1" is based on semantic segmentation of the bubble structure to determine the position of the feedback. The second "MF" predicts the fraction of mass coming from stellar feedback to estimate bubble masses. The ME1 model was trained to predict the pixel-by-pixel bubble position using the mean squared error (MSE) between the predicted output ($\mathrm{Y_{pred}}$) and the ground-truth tracer $\mathrm{Y_{tracer}}$ as the loss function: $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Y_{pred}} - \mathrm{Y_{tracer}})^2$. Both strategies were trained on synthetic observations derived from the ORION simulation (S. S. R. Offner and Arce, 2015), which models the formation of stars in a $5 \times 5 \times 5$ pc region of a turbulent molecular cloud. The simulated gas density fields were post-processed using RADMC-3D to generate ^{13}CO(1-0) emission cubes, which served as input to the neural network. Ground-truth cubes representing stellar feedback bubbles were created from the emission in pixels where $\geq 2\%$ mass was affected by the stellar feedback and the gas temperature was $\geq$ 12K. To reduce false positives, negative training cubes were included to prevent CASI-3D from identifying large scale shells and bubbles that could potentially be due to turbulence. The trained models were subsequently applied to the observational data of the Taurus molecular cloud to identify wind-driven bubbles.

We applied CASI-3D (ME1) to SEDIGISM data to predict stellar wind bubbles in the well studied feedback region RCW 120 (Sect. 2.2.3). This was a proof-of-concept study to assess whether CASI-3D can detect stellar wind signatures in PPV cubes that differ substantially from its training set. Figure 8.1 shows that the model identifies a small bubble as a wind-driven structure, but does not recover the entire shell of RCW 120. This outcome is expected, since ME1 was trained to detect small bubbles within Taurus molecular cloud rather than large-scale structures (D. Xu, S. S. R. Offner et al., 2020b). Despite the differences in resolution, sensitivity, and physical environment between the training data and SEDIGISM, the model highlights the localised regions of interest. The results demonstrate the ability of CASI-3D to identify bubbles in unseen regions of the SEDIGISM dataset. This is critical for this work, since although STARFORGE simulations can produce MCs similar to observations, they are not able to reproduce the entire diversity of MCs in observations (Chapter 6).

8.3.2 Model ME$_{star}$

To identify large scale feedback bubbles, we trained CASI-3D on synthetic observations derived from the STARFORGE simulations. We used the M2e4a1 simulation set, which follows the evolution of a 20 000 $\mathrm{M_\odot}$ GMC initialised with a virial parameter of 1 (detailed description in Guszejnov, Grudić, Hopkins et al., 2021). The input training cubes with ^{13}CO(2-1) emission were generated using the

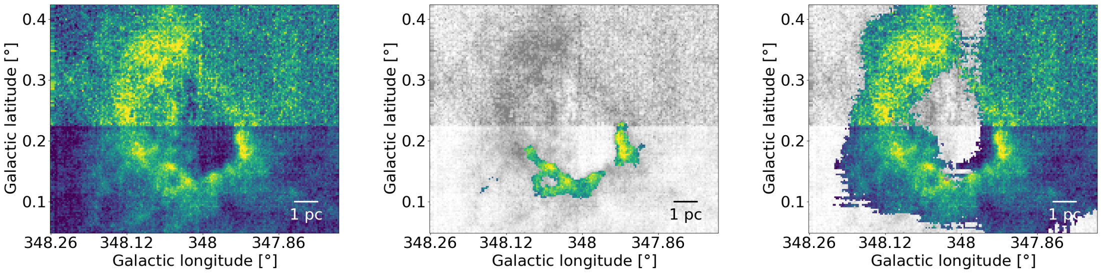


Figure 8.1: Left: The bubble RCW 120 in the SEDIGISM survey. Centre: Stellar wind prediction from model ME1. Right: Stellar wind prediction from model ME_{star}.

radiative transfer pipeline described in Chapter 6. The corresponding ground-truth PPV cubes were constructed by masking pixels that do not contain gas affected by stellar winds. This was enabled by STARFORGE tracking the fraction of mass influenced by outflows, stellar winds, and supernovae in each simulation cell (Chapter 5). The resulting ground truth cubes therefore represent regions affected by stellar winds (feedback fraction > 0; Sect. 3.2.2) and even marginally feedback-affected regions are included in the ground truth. The CASI-3D model trained on these data is hereafter referred to as ME_{star}. Figure 8.1 (right) presents the prediction on RCW 120 using ME_{star}. The algorithm successfully recovers the entire shell of the bubble, but also predicts the presence of wind affected gas in the surrounding noisy regions. This is a consequence of the low threshold on stellar wind feedback fraction when constructing the ground truth, which identifies even weakly feedback-affected regions. However, as a proof-of-concept test, the results demonstrate that CASI-3D can be trained on synthetic observations from STARFORGE to identify large-scale feedback bubbles.

8.4 Current limitations and future plans

Following preliminary analyses, we constructed new ground-truth cubes by adopting a higher threshold for the stellar wind feedback fraction[1]. However, because of the rapid mixing of gas in STARFORGE simulations, individual pixels can lie just above the threshold, while neighbouring pixels fall just below it. This creates artificial fragmented features that do not represent the feedback affected MCs in the observational data (Fig. 8.2). Despite extensive hyperparameter tuning and architectural modifications, CASI-3D fails to train on these artificial features. To address this limitation and identify realistic MCs affected by feedback, a more physically motivated constraint is needed while creating ground-truth cubes. In the future, ground-truth cubes will be constructed using feedback-affected MCs identified in Chapter 6. These MCs identified using dendrograms will provide clearer features for CASI-3D to learn and enable effective training.

The use of CASI-3D on the entire SEDIGISM data will result in an extensive and homogeneous catalogue of clouds, clumps, and cores affected by the individual feedback mechanism. This will help us answer how distinct feedback mechanisms contribute to structure formations and cloud

---

[1] Feedback fractions are defined in Chapter 5

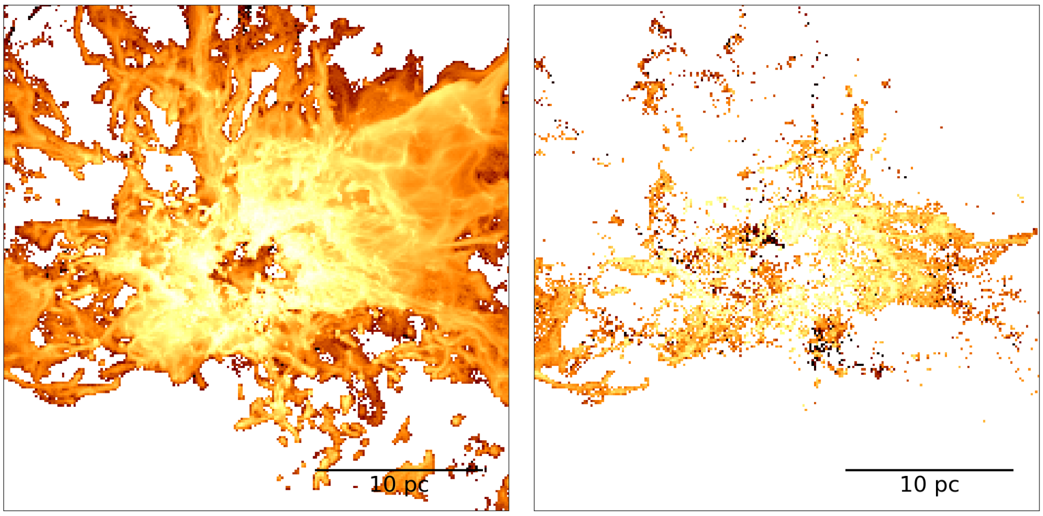


Figure 8.2: Projected gas density for a STARFORGE m2e4a2 simulation snapshot for all pixels (left) and for wind affected pixels using a wind feedback fraction threshold of 0.0001 (right).

evolution, which is challenging to disentangle in observational data due to overlapping effects from other processes such as turbulence. In addition, a range of structural properties will be analysed using the catalogue, including densities, opacities, luminosity-to-mass (L/M) ratios, line profiles, and excitation temperatures. These properties are used in observational studies to connect to different physical phenomena. For example, an increase in linewidth in a region is considered a sign of feedback, whereas L/M ratios are indicators of how evolved a gas structure is. Providing ground-truth labels of feedback versus non-feedback regions from simulations will therefore allow us to validate the results against observational findings, such as increased fragmentation in feedback regions (Mazumdar et al., 2021). Subsequently, the model will be used to identify bubbles in the outer Galaxy using the OGHReS (Outer Galaxy High Resolution Survey Colombo, König et al., 2021). In addition, the algorithm would be trained to detect signatures of other feedback mechanisms, such as bipolar jets from protostars and supernovae remnants.

CHAPTER 9

Summary and Outlooks

9.1 Summary and conclusions

This thesis investigates the impact of various stellar feedback mechanisms on the structure, properties, and evolution of molecular gas structures at different spatial scales. We begin by analysing structures in the simulation domain using molecular (H_2) gas density position-position-position cubes. Subsequently, we produce synthetic observations from the simulations and identify clouds and clumps within them. We study the evolution of clouds by associating them with a snapshot number (corresponding to a particular evolutionary time), whereas the clumps are tracked over multiple snapshots using a self-developed algorithm. Finally, we present preliminary work using machine-learning techniques to identify feedback affected structures in observations by training on synthetic datasets. Together, these studies pave the way towards a complete understanding of the effects of feedback on molecular gas in simulations and extending this knowledge to observations.

Studying the evolution of molecular clouds solely through observations is extremely challenging, as it cannot be followed in real time. Moreover, it is difficult to isolate the effects of individual stellar feedback mechanisms on the molecular gas in observations, since multiple feedback processes act together and produce similar signatures. For this reason, we use STARFORGE simulations that model the evolution of a giant molecular cloud (GMC), while including gravity, turbulence, magnetic field, and thermodynamics. These simulations resolve GMCs to scales of up to a few AU, enabling an extremely detailed study of clouds, clumps, and cores. In addition, they implement protostellar outflows, stellar winds, radiation, and supernovae, and track the gas affected by these mechanisms throughout the simulation at all times. Over the course of this work, molecular gas structures are identified using the hierarchical clustering algorithm `astrodendro`, with minor modifications to the source code as necessary. This allows us to investigate the impact of feedback on molecular gas from core scales, where feedback is generated, to cloud scales that represent the largest molecular gas structures, thereby improving our interpretation of observed feedback signatures in galaxies.

In Chapter 5, we leverage the high resolution of the simulations to extract sub-parsec molecular gas structures, i.e., cores, from the molecular gas density maps of the GMC. Following this, we quantify the impact of protostellar outflows, stellar winds, and supernovae on these cores by propagating the feedback fractions from individual simulation cells to the cores. We bin the cores based on the degree to which they are affected by individual feedback mechanisms, and create separate bins to analyse the general effects of stellar feedback. This enables the analysis of core radius, mass, velocity dispersion, and virial parameter across populations that are weakly, moderately, or strongly influenced by different stellar feedback mechanisms. The main findings of this work are as follows:

- The primary impact of stellar feedback on cores appears in their velocity dispersion and virial parameter distributions.
- The core populations associated with higher feedback fractions exhibit higher velocity dispersion. This occurs for each feedback mechanism including their combined effects (global feedback). We attribute this increase to the momentum injected into the ISM by various feedback mechanisms.
- The higher virial parameter for feedback-affected cores suggests that these cores are generally not gravitationally bound but rather pressure-confined structures.
- Individual feedback mechanisms influence the size and mass of the cores differently.
- On average, cores affected by protostellar outflows and stellar winds appear smaller, which is likely due to feedback-driven compression and dispersion of molecular gas. In contrast, supernovae-affected cores are larger, as a result of getting expanded in supernovae-driven shells. This behaviour is also a consequence of supernovae occurring at a late stage in STARFORGE, after most of the GMC is already dispersed by feedback.
- The mass of cores is not significantly affected by the various feedback mechanisms, with the mass-feedback trends being consistent with those of core radii.
- Pristine cores show a tighter fit to the Larson and Heyer scaling relations compared to their feedback-affected counterparts, indicating that feedback contributes significantly to the scatter in these relations.

Although simulations allow us to study the properties of MCs, these properties are not directly accessible by line- and continuum-emission observations of real MCs. In Chapter 6, we use the RADMC-3D radiative transfer code to generate synthetic observations from the STARFORGE simulations. These ^{13}CO(2-1) cubes are generated using the SEDIGISM survey as an observational benchmark. From these synthetic datacubes at different snapshots, we identify MCs using hierarchical clustering and analyse their observable properties as they evolve. After comparing the flux distributions of the two datasets to check if they are in agreement, we study the evolution of the synthetic MCs over multiple snapshots. Our main findings from this project are as follows.

- Synthetic MCs show physical properties comparable to those of their counterparts in SEDIGISM , validating the realism of the synthetic data. This agreement is supported by both samples occupying the same parameter space on Larson and Heyer scaling relations.
- MCs at different evolutionary stages occupy distinct regions of the scaling-relation parameter spaces, suggesting that the ages of MCs play a significant role in driving the observed scatter in these relations.
- The early evolution of MCs is primarily governed by gravity and turbulence. In the simulations, MCs form as small, diffuse, starless structures generated by turbulent fluctuations that grow through gravitational accretion.

- As evolution proceeds, MCs grow into dense, filamentary structures with embedded structures and active star formation. These correspond to most of the observed MCs in galaxies.
- The formation of stars gives rise to various stellar feedback processes. Once feedback becomes dominant, it drives cavities into MCs, making them appear as 3D bubbles, and eventually dispersed them.
- Overall, MCs evolve from small, diffuse structures to dense filamentary structures and finally to bubble-like structures that are eventually dispersed. The major changes in MC properties, morphology, and fragmentation strongly correlate with the onset of stellar feedback.

In Chapter 7, we exploit the ability of hierarchical clustering to identify structures across different spatial scales, and investigate the evolution of clumps embedded within the synthetic MCs. To track clump evolution, we have developed an algorithm, *CHANDRAST* (CHAsiNg DendRogrAm Structures over Time), which matches clumps to their counterparts in neighbouring simulation snapshots while accounting for fragmentation and mergers. Using this, we create hierarchical tree diagrams of clumps and their future counterparts (hyperclumps); and study the evolution of their properties. Averaging the properties of clumps over their evolution, we study the trends in hyperclump properties as a function of their lifetimes. We also study the effects of individual stellar feedback mechanisms on the lifetime of hyperclumps. The main results of this project are as follows.

- The radius and mass of the clumps show monotonic decrement over time, except during mergers and fragmentations. These properties represent the global structure of clumps and generally decrease as they collapse and are eroded by the combined effects of gravity, turbulence, and stellar feedback.
- The luminosity and surface density show more fluctuations in their temporal distribution as they are more strongly affected by the observational biases, such as the presence of CO-dark gas.
- The velocity dispersion and virial parameters of the clumps show rapid changes between successive snapshots. These fluctuations reflect constant changes in the internal gas dynamics of clumps, driven by momentum injection from stellar feedback as well as by turbulent gas motions.
- Long-lived clumps are typically larger, more massive, and have high velocity dispersions. This is expected as these clumps can self-shield more effectively and are resistant to erosion by turbulence and feedback. They are also associated with protostellar outflows, which is consistent with the formation of protostars in large clumps over time.
- Short-lived clumps are small and low-mass structures that evolve and disperse rapidly. They are linked to strong infrared emission, suggesting that radiation-dominated environments efficiently disperse molecular gas compared to quiescent regions.

Stellar feedback produces distinct morphological features in molecular gas, such as shells and bubbles driven by stellar winds. In Chapter 8, we explore the possibility of using the convolutional neural network CASI-3D to identify stellar-wind-driven bubbles in ^{13}CO(2-1) emission maps from the SEDIGISM survey. We first apply a model previously trained on Taurus-like molecular clouds (D. Xu,

S. S. R. Offner et al., 2020b) on the RCW-120 region, where it successfully detects a small-scale bubble. Subsequently, we train CASI-3D on synthetic observations generated from the STARFORGE simulations. This model identifies the complete shell of RCW 120 as wind-affected shell, but does not produce an entirely precise classification. Our results demonstrate the ability of CASI-3D to detect bubbles in previously unseen data. However, since this analysis is a preliminary work, the results are not yet suitable for scientific interpretation and will be further developed in future studies.

In summary, we find that stellar feedback is a key component that influences the structures, properties, and evolution of molecular clouds, clumps, and cores. On core scales (≤ 0.1 pc), stellar feedback increases velocity dispersion and the virial parameter of gas structures. Protostellar outflows and stellar winds compress and disperse gas, resulting in smaller cores, whereas supernova-affected cores appear larger. On clump scales ($\sim$ 1 pc), feedback contributes to a reduction in size and mass, while the injected momentum presents itself as variations in internal gas motions. On cloud scales (≥ 10 pc), feedback creates cavities, drives fragmentation, and governs their large-scale morphological evolution. Although the different feedback mechanisms act together, their effects on molecular gas can be analysed separately to some extent, providing insight into their individual roles in shaping gas structures.

9.2 Outlooks

Extension to other simulation suites

The primary analyses in this work are based on a single suite of STARFORGE simulations, which simulate an isolated GMC with a given set of initial conditions. Since the completion of this work, multiple simulation suites from STARFORGE have become available, including larger and more massive GMCs. Using these will allow us to characterise a broader range of structures using dendrogram-based methods, allowing direct comparison with the large diversity of clouds observed in surveys. Although computationally expensive, the use of multiple simulations would allow for a better exploration of multiple environments, which is necessary to understand the full diversity of MCs found in galaxies.

Generating synthetic observations from multiple simulations would also provide stronger constrains on MC evolution.In Chapter B, we study the evolution of MCs but do not separate them into distinct evolutionary stages, since such a classification is intrinsically simulation-dependent. A diverse set of synthetic observations will trace a wider range of physical conditions and evolutionary scenarios, enabling a realistic definition of the evolutionary stages of MCs. A large data set covering a wide range of galactic environments would also make it possible to predict the evolutionary stages of observed MCs by training a CNN on the synthetic observations.

Multi-scale and multi-tracer synthetic observations

In Chapters 6 and 7, we generate synthetic observations in the likelihood of the SEDIGISM survey. While this enables a detailed study of the parsec-scale clouds and clumps, it excludes smaller scale structures, especially dense cores that are progenitors of individual stars. As a result, these studies cannot fully exploit the high spatial resolution of the STARFORGE simulations. In addition, the ^{13}CO(2-1) transition used in these works traces moderately dense gas in galaxies. Although this is well suited for studying MCs, denser gas tracers such as HCN and N_2H^+ are better for probing denser regions such as cores. The modelling of multiple molecular line tracers thus enables different gas phases to be probed consistently across these structures.

Producing synthetic observations that span a range of angular resolutions, noise levels, and tracers (modelled after different surveys) would make full use of the simulation fidelity. It would enable the simultaneous study of clouds, clumps, and cores in the same physical region, providing a continuous view of gas structures across spatial scales and physical conditions (e.g. densities, temperatures).

The ability to track structures across scales would significantly improve studies of fragmentation without being constrained by a fixed lower limit on size. This would enable for an analysis of the relative effects of gravity, turbulence, and stellar feedback on different scales within the same region. By doing so, it might be possible to understand which physical processes dominate at particular spatial scales and whether smaller structures dynamically decouple from their parent clouds. In addition, it would allow for compairing the effects of individual feedback mechanisms across spatial scales by providing a continuous scaling of properties from clouds to cores.

Automated identification of evolutionary stages of molecular gas structures

In Chap. 7, we use the CHANDRAST algorithm to follow the evolution of individual clumps in synthetic observations generated from the STARFORGE simulations. Although the algorithm enables a systematic study of clump lifetimes and evolutionary trends, it still requires multiple upgrades. Future improvements to CHANDRAST will focus on refining the clump-matching procedure by incorporating multiple physical properties, such as mass and velocity dispersion, into the similarity metric. To remove pseudo merger and fragmentation events, we will check their persistance across multiple snapshots before identifying an event as a merger or fragmentation.

Beyond the improvements, combining CHANDRAST with `astrodendro` will enable a 4-dimensional dendrogram analysis of position-position-velocity-time synthetic observation cubes and position-position-position-time cubes in simulations. This will extend the identification of the structure from spatial connectivity to connectivity in both space and time, yielding unique evolutionary paths for each hyperclump and allowing individual clumps to be followed forward and backward in time. The PPV cubes of our synthetic clumps are comparable to those of the clumps in the observations. Identifying synthetic clumps that closely resemble observed counterparts can provide constraints on the evolutionary stage of observed clumps and shed light on their likely past and future evolution.

Although manual comparison of the observed structures with their synthetic counterparts is possible, it is often time-consuming and subjective. Machine learning techniques provide a better alternative for such comparisons. Simple methods, such as nearest-neighbour (sklearn), enable matching an observed clump with synthetic counterparts on the basis of their properties. Complex methods such as Bayesian convolutional neural networks (BayesCNN, Shridhar, Laumann and Liwicki, 2019) and self-organising maps (Guérin, Chauvet and Saubion, 2024) can provide a more accurate evolutionary classification of clumps based on their PPV cubes. Convolutional neural networks trained on PPV cubes of synthetic clumps as input and the evolutionary time as output (ground truth) can be used to predict the evolutionary stages of these structures from observational line emission maps. Our previous studies on the evolution of synthetic MCs highlight the role of stellar feedback in shaping the MCs structures. Therefore, applying transfer learning (Zhuang et al., 2021) from models trained to identify stellar feedback features (Sect. 8) would significantly improve the accuracy of evolutionary stage predictions.

Together, these works are essential for interpreting observed trends in the molecular cloud properties and studying the time evolution of molecular gas structures in galaxies. They are timely and are more critical now than ever before because gas structures are being revealed with unprecedented detail via ALMA and JWST observations. Moreover, cutting-edge simulations like STARFORGE are able to model complex processes in galaxies with remarkable accuracy and high resolution. These advancements offer an unparalleled opportunity to uncover the processes shaping ISM across cosmic time.

Bibliography

Agertz, O. and A. V. Kravtsov (2015), *On the Interplay between Star Formation and Feedback in Galaxy Formation Simulations*, ApJ **804**, 18 18, arXiv: `1404.2613 [astro-ph.GA]`.

Ahmad, A. et al. (2023), *The birth and early evolution of a low-mass protostar*, A&A **680**, A23 A23, arXiv: `2310.01516 [astro-ph.SR]`.

Ahmadi, A. et al. (2023), *Kinematics and stability of high-mass protostellar disk candidates at sub-arcsecond resolution. Insights from the IRAM NOEMA large programme CORE*, A&A **677**, A171 A171, arXiv: `2305.00020 [astro-ph.GA]`.

Akshaya, M. S. et al. (2019), *Components of the diffuse ultraviolet radiation at high latitudes*, MNRAS **489** 1120, arXiv: `1908.02260 [astro-ph.GA]`.

Ali, A. A., T. J. R. Bending and C. L. Dobbs (2022), *Stellar winds and photoionization in a spiral arm*, MNRAS **510** 5592, arXiv: `2201.04141 [astro-ph.GA]`.

Anderson, L. D., T. M. Bania et al. (2014), *The WISE Catalog of Galactic H II Regions*, ApJS **212**, 1 1, arXiv: `1312.6202 [astro-ph.GA]`.

Anderson, L. D., L. Deharveng et al. (2015), *Mopra CO Observations of the Bubble H II Region RCW 120*, ApJ **800**, 101 101, arXiv: `1412.6470 [astro-ph.GA]`.

André, P., J. Di Francesco, D. Ward-Thompson, S. .-. Inutsuka et al. (2014), “From Filamentary Networks to Dense Cores in Molecular Clouds: Toward a New Paradigm for Star Formation”, *Protostars and Planets VI*, ed. by H. Beuther et al. 27, arXiv: `1312.6232 [astro-ph.GA]`.

– (2014), “From Filamentary Networks to Dense Cores in Molecular Clouds: Toward a New Paradigm for Star Formation”, *Protostars and Planets VI*, ed. by H. Beuther et al. 27, arXiv: `1312.6232 [astro-ph.GA]`.

André, P., A. Men’shchikov et al. (2010), *From filamentary clouds to prestellar cores to the stellar IMF: Initial highlights from the Herschel Gould Belt Survey*, A&A **518**, L102 L102, arXiv: `1005.2618 [astro-ph.GA]`.

Appel, S. M. et al. (2025), *Protostellar Jets in Star Cluster Formation and Evolution: I. Implementation and Initial Results*, arXiv e-prints, arXiv:2509.15311 arXiv:2509.15311, arXiv: `2509.15311 [astro-ph.GA]`.

Arce, H. G., D. Shepherd et al. (2007), “Molecular Outflows in Low- and High-Mass Star-forming Regions”, *Protostars and Planets V*, ed. by B. Reipurth, D. Jewitt and K. Keil 245, arXiv: `astro-ph/0603071 [astro-ph]`.

Arce, H. G., M. A. Borkin, A. A. Goodman, J. E. Pineda and C. N. Beaumont (2011), *A Bubbling Nearby Molecular Cloud: COMPLETE Shells in Perseus*, ApJ **742**, 105 105, arXiv: `1109.3368 [astro-ph.SR]`.

Arce, H. G., M. A. Borkin, A. A. Goodman, J. E. Pineda and M. W. Halle (2010), *The COMPLETE Survey of Outflows in Perseus*, ApJ **715** 1170, arXiv: `1005.1714 [astro-ph.SR]`.

Arce, H. G. and A. A. Goodman (2001), *The Episodic, Precessing Giant Molecular Outflow from IRAS 04239+2436 (HH 300)*, ApJ **554** 132, arXiv: `astro-ph/0103482 [astro-ph]`.

Arzoumanian, D., P. André, P. Didelon et al. (2011), *Characterizing interstellar filaments with Herschel in IC 5146*, A&A **529**, L6 L6, arXiv: `1103.0201 [astro-ph.GA]`.

Arzoumanian, D., P. André, V. Könyves et al. (2019), *Characterizing the properties of nearby molecular filaments observed with Herschel*, A&A **621**, A42 A42, arXiv: `1810.00721 [astro-ph.GA]`.

Asogwa, M. O. et al. (2026), *The Influential Roles of Gravity, Turbulence, and Magnetic Fields in Shaping the Physical Evolution of Dense Massive Clumps*, Research in Astronomy and Astrophysics **26**, 035001 035001, arXiv: `2512.11409 [astro-ph.GA]`.

Bachiller, R. (1996), *Bipolar Molecular Outflows from Young Stars and Protostars*, ARA&A **34** 111.

Ballesteros-Paredes, J., P. André et al. (2020), *From Diffuse Gas to Dense Molecular Cloud Cores*, Space Sci. Rev. **216**, 76 76, arXiv: `2006.01326 [astro-ph.GA]`.

Ballesteros-Paredes, J., L. W. Hartmann et al. (2011a), *Gravity or turbulence? Velocity dispersion-size relation*, MNRAS **411** 65, arXiv: `1009.1583 [astro-ph.GA]`.

– (2011b), *Gravity or turbulence? Velocity dispersion-size relation*, Monthly Notices of the RAS **411** 65, arXiv: `1009.1583 [astro-ph.GA]`.

Bally, J., B. Reipurth and C. J. Davis (2007), "Observations of Jets and Outflows from Young Stars", *Protostars and Planets V*, ed. by B. Reipurth, D. Jewitt and K. Keil 215.

Bally, J. (2016), *Protostellar Outflows*, ARA&A **54** 491.

– (2024), *Jets, Outflows, and Explosions in Massive Star Formation*, arXiv e-prints, arXiv:2401.05623 arXiv:2401.05623, arXiv: `2401.05623 [astro-ph.GA]`.

Barnes, P. J. et al. (2018), *The Galactic Census of High- and Medium-mass Protostars. IV. Molecular Clump Radiative Transfer, Mass Distributions, Kinematics, and Dynamical Evolution*, ApJ **866**, 19 19, arXiv: `1806.00492 [astro-ph.GA]`.

Bate, M. R., T. S. Tricco and D. J. Price (2014), *Collapse of a molecular cloud core to stellar densities: stellar-core and outflow formation in radiation magnetohydrodynamic simulations*, MNRAS **437** 77, arXiv: `1310.1092 [astro-ph.SR]`.

Battersby, C. et al. (2025), *3D CMZ. I. Central Molecular Zone Overview*, ApJ **984**, 156 156, arXiv: `2410.17334 [astro-ph.GA]`.

Battisti, A. J. and M. H. Heyer (2014), *The Dense Gas Mass Fraction of Molecular Clouds in the Milky Way*, ApJ **780**, 173 173, arXiv: `1312.0643 [astro-ph.GA]`.

Beaumont, C. N., A. A. Goodman et al. (2014), *The Milky Way Project: Leveraging Citizen Science and Machine Learning to Detect Interstellar Bubbles*, ApJS **214**, 3 3, arXiv: `1406.2692 [astro-ph.GA]`.

Beaumont, C. N., J. P. Williams and A. A. Goodman (2011), *Classifying Structures in the Interstellar Medium with Support Vector Machines: The G16.05-0.57 Supernova Remnant*, ApJ **741**, 14 14, arXiv: `1107.5584 [astro-ph.GA]`.

Benedettini, M. et al. (2020), *The Forgotten Quadrant Survey. ^{12}CO and ^{13}CO (1-0) survey of the Galactic plane in the range 220° <l <240° -2.°5 < b < 0°*, A&A **633**, A147 A147, arXiv: `1911.11545 [astro-ph.GA]`.

Berezhko, E. G. and H. J. Völk (2007), *Spectrum of Cosmic Rays Produced in Supernova Remnants*, ApJ **661** L175, arXiv: `0704.1715 [astro-ph]`.

Bergin, E. A. and M. Tafalla (2007), *Cold Dark Clouds: The Initial Conditions for Star Formation*, ARA&A **45** 339, arXiv: `0705.3765 [astro-ph]`.

Bertoldi, F. and C. F. McKee (1992), *Pressure-confined Clumps in Magnetized Molecular Clouds*, ApJ **395** 140.

Beuther, H., S. Bihr et al. (2016), *The HI/OH/Recombination line survey of the inner Milky Way (THOR). Survey overview and data release 1*, A&A **595**, A32 A32, arXiv: `1609.03329 [astro-ph.GA]`.

Beuther, H., R. Kuiper and M. Tafalla (2025), *Star Formation from Low to High Mass: A Comparative View*, ARA&A **63** 1, arXiv: `2501.16866 [astro-ph.GA]`.

Beuther, H., S. E. Ragan et al. (2015), *Filament fragmentation in high-mass star formation*, A&A **584**, A67 A67, arXiv: `1510.07063 [astro-ph.GA]`.

Bialy, S. and A. Sternberg (2019), *Thermal Phases of the Neutral Atomic Interstellar Medium from Solar Metallicity to Primordial Gas*, ApJ **881**, 160 160, arXiv: `1902.06764 [astro-ph.GA]`.

Bianchi, S. (2024), *Revisiting the local interstellar radiation field using Gaia DR3*, A&A **691**, A330 A330, arXiv: `2409.15778 [astro-ph.GA]`.

Bolatto, A. D., M. Wolfire and A. K. Leroy (2013a), *The CO-to-H_2 Conversion Factor*, Annual Review of Astron and Astrophys **51** 207, arXiv: `1301.3498 [astro-ph.GA]`.

– (2013b), *The CO-to-H_2 Conversion Factor*, ARA&A **51** 207, arXiv: `1301.3498 [astro-ph.GA]`.

Bolya, D. et al. (2019), "YOLACT: Real-Time Instance Segmentation", *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* 9156.

Bonnor, W. B. (1956), *Boyle's Law and gravitational instability*, MNRAS **116** 351.

Bresnahan, D. et al. (2018), *The dense cores and filamentary structure of the molecular cloud in Corona Australis: Herschel SPIRE and PACS observations from the Herschel Gould Belt Survey*, A&A **615**, A125 A125, arXiv: `1801.07805 [astro-ph.GA]`.

Burkhardt, A. M. et al. (2019), *Modeling C-shock Chemistry in Isolated Molecular Outflows*, ApJ **881**, 32 32, arXiv: `1906.08774 [astro-ph.GA]`.

Burkhart, B., K. Stalpes and D. C. Collins (2017), *The Razor's Edge of Collapse: The Transition Point from Lognormal to Power-Law Distributions in Molecular Clouds*, ApJ **834**, L1 L1, arXiv: `1609.04409 [astro-ph.GA]`.

Calura, F. (2025), *Interstellar dust production, destruction and effects of dust depletion in galaxies*, arXiv e-prints, arXiv:2506.13851 arXiv:2506.13851, arXiv: `2506.13851 [astro-ph.GA]`.

Camenzind, M. (1990), *Magnetized Disk-Winds and the Origin of Bipolar Outflows.*, Reviews in Modern Astronomy **3** 234.

Carey, D. (2016), *Via Lactea Galactic Plane bubble Catalogue*, URL: `https://vialactea.iaps.inaf.it/vialactea/eng/products.php`.

Caselli, P., S. Spezzano et al. (2025), *Hunting pre-stellar cores with APEX: Overview*, A&A **703**, A77 A77, arXiv: `2509.00864 [astro-ph.GA]`.

Caselli, P., C. M. Walmsley et al. (1999), *CO Depletion in the Starless Cloud Core L1544*, ApJ **523** L165.

Casoli, F. and F. Combes (1982), *Can giant molecular clouds form in spiral arms*, Astronomy and Astrophysics **110** 287.

Castor, J. I., D. C. Abbott and R. I. Klein (1975), *Radiation-driven winds in Of stars.*, ApJ **195** 157.

Čemeljić, M., H. Shang and T.-Y. Chiang (2013), *Magnetospheric Accretion and Ejection of Matter in Resistive Magnetohydrodynamic Simulations*, ApJ **768**, 5 5, arXiv: `1112.6226 [astro-ph.SR]`.

Chandrasekhar, S. (1931), *The Maximum Mass of Ideal White Dwarfs*, ApJ **74** 81.

Chapman, N. L. et al. (2011), *The Magnetic Field in Taurus Probed by Infrared Polarization*, ApJ **741**, 21 21, arXiv: `1108.0410 [astro-ph.GA]`.

Chen, C.-Y., L. G. Mundy et al. (2020), *Self-gravitating filament formation from shocked flows: velocity gradients across filaments*, MNRAS **494** 3675, arXiv: `2004.02898 [astro-ph.GA]`.

Chen, C.-Y. and E. C. Ostriker (2014), *Formation of Magnetized Prestellar Cores with Ambipolar Diffusion and Turbulence*, ApJ **785**, 69 69, arXiv: `1403.0582 [astro-ph.SR]`.

Chen, H. H.-H. et al. (2019), *Droplets. I. Pressure-dominated Coherent Structures in L1688 and B18*, ApJ **877**, 93 93, arXiv: `1809.10223 [astro-ph.GA]`.

Chen, L. et al. (2022), *Simple Baselines for Image Restoration*, arXiv e-prints, arXiv:2204.04676 arXiv:2204.04676, arXiv: `2204.04676 [cs.CV]`.

Chevalier, R. A. (1977), *The interaction of supernovae with the interstellar medium.*, ARA&A **15** 175.

Chevance, M., M. R. Krumholz et al. (2023), "The Life and Times of Giant Molecular Clouds", *Protostars and Planets VII*, ed. by S. Inutsuka et al., vol. 534, Astronomical Society of the Pacific Conference Series 1, arXiv: `2203.09570 [astro-ph.GA]`.

Chevance, M., J. M. D. Kruijssen et al. (2022), *Pre-supernova feedback mechanisms drive the destruction of molecular clouds in nearby star-forming disc galaxies*, MNRAS **509** 272, arXiv: `2010.13788 [astro-ph.GA]`.

Chung, E. J. et al. (2021), *TRAO Survey of the Nearby Filamentary Molecular Clouds, the Universal Nursery of Stars (TRAO FUNS). II. Filaments and Dense Cores in IC 5146*, ApJ **919**, 3 3, arXiv: `2106.03897 [astro-ph.GA]`.

Churchwell, E., M. S. Povich et al. (2006), *The Bubbling Galactic Disk*, Astrophysical Journal **649** 759.

Churchwell, E., D. F. Watson et al. (2007), *The Bubbling Galactic Disk. II. The Inner 20°*, Astrophysical Journal **670** 428.

Churchwell, E., B. L. Babler et al. (2009a), *The Spitzer/GLIMPSE Surveys: A New View of the Milky Way*, PASP **121** 213.

– (2009b), *The Spitzer/GLIMPSE Surveys: A New View of the Milky Way*, PASP **121** 213.

Clark, P. C., R. S. Klessen and I. A. Bonnell (2007), *Clump lifetimes and the initial mass function*, MNRAS **379** 57, arXiv: `0704.2837 [astro-ph]`.

Clark, S. E., J. E. G. Peek and M. E. Putman (2014), *Magnetically Aligned H I Fibers and the Rolling Hough Transform*, ApJ **789**, 82 82, arXiv: `1312.1338 [astro-ph.GA]`.

Clarke, S. D., S. E. Jaffa and A. P. Whitworth (2022), *RJ-plots: An improved method to classify structures objectively*, MNRAS **516** 2782, arXiv: `2208.07509 [astro-ph.GA]`.

Coletta, A. et al. (2025), *ALMAGAL: III. Compact source catalog: Fragmentation statistics and physical evolution of the core population*, A&A **696**, A151 A151, arXiv: `2503.05663 [astro-ph.GA]`.

Colín, P., E. Vázquez-Semadeni and G. C. Gómez (2013), *Molecular cloud evolution - V. Cloud destruction by stellar feedback*, MNRAS **435** 1701, arXiv: `1304.1570 [astro-ph.GA]`.

Collacchioni, F. et al. (2018), *Semi-analytic galaxies - III. The impact of supernova feedback on the mass-metallicity relation*, MNRAS **481** 954, arXiv: `1808.09089 [astro-ph.GA]`.

Colman, T. et al. (2024), *Cloud properties across spatial scales in simulations of the interstellar medium*, A&A **686**, A155 A155, arXiv: `2403.00512 [astro-ph.GA]`.

Colombo, D., A. Duarte-Cabral et al. (2022), *The SEDIGISM survey: The influence of spiral arms on the molecular gas distribution of the inner Milky Way*, A&A **658**, A54 A54, arXiv: `2110.06071 [astro-ph.GA]`.

Colombo, D., C. König et al. (2021), *OGHReS: Large-scale filaments in the outer Galaxy*, A&A **655**, L2 L2, arXiv: `2111.02768 [astro-ph.GA]`.

Colombo, D., E. Rosolowsky, A. Duarte-Cabral et al. (2019), *The integrated properties of the molecular clouds from the JCMT CO(3-2) High-Resolution Survey*, MNRAS **483** 4291, arXiv: `1812.04688 [astro-ph.GA]`.

Colombo, D., E. Rosolowsky, A. Ginsburg et al. (2015), *Graph-based interpretation of the molecular interstellar medium segmentation*, MNRAS **454** 2067, arXiv: `1510.04253 [astro-ph.GA]`.

Colombo, D., A. Hughes et al. (2014), *The PdBI Arcsecond Whirlpool Survey (PAWS): Environmental Dependence of Giant Molecular Cloud Properties in M51*, ApJ **784**, 3 3, arXiv: `1401.1505 [astro-ph.GA]`.

Colombo, D., E. Schinnerer et al. (2013), "The PdBI Arcsecond Whirlpool Survey (PAWS) II. Resolved properties of Giant Molecular Clouds in M51 and Local Group galaxies", *American Astronomical Society Meeting Abstracts #221*, vol. 221, American Astronomical Society Meeting Abstracts 349.16 349.16.

Contreras, Y. et al. (2013a), *ATLASGAL - compact source catalogue: 330° < ℓ < 21°*, Astronomy and Astrophysics **549**, A45 A45, arXiv: `1211.0741 [astro-ph.GA]`.

– (2013b), *ATLASGAL - compact source catalogue: 330° < ℓ < 21°*, A&A **549**, A45 A45, arXiv: `1211.0741 [astro-ph.GA]`.

Cordes, J. M. (2004), "NE2001: A New Model for the Galactic Electron Density and its Fluctuations", *Milky Way Surveys: The Structure and Evolution of our Galaxy*, ed. by D. Clemens, R. Shah and T. Brainerd, vol. 317, Astronomical Society of the Pacific Conference Series 211.

Cox, D. P. (2005), *The Three-Phase Interstellar Medium Revisited*, ARA&A **43** 337.

Crowther, P. A. et al. (2016), *The R136 star cluster dissected with Hubble Space Telescope/STIS. I. Far-ultraviolet spectroscopic census and the origin of He II λ1640 in young star clusters*, MNRAS **458** 624, arXiv: `1603.04994 [astro-ph.SR]`.

Crutcher, R. M. (2012), *Magnetic Fields in Molecular Clouds*, Annual Review of Astron and Astrophys **50** 29.

Cunningham, A. J. et al. (2011), *Radiation-hydrodynamic Simulations of Massive Star Formation with Protostellar Outflows*, ApJ **740**, 107 107, arXiv: `1104.1218 [astro-ph.SR]`.

Dale, J. E., B. Ercolano and I. A. Bonnell (2012), *Ionizing feedback from massive stars in massive clusters - II. Disruption of bound clusters by photoionization*, MNRAS **424** 377, arXiv: `1205.0360 [astro-ph.GA]`.

Dame, T. M., D. Hartmann and P. Thaddeus (2001), *The Milky Way in Molecular Clouds: A New Complete CO Survey*, ApJ **547** 792, arXiv: `astro-ph/0009217 [astro-ph]`.

Deharveng, L., F. Schuller et al. (2010a), *A gallery of bubbles. The nature of the bubbles observed by Spitzer and what ATLASGAL tells us about the surrounding neutral material*, Astronomy and Astrophysics **523**, A6 A6, arXiv: `1008.0926 [astro-ph.GA]`.

– (2010b), *A gallery of bubbles. The nature of the bubbles observed by Spitzer and what ATLASGAL tells us about the surrounding neutral material*, A&A **523**, A6 A6, arXiv: `1008.0926 [astro-ph.GA]`.

Deharveng, L., A. Zavagno et al. (2009a), *Star formation around RCW 120, the perfect bubble*, Astronomy and Astrophysics **496** 177, arXiv: `0902.0903 [astro-ph.GA]`.

– (2009b), *Star formation around RCW 120, the perfect bubble*, A&A **496** 177, arXiv: `0902.0903 [astro-ph.GA]`.

Dessart, L. and S. P. Owocki (2005), *2D simulations of the line-driven instability in hot-star winds. II. Approximations for the 2D radiation force*, A&A **437** 657, arXiv: `astro-ph/0503514 [astro-ph]`.

Di Francesco, J. et al. (2001), *Infall, Outflow, Rotation, and Turbulent Motions of Dense Gas within NGC 1333 IRAS 4*, ApJ **562** 770, arXiv: `astro-ph/0108022 [astro-ph]`.

Dobbs, C. L. (2008), *GMC formation by agglomeration and self gravity*, Monthly Notices of the RAS **391** 844, arXiv: `0809.1942 [astro-ph]`.

Dobbs, C. L. and I. A. Bonnell (2006), *Spurs and feathering in spiral galaxies*, MNRAS **367** 873, arXiv: `astro-ph/0602100 [astro-ph]`.

Dobbs, C. L., M. R. Krumholz et al. (2014a), "Formation of Molecular Clouds and Global Conditions for Star Formation", *Protostars and Planets VI*, ed. by H. Beuther et al. 3, arXiv: `1312.3223 [astro-ph.GA]`.

– (2014b), "Formation of Molecular Clouds and Global Conditions for Star Formation", *Protostars and Planets VI*, ed. by H. Beuther et al. 3, arXiv: `1312.3223 [astro-ph.GA]`.

Dobbs, C. L. and J. E. Pringle (2013a), *The exciting lives of giant molecular clouds*, Monthly Notices of the RAS **432** 653, arXiv: `1303.4995 [astro-ph.GA]`.

– (2013b), *The exciting lives of giant molecular clouds*, MNRAS **432** 653, arXiv: `1303.4995 [astro-ph.GA]`.

Dobbs, C. L., J. E. Pringle and A. Burkert (2012), *Giant molecular clouds: what are they made from, and how do they get there?*, Monthly Notices of the RAS **425** 2157, arXiv: `1206.4904 [astro-ph.GA]`.

Dokara, R. et al. (2023), *A global view on star formation: The GLOSTAR Galactic plane survey. VII. Supernova remnants in the Galactic longitude range 28° < l < 36°*, A&A **671**, A145 A145, arXiv: `2211.13811 [astro-ph.GA]`.

Draine, B. T. (2003), *Interstellar Dust Grains*, ARA&A **41** 241, arXiv: `astro-ph/0304489 [astro-ph]`.

– (2011a), *On Radiation Pressure in Static, Dusty H II Regions*, ApJ **732**, 100 100, arXiv: `1003.0474 [astro-ph.GA]`.

Draine, B. T. (2011b), *Physics of the Interstellar and Intergalactic Medium*.

Duarte-Cabral, A., D. Colombo et al. (2021), *The SEDIGISM survey: molecular clouds in the inner Galaxy*, MNRAS **500** 3027, arXiv: `2012.01502 [astro-ph.GA]`.

Duarte-Cabral, A. and C. L. Dobbs (2017), *The evolution of giant molecular filaments*, MNRAS **470** 4261, arXiv: `1706.05421 [astro-ph.GA]`.

Dullemond, C. P. et al. (2012), *RADMC-3D: A multi-purpose radiative transfer tool*, Astrophysics Source Code Library, record ascl:1202.015.

Ebert, R. (1955), *Über die Verdichtung von H I-Gebieten. Mit 5 Textabbildungen*, ZAp **37** 217.

Eden, D. J., T. J. T. Moore, R. Plume, A. J. Rigby et al. (2021), *Characteristic scale of star formation - I. Clump formation efficiency on local scales*, MNRAS **500** 191, arXiv: `2007.00032 [astro-ph.GA]`.

Eden, D. J., T. J. T. Moore, R. Plume, J. S. Urquhart et al. (2017), *The JCMT Plane Survey: first complete data release - emission maps and compact source catalogue*, MNRAS **469** 2163, arXiv: `1704.02982 [astro-ph.GA]`.

Elmegreen, B. G. (2011), "Triggered Star Formation", *EAS Publications Series*, ed. by C. Charbonnel and T. Montmerle, vol. 51, EAS Publications Series, EDP 45, arXiv: `1101.3112 [astro-ph.GA]`.

Elmegreen, B. G. and C. J. Lada (1977), *Sequential formation of subgroups in OB associations.*, Astrophysical Journal **214** 725.

Eswaraiah, C. et al. (2021), *The JCMT BISTRO Survey: Revealing the Diverse Magnetic Field Morphologies in Taurus Dense Cores with Sensitive Submillimeter Polarimetry*, ApJ **912**, L27 L27, arXiv: `2103.02219 [astro-ph.GA]`.

Evans II, N. J. et al. (2009), *The Spitzer c2d Legacy Results: Star-Formation Rates and Efficiencies; Evolution and Lifetimes*, ApJS **181** 321, arXiv: `0811.1059 [astro-ph]`.

Fall, S. M., M. R. Krumholz and C. D. Matzner (2010), *Stellar Feedback in Molecular Clouds and its Influence on the Mass Function of Young Star Clusters*, ApJ **710** L142, arXiv: `0910.2238 [astro-ph.SR]`.

Federrath, C. (2020), "Star formation in cloud cores — simulations and observations of dense molecular cores and the formation of solar mass stars", *Origins: From the Protosun to the First Steps of Life*, ed. by B. G. Elmegreen, L. V. Tóth and M. Güdel, vol. 345, IAU Symposium 43, arXiv: `2002.04224 [astro-ph.SR]`.

Federrath, C. et al. (2021), *The sonic scale of interstellar turbulence*, Nature Astronomy **5** 365, arXiv: `2011.06238 [astro-ph.GA]`.

Feldmeier, A., R.-P. Kudritzki et al. (1997), *The X-ray emission from shock cooling zones in O star winds.*, A&A **320** 899.

Feldmeier, A., J. Puls and A. W. A. Pauldrach (1997), *A possible origin for X-rays from O stars.*, A&A **322** 878.

Fendt, C. (2009), *Formation of Protostellar Jets as Two-Component Outflows from Star-Disk Magnetospheres*, ApJ **692** 346, arXiv: `0810.4154 [astro-ph]`.

Fendt, C. and M. Čemeljić (2002), *Formation of protostellar jets - effects of magnetic diffusion*, A&A **395** 1045, arXiv: `astro-ph/0210082 [astro-ph]`.

Ferreira, J. (1997), *Magnetically-driven jets from Keplerian accretion discs.*, A&A **319** 340, arXiv: `astro-ph/9607057 [astro-ph]`.

Field, G. B., E. G. Blackman and E. R. Keto (2011), *Does external pressure explain recent results for molecular clouds?*, Monthly Notices of the RAS **416** 710, arXiv: `1106.3017 [astro-ph.GA]`.

Figueira, M., L. Bronfman et al. (2018), *ALMA observations of RCW 120 Fragmentation at 0.01 pc scale&star*, A&A **616**, L10 L10, arXiv: `1807.11860 [astro-ph.GA]`.

Figueira, M., A. Zavagno, L. Bronfman et al. (2020), *APEX CO observations towards the photodissociation region of RCW 120*, A&A **639**, A93 A93, arXiv: `2005.08810 [astro-ph.GA]`.

Figueira, M., A. Zavagno, L. Deharveng et al. (2017), *Star formation towards the Galactic H II region RCW 120. Herschel observations of compact sources*, A&A **600**, A93 A93, arXiv: `1612.08862 [astro-ph.GA]`.

Flower, D. R. and G. Pineau Des Forêts (2010), *Excitation and emission of H2, CO and H2O molecules in interstellar shock waves*, MNRAS **406** 1745.

Frank, A. et al. (2014), "Jets and Outflows from Star to Cloud: Observations Confront Theory", *Protostars and Planets VI*, ed. by H. Beuther et al. 451, arXiv: `1402.3553 [astro-ph.SR]`.

Freytag, B. and S. Höfner (2023), *Global 3D radiation-hydrodynamical models of AGB stars with dust-driven winds*, A&A **669**, A155 A155, arXiv: `2301.11836 [astro-ph.SR]`.

Fujimoto, Y., K. Fukushima and K. Murase (2021), *Extensive studies of the neutron star equation of state from the deep learning inference with the observational data augmentation*, Journal of High Energy Physics **2021**, 273 273, arXiv: `2101.08156 [nucl-th]`.

Fukui, Y., A. Kawamura et al. (2008), *The Second Survey of the Molecular Clouds in the Large Magellanic Cloud by NANTEN. I. Catalog of Molecular Clouds*, Astrophysical Journals **178** 56, arXiv: `0804.1458 [astro-ph]`.

Fukui, Y., T. Inoue et al. (2021), *Rapid and efficient mass collection by a supersonic cloud-cloud collision as a major mechanism of high-mass star formation*, PASJ **73** S405, arXiv: `1909.08202 [astro-ph.GA]`.

Gaia Collaboration et al. (2018), *Gaia Data Release 2. Mapping the Milky Way disc kinematics*, A&A **616**, A11 A11, arXiv: `1804.09380 [astro-ph.GA]`.

Galliano, F., M. Galametz and A. P. Jones (2018), *The Interstellar Dust Properties of Nearby Galaxies*, ARA&A **56** 673, arXiv: `1711.07434 [astro-ph.GA]`.

Gatto, A. et al. (2017), *The SILCC project - III. Regulation of star formation and outflows by stellar winds and supernovae*, MNRAS **466** 1903, arXiv: `1606.05346 [astro-ph.GA]`.

Geen, S., P. Agrawal et al. (2023), *Bringing Stellar Evolution & Feedback Together: Summary of proposals from the Lorentz Center Workshop, 2022*, arXiv e-prints, arXiv:2301.13611 arXiv:2301.13611, arXiv: `2301.13611 [astro-ph.SR]`.

Geen, S., J. Rosdahl et al. (2015), *A detailed study of feedback from a massive star*, MNRAS **448** 3248, arXiv: `1412.0484 [astro-ph.GA]`.

Geen, S., J. D. Soler and P. Hennebelle (2017), *Interpreting the star formation efficiency of nearby molecular clouds with ionizing radiation*, MNRAS **471** 4844, arXiv: `1703.10071 [astro-ph.GA]`.

Geron, A. (2019), *Hands-on machine learning with scikit-learn, keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*, 2nd ed., O'Reilly Media.

Gholamalinezhad, H. and H. Khosravi (2020), *Pooling Methods in Deep Neural Networks, a Review*, arXiv e-prints, arXiv:2009.07485 arXiv:2009.07485, arXiv: `2009.07485 [cs.CV]`.

Girichidis, P. et al. (2020), *Physical Processes in Star Formation*, Space Sci. Rev. **216**, 68 68, arXiv: `2005.06472 [astro-ph.GA]`.

Girshick, R. et al. (2014), "Rich feature hierarchies for accurate object detection and semantic segmentation", *Proceedings of the IEEE conference on computer vision and pattern recognition* 580.

Glassner, A. (2021), *Deep learning: a visual approach*, No Starch Press.

Glover, S. C. O. and P. C. Clark (2012a), *Is molecular gas necessary for star formation?*, Monthly Notices of the RAS **421** 9, arXiv: `1105.3073 [astro-ph.GA]`.

– (2012b), *Star formation in metal-poor gas clouds*, MNRAS **426** 377, arXiv: `1203.4251 [astro-ph.GA]`.

Gong, M. and E. C. Ostriker (2015), *Prestellar Core Formation, Evolution, and Accretion from Gravitational Fragmentation in Turbulent Converging Flows*, ApJ **806**, 31 31, arXiv: `1504.02140 [astro-ph.SR]`.

Goodfellow, I., Y. Bengio and A. Courville (2016), *Deep Learning*, `http://www.deeplearningbook.org`, MIT Press.

Gorski, M. D. and L. Murchikova (2025), *The CO-to-H_2 Conversion Factor in the Milky Way's Central Parsec*, ApJ **995**, L51 L51.

Grishunin, K. et al. (2024), *Observing the LMC with APEX: Signatures of large-scale feedback in the molecular clouds of 30 Doradus*, A&A **682**, A137 A137, arXiv: `2310.20701 [astro-ph.GA]`.

Groves, B. and E. Schinnerer (2015), "Measuring the molecular gas content and XCO variation in nearby galaxies", *IAU General Assembly*, vol. 29 2256331 2256331.

Grudić, M. Y., D. Guszejnov, P. F. Hopkins et al. (2021), *STARFORGE: Towards a comprehensive numerical model of star cluster formation and feedback*, MNRAS **506** 2199, arXiv: `2010.11254 [astro-ph.IM]`.

Grudić, M. Y., D. Guszejnov, S. S. R. Offner et al. (2022), *The dynamics and outcome of star formation with jets, radiation, winds, and supernovae in concert*, MNRAS **512** 216, arXiv: `2201.00882 [astro-ph.GA]`.

Grudić, M. Y., P. F. Hopkins et al. (2019), *On the nature of variations in the measured star formation efficiency of molecular clouds*, MNRAS **488** 1501, arXiv: `1809.08348 [astro-ph.GA]`.

Guérin, A., P. Chauvet and F. Saubion (2024), *A Survey on Recent Advances in Self-Organizing Maps*, arXiv e-prints, arXiv:2501.08416 arXiv:2501.08416, arXiv: `2501.08416 [cs.NE]`.

Guo, H. .-. et al. (2020), *Three-Dimensional Distribution of the Interstellar Dust in the Milky Way*, arXiv e-prints, arXiv:2010.14092 arXiv:2010.14092, arXiv: `2010.14092 [astro-ph.GA]`.

Guo, M.-H. et al. (2022), *SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation*, ArXiv **abs/2209.08575**, URL: `https://api.semanticscholar.org/CorpusID:252367800`.

Güsten, R. et al. (2006), *The Atacama Pathfinder EXperiment (APEX) - a new submillimeter facility for southern skies -*, A&A **454** L13.

Guszejnov, D., M. Y. Grudić, P. F. Hopkins et al. (2021), *STARFORGE: the effects of protostellar outflows on the IMF*, MNRAS **502** 3646, arXiv: `2010.11249 [astro-ph.GA]`.

Guszejnov, D., M. Y. Grudić, S. S. R. Offner et al. (2022), *Effects of the environment and feedback physics on the initial mass function of stars in the STARFORGE simulations*, MNRAS **515** 4929, arXiv: `2205.10413 [astro-ph.GA]`.

Guszejnov, D., P. F. Hopkins and M. Y. Grudić (2018), *Universal scaling relations in scale-free structure formation*, MNRAS **477** 5139, arXiv: `1707.05799 [astro-ph.GA]`.

Habart, E. et al. (2005), *Molecular Hydrogen*, Space Sci. Rev. **119** 71.

Habing, H. J., F. P. Israel and T. de Jong (1972), *On the coincidence of OH emission sources and bright knots in H II regions.*, Astronomy and Astrophysics **17** 329.

Hacar, A., S. E. Clark et al. (2023), "Initial Conditions for Star Formation: a Physical Description of the Filamentary ISM", *Protostars and Planets VII*, ed. by S. Inutsuka et al., vol. 534, Astronomical Society of the Pacific Conference Series 153, arXiv: `2203.09562 [astro-ph.GA]`.

Hacar, A., M. Tafalla, J. Kauffmann and A. Kovács (2013), *Cores, filaments, and bundles: hierarchical core formation in the L1495/B213 Taurus region*, A&A **554**, A55 A55, arXiv: `1303.2118 [astro-ph.GA]`.

Hacar, A., M. Tafalla, J. Kauffmann and A. Kovacs (2013), "Cores, Filaments, and Bundles: Hierarchical core formation in the B213 filament in Taurus", *Protostars and Planets VI Posters*.

Hacar, A. and E. van Dischoeck (2018), "Chemistry in velocity: Investigating gas chemical properties in 4D", *Astrochemistry VII: Through the Cosmos from Galaxies to Planets*, ed. by M. Cunningham, T. Millar and Y. Aikawa, vol. 332, IAU Symposium 330.

Haemmerlé, L. and T. Peters (2016), *Impact of initial models and variable accretion rates on the pre-main-sequence evolution of massive and intermediate-mass stars and the early evolution of H II regions*, MNRAS **458** 3299, arXiv: `1603.00332 [astro-ph.SR]`.

Haffner, L. M. et al. (2009), *The warm ionized medium in spiral galaxies*, Reviews of Modern Physics **81** 969, arXiv: `0901.0941 [astro-ph.GA]`.

Haro, G. (1952), *Herbig's Nebulous Objects Near NGC 1999.*, ApJ **115** 572.

Hartmann, L., J. Ballesteros-Paredes and E. A. Bergin (2001), *Rapid Formation of Molecular Clouds and Stars in the Solar Neighborhood*, Astrophysical Journal **562** 852, arXiv: `astro-ph/0108023 [astro-ph]`.

He, K. et al. (2016), “Deep Residual Learning for Image Recognition”, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1 1, arXiv: `1512.03385 [cs.CV]`.

Heitsch, F. et al. (2008), *Cooling, Gravity, and Geometry: Flow-driven Massive Core Formation*, Astrophysical Journal **674** 316, arXiv: `0709.2451 [astro-ph]`.

Hennebelle, P., R. Banerjee et al. (2008), *From the warm magnetized atomic medium to molecular clouds*, Astronomy and Astrophysics **486** L43, arXiv: `0805.1366 [astro-ph]`.

Hennebelle, P. and M. Y. Grudić (2024), *The Physical Origin of the Stellar Initial Mass Function*, ARA&A **62** 63, arXiv: `2404.07301 [astro-ph.GA]`.

Hennebelle, P. (2013), *On the origin of non-self-gravitating filaments in the ISM*, A&A **556**, A153 A153, arXiv: `1306.5452 [astro-ph.GA]`.

Herbig, G. H. (1951), *The Spectra of Two Nebulous Objects Near NGC 1999.*, ApJ **113** 697.

Heyer, M. and T. M. Dame (2015), *Molecular Clouds in the Milky Way*, ARA&A **53** 583.

Heyer, M., C. Krawczyk et al. (2009), *Re-Examining Larson’s Scaling Relationships in Galactic Molecular Clouds*, ApJ **699** 1092, arXiv: `0809.1397 [astro-ph]`.

Heyer, M. H. and C. M. Brunt (2004), *The Universality of Turbulence in Galactic Molecular Clouds*, ApJ **615** L45, arXiv: `astro-ph/0409420 [astro-ph]`.

Hezaveh, Y. D., L. Perreault Levasseur and P. J. Marshall (2017), *Fast automated analysis of strong gravitational lenses with convolutional neural networks*, Nature **548** 555, arXiv: `1708.08842 [astro-ph.IM]`.

Hirano, N. et al. (2010), *Extreme Active Molecular Jets in L1448C*, ApJ **717** 58, arXiv: `1005.0703 [astro-ph.SR]`.

Höfner, S. and H. Olofsson (2018), *Mass loss of stars on the asymptotic giant branch. Mechanisms, models and measurements*, A&A Rev. **26**, 1 1.

Holdship, J. et al. (2017), *UCLCHEM: A Gas-grain Chemical Code for Clouds, Cores, and C-Shocks*, AJ **154**, 38 38, arXiv: `1705.10677 [astro-ph.GA]`.

Hollenbach, D. and E. E. Salpeter (1971), *Surface Recombination of Hydrogen Molecules*, Astrophysical Journal **163** 155.

Hopkins, P. F. (2015), *A new class of accurate, mesh-free hydrodynamic simulation methods*, MNRAS **450** 53, arXiv: `1409.7395 [astro-ph.CO]`.

Hopkins, P. F. and M. Y. Grudić (2019), *Numerical problems in coupling photon momentum (radiation pressure) to gas*, MNRAS **483** 4187, arXiv: `1803.07573 [astro-ph.GA]`.

Hopkins, P. F., M. Y. Grudić et al. (2020), *Radiative stellar feedback in galaxy formation: Methods and physics*, MNRAS **491** 3702, arXiv: `1811.12462 [astro-ph.GA]`.

Hopkins, P. F., E. Quataert and N. Murray (2012), *The structure of the interstellar medium of star-forming galaxies*, MNRAS **421** 3488, arXiv: `1110.4636 [astro-ph.CO]`.

Hopkins, P. F. and M. J. Raives (2016), *Accurate, meshless methods for magnetohydrodynamics*, MNRAS **455** 51, arXiv: `1505.02783 [astro-ph.IM]`.

Hopkins, P. F., A. Wetzel et al. (2018), *FIRE-2 simulations: physics versus numerics in galaxy formation*, MNRAS **480** 800, arXiv: `1702.06148 [astro-ph.GA]`.

Hosokawa, T., H. W. Yorke and K. Omukai (2010), *Evolution of Massive Protostars Via Disk Accretion*, ApJ **721** 478, arXiv: `1005.2827 [astro-ph.SR]`.

Hu, C.-Y. et al. (2022), *Dependence of* X_{CO} *on Metallicity, Intensity, and Spatial Scale in a Self-regulated Interstellar Medium*, ApJ **931**, 28 28, arXiv: `2201.03885 [astro-ph.GA]`.

Hubble, E. P. (1926), *Extragalactic nebulae.*, Astrophysical Journal **64** 321.

Iani, E. et al. (2021), *Stellar feedback in a clumpy galaxy at z 3.4*, MNRAS **507** 3830, arXiv: `2109.06898 [astro-ph.GA]`.

Ibrahim, D. and C. Kobayashi (2025), *The impact of supernova feedback on metallicity-gradient evolution in cosmological simulations*, MNRAS **544** 815, arXiv: `2501.11209 [astro-ph.GA]`.

Inoue, T. and Y. Fukui (2013), *Formation of Massive Molecular Cloud Cores by Cloud-Cloud Collision*, ApJ **774**, L31 L31, arXiv: `1305.4655 [astro-ph.GA]`.

Inoue, T. and S.-i. Inutsuka (2009), *Two-Fluid Magnetohydrodynamics Simulations of Converging H I Flows in the Interstellar Medium. II. Are Molecular Clouds Generated Directly from a Warm Neutral Medium?*, Astrophysical Journal **704** 161, arXiv: `0908.3701 [astro-ph.GA]`.

Ishihara, K. et al. (2024), *Digging into the Interior of Hot Cores with ALMA (DIHCA). IV. Fragmentation in High-mass Star-forming Clumps*, ApJ **974**, 95 95, arXiv: `2407.06845 [astro-ph.GA]`.

Jackson, J. M. et al. (2006), *The Boston University-Five College Radio Astronomy Observatory Galactic Ring Survey*, ApJS **163** 145, arXiv: `astro-ph/0602160 [astro-ph]`.

Jacob, A. M., D. A. Neufeld et al. (2022), *HyGAL: Characterizing the Galactic Interstellar Medium with Observations of Hydrides and Other Small Molecules. I. Survey Description and a First Look Toward W3(OH), W3 IRS5, and NGC 7538 IRS1*, ApJ **930**, 141 141, arXiv: `2202.05046 [astro-ph.GA]`.

Jacob, A. M., K. M. Menten et al. (2020), *First detection of ^{13}CH in the interstellar medium*, A&A **640**, A125 A125, arXiv: `2007.01190 [astro-ph.GA]`.

Jaffa, S. E. et al. (2018), *J plots: a new method for characterizing structures in the interstellar medium*, MNRAS **477** 1940, arXiv: `1803.01640 [astro-ph.GA]`.

Jaupart, E. and G. Chabrier (2020), *Evolution of the Density PDF in Star-forming Clouds: The Role of Gravity*, ApJ **903**, L2 L2, arXiv: `2010.00603 [astro-ph.GA]`.

Jayasinghe, T. et al. (2019), *The Milky Way Project second data release: bubbles and bow shocks*, MNRAS **488** 1141, arXiv: `1905.12625 [astro-ph.GA]`.

Jerkstrand, A., D. Milisavljevic and B. Müller (2026), “Core-collapse supernovae”, *Encyclopedia of Astrophysics, Volume 2*, vol. 2 639, arXiv: `2503.01321 [astro-ph.HE]`.

Karska, A. et al. (2025), *SOFIA FIFI-LS spectroscopy of DR21 Main: Energetics of the spatially resolved outflow from a high-mass protostar*, A&A **697**, A186 A186, arXiv: `2503.15059 [astro-ph.GA]`.

Kauffmann, J., T. Pillai et al. (2010), *The Mass-size Relation from Clouds to Cores. II. Solar Neighborhood Clouds*, ApJ **716** 433, arXiv: `1004.1170 [astro-ph.GA]`.

Kauffmann, J. and T. Pillai (2010), *How Many Infrared Dark Clouds Can form Massive Stars and Clusters?*, ApJ **723** L7, arXiv: `1009.1617 [astro-ph.GA]`.

Kee, N. D. et al. (2021), *Analytic, dust-independent mass-loss rates for red supergiant winds initiated by turbulent pressure*, A&A **646**, A180 A180, arXiv: `2101.03070 [astro-ph.SR]`.

Kennicutt, R. C. and N. J. Evans (2012), *Star Formation in the Milky Way and Nearby Galaxies*, ARA&A **50** 531, arXiv: `1204.3552 [astro-ph.GA]`.

Khan, S. et al. (2026), *Exploring the interplay between molecular and ionized gas in HII regions*, arXiv e-prints, arXiv:2601.01748 arXiv:2601.01748, arXiv: `2601.01748 [astro-ph.GA]`.

Khullar, S. et al. (2021), *The density structure of supersonic self-gravitating turbulence*, MNRAS **507** 4335, arXiv: `2107.00725 [astro-ph.GA]`.

Kim, C.-G. and E. C. Ostriker (2017), *Three-phase Interstellar Medium in Galaxies Resolving Evolution with Star Formation and Supernova Feedback (TIGRESS): Algorithms, Fiducial Model, and Convergence*, ApJ **846**, 133 133, arXiv: `1612.03918 [astro-ph.GA]`.

Kim, D. A. et al. (2023), *The kinematic structure of magnetically aligned H I filaments*, MNRAS **526** 4345, arXiv: `2309.10777 [astro-ph.GA]`.

Kim, J.-G., W.-T. Kim and E. C. Ostriker (2016), *Disruption of Molecular Clouds by Expansion of Dusty H II Regions*, ApJ **819**, 137 137, arXiv: `1601.03035 [astro-ph.GA]`.

– (2018), *Modeling UV Radiation Feedback from Massive Stars. II. Dispersal of Star-forming Giant Molecular Clouds by Photoionization and Radiation Pressure*, ApJ **859**, 68 68, arXiv: `1804.04664 [astro-ph.GA]`.

Kim, W.-T. and E. C. Ostriker (2006), *Formation of Spiral-Arm Spurs and Bound Clouds in Vertically Stratified Galactic Gas Disks*, Astrophysical Journal **646** 213, arXiv: `astro-ph/0603751 [astro-ph]`.

Kirk, H. et al. (2017), *The Green Bank Ammonia Survey: Dense Cores under Pressure in Orion A*, ApJ **846**, 144 144, arXiv: `1708.05426 [astro-ph.GA]`.

Kirk, J. M., D. Ward-Thompson and P. André (2005), *The initial conditions of isolated star formation - VI. SCUBA mappingof pre-stellar cores*, MNRAS **360** 1506, arXiv: `astro-ph/0505190 [astro-ph]`.

Kirk, J. M., D. Ward-Thompson, P. Palmeirim et al. (2013), *First results from the Herschel Gould Belt Survey in Taurus*, MNRAS **432** 1424, arXiv: `1304.4098 [astro-ph.GA]`.

Kirsanova, M. S., Y. N. Pavlyuchenkov, A. O. H. Olofsson et al. (2023), *The shocked molecular layer in RCW 120*, MNRAS **520** 751, arXiv: `2212.08702 [astro-ph.GA]`.

Kirsanova, M. S., Y. N. Pavlyuchenkov, D. S. Wiebe et al. (2019), *Molecular envelope around the HII region RCW 120*, MNRAS **488** 5641, arXiv: `1908.05394 [astro-ph.SR]`.

Klessen, R. S. and A. Burkert (2000), *The Formation of Stellar Clusters: Gaussian Cloud Conditions. I.*, Astrophysical Journals **128** 287, arXiv: `astro-ph/9904090 [astro-ph]`.

Klessen, R. S. and S. C. O. Glover (2016), *Physical Processes in the Interstellar Medium*, Saas-Fee Advanced Course **43** 85, arXiv: `1412.5182 [astro-ph.GA]`.

Klos, K. S., I. A. Bonnell and R. J. Smith (2025), *The role of magnetic fields in the formation of high-mass star-forming cores*, MNRAS **539** 2307, arXiv: `2504.00864 [astro-ph.GA]`.

Koch, E. W. and E. W. Rosolowsky (2015), *Filament identification through mathematical morphology*, MNRAS **452** 3435, arXiv: `1507.02289 [astro-ph.GA]`.

König, C. et al. (2017), *ATLASGAL-selected massive clumps in the inner Galaxy. III. Dust continuum characterization of an evolutionary sample*, A&A **599**, A139 A139, arXiv: `1610.09055 [astro-ph.GA]`.

Könyves, V. et al. (2020), *Properties of the dense core population in Orion B as seen by the Herschel Gould Belt survey*, A&A **635**, A34 A34, arXiv: `1910.04053 [astro-ph.SR]`.

Korpela, E. J., S. Bowyer and J. Edelstein (1998), *Spectral Observations of Diffuse Far-Ultraviolet Emission from the Hot Phase of the Interstellar Medium with the Diffuse Ultraviolet Experiment*, ApJ **495** 317, arXiv: `astro-ph/9709291 [astro-ph]`.

Körtgen, B. et al. (2018), *The origin of filamentary star forming clouds in magnetized galaxies*, MNRAS **479** L40, arXiv: `1805.08509 [astro-ph.GA]`.

Kosiński, R. and M. Hanasz (2007), *On the influence of cooling and heating processes on the Parker instability - II. Numerical simulations*, Monthly Notices of the RAS **376** 861.

Kristensen, L. E. et al. (2011), *Water in low-mass star-forming regions with Herschel (WISH-LM). High-velocity H_2O bullets in L1448-MM observed with HIFI*, A&A **531**, L1 L1, arXiv: `1105.4884 [astro-ph.SR]`.

Krizhevsky, A., I. Sutskever and G. E. Hinton (2012), “ImageNet Classification with Deep Convolutional Neural Networks”, *Advances in Neural Information Processing Systems*, ed. by F. Pereira et al., vol. 25, Curran Associates, Inc., URL: `https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf`.

Kroupa, P. (2002), *The Initial Mass Function of Stars: Evidence for Uniformity in Variable Systems*, Science **295** 82, arXiv: `astro-ph/0201098 [astro-ph]`.

Kroupa, P. et al. (2026), “The initial mass function of stars”, *Encyclopedia of Astrophysics, Volume 2*, vol. 2 173, arXiv: `2410.07311 [astro-ph.GA]`.

Kruijssen, J. M. D. et al. (2019), *Fast and inefficient star formation due to short-lived molecular clouds and rapid feedback*, Nature **569** 519, arXiv: `1905.08801 [astro-ph.GA]`.

Krumholz, M. R., M. R. Bate et al. (2014), “Star Cluster Formation and Feedback”, *Protostars and Planets VI*, ed. by H. Beuther et al. 243, arXiv: `1401.2473 [astro-ph.GA]`.

Krumholz, M. R. (2014), *DESPOTIC - a new software library to Derive the Energetics and SPectra of Optically Thick Interstellar Clouds*, MNRAS **437** 1662, arXiv: `1304.2404 [astro-ph.IM]`.

Krumholz, M. R., R. I. Klein and C. F. McKee (2005), “Radiation pressure in massive star formation”, *Massive Star Birth: A Crossroads of Astrophysics*, ed. by R. Cesaroni et al., vol. 227, IAU Symposium 231, arXiv: `astro-ph/0510432 [astro-ph]`.

Krumholz, M. R., R. I. Klein, C. F. McKee et al. (2009), *The Formation of Massive Star Systems by Accretion*, Science **323** 754, arXiv: `0901.3157 [astro-ph.SR]`.

Krumholz, M. R., A. K. Leroy and C. F. McKee (2011), *Which Phase of the Interstellar Medium Correlates with the Star Formation Rate?*, Astrophysical Journal **731**, 25 25, arXiv: `1101.1296 [astro-ph.CO]`.

Kulkarni, S. R. (2022), *The Far Ultra-violet Background*, PASP **134**, 084302 084302, arXiv: `2107.09585 [astro-ph.GA]`.

Kumar, M. S. N. et al. (2020), *Unifying low- and high-mass star formation through density-amplified hubs of filaments. The highest mass stars (>100 $M_{\odot}$) form only in hubs*, A&A **642**, A87 A87, arXiv: `2008.00295 [astro-ph.GA]`.

Kundu, E. (2025), *Radio Supernovae*, arXiv e-prints, arXiv:2512.22117 arXiv:2512.22117, arXiv: `2512.22117 [astro-ph.HE]`.

Kwan, J. and N. Scoville (1976), *The nature of the broad molecular line emission at the Kleinmann-Low nebula.*, ApJ **210** L39.

Kwan, J. and F. Valdes (1983), *Spiral gravitational potentials and the mass growth of molecular clouds*, Astrophysical Journal **271** 604.

Lamers, H. J. G. L. M., T. P. Snow and D. M. Lindholm (1995), *Terminal Velocities and the Bistability of Stellar Winds*, ApJ **455** 269.

Langer, N. et al. (1994), *Towards an understanding of very massive stars. A new evolutionary scenario relating O stars, LBVs and Wolf-Rayet stars.*, A&A **290** 819.

Larson, R. B. (1981a), *Turbulence and star formation in molecular clouds.*, MNRAS **194** 809.

– (1981b), *Turbulence and star formation in molecular clouds.*, Monthly Notices of the RAS **194** 809.

LeCun, Y. et al. (1989), *Backpropagation applied to handwritten zip code recognition*, Neural computation **1** 541.

Lee, M.-Y. et al. (2012), *A High-resolution Study of the H I-H_2 Transition across the Perseus Molecular Cloud*, Astrophysical Journal **748**, 75 75, arXiv: `1110.2745 [astro-ph.GA]`.

Li, G.-X. et al. (2016), *ATLASGAL: A Galaxy-wide sample of dense filamentary structures*, Astronomy and Astrophysics **591**, A5 A5, arXiv: `1604.00544 [astro-ph.SR]`.

Li, Y., M.-M. Mac Low and R. S. Klessen (2005), *Star Formation in Isolated Disk Galaxies. I. Models and Characteristics of Nonlinear Gravitational Collapse*, Astrophysical Journal **626** 823, arXiv: `astro-ph/0501022 [astro-ph]`.

Lin, Y., F. Wyrowski, H. B. Liu, Y. Gong, O. Sipilä, A. Izquierdo et al. (2024), *Massive clumps in W43-main: Structure formation in an extensively shocked molecular cloud*, A&A **685**, A101 A101, arXiv: `2401.17192 [astro-ph.GA]`.

Lin, Y., F. Wyrowski, H. B. Liu, Y. Gong, O. Sipilä, A. F. Izquierdo et al. (2024), *Massive clumps in W43-main: Structure formation in an extensively shocked molecular cloud*, arXiv e-prints, arXiv:2401.17192 arXiv:2401.17192, arXiv: `2401.17192 [astro-ph.GA]`.

Lippok, N. et al. (2013), *Gas-phase CO depletion and N_2H^+ abundances in starless cores*, A&A **560**, A41 A41, arXiv: `1308.2958 [astro-ph.SR]`.

Liu, X. and K. M. Goh (2025), *ResNet: Enabling Deep Convolutional Neural Networks through Residual Learning*, arXiv e-prints, arXiv:2510.24036 arXiv:2510.24036, arXiv: `2510.24036 [cs.CV]`.

López-Vázquez, J. A. et al. (2025), *Erosion of a dense molecular core by a strong outflow from a massive protostar*, A&A **695**, A236 A236, arXiv: `2502.17786 [astro-ph.SR]`.

Louvet, F. et al. (2024), *ALMA-IMF: XV. Core mass function in the high-mass star formation regime*, A&A **690**, A33 A33, arXiv: `2407.18719 [astro-ph.GA]`.

Lucy, L. B. and P. M. Solomon (1970), *Mass Loss by Hot Stars*, ApJ **159** 879.

Maciel, W. J. and R. D. D. Costa (2010), "Metallicity gradients in the Milky Way", *Chemical Abundances in the Universe: Connecting First Stars to Planets*, ed. by K. Cunha, M. Spite and B. Barbuy, vol. 265, IAU Symposium 317, arXiv: `0911.3763 [astro-ph.GA]`.

Madden, S. C. et al. (2020), *Tracing the total molecular gas in galaxies: [CII] and the CO-dark gas*, A&A **643**, A141 A141, arXiv: `2009.00649 [astro-ph.GA]`.

Maret, S. et al. (2009), *Spitzer Mapping of Molecular Hydrogen Pure Rotational Lines in NGC 1333: A Detailed Study of Feedback in Star Formation*, ApJ **698** 1244, arXiv: `0904.0603 [astro-ph.SR]`.

Martínez-Núñez, S. et al. (2017), *Towards a Unified View of Inhomogeneous Stellar Winds in Isolated Supergiant Stars and Supergiant High Mass X-Ray Binaries*, Space Sci. Rev. **212** 59, arXiv: `1701.08618 [astro-ph.HE]`.

Maruta, H. et al. (2010), *Physical Properties of Dense Cores in the ρ Ophiuchi Main Cloud and a Significant Role of External Pressures in Clustered Star Formation*, ApJ **714** 680, arXiv: `0907.2558 [astro-ph.SR]`.

Mattern, M. et al. (2018), *Structure and fragmentation of a high line-mass filament: Nessie*, Astronomy and Astrophysics **616**, A78 A78, arXiv: `1804.02256 [astro-ph.GA]`.

Matzner, C. D. and C. F. McKee (2000), *Efficiencies of Low-Mass Star and Star Cluster Formation*, ApJ **545** 364, arXiv: `astro-ph/0007383 [astro-ph]`.

Maud, L. T. et al. (2015), *A distance-limited sample of massive molecular outflows*, MNRAS **453** 645, arXiv: `1509.00199 [astro-ph.SR]`.

Mazumdar, P. et al. (2021), *High resolution LAsMA ^{12}CO and ^{13}CO observation of the G305 giant molecular cloud complex. II. Effect of feedback on clump properties*, A&A **656**, A101 A101, arXiv: `2109.09615 [astro-ph.GA]`.

McConnachie, A. W. (2012), *The Observed Properties of Dwarf Galaxies in and around the Local Group*, AJ **144**, 4 4, arXiv: `1204.1562 [astro-ph.CO]`.

McKee, C. F. and J. P. Ostriker (1977), *A theory of the interstellar medium: three components regulated by supernova explosions in an inhomogeneous substrate.*, ApJ **218** 148.

Medina, S.-N. X. et al. (2019), *GLOSTAR: Radio Source Catalog I. 28° < ℓ < 36° and |b| < 1°*, A&A **627**, A175 A175, arXiv: `1905.09281 [astro-ph.GA]`.

Megeath, S. T., R. A. Gutermuth and M. A. Kounkel (2022), *Low Mass Stars as Tracers of Star and Cluster Formation*, PASP **134**, 042001 042001, arXiv: `2203.03655 [astro-ph.GA]`.

Meidt, S. E., A. Hughes et al. (2015), *Short GMC Lifetimes: An Observational Estimate with the PdBI Arcsecond Whirlpool Survey (PAWS)*, ApJ **806**, 72 72, arXiv: `1504.04528 [astro-ph.GA]`.

Meidt, S. E., E. Rosolowsky et al. (2023), *PHANGS-JWST First Results: Interstellar Medium Structure on the Turbulent Jeans Scale in Four Disk Galaxies Observed by JWST and the Atacama Large Millimeter/submillimeter Array*, ApJ **944**, L18 L18, arXiv: `2212.06434 [astro-ph.GA]`.

Men'shchikov, A. (2013), *A multi-scale filament extraction method: getfilaments*, A&A **560**, A63 A63, arXiv: `1309.2170 [astro-ph.GA]`.

– (2021), *Multicomponent, multiwavelength benchmarks for source- and filament-extraction methods*, A&A **654**, A78 A78, arXiv: `2108.05585 [astro-ph.IM]`.

Men'shchikov, A. et al. (2010), *Filamentary structures and compact objects in the Aquila and Polaris clouds observed by Herschel*, A&A **518**, L103 L103, arXiv: `1005.3115 [astro-ph.GA]`.

Menon, S. H., C. Federrath and M. R. Krumholz (2023), *Outflows driven by direct and reprocessed radiation pressure in massive star clusters*, MNRAS **521** 5160, arXiv: `2210.02818 [astro-ph.GA]`.

Mitofsky, A. M. (2018), *Direct Energy Conversion*.

Miville-Deschênes, M.-A., N. Murray and E. J. Lee (2017), *Physical Properties of Molecular Clouds for the Entire Milky Way Disk*, ApJ **834**, 57 57, arXiv: `1610.05918 [astro-ph.GA]`.

Miyama, S. M., S. Narita and C. Hayashi (1987), *Fragmentation of Isothermal Sheet-Like Clouds. I —Solutions of Linear and Second-Order Perturbation Equations—*, Progress of Theoretical Physics **78** 1051.

Molinari, S., P. Schilke et al. (2025), *ALMAGAL: I. The ALMA evolutionary study of high-mass protocluster formation in the Galaxy: Presentation of the survey and early results*, A&A **696**, A149 A149, arXiv: `2503.05555 [astro-ph.GA]`.

Molinari, S., E. Schisano et al. (2011), *Source extraction and photometry for the far-infrared and sub-millimeter continuum in the presence of complex backgrounds*, A&A **530**, A133 A133, arXiv: `1011.3946 [astro-ph.GA]`.

Molinari, S., B. Swinyard et al. (2010a), *Clouds, filaments, and protostars: The Herschel Hi-GAL Milky Way*, Astronomy and Astrophysics **518**, L100 L100, arXiv: `1005.3317 [astro-ph.GA]`.

– (2010b), *Clouds, filaments, and protostars: The Herschel Hi-GAL Milky Way*, A&A **518**, L100 L100, arXiv: `1005.3317 [astro-ph.GA]`.

Motte, F., S. Bontemps and F. Louvet (2018), *High-Mass Star and Massive Cluster Formation in the Milky Way*, ARA&A **56** 41, arXiv: `1706.00118 [astro-ph.GA]`.

Mouschovias, T. C., M. W. Kunz and D. A. Christie (2009), *Formation of interstellar clouds: Parker instability with phase transitions*, Monthly Notices of the RAS **397** 14, arXiv: `0901.0914 [astro-ph.GA]`.

Murray, N., E. Quataert and T. A. Thompson (2005), *On the Maximum Luminosity of Galaxies and Their Central Black Holes: Feedback from Momentum-driven Winds*, ApJ **618** 569, arXiv: `astro-ph/0406070 [astro-ph]`.

– (2010), *The Disruption of Giant Molecular Clouds by Radiation Pressure & the Efficiency of Star Formation in Galaxies*, ApJ **709** 191, arXiv: `0906.5358 [astro-ph.GA]`.

Nakamura, F. and Z.-Y. Li (2007), *Protostellar Turbulence Driven by Collimated Outflows*, ApJ **662** 395, arXiv: `astro-ph/0703152 [astro-ph]`.

Nakamura, F. and Z.-Y. Li (2008), *Magnetically Regulated Star Formation in Three Dimensions: The Case of the Taurus Molecular Cloud Complex*, Astrophysical Journal **687** 354, arXiv: `0804.4201 [astro-ph]`.

Nakamura, F., T. Miura et al. (2012), *Evidence for Cloud-Cloud Collision and Parsec-scale Stellar Feedback within the L1641-N Region*, ApJ **746**, 25 25, arXiv: `1110.6225 [astro-ph.SR]`.

Neralwar, K. R., D. Colombo, A. Duarte-Cabral et al. (2022), *The SEDIGISM survey: Molecular cloud morphology. I. Classification and star formation*, A&A **663**, A56 A56, arXiv: `2203.02504 [astro-ph.GA]`.

Neralwar, K. R., D. Colombo, S. Offner, A. Karska et al. (2025), *A multi-scale evolutionary study of molecular gas in STARFORGE: I. Synthetic observations of SEDIGISM-like molecular clouds*, A&A **704**, A38 A38, arXiv: `2510.07393 [astro-ph.GA]`.

Neralwar, K. R., D. Colombo, S. Offner, F. Wyrowski et al. (2024), *Effects of stellar feedback on cores in STARFORGE*, A&A **690**, A345 A345, arXiv: `2409.05949 [astro-ph.GA]`.

Neufeld, D. A., J. R. Goicoechea et al. (2010), *Herschel/HIFI observations of interstellar* OH^+ *and* H_2O^+ *towards W49N: a probe of diffuse clouds with a small molecular fraction*, A&A **521**, L10 L10, arXiv: `1007.0987 [astro-ph.GA]`.

Neufeld, D. A. and M. G. Wolfire (2016), *The Chemistry of Interstellar Argonium and Other Probes of the Molecular Fraction in Diffuse Clouds*, ApJ **826**, 183 183, arXiv: `1607.00375 [astro-ph.GA]`.

Neupert, T. et al. (2021), *Introduction to Machine Learning for the Sciences*, arXiv e-prints, arXiv:2102.04883 arXiv:2102.04883, arXiv: `2102.04883 [physics.comp-ph]`.

Nisini, B. et al. (2018), *Connection between jets, winds and accretion in T Tauri stars. The X-shooter view*, A&A **609**, A87 A87, arXiv: `1710.05587 [astro-ph.SR]`.

Nony, T. et al. (2023), *ALMA-IMF. V. Prestellar and protostellar core populations in the W43 cloud complex*, A&A **674**, A75 A75, arXiv: `2301.07238 [astro-ph.GA]`.

Ntampaka, M. et al. (2019), *A Deep Learning Approach to Galaxy Cluster X-Ray Masses*, ApJ **876**, 82 82, arXiv: `1810.07703 [astro-ph.CO]`.

Offner, S. S. R., P. C. Clark et al. (2014), “The Origin and Universality of the Stellar Initial Mass Function”, *Protostars and Planets VI*, ed. by H. Beuther et al. 53, arXiv: `1312.5326 [astro-ph.SR]`.

Offner, S. S. R. and H. G. Arce (2015), *Impact of Winds from Intermediate-mass Stars on Molecular Cloud Structure and Turbulence*, ApJ **811**, 146 146, arXiv: `1508.07008 [astro-ph.GA]`.

Offner, S. S. R., R. I. Klein et al. (2009), *The Effects of Radiative Transfer on Low-Mass Star Formation*, ApJ **703** 131, arXiv: `0904.2004 [astro-ph.SR]`.

Offner, S. S. R., J. Taylor and M. Y. Grudic (2025), *The Life and Times of Star-forming Cores: An Analysis of Dense Gas in the STARFORGE Simulations*, ApJ **982**, 138 138, arXiv: `2502.15057 [astro-ph.GA]`.

Offner, S. S. R., J. Taylor, C. Markey et al. (2022), *Turbulence, coherence, and collapse: Three phases for core evolution*, MNRAS **517** 885, arXiv: `2006.07325 [astro-ph.GA]`.

Oliva, A. and R. Kuiper (2023), *Modeling disks and magnetic outflows around a forming massive star. II. Dynamics of jets from massive protostars*, A&A **669**, A81 A81, arXiv: `2210.01492 [astro-ph.SR]`.

Olivier, G. M. et al. (2021), *Evolution of Stellar Feedback in H II Regions*, ApJ **908**, 68 68, arXiv: `2009.10079 [astro-ph.GA]`.

Olmi, L., J. Brand and D. Elia (2023), *Probing fragmentation with ALMA continuum and spectral line observations of the dense clumps in the* $\ell = 224°$ *region*, MNRAS **518** 1917.

Oskinova, L. M. and D. Schaerer (2022), *Ionization of He II in star-forming galaxies by X-rays from cluster winds and superbubbles*, A&A **661**, A67 A67, arXiv: `2203.04987 [astro-ph.GA]`.

Ossenkopf, V. (1997), *The Sobolev approximation in molecular clouds*, New A **2** 365.

Ostriker, E. C., C. F. McKee and A. K. Leroy (2010), *Regulation of Star Formation Rates in Multiphase Galactic Disks: A Thermal/Dynamical Equilibrium Model*, ApJ **721** 975, arXiv: `1008.0410 [astro-ph.CO]`.

Ostriker, E. C. and R. Shetty (2011), *Maximally Star-forming Galactic Disks. I. Starburst Regulation Via Feedback-driven Turbulence*, ApJ **731**, 41 41, arXiv: `1102.1446 [astro-ph.CO]`.

Owocki, S. P., J. I. Castor and G. B. Rybicki (1988), *Time-dependent Models of Radiatively Driven Stellar Winds. I. Nonlinear Evolution of Instabilities for a Pure Absorption Model*, ApJ **335** 914.

Owocki, S. P., K. G. Gayley and N. J. Shaviv (2004), *A Porosity-Length Formalism for Photon-Tiring-limited Mass Loss from Stars above the Eddington Limit*, ApJ **616** 525, arXiv: `astro-ph/0409573 [astro-ph]`.

Pabst, C. H. M. et al. (2020a), *Expanding bubbles in Orion A: [C II] observations of M 42, M 43, and NGC 1977*, A&A **639**, A2 A2, arXiv: `2005.03917 [astro-ph.GA]`.

– (2020b), *Expanding bubbles in Orion A: [C II] observations of M 42, M 43, and NGC 1977*, Astronomy and Astrophysics **639**, A2 A2, arXiv: `2005.03917 [astro-ph.GA]`.

Padoan, P., B. J. T. Jones and Å. P. Nordlund (1997), *Supersonic Turbulence in the Interstellar Medium: Stellar Extinction Determinations as Probes of the Structure and Dynamics of Dark Clouds*, ApJ **474** 730, arXiv: `astro-ph/9603061 [astro-ph]`.

Padoan, P., M. Juvela et al. (2001), *The Turbulent Shock Origin of Proto-Stellar Cores*, Astrophysical Journal **553** 227, arXiv: `astro-ph/0011122 [astro-ph]`.

Padoan, P. and Å. Nordlund (1999), *A Super-Alfvénic Model of Dark Clouds*, ApJ **526** 279, arXiv: `astro-ph/9901288 [astro-ph]`.

Padoan, P., Å. Nordlund et al. (2007), *Two Regimes of Turbulent Fragmentation and the Stellar Initial Mass Function from Primordial to Present-Day Star Formation*, Astrophysical Journal **661** 972, arXiv: `astro-ph/0701795 [astro-ph]`.

Padoan, P., L. Pan, T. Haugbølle et al. (2016), *Supernova Driving. I. The Origin of Molecular Cloud Turbulence*, ApJ **822**, 11 11, arXiv: `1509.04663 [astro-ph.GA]`.

Padoan, P., L. Pan, M. Juvela et al. (2020), *The Origin of Massive Stars: The Inertial-inflow Model*, ApJ **900**, 82 82, arXiv: `1911.04465 [astro-ph.GA]`.

Palla, F. and S. W. Stahler (1993), *The Pre-Main-Sequence Evolution of Intermediate-Mass Stars*, ApJ **418** 414.

Palmeirim, P., P. André et al. (2013), *Herschel view of the Taurus B211/3 filament and striations: evidence of filamentary growth?*, A&A **550**, A38 A38, arXiv: `1211.6360 [astro-ph.SR]`.

Palmeirim, P., A. Zavagno et al. (2017), *Spatial distribution of star formation related to ionized regions throughout the inner Galactic plane*, A&A **605**, A35 A35, arXiv: `1705.09877 [astro-ph.GA]`.

Panopoulou, G. V., I. Psaradaki and K. Tassis (2016), *The magnetic field and dust filaments in the Polaris Flare*, MNRAS **462** 1517, arXiv: `1607.00005 [astro-ph.GA]`.

Papadopoulos, P. P. et al. (2008), *High-excitation molecular gas in local luminous AGN hosts*, A&A **491** 483, arXiv: `0809.4971 [astro-ph]`.

Paradis, D. et al. (2023), *Toward a better understanding of the mid-infrared emission in the Large Magellanic Cloud*, A&A **674**, A141 A141, arXiv: `2304.09519 [astro-ph.GA]`.

Parker, E. N. (1958), *Dynamics of the Interplanetary Gas and Magnetic Fields.*, ApJ **128** 664.

Parrent, J., B. Friesen and M. Parthasarathy (2014), *A review of type Ia supernova spectra*, Ap&SS **351** 1, arXiv: `1402.6337 [astro-ph.HE]`.

Pattle, K., L. Fissel et al. (2023), "Magnetic Fields in Star Formation: from Clouds to Cores", *Protostars and Planets VII*, ed. by S. Inutsuka et al., vol. 534, Astronomical Society of the Pacific Conference Series 193, arXiv: `2203.11179 [astro-ph.GA]`.

Pattle, K., D. Ward-Thompson et al. (2015), *The JCMT Gould Belt Survey: first results from the SCUBA-2 observations of the Ophiuchus molecular cloud and a virial analysis of its prestellar core population*, MNRAS **450** 1094, arXiv: `1502.05858 [astro-ph.GA]`.

Pazukhin, A. G., I. I. Zinchenko and E. A. Trofimova (2025), *Study of the Physical and Chemical Properties of Dense Clumps at Different Evolutionary Stages in Several Regions of Massive Star and Stellar Cluster Formation*, Astronomy Reports **69** 87, arXiv: `2412.18506 [astro-ph.GA]`.

Pelkonen, V.-M. et al. (2021), *From the CMF to the IMF: beyond the core-collapse model*, MNRAS **504** 1219, arXiv: `2008.02192 [astro-ph.SR]`.

Perets, H. B. and O. Biham (2006), *Molecular hydrogen formation on porous dust grains*, MNRAS **365** 801, arXiv: `astro-ph/0506492 [astro-ph]`.

Peretto, N., P. André et al. (2012), *The Pipe Nebula as seen with Herschel: formation of filamentary structures by large-scale compression?*, A&A **541**, A63 A63, arXiv: `1203.3403 [astro-ph.GA]`.

Peretto, N., G. A. Fuller et al. (2013), *Global collapse of molecular clouds as a formation mechanism for the most massive stars*, A&A **555**, A112 A112, arXiv: `1307.2590 [astro-ph.GA]`.

Peretto, N., A. J. Rigby et al. (2023), *Star cluster progenitors are dynamically decoupled from their parent molecular clouds*, MNRAS **525** 2935, arXiv: `2305.02701 [astro-ph.GA]`.

Peters, T. et al. (2017), *The SILCC project - IV. Impact of dissociating and ionizing radiation on the interstellar medium and Hα emission as a tracer of the star formation rate*, MNRAS **466** 3293, arXiv: `1610.06569 [astro-ph.GA]`.

Pezzuto, S. et al. (2021), *Physical properties of the ambient medium and of dense cores in the Perseus star-forming region derived from Herschel Gould Belt Survey observations*, A&A **645**, A55 A55, arXiv: `2010.00006 [astro-ph.GA]`.

Pineda, J. E., D. Arzoumanian et al. (2023), "From Bubbles and Filaments to Cores and Disks: Gas Gathering and Growth of Structure Leading to the Formation of Stellar Systems", *Protostars and Planets VII*, ed. by S. Inutsuka et al., vol. 534, Astronomical Society of the Pacific Conference Series 233, arXiv: `2205.03935 [astro-ph.GA]`.

Pineda, J. E., A. A. Goodman et al. (2010), *Direct Observation of a Sharp Transition to Coherence in Dense Cores*, ApJ **712** L116, arXiv: `1002.2946 [astro-ph.GA]`.

Pontoppidan, K. M. et al. (2009), *A New Raytracer for Modeling AU-Scale Imaging of Lines from Protoplanetary Disks*, ApJ **704** 1482, arXiv: `0908.2997 [astro-ph.EP]`.

Portail, M. (2016), *Structure and Dynamics of the Galactic Bulge and Bar*, PhD thesis: Ludwig Maximilian University of Munich.

Pudritz, R. E. and N. K.-R. Kevlahan (2013), *Shock interactions, turbulence and the origin of the stellar mass spectrum*, Philosophical Transactions of the Royal Society of London Series A **371** 20120248, arXiv: `1201.2650 [astro-ph.GA]`.

Pudritz, R. E., R. Ouyed et al. (2007), "Disk Winds, Jets, and Outflows: Theoretical and Computational Foundations", *Protostars and Planets V*, ed. by B. Reipurth, D. Jewitt and K. Keil 277, arXiv: `astro-ph/0603592 [astro-ph]`.

Puls, J., A. Herrero and C. Allende Prieto (2024), *Stellar Atmospheres*, arXiv e-prints, arXiv:2409.03329 arXiv:2409.03329, arXiv: `2409.03329 [astro-ph.SR]`.

Puls, J., J. S. Vink and F. Najarro (2008), *Mass loss from hot massive stars*, A&A Rev. **16** 209, arXiv: 0811.0487 [astro-ph].

Purcell, C. R. et al. (2013), *The Coordinated Radio and Infrared Survey for High-mass Star Formation. II. Source Catalog*, ApJS **205**, 1 1, arXiv: 1211.7116 [astro-ph.GA].

Quillen, A. C. et al. (2005), *Turbulence Driven by Outflow-blown Cavities in the Molecular Cloud of NGC 1333*, ApJ **632** 941, arXiv: astro-ph/0503167 [astro-ph].

Rahner, D. et al. (2017), *Winds and radiation in unison: a new semi-analytic feedback model for cloud dissolution*, MNRAS **470** 4453, arXiv: 1704.04240 [astro-ph.GA].

Ramachandran, V. et al. (2018), *Stellar population of the superbubble N 206 in the LMC. II. Parameters of the OB and WR stars, and the total massive star feedback*, A&A **615**, A40 A40, arXiv: 1802.07494 [astro-ph.GA].

Ramos Padilla, A. F. et al. (2023), *Diagnosing the interstellar medium of galaxies with far-infrared emission lines. II. [C II], [O I], [O III], [N II], and [N III] up to $z = 6$*, A&A **679**, A131 A131, arXiv: 2205.11955 [astro-ph.GA].

Raskutti, S., E. C. Ostriker and M. A. Skinner (2016), *Numerical Simulations of Turbulent Molecular Clouds Regulated by Radiation Feedback Forces. I. Star Formation Rate and Efficiency*, ApJ **829**, 130 130, arXiv: 1608.04469 [astro-ph.GA].

Ray, T. et al. (2007), “Toward Resolving the Outflow Engine: An Observational Perspective”, *Protostars and Planets V*, ed. by B. Reipurth, D. Jewitt and K. Keil 231, arXiv: astro-ph/0605597 [astro-ph].

Ray, T. P. et al. (2023), *Outflows from the youngest stars are mostly molecular*, Nature **622** 48.

Rebolledo, D. et al. (2012), *Giant Molecular Clouds and Star Formation in the Non-grand Design Spiral Galaxy NGC 6946*, Astrophysical Journal **757**, 155 155, arXiv: 1208.5499 [astro-ph.GA].

Reissl, S. et al. (2018), *Spectral shifting strongly constrains molecular cloud disruption by radiation pressure on dust*, A&A **611**, A70 A70, arXiv: 1710.02854 [astro-ph.GA].

Rezaei Kh., S. et al. (2024), *3D structure of the Milky Way out to 10 kpc from the Sun: Catalogue of large molecular clouds in the Galactic Plane*, A&A **692**, A255 A255, arXiv: 2405.09634 [astro-ph.GA].

Rigby, A. J., T. J. T. Moore, D. J. Eden et al. (2019), *CHIMPS: physical properties of molecular clumps across the inner Galaxy*, A&A **632**, A58 A58, arXiv: 1909.04714 [astro-ph.GA].

Rigby, A. J., T. J. T. Moore, R. Plume et al. (2016), *CHIMPS: the $^{13}CO/C^{18}O$ ($J = 3 \rightarrow 2$) Heterodyne Inner Milky Way Plane Survey*, MNRAS **456** 2885, arXiv: 1512.08235 [astro-ph.GA].

Rodgers, A. W., C. T. Campbell and J. B. Whiteoak (1960), *A catalogue of Hα-emission regions in the southern Milky Way*, MNRAS **121** 103.

Roman-Duval, J. et al. (2010), *Physical Properties and Galactic Distribution of Molecular Clouds Identified in the Galactic Ring Survey*, Astrophysical Journal **723** 492, arXiv: 1010.2798 [astro-ph.GA].

Romanova, M. M. and S. P. Owocki (2015), *Accretion, Outflows, and Winds of Magnetized Stars*, Space Sci. Rev. **191** 339, arXiv: 1605.04979 [astro-ph.SR].

Ronneberger, O., P. Fischer and T. Brox (2015), *U-Net: Convolutional Networks for Biomedical Image Segmentation*, arXiv e-prints, arXiv:1505.04597 arXiv:1505.04597, arXiv: 1505.04597 [cs.CV].

Rosen, A. L., M. R. Krumholz et al. (2016), *An unstable truth: how massive stars get their mass*, MNRAS **463** 2553, arXiv: 1607.03117 [astro-ph.SR].

Rosen, A. L., S. S. R. Offner et al. (2020), *Zooming in on Individual Star Formation: Low- and High-Mass Stars*, Space Sci. Rev. **216**, 62 62, arXiv: 2005.07717 [astro-ph.SR].

Rosenblatt, F. et al. (1962), *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*, vol. 55, Spartan books Washington, DC.

Rosolowsky, E. W., J. E. Pineda, J. B. Foster et al. (2008), *An Ammonia Spectral Atlas of Dense Cores in Perseus*, ApJS **175** 509, arXiv: `0711.0231 [astro-ph]`.

Rosolowsky, E. W., J. E. Pineda, J. Kauffmann et al. (2008a), *Structural Analysis of Molecular Clouds: Dendrograms*, Astrophysical Journal **679** 1338, arXiv: `0802.2944 [astro-ph]`.

– (2008b), *Structural Analysis of Molecular Clouds: Dendrograms*, ApJ **679** 1338, arXiv: `0802.2944 [astro-ph]`.

Rosolowsky, E., A. Hughes et al. (2021), *Giant molecular cloud catalogues for PHANGS-ALMA: methods and initial results*, MNRAS **502** 1218, arXiv: `2101.04697 [astro-ph.GA]`.

Rosolowsky, E. and A. Leroy (2006), *Bias-free Measurement of Giant Molecular Cloud Properties*, PASP **118** 590, arXiv: `astro-ph/0601706 [astro-ph]`.

Roueff, A. et al. (2021), *$C^{18}O$, ^{13}CO, and ^{12}CO abundances and excitation temperatures in the Orion B molecular cloud. Analysis of the achievable precision in modeling spectral lines within the approximation of the local thermodynamic equilibrium*, A&A **645**, A26 A26, arXiv: `2005.08317 [astro-ph.GA]`.

Russeil, D. (2003), *Star-forming complexes and the spiral structure of our Galaxy*, A&A **397** 133.

Rybicki, G. B. and A. P. Lightman (1979), *Radiative processes in astrophysics*.

Saintonge, A. (2025), *The Interstellar Medium*, arXiv e-prints, arXiv:2504.01410 arXiv:2504.01410, arXiv: `2504.01410 [astro-ph.GA]`.

Sales, L. V. et al. (2014), *Stellar feedback by radiation pressure and photoionization*, MNRAS **439** 2990, arXiv: `1310.7572 [astro-ph.CO]`.

Samuel, J. (2020), *2020 Nobel Prize for Physics: Black holes and the Milky Way's darkest secret*, arXiv e-prints, arXiv:2011.06656 arXiv:2011.06656, arXiv: `2011.06656 [physics.pop-ph]`.

Schaerer, D., T. Fragos and Y. I. Izotov (2019), *X-ray binaries as the origin of nebular He II emission in low-metallicity star-forming galaxies*, A&A **622**, L10 L10, arXiv: `1902.10496 [astro-ph.GA]`.

Schinnerer, E. et al. (2017), *The PdBI Arcsecond Whirlpool Survey (PAWS): The Role of Spiral Arms in Cloud and Star Formation*, ApJ **836**, 62 62, arXiv: `1701.02184 [astro-ph.GA]`.

Schisano, E., S. Molinari, A. Coletta et al. (2025), *ALMAGAL VI. The spatial distribution of dense cores during the evolution of cluster-forming massive clump*, arXiv e-prints, arXiv:2512.05914 arXiv:2512.05914, arXiv: `2512.05914 [astro-ph.GA]`.

Schisano, E., K. L. J. Rygl et al. (2014), *The Identification of Filaments on Far-infrared and Submillimiter Images: Morphology, Physical Conditions and Relation with Star Formation of Filamentary Structure*, Astrophysical Journal **791**, 27 27, arXiv: `1406.4443 [astro-ph.SR]`.

Schisano, E., S. Molinari, D. Elia et al. (2020), *The Hi-GAL catalogue of dusty filamentary structures in the Galactic plane*, MNRAS **492** 5420, arXiv: `1912.04020 [astro-ph.GA]`.

Schneider, N., T. Csengeri et al. (2010), *Dynamic star formation in the massive DR21 filament*, A&A **520**, A49 A49, arXiv: `1003.4198 [astro-ph.GA]`.

Schneider, N., R. Simon et al. (2020), *FEEDBACK: a SOFIA Legacy Program to Study Stellar Feedback in Regions of Massive Star Formation*, PASP **132**, 104301 104301, arXiv: `2009.08730 [astro-ph.GA]`.

Schuller, F., T. Csengeri et al. (2017), *SEDIGISM: Structure, excitation, and dynamics of the inner Galactic interstellar medium*, A&A **601**, A124 A124, arXiv: `1701.04712 [astro-ph.GA]`.

Schuller, F., K. M. Menten et al. (2009), *ATLASGAL - The APEX telescope large area survey of the galaxy at 870 μm*, A&A **504** 415, arXiv: `0903.1369 [astro-ph.GA]`.

Schuller, F., J. S. Urquhart et al. (2021), *The SEDIGISM survey: First Data Release and overview of the Galactic structure*, MNRAS **500** 3064, arXiv: `2012.01527 [astro-ph.GA]`.

Schultz, K. et al. (2025), “Resolved stellar populations in the bubbles of NGC 628”, *American Astronomical Society Meeting Abstracts #245*, vol. 245, American Astronomical Society Meeting Abstracts 104.17 104.17.

Seifried, D. et al. (2020), *SILCC-Zoom: H_2 and CO-dark gas in molecular clouds - the impact of feedback and magnetic fields*, MNRAS **492** 1465, arXiv: `1906.01015 [astro-ph.GA]`.

Senchyna, P. et al. (2020), *High-mass X-ray binaries in nearby metal-poor galaxies: on the contribution to nebular He II emission*, MNRAS **494** 941, arXiv: `1909.10574 [astro-ph.GA]`.

Seta, A. and N. M. McClure-Griffiths (2025), *Magnetic fields in the multiphase interstellar medium of the Milky Way: turbulent kinetic and magnetic energy density relation*, MNRAS **539** 1024, arXiv: `2503.23634 [astro-ph.GA]`.

Shen, J. and X.-W. Zheng (2020), *The bar and spiral arms in the Milky Way: structure and kinematics*, Research in Astronomy and Astrophysics **20**, 159 159.

Shetty, R., D. C. Collins et al. (2010), *The Effect of Projection on Derived Mass-Size and Linewidth-Size Relationships*, ApJ **712** 1049, arXiv: `1001.4549 [astro-ph.GA]`.

Shetty, R., S. C. Glover et al. (2011), *Modelling CO emission - I. CO as a column density tracer and the X factor in molecular clouds*, MNRAS **412** 1686, arXiv: `1011.2019 [astro-ph.GA]`.

Shridhar, K., F. Laumann and M. Liwicki (2019), *A Comprehensive guide to Bayesian Convolutional Neural Network with Variational Inference*, arXiv e-prints, arXiv:1901.02731 arXiv:1901.02731, arXiv: `1901.02731 [cs.LG]`.

Simpson, R. J. et al. (2012), *The Milky Way Project First Data Release: a bubblier Galactic disc*, MNRAS **424** 2442, arXiv: `1201.6357 [astro-ph.GA]`.

Siringo, G. et al. (2009), *The Large APEX BOlometer CAmera LABOCA*, Astronomy and Astrophysics **497** 945, arXiv: `0903.1354 [astro-ph.IM]`.

Smith, M. C. et al. (2021), *Efficient early stellar feedback can suppress galactic outflows by reducing supernova clustering*, MNRAS **506** 3882, arXiv: `2009.11309 [astro-ph.GA]`.

Smith, N. (2006), *A census of the Carina Nebula - I. Cumulative energy input from massive stars*, MNRAS **367** 763, arXiv: `astro-ph/0601060 [astro-ph]`.

– (2014), *Mass Loss: Its Effect on the Evolution and Fate of High-Mass Stars*, ARA&A **52** 487, arXiv: `1402.1237 [astro-ph.SR]`.

Smith, N. and S. P. Owocki (2006), *On the Role of Continuum-driven Eruptions in the Evolution of Very Massive Stars and Population III Stars*, ApJ **645** L45, arXiv: `astro-ph/0606174 [astro-ph]`.

Smith, R. J., P. C. Clark and I. A. Bonnell (2009), *Fragmentation in molecular clouds and its connection to the IMF*, MNRAS **396** 830, arXiv: `0903.3240 [astro-ph.GA]`.

Smith, R. J., S. C. O. Glover et al. (2014), *CO-dark gas and molecular filaments in Milky Way-type galaxies*, MNRAS **441** 1628, arXiv: `1403.1589 [astro-ph.GA]`.

Smullen, R. A. et al. (2020), *The highly variable time evolution of star-forming cores identified with dendrograms*, MNRAS **497** 4517, arXiv: `2004.01263 [astro-ph.GA]`.

Sneppen, A. and D. Watson (2025), *The optical constants and grain sizes of interstellar dust measured directly using the dust-scattered X-ray halo of GRB221009A*, A&A **701**, A65 A65, arXiv: `2506.12125 [astro-ph.HE]`.

Sofue, Y. (2020), *Rotation Curve of the Milky Way and the Dark Matter Density*, Galaxies **8** 37, arXiv: `2004.11688 [astro-ph.GA]`.

Soker, N. (2024), *Supernovae in 2023 (review): possible breakthroughs by late observations*, The Open Journal of Astrophysics **7**, 31 31, arXiv: `2311.17732 [astro-ph.HE]`.

Soler, J. D., H. Beuther et al. (2019), *Histogram of oriented gradients: a technique for the study of molecular cloud formation*, A&A **622**, A166 A166, arXiv: `1809.08338 [astro-ph.GA]`.

Soler, J. D., C. Zucker et al. (2023), *A panoptic view of the Taurus molecular cloud. I. The cloud dynamics revealed by gas emission and 3D dust*, A&A **675**, A206 A206, arXiv: `2305.13180 [astro-ph.GA]`.

Solomon, P. M., A. R. Rivolo et al. (1987a), *Mass, Luminosity, and Line Width Relations of Galactic Molecular Clouds*, Astrophysical Journal **319** 730.

– (1987b), *Mass, Luminosity, and Line Width Relations of Galactic Molecular Clouds*, ApJ **319** 730.

Solomon, P. M. and N. C. Wickramasinghe (1969), *Molecular and Solid Hydrogen in Dense Interstellar Clouds*, Astrophysical Journal **158** 449.

Soltis, J. et al. (2025), *A Multiwavelength Technique for Estimating Galaxy Cluster Mass Accretion Rates*, ApJ **985**, 212 212, arXiv: `2412.05370 [astro-ph.CO]`.

Souropanis, D. et al. (2022), *Linking the properties of accreting white dwarfs with the ionization state of their ambient medium*, MNRAS **513** 2369, arXiv: `2203.16647 [astro-ph.SR]`.

Sousbie, T. (2011), *The persistent cosmic web and its filamentary structure - I. Theory and implementation*, MNRAS **414** 350, arXiv: `1009.4015 [astro-ph.CO]`.

Spitzer, L. (1998), *Physical Processes in the Interstellar Medium*.

Springel, V. (2010), *E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh*, MNRAS **401** 791, arXiv: `0901.4107 [astro-ph.CO]`.

Strömgren, B. (1939), *The Physical State of Interstellar Hydrogen.*, ApJ **89** 526.

Stutzki, J. (2014), *GAUSSCLUMPS: Gaussian-shaped clumping from a spectral map*, Astrophysics Source Code Library, record ascl:1406.018, ascl: `1406.018`.

Suin, P. et al. (2025), *The role of magnetic field and stellar feedback in the evolution of filamentary structures in collapsing star-forming clouds*, A&A **698**, A119 A119, arXiv: `2505.02903 [astro-ph.GA]`.

Suri, S. et al. (2019), *The CARMA-NRO Orion Survey. Filamentary structure as seen in $C^{18}O$ emission*, A&A **623**, A142 A142, arXiv: `1901.00176 [astro-ph.GA]`.

Tafalla, M. and A. Hacar (2015), *Chains of dense cores in the Taurus L1495/B213 complex*, A&A **574**, A104 A104, arXiv: `1412.1083 [astro-ph.GA]`.

Tafalla, M., J. Santiago et al. (2004), *A highly collimated, extremely high velocity outflow in Taurus*, A&A **423** L21, arXiv: `astro-ph/0406539 [astro-ph]`.

Tammann, G. A., W. Loeffler and A. Schroeder (1994), *The Galactic Supernova Rate*, ApJS **92** 487.

Tasker, E. J. and J. C. Tan (2009), *Star Formation in Disk Galaxies. I. Formation and Evolution of Giant Molecular Clouds via Gravitational Instability and Cloud Collisions*, Astrophysical Journal **700** 358, arXiv: `0811.0207 [astro-ph]`.

Taylor, J. H. and J. M. Cordes (1993), *Pulsar Distances and the Galactic Distribution of Free Electrons*, ApJ **411** 674.

Thompson, T. A. and M. R. Krumholz (2016), *Sub-Eddington star-forming regions are super-Eddington: momentum-driven outflows from supersonic turbulence*, MNRAS **455** 334, arXiv: `1411.1769 [astro-ph.GA]`.

Tobin, J. J. and P. D. Sheehan (2024), *An Observational View of Structure in Protostellar Systems*, ARA&A **62** 203, arXiv: `2403.15550 [astro-ph.SR]`.

Toledano-Juárez, I. et al. (2023), *Collision of molecular outflows in the L1448-C system*, MNRAS **522** 1591, arXiv: `2304.00766 [astro-ph.GA]`.

Tommasino, C. et al. (2023), *HoVer-UNet: Accelerating HoVerNet with UNet-based multi-class nuclei segmentation via knowledge distillation*, arXiv e-prints, arXiv:2311.12553 arXiv:2311.12553, arXiv: `2311.12553 [eess.IV]`.

Torii, K. et al. (2015), *Cloud-Cloud Collision as a Trigger of the High-mass Star Formation: a Molecular Line Study in RCW120*, ApJ **806**, 7 7, arXiv: `1503.00070 [astro-ph.GA]`.

Traficante, A., A. Duarte-Cabral et al. (2018), *Testing the Larson relations in massive clumps*, MNRAS **477** 2220, arXiv: `1803.08929 [astro-ph.GA]`.

Traficante, A., G. A. Fuller et al. (2015), *Hyper: Hybrid photometry and extraction routine*, A&A **574**, A119 A119, arXiv: `1410.7293 [astro-ph.IM]`.

Traficante, A., B. M. Jones et al. (2023), *The SQUALO project (Star formation in QUiescent And Luminous Objects) I: clump-fed accretion mechanism in high-mass star-forming objects*, MNRAS **520** 2306, arXiv: `2301.09917 [astro-ph.GA]`.

Tsang, B. T.-H. and M. Milosavljević (2018), *Radiation pressure in super star cluster formation*, MNRAS **478** 4142, arXiv: `1709.07539 [astro-ph.GA]`.

Turatto, M. (2003), "Classification of Supernovae", *Supernovae and Gamma-Ray Bursters*, ed. by K. Weiler, vol. 598 21.

Tychoniec, Ł. et al. (2019), *Chemical and kinematic structure of extremely high-velocity molecular jets in the Serpens Main star-forming region*, A&A **632**, A101 A101, arXiv: `1910.07857 [astro-ph.SR]`.

Umemoto, T. et al. (2017), *FOREST unbiased Galactic plane imaging survey with the Nobeyama 45 m telescope (FUGIN). I. Project overview and initial results*, PASJ **69**, 78 78, arXiv: `1707.05981 [astro-ph.GA]`.

Urquhart, J. S., C. König, D. Colombo, A. Karska, A. Giannetti et al. (2025), *OGHReS: star formation in the outer Galaxy II (ℓ = 180–280°)*, MNRAS **539** 3105, arXiv: `2505.13756 [astro-ph.GA]`.

Urquhart, J. S., T. Csengeri et al. (2014), *ATLASGAL - Complete compact source catalogue: 280°<ℓ< 60°*, A&A **568**, A41 A41, arXiv: `1406.5741 [astro-ph.GA]`.

Urquhart, J. S., C. Figura et al. (2021), *SEDIGISM-ATLASGAL: dense gas fraction and star formation efficiency across the Galactic disc*, MNRAS **500** 3050, arXiv: `2012.01464 [astro-ph.GA]`.

Urquhart, J. S., C. König, D. Colombo, A. Karska, F. Wyrowski et al. (2024), *OGHReS: star formation in the outer galaxy (ℓ = 250°-280°)*, MNRAS **528** 4746, arXiv: `2401.00808 [astro-ph.GA]`.

Urquhart, J. S., C. König, A. Giannetti et al. (2018), *ATLASGAL - properties of a complete sample of Galactic clumps*, MNRAS **473** 1059, arXiv: `1709.00392 [astro-ph.GA]`.

Urquhart, J. S., M. R. A. Wells et al. (2022), *ATLASGAL - evolutionary trends in high-mass star formation*, MNRAS **510** 3389, arXiv: `2111.12816 [astro-ph.GA]`.

Valeille-Manet, M. et al. (2025), *ALMA-IMF: XVII. Census and lifetime of high-mass prestellar cores in 14 massive protoclusters*, A&A **696**, A11 A11, arXiv: `2502.09426 [astro-ph.GA]`.

Van Loo, S., E. Keto and Q. Zhang (2014), *Core and Filament Formation in Magnetized, Self-gravitating Isothermal Layers*, ApJ **789**, 37 37, arXiv: `1405.1013 [astro-ph.SR]`.

van Marle, A. J., S. P. Owocki and N. J. Shaviv (2008), "Continuum-Driven Winds from Super-Eddington Stars: A Tale of Two Limits", *First Stars III*, ed. by B. W. O'Shea and A. Heger, vol. 990, American Institute of Physics Conference Series, AIP 250, arXiv: `0708.4207 [astro-ph]`.

Van Oort, C. M. et al. (2019), *CASI: A Convolutional Neural Network Approach for Shell Identification*, ApJ **880**, 83 83, arXiv: `1905.09310 [astro-ph.IM]`.

Vasiliev, E. O. (2025), *Destruction of the interstellar dust by a supernova*, arXiv e-prints, arXiv:2512.24677 arXiv:2512.24677, arXiv: `2512.24677 [astro-ph.GA]`.

Vázquez-Semadeni, E., R. Banerjee et al. (2011a), *Molecular cloud evolution - IV. Magnetic fields, ambipolar diffusion and the star formation efficiency*, Monthly Notices of the RAS **414** 2511, arXiv: `1101.3384 [astro-ph.GA]`.

– (2011b), *Molecular cloud evolution - IV. Magnetic fields, ambipolar diffusion and the star formation efficiency*, Monthly Notices of the RAS **414** 2511, arXiv: `1101.3384 [astro-ph.GA]`.

Vázquez-Semadeni, E., G. C. Gómez et al. (2007), *Molecular Cloud Evolution. II. From Cloud Formation to the Early Stages of Star Formation in Decaying Conditions*, ApJ **657** 870, arXiv: `astro-ph/0608375 [astro-ph]`.

Vázquez-Semadeni, E., A. González-Samaniego and P. Colín (2017), *Hierarchical star cluster assembly in globally collapsing molecular clouds*, MNRAS **467** 1313, arXiv: `1611.00088 [astro-ph.GA]`.

Vázquez-Semadeni, E., A. Palau, J. Ballesteros-Paredes et al. (2019), *Global hierarchical collapse in molecular clouds. Towards a comprehensive scenario*, MNRAS **490** 3061, arXiv: `1903.11247 [astro-ph.GA]`.

Vázquez-Semadeni, E., A. Palau, G. C. Gómez et al. (2024), *The Turbulent Support (TS) and Global Hierarchical Collapse (GHC) models for molecular clouds compared. Differences, convergence, and myths*, arXiv e-prints, arXiv:2408.10406 arXiv:2408.10406, arXiv: `2408.10406 [astro-ph.GA]`.

Vink, J. S. (2011), *The theory of stellar winds*, Ap&SS **336** 163, arXiv: `1112.0952 [astro-ph.SR]`.

– (2022), *Theory and Diagnostics of Hot Star Mass Loss*, ARA&A **60** 203, arXiv: `2109.08164 [astro-ph.SR]`.

– (2024), *Stellar Winds*, arXiv e-prints, arXiv:2406.16517 arXiv:2406.16517, arXiv: `2406.16517 [astro-ph.SR]`.

Wakelam, V. et al. (2017), *H_2 formation on interstellar dust grains: The viewpoints of theory, experiments, models and observations*, Molecular Astrophysics **9** 1, arXiv: `1711.10568 [astro-ph.GA]`.

Walch, S., P. Girichidis et al. (2015), *The SILCC (SImulating the LifeCycle of molecular Clouds) project - I. Chemical evolution of the supernova-driven ISM*, MNRAS **454** 238, arXiv: `1412.2749 [astro-ph.GA]`.

Walch, S., A. P. Whitworth et al. (2015), *Comparing simulations of ionization triggered star formation and observations in RCW 120*, MNRAS **452** 2794, arXiv: `1109.3478 [astro-ph.GA]`.

Walch, S. and T. Naab (2015), *The energy and momentum input of supernova explosions in structured and ionized molecular clouds*, MNRAS **451** 2757, arXiv: `1410.0011 [astro-ph.GA]`.

Ward-Thompson, D. et al. (2016), *The JCMT and Herschel Gould Belt Surveys: a comparison of SCUBA-2 and Herschel data of dense cores in the Taurus dark cloud L1495*, MNRAS **463** 1008, arXiv: `1608.04353 [astro-ph.GA]`.

Watkins, E. J. et al. (2023), *PHANGS-JWST First Results: A Statistical View on Bubble Evolution in NGC 628*, ApJ **944**, L24 L24, arXiv: `2212.00811 [astro-ph.GA]`.

Weaver, R. et al. (1977), *Interstellar bubbles. II. Structure and evolution.*, ApJ **218** 377.

Weber, E. J. and L. Davis Jr. (1967), *The Angular Momentum of the Solar Wind*, ApJ **148** 217.

Wilson, R. W., K. B. Jefferts and A. A. Penzias (1970), *Carbon Monoxide in the Orion Nebula*, ApJ **161** L43.

Wise, J. H. et al. (2012), *The birth of a galaxy - II. The role of radiation pressure*, MNRAS **427** 311, arXiv: `1206.1043 [astro-ph.CO]`.

Witt, A. N. (2000), *Small and very small interstellar grains*, J. Geophys. Res. **105** 10299, arXiv: `astro-ph/9910259 [astro-ph]`.

Woltjer, L. (1972), *Supernova Remnants*, ARA&A **10** 129.

Wyrowski, F. et al. (2016), *Infall through the evolution of high-mass star-forming clumps*, A&A **585**, A149 A149, arXiv: `1510.08374 [astro-ph.SR]`.

Xu, D. and Y. Zhu (2024), *Surveying image segmentation approaches in astronomy*, Astronomy and Computing **48**, 100838 100838, arXiv: `2405.14238 [astro-ph.IM]`.

Xu, D., S. S. R. Offner et al. (2020a), *Application of Convolutional Neural Networks to Identify Protostellar Outflows in CO Emission*, ApJ **905**, 172 172, arXiv: `2010.12525 [astro-ph.GA]`.

– (2020b), *Application of Convolutional Neural Networks to Identify Stellar Feedback Bubbles in CO Emission*, ApJ **890**, 64 64, arXiv: `2001.04506 [astro-ph.GA]`.

Xu, S., S. Ji and A. Lazarian (2019), *On the Formation of Density Filaments in the Turbulent Interstellar Medium*, ApJ **878**, 157 157, arXiv: `1905.06341 [astro-ph.GA]`.

Xu, Z. et al. (2024), "Multi-Class Segmentation Based on the B-UNet Model", *2024 17th International Conference on Advanced Computer Theory and Engineering (ICACTE)* 257.

Yang, K., K. Qiu and X. Pan (2024), *Surveys of clumps, cores, and condensations in Cygnus-X. SMA observations of SiO (5–4)*, A&A **684**, A140 A140, arXiv: `2312.04880 [astro-ph.GA]`.

Yasuda, H., S.-H. Lee and K. Maeda (2021), *Dark Age of Type II Supernova Remnants*, ApJ **919**, L16 L16, arXiv: `2109.04032 [astro-ph.HE]`.

Yu, Q. et al. (2021), "Mask Guided Matting via Progressive Refinement Network", *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 1154.

Yuan, L. et al. (2021), *A Morphological Classification of 18,190 Molecular Clouds Identified in ^{12}CO Data from the MWISP Survey*, ApJS **257**, 51 51, arXiv: `2108.12110 [astro-ph.GA]`.

Zapata, L. A. et al. (2018), *ALMA Reveals a Collision between Protostellar Outflows in BHR 71*, AJ **156**, 239 239, arXiv: `1804.00625 [astro-ph.SR]`.

Zavagno, A., L. D. Anderson et al. (2010), *Star formation triggered by H II regions in our Galaxy. First results for N49 from the Herschel infrared survey of the Galactic plane*, Astronomy and Astrophysics **518**, L101 L101, arXiv: `1005.1591 [astro-ph.GA]`.

Zavagno, A., L. Deharveng et al. (2006), *Triggered massive-star formation on the borders of Galactic H II regions. II. Evidence for the collect and collapse process around RCW 79*, Astronomy and Astrophysics **446** 171, arXiv: `astro-ph/0509289 [astro-ph]`.

Zavagno, A., M. Pomarès et al. (2007), *Triggered star formation on the borders of the Galactic H ii region RCW 120*, A&A **472** 835, arXiv: `0707.1185 [astro-ph]`.

Zavagno, A., D. Russeil et al. (2010), *Star formation triggered by the Galactic H II region RCW 120. First results from the Herschel Space Observatory*, A&A **518**, L81 L81, arXiv: `1005.1615 [astro-ph.GA]`.

Zetterlund, E., J. Glenn and E. Rosolowsky (2018), *Northern Galactic molecular cloud clumps in Hi-GAL: dense gas map and environmental trends*, MNRAS **480** 893, arXiv: `1809.09233 [astro-ph.GA]`.

Zhang, X. et al. (2024), *Surveys of clumps, cores, and condensations in Cygnus X. Temperature and nonthermal velocity dispersion revealed by VLA NH_3 observations*, A&A **684**, A142 A142, arXiv: `2403.03845 [astro-ph.GA]`.

Zhuang, F. et al. (2021), *A Comprehensive Survey on Transfer Learning: This survey provides a comprehensive understanding of transfer learning from the perspectives of data and model.*, IEEE Proceedings **109** 43, arXiv: `1911.02685 [cs.LG]`.

Zinnecker, H. and H. W. Yorke (2007), *Toward Understanding Massive Star Formation*, ARA&A **45** 481, arXiv: `0707.1279 [astro-ph]`.

Zucker, C., C. Battersby and A. Goodman (2018), *Physical Properties of Large-scale Galactic Filaments*, Astrophysical Journal **864**, 153 153, arXiv: `1712.09655 [astro-ph.GA]`.

APPENDIX A

Effects of stellar feedback on molecular gas cores paper

A&A, 690, A345 (2024)
https://doi.org/10.1051/0004-6361/202451156

Astronomy & Astrophysics

Effects of stellar feedback on cores in STARFORGE

K. R. Neralwar[1,⋆,⋆⋆], D. Colombo[2], S. Offner[3], F. Wyrowski[1], K. M. Menten[1], A. Karska[1,2,4],
M. Y. Grudić[5], and S. Neupane[1]

[1] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[2] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[3] Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA
[4] Institute of Astronomy, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, ul. Grudziądzka 5, 87-100 Toruń, Poland
[5] Carnegie Observatories, 813 Santa Barbara St, Pasadena, CA 91101, USA

Received 18 June 2024 / Accepted 9 September 2024

ABSTRACT

Stars form in dense cores within molecular clouds, and newly formed stars influence their natal environments. How stellar feedback impacts core properties and evolution has been the subject of extensive investigation. We performed a hierarchical clustering (dendrogram) analysis of a STARFORGE (STAR FORmation in Gaseous Environments) simulation, modelling a giant molecular cloud to identify gas overdensities (cores) and study changes in their radius, mass, velocity dispersion, and virial parameter with respect to stellar feedback. We binned these cores on the basis of the fraction of gas affected by protostellar outflows, stellar winds, and supernovae and analysed the property distributions for each feedback bin. We find that cores that experience more feedback influence are smaller. Feedback notably enhances the velocity dispersion and virial parameter of the cores, more so than it reduces their radius. This is also evident in the linewidth–size relation, according to which cores in higher-feedback bins exhibit higher velocities than their similarly sized pristine counterparts. We conclude that stellar feedback mechanisms, which impart momentum to the molecular cloud, simultaneously compress and disperse the dense molecular gas.

Key words. supernovae: general – stars: winds, outflows – ISM: jets and outflows – ISM: structure

1. Introduction

Stellar feedback is one of the most important processes that regulate star formation in a galaxy (Krumholz et al. 2014; Agertz & Kravtsov 2015; Gatto et al. 2017; Peters et al. 2017; Kruijssen et al. 2019; Guszejnov et al. 2022). It can have a positive as well as a negative impact on star formation rates: it can trigger the formation of a new star by compressing gas or limit star formation by destroying the natal cloud. Simulations show that in the absence of stellar feedback, the gas cools rapidly, forming dense structures; this does not lead to the multi-phase interstellar medium (ISM) seen in a multitude of observations (Schuller et al. 2009; Colombo et al. 2019; Duarte-Cabral et al. 2021).

Stellar feedback mechanisms include protostellar outflows, stellar winds, radiation pressure, photoionisation, and supernova explosions (Krumholz et al. 2014; Girichidis et al. 2020). These different mechanisms operate at different spatial scales and are strongly linked to the evolutionary stages of stars (Hopkins et al. 2012). Feedback from massive stars provides an endpoint for star formation (Geen et al. 2015; Grudić et al. 2022).

Protostars expel gas in the form of bipolar jets (Fendt & Čemeljić 2002; Wu et al. 2004; Bally 2016). Expelled material entrains molecular gas from the surrounding core and accelerates it to high velocities, thus creating molecular outflows. These outflows have complex morphologies and clumpy structures (Arce & Goodman 2001) that evolve over time (Lee et al. 2002; Offner et al. 2011). They inject momentum and energy into the ISM on scales ranging from a few AU to tens of parsecs (Frank et al. 2014; Bally 2016). Protostellar outflows reduce stellar mass by ejecting accreting gas. They also remove excess angular momentum and facilitate reduced accretion rates during the protostellar phase of young stellar objects (Arce et al. 2007; Federrath et al. 2014; Bally 2016; Offner & Chaban 2017; Guszejnov et al. 2021, 2022). Outflows unbind the natal core gas, which can reduce the star formation efficiency of the dense gas (Offner & Chaban 2017).

Stellar winds are generated by the ejection of matter from the stellar surface due to the impact of radiation pressure on the gas in the stellar atmosphere (Arce et al. 2011; Pabst et al. 2020; Geen et al. 2023). Winds from massive stars profoundly affect the ISM and lead to complex molecular gas structures (Zinnecker & Yorke 2007; Xu et al. 2020b). They often manifest as shells or bubbles around individual stars that are often associated with HII regions (Deharveng et al. 2010; Schneider et al. 2020; Kirsanova & Pavlyuchenkov 2023). Stellar winds do not strongly affect the global cloud evolution but aid in preventing the runaway accretion of massive stars (Guszejnov et al. 2022).

Supernovae mark the endpoints of stellar evolution for massive stars and are a key ingredient in the study of interstellar gas (McKee & Ostriker 1977; Dubner & Giacani 2015). They act on scales of up to ∼100 pc and inject a large amount of energy (∼10^{51} ergs), which drives turbulence into the ISM; they also destroy molecular clouds (MCs) and regulate the chemical composition of the ISM by producing metals and dust (Wesson & Bevan 2021; Kirchschlager et al. 2024). These explosions inject

---

[⋆] Member of the International Max Planck Research School (IMPRS) for Astronomy and Astrophysics at the Universities of Bonn and Cologne.

[⋆⋆] Corresponding author; `kneralwar@mpifr-bonn.mpg.de` material and drive shocks into the ISM, and create supernova remnants (Dubner & Giacani 2015; Dokara et al. 2023).

The various stellar feedback mechanisms play an important role in shaping the internal structure of giant molecular clouds (GMCs) and constraining their lifetimes (Williams & McKee 1997; Hopkins et al. 2012; Chevance et al. 2023). The internal structure of MCs is often hierarchically classified according to their size and mass. GMCs are the largest molecular gas structures in the galaxy, and they form the denser molecular ISM in the Galaxy (Blitz 1993; Rosolowsky et al. 2021; Chevance et al. 2023). Cores are gas structures with sizes of around 0.1 pc or less (Ballesteros-Paredes et al. 2020). Some dense cores accrete gas and become massive enough to collapse under self-gravity (Offner et al. 2022), leading to the formation of stars.

New observational capabilities, in particular the high angular resolution combined with high sensitivity afforded by the Atacama Large Millimeter/submillimeter Array (ALMA), have led to observations of large samples of cores in the Galaxy (Kramer et al. 2024; Nony et al. 2023; Olmi et al. 2023), but how stellar feedback affects them is still not well understood. Simulations allow us to probe sub-core spatial scales and investigate how feedback operates on cores and shapes their properties (Offner & Chaban 2017).

In this study we examined the effect of outflows, winds, and supernovae on cores based on a STARFORGE (STAR FORmation in Gaseous Environments) simulation of a GMC that resolves the formation of individual stars and follows the star formation process until it is self-consistently halted by stellar feedback. We explored how different feedback mechanisms influence the physical properties of cores (i.e. the radius, mass, velocity dispersion, and virial parameter). Section 2 describes the STARFORGE simulation that we analysed. We performed a dendrogram analysis on the simulation snapshots to obtain the 'cores'. We discuss this analysis along with the methods for obtaining the core properties in Sect. 3. We present and discuss the effects of the three feedback mechanisms on the core properties in Sect. 4, where we also address the well-known scaling relations from the star formation literature. Finally, we summarise our findings in Sect. 5.

2. STARFORGE

The STARFORGE[1] simulations are 3D radiation magnetohydrodynamic simulations that follow the life of GMCs with a typical maximum spatial resolution of (∼10 AU). The framework is implemented in the GIZMO code (Hopkins 2015), which employs a Lagrangian mesh-less finite-mass method to solve the magnetohydrodynamic equations (Hopkins & Raives 2016). The simulations follow the formation, accretion, evolution, and dynamics of individual stars in a GMC and include all feedback mechanisms: jets, radiation, stellar winds, and supernovae. A detailed description of the numerical methods and relevant tests can be found in Grudić et al. (2021).

Jets, stellar winds, and supernovae are introduced as mass injection events in the simulations. The mass is injected from star particles into the simulation domain by two processes: local injection and cell spawning. Local injection refers to the weighted distribution of fluxes (mass, momentum, energy) to neighbouring cells. This method is used for photon injection and to propagate stellar winds when the free-expansion radius is unresolved. Cell-spawning refers to the process of creating new finite-mass gas cells at a certain rate, directly resolving the flow of feedback around the star. This is performed for protostellar jets and for winds when the free-expansion radius can be resolved at the given mass resolution.

In STARFORGE, protostellar outflows are modelled based on the prescription in Cunningham et al. (2011), using three parameters: (i) the fraction of mass accreted by the disc that is diverted to the jet ($f_w = 30\%$), (ii) the fraction of Keplerian velocity (v_{jet}) at the protostellar radius, and (iii) the collimation angle θ_0, which sets the angular distribution of the injected jet and outflow momentum. Main-sequence stars with mass $M > 2\ M_\odot$ generate stellar winds with a mass loss rate and wind velocity given by Eqs. (44) and (45) in Grudić et al. (2021). All stars with mass $M > 8\ M_\odot$ are modelled to end as supernova at the end of their lifetimes (following Eq. (47) in Grudić et al. 2021).A detailed description of how the feedback mechanisms are modelled in STARFORGE is given in Sect. 4 of Grudić et al. (2021).

In this work we analysed the fiducial `M2e4` simulation described in Table 1 of Guszejnov et al. (2022). This particular suite of simulations follows the evolution of an initially spherical cloud of $2 \times 10^4\ M_\odot$ with an initial radius ($R_{\rm cloud}$) of 10 pc, discretised with a mass resolution of $\delta m = 10^{-3} M_\odot$. The simulation contains 454 snapshots with an equal spacing of 24.7 kyr. The initial uniform magnetic field is in the $\hat{z}$ direction and corresponds to a mass-to-magnetic flux ratio (μ) of 4.2. The cloud is initialised with a Gaussian random velocity field scaled for an initial virial parameter $\alpha = 2$. The first stars form and launch protostellar outflows at 0.8 Myr. The first massive stars join the main-sequence star and begin producing stellar winds at 3.6 Myr. It is important to note that winds and radiation halt star formation before the first supernova occurs at 9.8 Myr (Grudić et al. 2022; Guszejnov et al. 2022).

The STARFORGE simulation tracks gas originating from jets (protostellar outflows), stellar winds, and supernovae using tracer fields. We refer to the mass of the feedback material contained in a cell relative to the mass of the non-feedback gas in the same cell as the 'feedback fraction'. We defined the gas associated with a particular feedback process as 'feedback gas'. We used the outflow, wind, and supernova feedback fractions as a measure of how much a particular type of feedback has affected a core; for example, the cores in low-feedback bins have less influence from feedback mechanisms compared to those in high-feedback bins. However, it is important to note that the feedback fractions track only the mass of the gas launched by the different feedback mechanisms. Thus, we miss regions, such as shocks produced by photoionisation, that are influenced by radiative feedback, which does not inject mass. STARFORGE implements feedback by injecting new cells (see Grudić et al. 2021, for details). The different feedback fractions are not directly comparable to each other, because the mass fraction does not fully encapsulate the relative significance of the feedback as it does not reflect velocity information. Thus, we binned the cores based on these feedback fractions (Sect. 3.3). These bins indicate how much a feedback has influenced a core and are therefore a better tool for comparing the influence of different feedback mechanisms on cores. For example, cores in high outflow bins are the ones most affected by outflows, and cores in high-wind bins are the ones most affected by winds. We therefore compared these cores most affected and least affected by different feedback mechanisms using feedback bins.

---

[1] `https://users.flatironinstitute.org/~mgrudic/starforge/`

3. Methodology

3.1. Dendrograms and cores

We used a dendrogram analysis (e.g. Rosolowsky et al. 2008) to identify the cores. Dendrograms are widely used to decompose molecular emission distributions into discrete structures such as clouds, clumps, and cores in simulations (Smullen et al. 2020; Offner et al. 2022, and references therein) and observations (Seo et al. 2015; Friesen et al. 2016; Keown et al. 2017; Colombo et al. 2019; Duarte-Cabral et al. 2021; O'Neill et al. 2021). In our case, the dendrogram is an abstraction of the hierarchical structure of the simulated MC gas density. In essence, a dendrogram describes how isosurfaces (or 3D contours) are nested within each other in the position-position-position datacube. Dendrograms are composed of three classes of structures: trunks, branches, and leaves. Leaves are 3D structures formed by single local maxima in the molecular gas distribution. The local maxima are identified through a nearest neighbour search. Here, we did not flatten the simulation data to a grid but instead used the SciPy KDTREE method within the dendrogram algorithm to identify the six closest neighbours to each cell. This allowed us to construct the dendrogram using the native simulation resolution. We defined the cores as the leaves of the dendrogram structure.

We adopted dendrogram parameters informed by the properties of typical observed cores, which are estimated to have densities $\geq 10^4$ and sizes of ~0.1 pc. In the implementation of the dendrogram algorithm used here (ASTRODENDRO), leaves must contain more than a minimum number of volumetric pixels[2] (min_npix). In addition, min_value is the minimum value considered in the dataset when constructing the dendrogram, and min_delta sets the significance level at which a leaf is considered to be an independent structure[3]. We set min_npix to 100, min_value to an H_2 number density of 10^4 cm^{-3} and min_delta to 10^4 cm^{-3}, following Offner et al. (2022). The min_npix = 100 corresponds to a minimum leaf mass of 0.1 M$_\odot$, which is comparable to the completeness limits of the core surveys (e.g. Sokol et al. 2019). The chosen peak threshold filters out low-density structures that would otherwise not be identified as cores in the observational data (Chen et al. 2019). While GIZMO estimates the fraction of mass in each cell that is ionised, neutral, and molecular, we made no explicit assumption about the phase or temperature of the gas at these densities. All gas in a given cell that satisfies the density threshold was included in the analysis.

We aimed to identify the causation between the core properties and the presence of feedback, and not just the correlation. For example, if a structure with a significant amount of directly launched outflow material is included in the core identification, such a core will naturally tend to have a higher velocity dispersion, even if most of the core is largely unaffected. We instead aimed to explore the properties of cores affected by the feedback as seen through molecular line observations, for example by excluding hot outflow and wind material that is traced by the atomic and ionised gas. Our cores mostly contain gas that is traced by various molecular lines such as ^{12}CO, ^{13}CO, C^{18}O, and NH$_3$, where the relatively dense and cold ISM varies in feedback influence. To exclude raw (recently injected) feedback material, we mask any pure feedback cells that, by construction, have masses below the fiducial gas mass resolution of 10^{-3} M$_\odot$ (see Grudić et al. 2021). Some feedback gas may otherwise be included by the above density threshold criterion, since the outflow material launched by the protostar sub-grid model is often directly injected into the relatively dense core environment. Moreover, if we include these pure feedback cells when constructing leaves, we retrieve some cores with a mass less than 0.1 M$_\odot$, below typical observational completeness limits. Our dendrogram parameter *min_npix* acts as a minimum mass threshold of $min\_npix \times \Delta m$, given $\Delta m \geq 10^{-3}$ M$_\odot$. Consequently, our dendrogram structures include only well-mixed feedback material, which has already interacted (merged) with the surrounding envelope, similar to observational expectations.

3.2. Core properties

We used the properties of simulation cells grouped as a leaf by dendrogram analysis to obtain four independent properties of the cores[4]: the radius, velocity dispersion, mass, and virial parameter.

We defined the core radius as

$$r = \sqrt{\frac{5}{2}\frac{\sum r_1^2 m_{\rm cell}}{\sum m_{\rm cell}}}, \tag{1}$$

where $r_1 = \sqrt{dx^2 + dy^2 + dz^2}$ and $d_i = i_{\rm peak}$[5] - $i_{\rm cell}$ ($i = x, y, z$). The radius (Eq. (1)) was obtained by assuming the cores to be spherical and using the moment of inertia relation for a sphere. We defined the 1D velocity dispersion for each leaf as

$$\sigma_{1d} = \sqrt{\frac{\sum (v_i^{\rm cell} - \bar{v}_i^{\rm cell}) m_{\rm cell}}{3 \sum m_{\rm cell}}}, \tag{2}$$

where $v_i^{\rm cell}$ represent the cell velocities in the x, y, z directions. We defined the core mass as the sum of the molecular gas mass of the leaf cells, where we used the neutral hydrogen abundance ($f_{\rm neutral}$) and the molecular mass fraction ($f_{\rm mol}$) to determine the amount of molecular material in each cell:

$$m_{\rm core} = \sum m_{\rm cell} \times f_{\rm neutral} \times f_{\rm mol}. \tag{3}$$

This mass is comparable to the total gas mass in the cells, since most of the identified cores are composed of predominantly cold and molecular gas.

The core virial parameter $\alpha_{\rm vir}$ is the ratio of the kinetic energy to the potential energy. We defined the core kinetic energy as

$$K.E. = \sum \frac{m_{\rm cell}(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)}{2}. \tag{4}$$

The potential energy was obtained using the gravitational potential calculation function[6] (Grudić & Gurvich 2021).

We used the feedback fractions of the cells, $f_{\rm o}^{\rm cell}$, $f_{\rm w}^{\rm cell}$, and $f_{\rm s}^{\rm cell}$, to calculate the gas mass affected by each type of feedback for each cell ($m_{\rm cell,o}$, $m_{\rm cell,w}$ and $m_{\rm cell,s}$). The sum of these masses for the leaf cells gives the total feedback mass for the core.

---

[2] Gas cells in STARFORGE .

[3] A detailed description of these dendrogram variables can be found here: `https://dendrograms.readthedocs.io/en/stable/using.html`

[4] We use the terms leaf and core interchangeably throughout the paper. Hereafter, we refer to cells that belong to a particular leaf of the dendrogram as 'leaf cells'.

[5] The leaf peak cell ($i_{\rm peak}$) refers to the cell in the leaf with the maximum density as determined by the dendrogram.

[6] `https://github.com/mikegrudic/pytreegrav/blob/aa88a552206d22c194d0cdc3399c4b5d256c0990/src/pytreegrav/frontend.py#L68`

Table 1. 25th and 75th percentile values of the feedback fraction distributions.

Feedback	25% pth	75% pth
Outflow	1.1e-03	2.7e-02
Wind	7.1e-10	3.8e-06
Supernova	1.1e-11	2.6e-06

Notes. We use these values to define the feedback bins in Sect. 3.3.

The difference between the core mass and the mass affected by different feedback gives the non-feedback core mass. The ratio of the core feedback mass to the non-feedback mass gives the respective feedback fractions (f_o, f_w, and f_s) for the cores:

$$f_x = \sum \frac{m_{\mathrm{cell},x}}{m_{\mathrm{cell}} - m_{\mathrm{cell},y} - m_{\mathrm{cell},z}}, \tag{5}$$

where $x, y, z = \{\mathrm{o, w, s}\}$ corresponds to outflows, winds, and supernovae, respectively.

3.3. Outflow, wind, and supernova bins

Based on the outflow, wind, and supernova feedback fractions, f_o, f_w, and f_s, we binned the cores (Fig. 1) and analysed their properties. The 'no feedback' bin contains pristine cores that have zero feedback (f_o= 0, f_w = 0 and f_s = 0). This bin remains the same for outflow, wind, supernova, and global feedback. The other three bins are based on the percentile cuts of the respective feedback fractions (Table 1). The 'low' feedback bin contains cores containing 0–25% percentile feedback gas, the 'moderate' feedback bins contain cores with 25–75% percentile feedback gas and the 'high' feedback bins contain cores with >75% percentile feedback gas. These bins differ for outflows, winds, and supernovae as they are calculated using the percentile cuts on f_o, f_w, and f_s, respectively. We placed an additional constraint on the cores in each bin to ensure that the respective feedback is the most prominent. For example, the low-outflow bin contains only cores with $f_\mathrm{w} = f_\mathrm{s} = 0$. The moderate outflow bin has cores with $f_\mathrm{w} < 25\%$ and $f_\mathrm{s} < 25\%$. The high outflow bin contains only cores that have $f_\mathrm{w} < 75\%$ and $f_\mathrm{s} < 75\%$. The wind and supernova bins follow the same criterion; for example, the low-wind bin contains only cores with $f_\mathrm{o} = f_\mathrm{s} = 0$. This definition ensures that trends in the properties associated with a particular bin are likely the result of the respective feedback. For example, since the high-outflow bin does not have any cores with high winds and high supernovae, we can assume that the general trends in the property distributions (e.g. high velocity dispersion) of cores in these bins are a result of outflows.

3.4. Global feedback bins

We defined global feedback bins to study the combined effects of the three types of feedback. This ensured that our results can be compared with observational data, for which it is not possible to discriminate between the effects of the different mechanisms. The 'no global' bin is identical to the 'no feedback' bin. The other three bins are based on combinations of the outflow, wind, and supernova bins. The 'low global' bin contains cores that belong to at least two 'low-feedback' bins (e.g. low outflow and low wind), with the third feedback (supernova, in this case) being empty. For example, all the cores that have $0 < f_\mathrm{o} < 25\%$, $0 < f_\mathrm{w} < 25\%$ and $f_\mathrm{s} = 0$ belong to this bin. Similarly, cores with low wind, low supernova, and no outflows also belong to the low global bin. The cores in the 'moderate global' bins must belong to at least two moderate-feedback bins, the third feedback being < 25%. The 'high global' bin contains cores that are in at least two high bins, the third feedback being < 75%. The binning helps us combine the cores most affected by feedback and least affected by feedback.

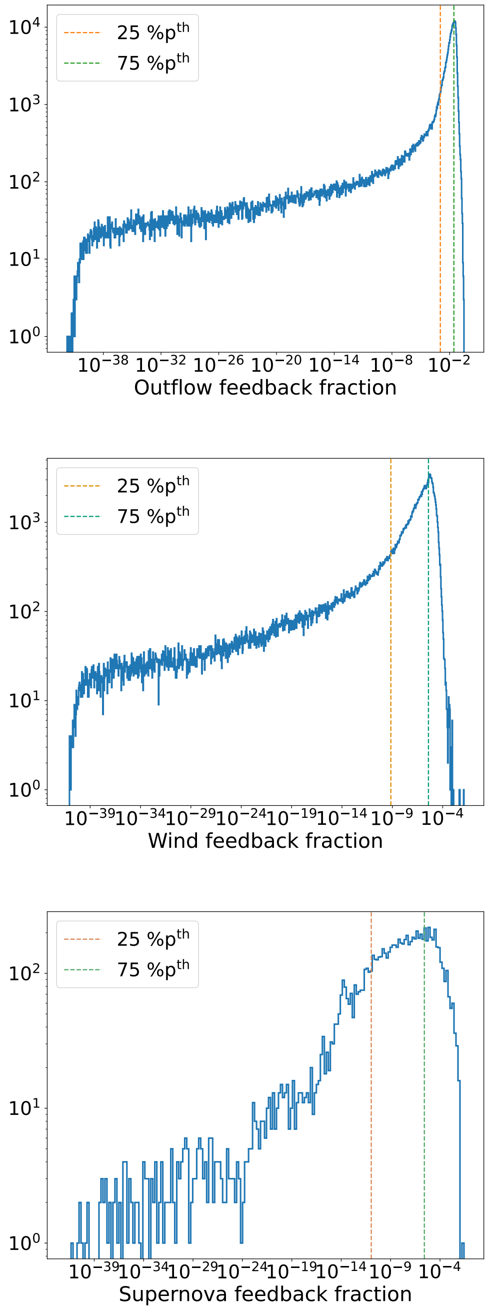


Fig. 1. Histogram of the outflow (top), wind (middle), and supernova (bottom) feedback fractions. We mark the 25th percentile and the 75th percentile limits for the feedback fractions.

4. Results

Figure 2 shows the structure of the cores as contours of the dendrogram leaves (cores) plotted on top of the GMC from which they were derived, for four different snapshots. The grey scale represents the projected H_2 density for the GMC. The contour

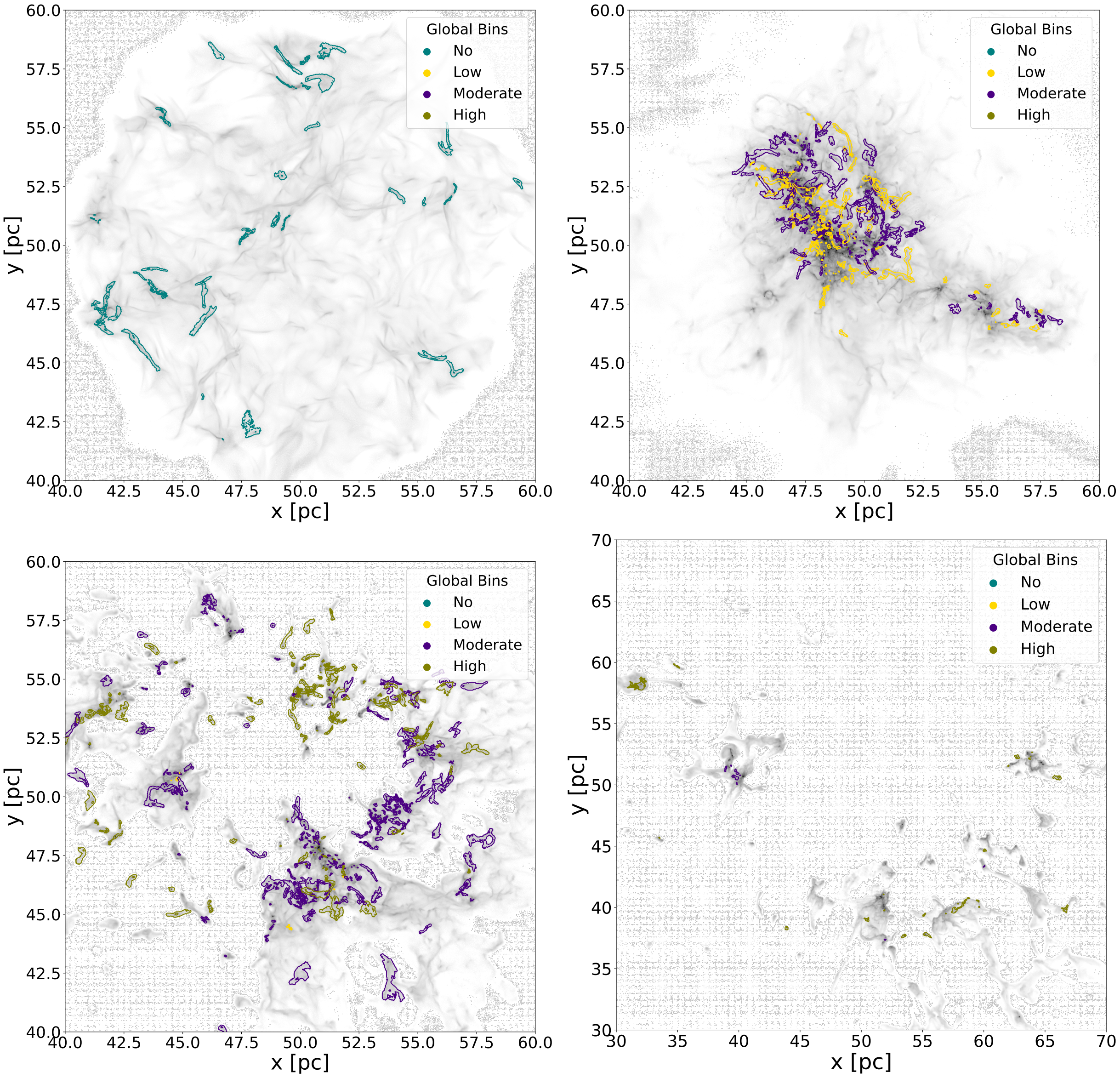


Fig. 2. Visual representation of cores in four different snapshots that correspond to 0.4 Myr (top left), 4.9 Myr (top right), 7.5 Myr (bottom left), and 9.9 Myr (bottom right) after the start of the simulation. The grey background shows the GMC (H_2 density) projected along the z axis. The contours represent the cores in different global feedback bins (Sect. 3.4).

colours represent the cores based on the global feedback bins (Sect. 3.4). We also visualised the changes in the properties of the cores by plotting them as spheres with their radius as the various properties (Fig. 3). By analysing the cores from multiple snapshots, we also followed the evolution of the simulated GMC to understand the effects of the evolutionary stage of the GMC on the core properties. Figure 2 shows the spherical GMC (top left) that appears to contract under self-gravity (top right) to form dense structures that lead to the formation of protostars. The GMC further evolves to have main-sequence stars (Fig. 2, bottom left) with stellar winds, which, along with other feedback mechanisms, leads to a more scattered configuration of the cores. Figure 2 (bottom right) shows the GMC in which star formation has halted and supernova explosions have started. Here, most of the molecular gas is dispersed by the various feedback mechanisms.

Most of the cores appear to be elongated filamentary structures (Fig. 2). These are projected images of the GMC and 3D plots might reveal different shapes for cores. However, it is beyond the scope of this paper. Figure 3 shows that most of the cores at later times belong to moderate- and high-feedback bins and have high velocity dispersions and virial parameters. This agrees with the results of Xu et al. (2020a) that the outflows driven by older sources have a higher velocity dispersion. Moreover, stellar winds and supernova feedback, which start at later snapshots, also contribute to these velocity dispersion values. The moderate and high-feedback bins also contain a few massive cores. These are bound cores with low-velocity dispersions that are likely sites for star formation. The cores in the later snapshots appear more dispersed and separated. This is a result of the evolution and expansion of the GMC as seen in Fig. 2.

4.1. Core properties versus mechanical feedback

In this section we compare the properties of the cores with the feedback fractions of protostellar outflow, stellar wind, and

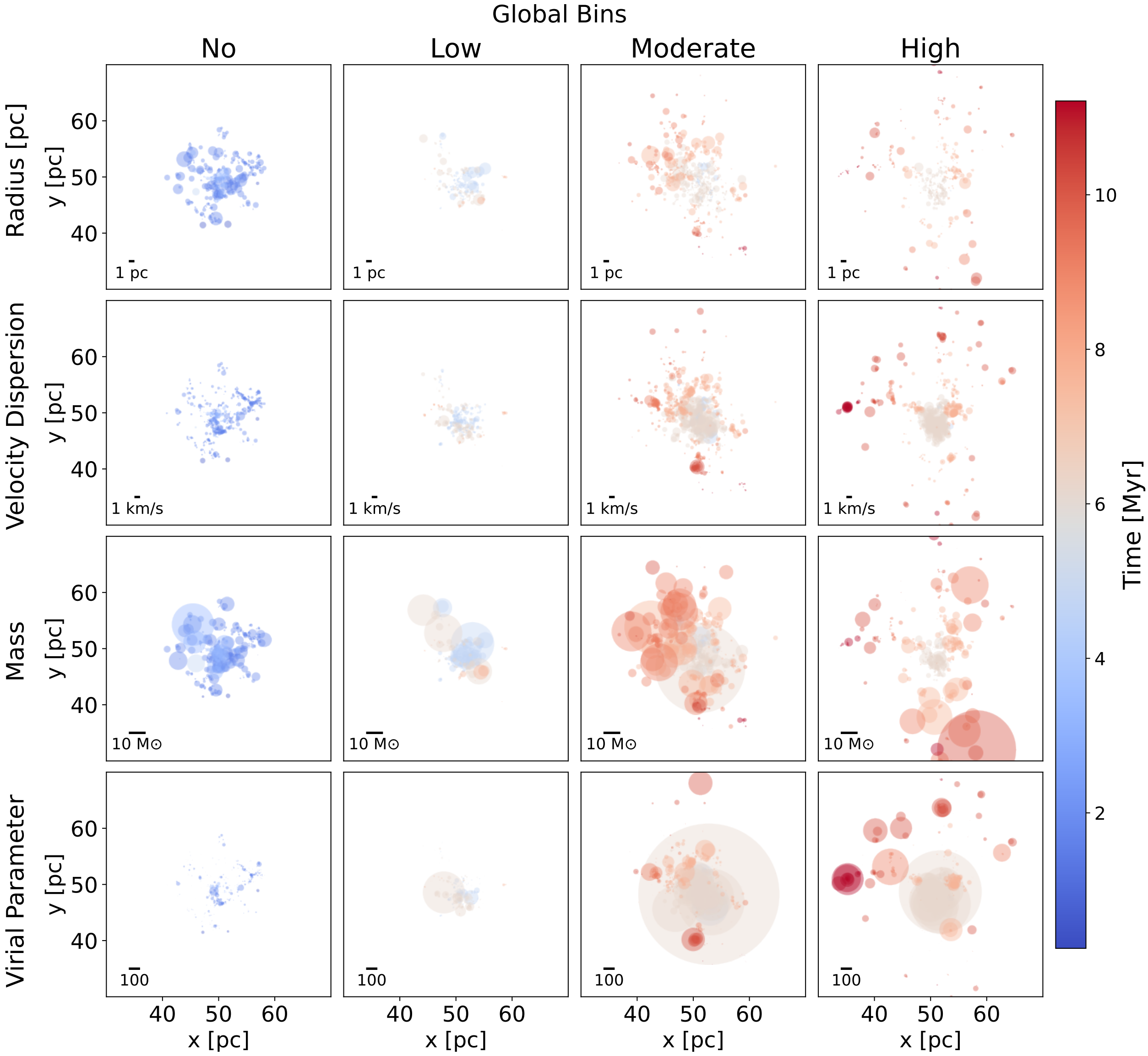


Fig. 3. Distribution of cores in different snapshots along with their properties in various global feedback bins. The positions of the cores represent their locations in the simulation box. The cores are are assumed to be circular, and the core properties (Sect. 3) are scaled with the size of these circles. First row: Radius of the core. Second and third rows: Velocity dispersion and mass of the cores, respectively. Bottom row: Virial parameter. The colours represent the time elapsed (in Myr) since the start of the simulation.

supernova. The distribution of core properties with feedback fractions $< 10^{-15}$ (~10%) is similar to the property distribution with feedback fractions around 10^{-15}. While the feedback fraction is captured down to machine precision, there is little change in the core properties, and the cores that have very low feedback fractions are not noticeably affected by the feedback. The turbulent eddy mixing model implemented in GIZMO rapidly distributes the feedback gas throughout the domain as soon as feedback processes are initiated. Therefore, we truncate Fig. 4 to 10^{-15} to focus on the trends associated with the higher feedback fractions. The cores in the very late snapshots do not have the highest feedback fractions. This suggests that the feedback is not entirely cumulative. These later timescales represent the epoch when star formation has nearly halted, so the rate of feedback injection from new stars is significantly reduced (star formation history of the simulated GMC; Guszejnov et al. 2022, as shown in their Fig. 2).

Figure 4 shows an increase in the extreme values of the core radius[7] with the feedback fraction for both protostellar outflows and stellar winds. We find two competing effects. The cores with the largest radii belong to the late-time snapshots, where a supernova explosion has already occurred. The increase in the core radii is most likely due to the momentum injected by the various feedback mechanisms. Due to the injected momentum, there could be expansion, rotation and shear motions in the cores, which increase their size. Figure 2 shows most cores as elongated or irregular structures rather than spheres. At the same time, feedback also compresses the cores and disperses the molecular gas, producing smaller cores. The smaller high-feedback cores could also be due to the evolving conditions in the cloud as the

---

[7] Core radii are calculated with an assumption that the cores are spherical; however, most of our cores appear to be filamentary. Thus, an elongated core also leads to a large effective core radius.

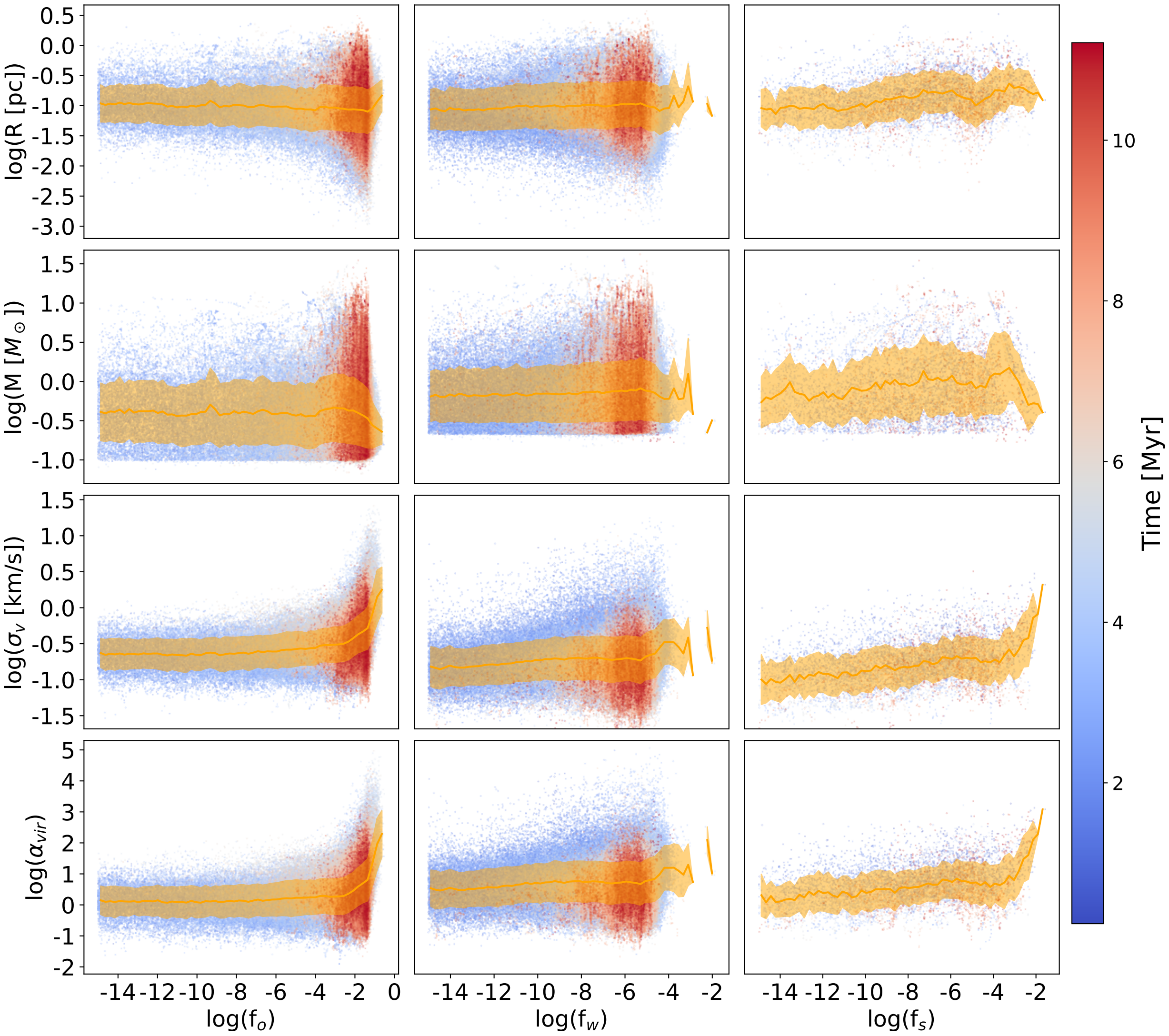


Fig. 4. Core properties versus feedback fraction from outflows (left), winds (centre), and supernovae (right). The rows represent the radius of the core, the velocity dispersion, the mass, and the virial parameter, respectively. The orange filled area represents the average distribution of all cores, obtained by binning over the feedback fractions to get the 25th and 75th percentiles. The point colours indicate the simulation time when the core is identified, as time elapsed in Myr.

GMC collapses, which produces smaller cores while simultaneously increasing the number of stars and the amount of stellar feedback, which leads to higher feedback fractions in the cores. Despite the presence of large cores with high feedback, our overall analysis shows that outflows and winds lead to a decrease in the average core radius.

The cores with the largest velocity dispersions are in regions with significant stellar feedback. Figure 4 shows a gradual increase in the average velocity dispersion of the cores as the feedback increases, and a rapid increase in the cores with high feedback fractions. The increase in velocity dispersion for cores with high feedback is likely due to the momentum injected by the feedback mechanisms (Offner & Chaban 2017). The most turbulent and least turbulent cores have high wind feedback fractions (f_w; Fig. 4; central plot in the second row). The cores with small velocity dispersions belong to the evolved GMC where star formation has halted and the amount of gas from stellar winds is lower compared to the peak star formation phase.

The third row of panels in Fig. 4 shows that the average mass distribution is generally constant and only decreases at high feedback fractions (f_o, f_w, and f_s). The decrease in mass at high feedback fractions might be the result of feedback destroying cores and splitting them into less massive cores. However, the most massive cores have relatively high feedback fractions and belong to late-time snapshots. These cores are most likely the result of winds and supernova compressing the gas.

The bottom row of Fig. 4 shows a gradual increase in the average virial parameter with increasing feedback fraction, with a strong increase at the highest feedback fractions. It is clear that this is driven by higher velocity dispersions, as the core mass and radius do not change significantly.

4.2. Violin plots

In this section we use double violin plots (Fig. 5) to visualise the distributions of core properties in the various feedback bins defined in Sects. 3.3 and 3.4. Binning the data using percentiles based on the feedback fraction allows us to study the changes in the properties for the cores that are actually influenced by stellar

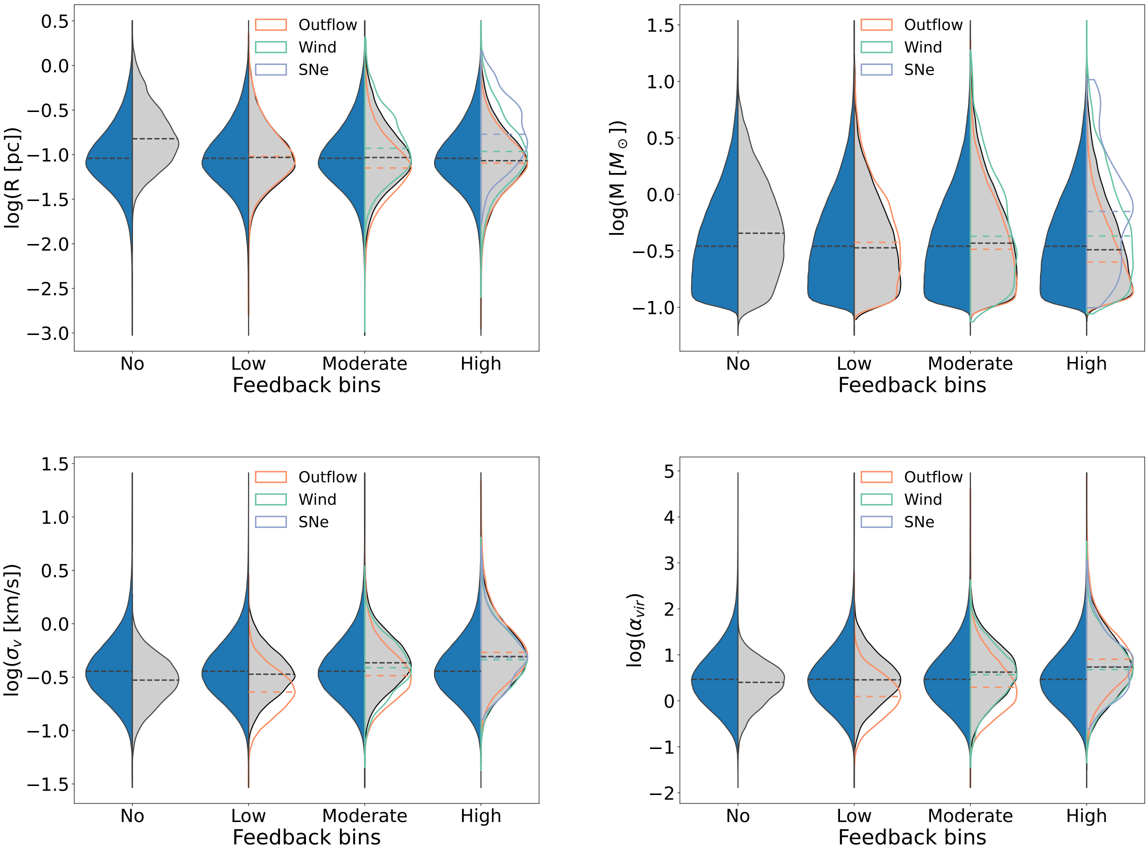


Fig. 5. Distributions of the core radius (top left), mass (top right), velocity dispersion (bottom left), and virial parameter (bottom right) in the different feedback bins. The split violin plot presents the total distribution (blue) and the distribution of cores in the global feedback bins (grey) for each axis. The colours representing the outflow, wind, and supernova bins are marked in the legends. The dashed lines represent the medians of the different distributions. Violin plots present the density of the data at different values, which were smoothed using the kernel density estimator.

Table 2. Number of cores in each feedback bin.

Feedback	Low	Moderate	High
Global	16 994	83 506	46 358
Outflow	65 752	75 297	41 751
Wind	0	5326	15 149
Supernova	0	1	779

Notes. We obtain 391 261 cores through the dendrogram analysis, of which 388 958 are distributed in the various feedback bins. The no-feedback bin contains 38 045 cores.

feedback. The low wind, low supernova, and moderate supernova bins contain almost no cores (Table 2) and are therefore not included in the violin plots. We used the Wasserstein distance (Appendix A) to statistically compare the different distributions.

Figure 5 shows a decrease in the average radius and mass from the no-feedback bin to the low-global feedback bin. However, the radius (Fig. 4) shows only a slight change from low to high-global feedback. The mass distributions show negligible changes over the feedback bins. The cores in the low global bin are most likely the star-forming cores that are gravitationally collapsing and therefore are smaller (Fig. 2). This suggests that the main impact of the feedback occurs via turbulence driving or energy injection. The lower average values of radius in the outflow bins compared to the respective wind and supernova bins point to the disruptive effects of these mechanisms on the cores (Fuente et al. 2002; Maret et al. 2009; Narayanan et al. 2012). However, flows that play a very important role in core dispersal may not be the main factor in GMC disruption (Maret et al. 2009; Guszejnov et al. 2022). The nearly constant radius and mass over the wind bins suggest that stellar winds do not affect the radius of the cores as strongly as other mechanisms or that the expansion of the cores due to stellar winds might be suppressed due to the dissipation of gas by the various feedback mechanisms (including winds themselves; e.g. Pabst et al. 2019). The high supernova bin shows an average mass and radius higher than any other bins, including the pristine cores (grey violins in Fig. 5) and the total core distribution (blue violins in Fig. 5). These cores are likely due to gas compression along the shock front of the supernova explosion, leading to elongated massive structures.

The lower panels of Figure 5 show the velocity dispersion and virial parameter distributions, which both monotonically increase from no- to high-feedback bins. The increase in core velocity dispersion from low- to high-feedback bins for all feedback mechanisms (Fig. 5) is consistent with the standard picture of momentum and energy injection via stellar feedback (Arce et al. 2007; Offner & Chaban 2017; Offner & Liu 2018; Bieri et al. 2023; Geen et al. 2023). Observations of MCs have also revealed an increase in the linewidths of structures in feedback-dominated regions (Wong et al. 2022; Grishunin et al. 2024). The low-outflow bin velocity dispersion distribution is significantly different from the distribution in the low-global feedback bin. This difference becomes smaller in the moderate

Table 3. Larson relation: slopes and widths of the PCA ellipse.

Feedback bins	Slope	σ_{slope}	Scatter	σ_{scatter}
No	0.54	<0.01	0.3	<0.01
Global				
Low	0.48	0.01	0.34	<0.01
Moderate	0.38	<0.01	0.46	<0.01
High	0.35	0.01	0.58	<0.01
Outflow				
Low	0.45	<0.01	0.32	<0.01
Moderate	0.31	0.01	0.48	<0.01
High	0.56	0.02	0.56	<0.01
Wind				
Moderate	0.53	0.01	0.36	<0.01
High	0.36	0.01	0.48	<0.01
Supernova				
High	0.55	0.04	0.42	0.01

Notes. All values are rounded to two decimal places. The lines of constant surface densities in Fig. 6 have a slope of 1.

Table 4. Heyer relation: slopes and widths of the PCA ellipse.

Feedback bins	Slope	σ_{slope}	Scatter	σ_{scatter}
No	0.15	0.01	0.67	<0.01
Global				
Low	0.31	0.01	0.74	<0.01
Moderate	0.71	0.01	0.89	<0.01
High	1.18	0.02	1.04	<0.01
Outflow				
Low	0.4	0.01	0.66	<0.01
Moderate	0.8	0.01	0.87	<0.01
High	1.06	0.02	1.03	<0.01
Wind				
Moderate	0.02	0.04	0.81	0.01
High	0.89	0.02	0.93	0.01
Supernova				
High	–1.84	11.57	0.89	0.03

Notes. The lines of constant virial parameters in Fig. 7 have a slope of 1. The symbols and conventions follow Table 3.

and high-feedback bins. This large increase in velocity dispersion is likely a result of outflows injecting significant momentum and energy, thereby becoming a major source of turbulence in the cores (Zhang et al. 2005; Arce et al. 2007; Offner & Chaban 2017; Davis et al. 2008; Duarte-Cabral et al. 2012; Bally 2016; Verliat et al. 2022). In comparison to outflows, stellar winds appear to have less influence on velocity dispersion, which increases less in cores dominated by wind feedback (Sect. 4.2). Although the total integrated wind energy injection may be higher in GMCs like the one we model (Grudić et al. 2022), the wind energy and momentum are injected at later times by massive stars into lower-density gas, occurring after much of the dense core envelope has already been accreted or dispersed. Previous studies also show that winds deposit energy and momentum locally (Fichtner et al. 2022) but do not offset global turbulence in a MC (Offner & Arce 2015; Guszejnov et al. 2022).

Table 5. Size-mass relation: slopes and widths of the PCA ellipse.

Feedback bins	Slope	σ_{slope}	Scatter	σ_{scatter}
No	1.29	<0.01	0.34	<0.01
Global				
Low	1.28	0.01	0.4	<0.01
Moderate	1.25	0.01	0.48	<0.01
High	1.16	0.01	0.52	<0.01
Outflow				
Low	1.28	<0.01	0.39	<0.01
Moderate	1.22	0.01	0.51	<0.01
High	0.87	0.01	0.49	<0.01
Wind				
Moderate	1.35	0.02	0.4	<0.01
High	1.25	0.01	0.46	<0.01
Supernova				
High	1.54	0.04	0.38	0.01

Notes. The lines of constant surface mass densities in Fig. 8 have a slope of 2. The symbols and conventions follow those of Table 3.

Figure 5 (bottom right) shows that the core virial parameters span three orders of magnitude. Although the typical core is close to virial equilibrium, many cores have significantly higher virial parameters. The large spread in Fig. 5 (bottom right) is not unexpected, since we did not require boundedness for core identification and since many of the identified cores likely do not go on to form stars (e.g. Offner et al. 2022). We expect that a larger fraction of cores will appear bound for choices of higher min_value. However, these results are not inconsistent with what we find for observed cores and clumps, which exhibit a broad range of virial parameters ranging from 0.03 to a few hundred (Kauffmann et al. 2013). Here, the cores with high virial parameters could be pressure-dominated structures such as low-density thermal cores (Lada et al. 2008), droplets (Chen et al. 2019), or overdensities in shocked regions.

4.3. Scaling relations

In this section we analyse the correlations between the core properties for various feedback bins and compare them to three scaling relations from the star formation literature. The core distributions in scaling relation plots are shown with 1-sigma kernel density estimators[8]. We also used principal component analysis (PCA), following Colombo et al. (2019) and Neralwar et al. (2022), to obtain the slopes of the distributions and their scatter.

Principal component analysis constructs the covariance matrix between the two core properties and calculates the eigenvectors of the matrix. The inclination of the largest eigenvector (major axis of the PCA ellipse) with respect to the x-axis gives the slope of the distribution. The eigenvector perpendicular to the major axis is a measure of the scatter in the data. The errors on slopes and scatter (Tables 3–5) were obtained using the bootstrap technique, where 1000 new samples are generated from the main sample by drawing with replacement. The slopes and

---

[8] We used the seaborn kernel density estimator plots `https://seaborn.pydata.org/generated/seaborn.kdeplot.html` with level = 0.0027, which corresponds to 1-sigma contours.

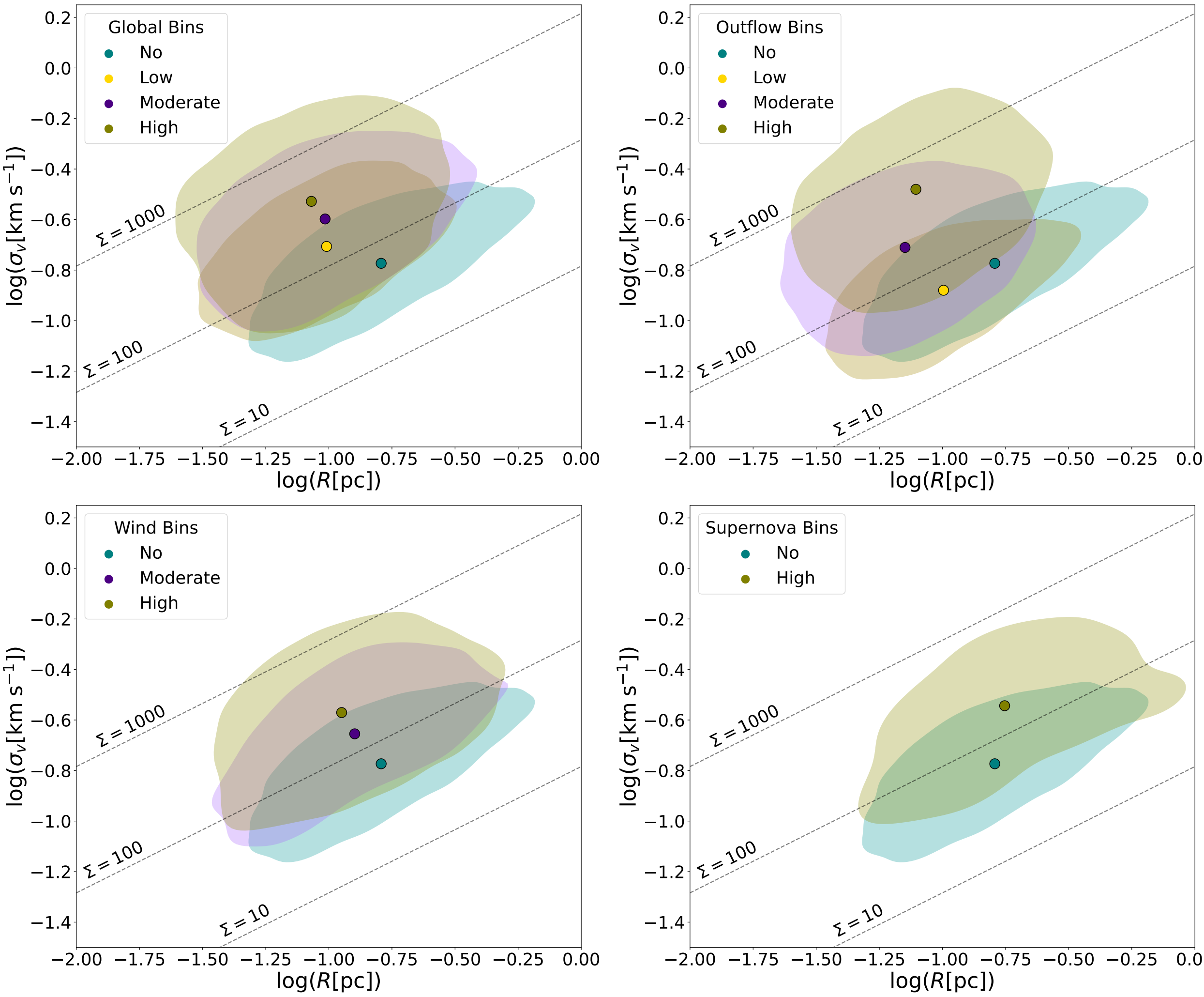


Fig. 6. Size-linewidth relation (σ_v versus R) for cores in different feedback bins. Top left: Distribution in the four global feedback bins. Top right: distributions in the outflow bins. Lower left and lower right: wind and supernova bins, respectively. The dashed lines represent Larson's first relation (slope of 0.5; Larson 1981; Solomon & Wickramasinghe 1969) for different surface mass densities (units of $M_\odot$ pc^{-2}) for virialised structures ($\alpha_{\rm vir}$ = 1). The different colours represent the distribution of cores in the various feedback bins, as noted in the legend.

widths in the table represent the respective mean values of the bootstrapped samples. $\sigma_{\rm slope}$ and $\sigma_{\rm width}$ provide the errors on the slopes and widths, respectively; calculated as the standard deviation of the respective values on the bootstrapped samples. While scaling relations have been measured extensively in observations of MCs, here we studied cores in simulations. Moreover, we obtained the properties of the cores using methods different from those that observers generally use. Observed structures, particularly those defined in optically thick molecular line tracers like ^{12}CO and ^{13}CO, are subject to projection and opacity effects that may significantly impact inferred masses, velocity dispersions and virial parameters (Beaumont et al. 2013; Mairs et al. 2014). We caution the reader that this may lead to discrepancies in the slopes for the different distributions.

Larson's size-linewidth relation $\sigma_v = 1.10\,L^{0.38}$ has been used to study and constrain the fundamental properties of MCs for the past 40 years (Larson 1981). It was obtained by studying MCs of sizes ranging from sub-parsec scales to ∼ 100 pc. Observations of dense cores (Myers 1983), and MCs (Solomon et al. 1987) suggest a slope of 0.5, which is consistent with the latest observations (e.g. Colombo et al. 2019; Neralwar et al. 2022).

Figure 6 presents Larson's law for cores in different feedback bins. A general observation is that the cores in each bin follow Kolmogorov's turbulence law: larger cores have a higher velocity dispersion. All sets of cores follow the expected observational trend; however, the cores in the higher-feedback bins have a higher velocity dispersion compared to the similar-sized cores in the lower-feedback bins. In addition, the cores in the no-feedback bin show a tighter correlation (Table 3).

Figure 6 (top right) shows a shift in the distribution from no- to low-outflow bins towards smaller cores with low-velocity dispersions. These low-outflow cores are detected in the early stages of GMC evolution, when denser structures are still forming, and parts of the GMC are collapsing under self-gravity. The moderate- and high-outflow bins represent a distribution of cores with strong outflows leading to gas dispersal and high-velocity dispersion. The distribution of high supernova feedback cores (Fig. 6; bottom right) contains some of the largest cores in the simulation, which is likely due to the momentum from the supernova leading to core expansion and/or elongation. Many of these cores are identified in the large, high-velocity shell of entrained gas created by the supernova shock. However, the cores with the

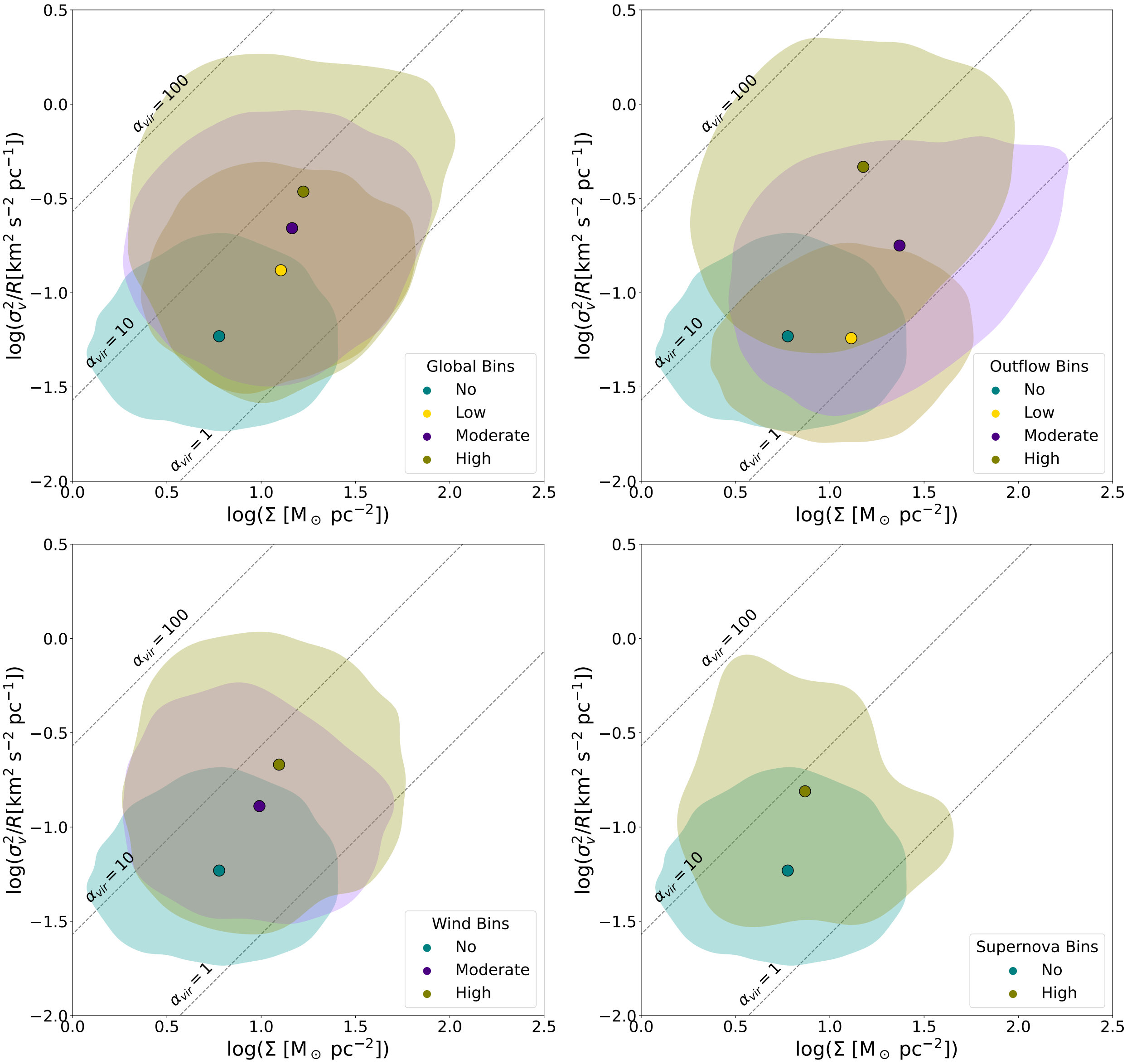


Fig. 7. Scaling relation between σ_v^2/R and the gas surface density (Σ) for the cores in different outflow bins. The dashed lines are isocontours of virial parameters. The symbols and conventions are the same as in Fig. 6.

highest velocity dispersion are present in the high-outflow and high-wind bins rather than the high-supernova bin. This suggests that a supernova may be less efficient in driving the turbulence within the dense gas. This agrees with the notion that outflows and winds deposit energy more locally, directly into the dense gas where the stars are forming.

The Heyer relation (Heyer et al. 2009) ($\sigma_v^2/R \propto \Sigma$) compares the surface density of structures with their scaling parameter (radius and velocity dispersion). Figure 7 shows that the cores impacted by any type of feedback have higher typical surface densities compared to the pristine cores. This is consistent with significant feedback injection occurring while stars are still embedded. Supernova and stellar wind feedback may also compress gas, leading to higher densities. The cores in low- and moderate-outflow bins have a higher surface mass density but comparable virial parameters to pristine cores. The collapse of the GMC under self-gravity in early times with low feedback leads to these dense bound cores. The structures in higher-feedback bins have higher virial parameters (i.e. are more unbound). As shown in Sect. 4.2, this is largely because stellar feedback increases velocity dispersion in these cores.

Our cores have relatively low surface mass densities compared to dense cores found in observations (Ballesteros-Paredes et al. 2020). However, our surface densities are not obtained in a fashion similar to those obtained by observations. We calculated the mass and the area of the cores based on their actual 3D structure and not the ppv cubes used in observational studies that provide only the projected images of clouds. Since we identified cores using the 3D density field, our estimated cores are free from projection effects and line-of-sight contaminations (Beaumont et al. 2013). Moreover, we analysed the total gas distribution, unlike most observational results on GMCs. Observations are often based on detection using molecular tracers (e.g. ^{12}CO and ^{13}CO) and dust emission, but do not probe the most diffuse molecular gas (Duarte-Cabral & Dobbs 2016).

Figure 8 compares the mass and size of the cores. We observe trends similar to the above scaling relations, with stellar feedback

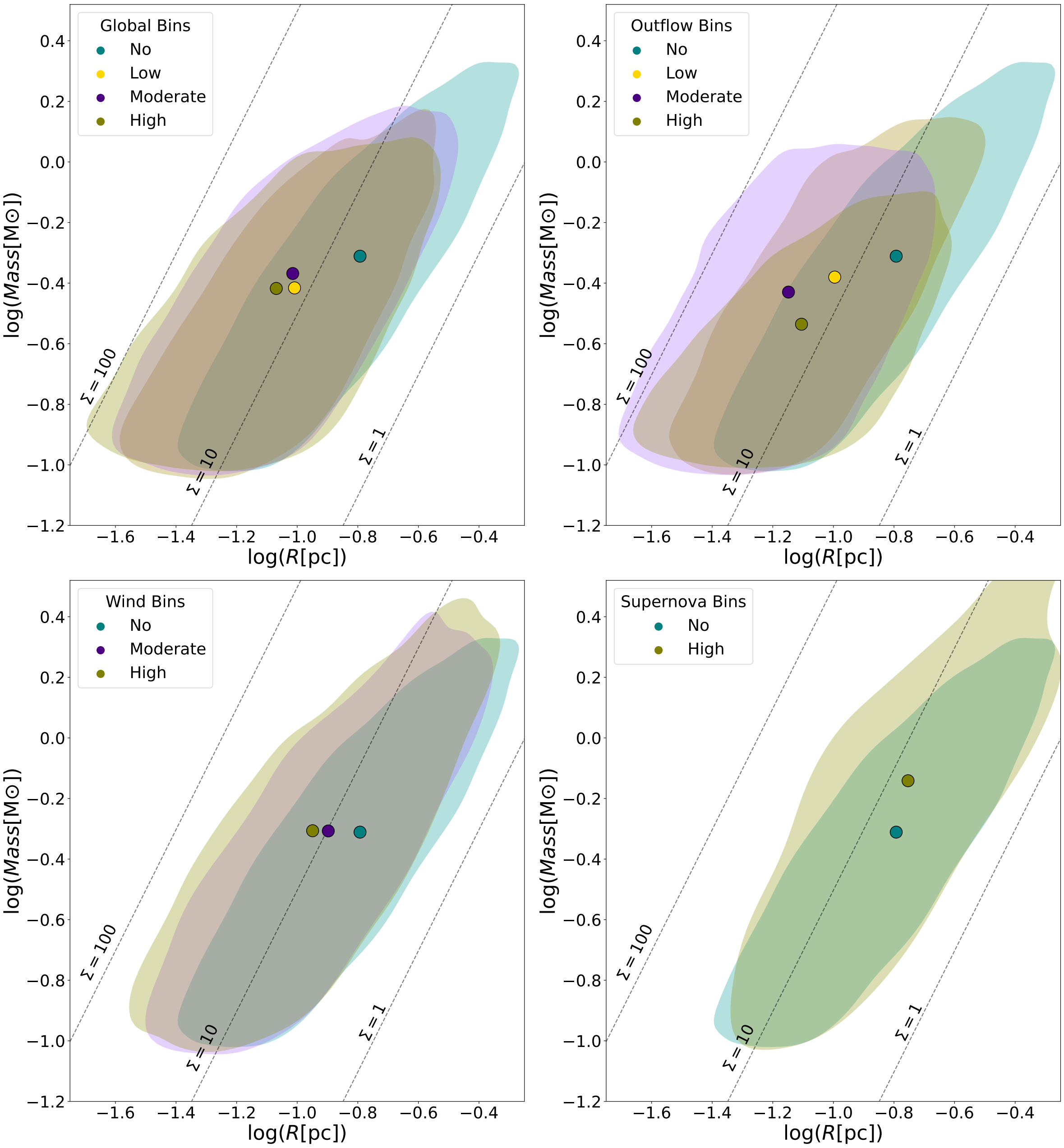


Fig. 8. Scaling relation between mass and size (*M* versus *R*) of cores in the outflow bins. The dashed lines represent the constant surface mass densities (in $M_\odot$ pc^{-2}). The symbols and conventions are the same as in Fig. 6.

leading to smaller cores. This effect is most distinct for outflows. However, the differences in various distributions are not as significant as in other scaling relations. The similar average mass and radius of the cores indicate that core selection is stable despite environmental variations, and the differences in distributions are due to stellar feedback. The slopes of our mass-size ellipses are ∼1–1.5 (Table 5). They are different from the slopes of the dense cores observed in the Galaxy, which are typically ∼2 or higher (Lada et al. 2008).

5. Summary and conclusions

In this paper we have explored the effects of different stellar feedback mechanisms on molecular cores. We performed a dendrogram analysis on the gas density modelled in a 20 000 $M_\odot$ STARFORGE simulation. We then analysed the properties of the dendrogram leaves or cores to understand how these structures are affected by protostellar outflows, stellar winds, and supernovae.

Our main conclusions are the following:

1. Cores strongly affected by feedback have a higher velocity dispersion on average than cores with less feedback, especially when comparing similar-sized structures. This is observed for all three types of feedback mechanisms. We attribute this to the injection of momentum into the dense gas via these feedback mechanisms;
2. Cores influenced by feedback tend to be smaller and denser on average compared to the pristine (no-feedback bin) cores. This could be the result of compression of the molecular gas as a result of stellar feedback or because these cores are

more evolved and are gravitationally collapsing. There is no significant change in the radius of the cores from low- to high-feedback bins; however, the high-supernova bin has the largest cores on average;
3. Feedback does not significantly affect the core mass. The slight changes in the mass distributions that occur for different amounts of feedback follow the same trends as for the core radius;
4. The amount of feedback influence correlates with the virial parameter: cores that are characterised by the presence of more feedback are more gravitationally unbound. We see this trend for all types of feedback;
5. The scaling relations suggest a scenario in which outflow and wind feedback leads to smaller cores, whereas supernova feedback tends to increase the core size. Outflows likely disperse molecular gas, whereas winds compress the cores and disperse the gas. The shocks produced by the supernova explosion lead to large cores;
6. The cores with less feedback adhere more closely to the traditional star formation scaling relations (e.g. Larson's laws), as evidenced by their reduced scatter.

Acknowledgements. The authors thank the anonymous referee for a thorough and constructive report, which has improved the quality of the manuscript. This research made use of astrodendro, a Python package to compute dendrograms of Astronomical data (`http://www.dendrograms.org/`). Support for MYG was provided by NASA through the NASA Hubble Fellowship grant #HST-HF2-51479 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. This research is part of the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by National Science Foundation award OAC-1818253.

References

Agertz, O., & Kravtsov, A. V. 2015, ApJ, 804, 18
Arce, H. G., & Goodman, A. A. 2001, ApJ, 554, 132
Arce, H. G., Shepherd, D., Gueth, F., et al. 2007, in Protostars and Planets V, eds. B. Reipurth, D. Jewitt, & K. Keil, 245
Arce, H. G., Borkin, M. A., Goodman, A. A., Pineda, J. E., & Beaumont, C. N. 2011, ApJ, 742, 105
Ballesteros-Paredes, J., André, P., Hennebelle, P., et al. 2020, Space Sci. Rev., 216, 76
Bally, J. 2016, ARA&A, 54, 491
Beaumont, C. N., Offner, S. S. R., Shetty, R., Glover, S. C. O., & Goodman, A. A. 2013, ApJ, 777, 173
Bieri, R., Naab, T., Geen, S., et al. 2023, MNRAS, 523, 6336
Blitz, L. 1993, in Protostars and Planets III, eds. E. H. Levy, & J. I. Lunine, 125
Chen, H. H.-H., Pineda, J. E., Goodman, A. A., et al. 2019, ApJ, 877, 93
Chevance, M., Krumholz, M. R., McLeod, A. F., et al. 2023, in Astronomical Society of the Pacific Conference Series, 534, Protostars and Planets VII, eds. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 1
Colombo, D., Rosolowsky, E., Duarte-Cabral, A., et al. 2019, MNRAS, 483, 4291
Cunningham, A. J., Klein, R. I., Krumholz, M. R., & McKee, C. F. 2011, ApJ, 740, 107
Davis, C. J., Scholz, P., Lucas, P., Smith, M. D., & Adamson, A. 2008, MNRAS, 387, 954
Deharveng, L., Schuller, F., Anderson, L. D., et al. 2010, A&A, 523, A6
Dokara, R., Gong, Y., Reich, W., et al. 2023, A&A, 671, A145
Duarte-Cabral, A., & Dobbs, C. L. 2016, MNRAS, 458, 3667
Duarte-Cabral, A., Chrysostomou, A., Peretto, N., et al. 2012, A&A, 543, A140
Duarte-Cabral, A., Colombo, D., Urquhart, J. S., et al. 2021, MNRAS, 500, 3027
Dubner, G., & Giacani, E. 2015, A&A Rev., 23, 3
Federrath, C., Schrön, M., Banerjee, R., & Klessen, R. S. 2014, ApJ, 790, 128
Fendt, C., & Čemeljić, M. 2002, A&A, 395, 1045
Fichtner, Y. A., Grassitelli, L., Romano-Díaz, E., & Porciani, C. 2022, MNRAS, 512, 4573
Frank, A., Ray, T. P., Cabrit, S., et al. 2014, in Protostars and Planets VI, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 451
Friesen, R. K., Bourke, T. L., Di Francesco, J., Gutermuth, R., & Myers, P. C. 2016, ApJ, 833, 204
Fuente, A., Martın-Pintado, J., Bachiller, R., Rodrıguez-Franco, A., & Palla, F. 2002, A&A, 387, 977
Gatto, A., Walch, S., Naab, T., et al. 2017, MNRAS, 466, 1903
Geen, S., Rosdahl, J., Blaizot, J., Devriendt, J., & Slyz, A. 2015, MNRAS, 448, 3248
Geen, S., Agrawal, P., Crowther, P. A., et al. 2023, PASP, 135, 021001
Girichidis, P., Offner, S. S. R., Kritsuk, A. G., et al. 2020, Space Sci. Rev., 216, 68
Grishunin, K., Weiss, A., Colombo, D., et al. 2024, A&A, 682, A137
Grudić, M., & Gurvich, A. 2021, J. Open Source Softw., 6, 3675
Grudić, M. Y., Guszejnov, D., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2021, MNRAS, 506, 2199
Grudić, M. Y., Guszejnov, D., Offner, S. S. R., et al. 2022, MNRAS, 512, 216
Guszejnov, D., Grudić, M. Y., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2021, MNRAS, 502, 3646
Guszejnov, D., Grudić, M. Y., Offner, S. S. R., et al. 2022, MNRAS, 515, 4929
Heyer, M., Krawczyk, C., Duval, J., & Jackson, J. M. 2009, ApJ, 699, 1092
Hopkins, P. F. 2015, MNRAS, 450, 53
Hopkins, P. F., & Raives, M. J. 2016, MNRAS, 455, 51
Hopkins, P. F., Quataert, E., & Murray, N. 2012, MNRAS, 421, 3488
Kauffmann, J., Pillai, T., & Goldsmith, P. F. 2013, ApJ, 779, 185
Keown, J., Di Francesco, J., Kirk, H., et al. 2017, ApJ, 850, 3
Kirchschlager, F., Sartorio, N. S., De Looze, I., et al. 2024, MNRAS, 528, 5364
Kirsanova, M. S., & Pavlyuchenkov, Y. N. 2023, in The Predictive Power of Computational Astrophysics as a Discover Tool, 362, eds. D. Bisikalo, D. Wiebe, & C. Boily, 268
Kramer, C., Adam, R., Ade, P., et al. 2024, in European Physical Journal Web of Conferences, 293, mm Universe 2023 – Observing the Universe at mm Wavelengths, 00027
Kruijssen, J. M. D., Schruba, A., Chevance, M., et al. 2019, Nature, 569, 519
Krumholz, M. R., Bate, M. R., Arce, H. G., et al. 2014, in Protostars and Planets VI, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 243
Lada, C. J., Muench, A. A., Rathborne, J., Alves, J. F., & Lombardi, M. 2008, ApJ, 672, 410
Larson, R. B. 1981, MNRAS, 194, 809
Lee, C.-F., Mundy, L. G., Stone, J. M., & Ostriker, E. C. 2002, ApJ, 576, 294
Mairs, S., Johnstone, D., Offner, S. S. R., & Schnee, S. 2014, ApJ, 783, 60
Maret, S., Bergin, E. A., Neufeld, D. A., et al. 2009, ApJ, 698, 1244
McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
Myers, P. C. 1983, ApJ, 270, 105
Narayanan, G., Snell, R., & Bemis, A. 2012, MNRAS, 425, 2641
Neralwar, K. R., Colombo, D., Duarte-Cabral, A., et al. 2022, A&A, 664, A84
Nony, T., Galván-Madrid, R., Motte, F., et al. 2023, A&A, 674, A75
Offner, S. S. R., & Arce, H. G. 2015, ApJ, 811, 146
Offner, S. S. R., & Chaban, J. 2017, ApJ, 847, 104
Offner, S. S. R., & Liu, Y. 2018, Nat. Astron., 2, 896
Offner, S. S. R., Lee, E. J., Goodman, A. A., & Arce, H. 2011, ApJ, 743, 91
Offner, S. S. R., Taylor, J., Markey, C., et al. 2022, MNRAS, 517, 885
Olmi, L., Brand, J., & Elia, D. 2023, MNRAS, 518, 1917
O'Neill, T., Cosentino, G., Tan, J. C., Cheng, Y., & Liu, M. 2021, in American Astronomical Society Meeting Abstracts, 53, 112.03
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
Peters, T., Naab, T., Walch, S., et al. 2017, MNRAS, 466, 3293
Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, ApJ, 679, 1338
Rosolowsky, E., Hughes, A., Leroy, A. K., et al. 2021, MNRAS, 502, 1218
Schneider, N., Simon, R., Guevara, C., et al. 2020, PASP, 132, 104301
Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, A&A, 504, 415
Seo, Y. M., Shirley, Y. L., Goldsmith, P., et al. 2015, ApJ, 805, 185
Smullen, R. A., Kratter, K. M., Offner, S. S. R., Lee, A. T., & Chen, H. H.-H. 2020, MNRAS, 497, 4517
Sokol, A. D., Gutermuth, R. A., Pokhrel, R., et al. 2019, MNRAS, 483, 407
Solomon, P. M., & Wickramasinghe, N. C. 1969, ApJ, 158, 449
Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730
Verliat, A., Hennebelle, P., González, M., Lee, Y.-N., & Geen, S. 2022, A&A, 663, A6
Wesson, R., & Bevan, A. 2021, ApJ, 923, 148
Williams, J. P., & McKee, C. F. 1997, ApJ, 476, 166
Wong, T., Oudshoorn, L., Sofovich, E., et al. 2022, ApJ, 932, 47
Wu, Y., Wei, Y., Zhao, M., et al. 2004, A&A, 426, 503
Xu, D., Offner, S. S. R., Gutermuth, R., & Oort, C. V. 2020a, ApJ, 905, 172
Xu, D., Offner, S. S. R., Gutermuth, R., & Oort, C. V. 2020b, ApJ, 890, 64
Zhang, Q., Hunter, T. R., Brand, J., et al. 2005, ApJ, 625, 864
Zinnecker, H., & Yorke, H. W. 2007, ARA&A, 45, 481

Appendix A: Wasserstein distance

The Wasserstein distance is a similarity metric between two probability distributions. It can be understood as the cost of transport or the work done while converting one distribution to another in the optimal way possible. It is also called the Earth mover's distance. We used the SciPy implementation [9] of the Wasserstein distance to calculate the difference (drift) between the distributions of properties in the various feedback bins and report them in the following tables. The drift represents the level of dissimilarity between two distributions and can vary between zero and infinity. We considered two distributions to be dissimilar if they have a drift > 0.1. It is important to note that the Wasserstein distance is only useful for comparing the same properties, for example the core radius in different bins. Therefore, the drifts obtained for different properties should not be compared directly.

Table A.1. Drifts for the distributions of radius (above the diagonal) and mass (below the diagonal) for the global feedback bins obtained using the Wasserstein distance.

	All	No	Low	Moderate	High
All	-	0.23	0.02	0.01	0.04
No	0.09	-	0.22	0.22	0.28
Low	0.02	0.11	-	0.03	0.06
Mid	0.03	0.07	0.05	-	0.05
High	0.02	0.11	0.02	0.05	-

Notes. We consider two distributions to be different if they have a drift > 0.1.

Table A.2. Drifts for the distributions of radius (above the diagonal) and mass (below the diagonal) for the low-feedback bins obtained using the Wasserstein distance.

	All	Global	Out
All	-	0.02	0.03
Global	0.02	-	0.01
Out	0.02	0.04	-

Table A.3. Drifts for the distributions of radius (above the diagonal) and mass (below the diagonal) for the moderate-feedback bins obtained using the Wasserstein distance.

	All	Global	Out	Wind
All	-	0.01	0.12	0.13
Global	0.03	-	0.13	0.12
Out	0.03	0.06	-	0.25
Wind	0.09	0.06	0.13	-

Table A.4. Drifts for the distributions of radius (above the diagonal) and mass (below the diagonal) for the high-feedback bins obtained using the Wasserstein distance.

	All	Global	Out	Wind	Supernova
All	-	0.04	0.08	0.08	0.27
Global	0.02	-	0.04	0.12	0.32
Out	0.14	0.12	-	0.16	0.35
Wind	0.09	0.11	0.23	-	0.2
Supernova	0.26	0.28	0.39	0.17	-

Table A.5. Drifts for the distributions of velocity dispersion (above the diagonal) and virial parameter (below the diagonal) for the global feedback bins obtained using the Wasserstein distance.

	All	No	Low	Moderate	High
All	-	0.11	0.05	0.07	0.14
No	0.15	-	0.07	0.18	0.25
Low	0.09	0.08	-	0.11	0.18
Mid	0.14	0.25	0.17	-	0.07
High	0.27	0.4	0.32	0.15	-

Table A.6. Drifts for the distributions of velocity dispersion (above the diagonal) and virial parameter (below the diagonal) for the low-feedback bins obtained using the Wasserstein distance.

	All	Global	Out
All	-	0.05	0.21
Global	0.09	-	0.17
Out	0.41	0.37	-

Table A.7. Drifts for the distributions of velocity dispersion (above the diagonal) and virial parameter (below the diagonal) for the moderate-feedback bins obtained using the Wasserstein distance.

	All	Global	Out	Wind
All	-	0.07	0.04	0.04
Global	0.14	-	0.11	0.06
Out	0.16	0.29	-	0.06
Wind	0.1	0.07	0.22	-

Table A.8. Drifts for the distributions of velocity dispersion (above the diagonal) and virial parameter (below the diagonal) for the high-feedback bins obtained using the Wasserstein distance.

	All	Global	Out	Wind	Supernova
All	-	0.14	0.19	0.1	0.12
Global	0.27	-	0.05	0.04	0.03
Out	0.45	0.18	-	0.09	0.07
Wind	0.19	0.09	0.27	-	0.03
Supernova	0.26	0.07	0.19	0.08	-

---

[9] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wasserstein_distance.html

APPENDIX B

An evolutionary study of synthetic molecular clouds in STARFORGE paper

A&A, 704, A38 (2025)
https://doi.org/10.1051/0004-6361/202557095

Astronomy & Astrophysics

A multi-scale evolutionary study of molecular gas in STARFORGE

I. Synthetic observations of SEDIGISM-like molecular clouds

K. R. Neralwar[1,⋆,⋆⋆], D. Colombo[2], S. Offner[3], A. Karska[4,1], M. Figueira[5,1], F. Wyrowski[1], S. Neupane[1], J. S. Urquhart[6], and A. Duarte-Cabral[7]

[1] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[2] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[3] Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA
[4] Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University in Torun, Wileńska 4, 87-100 Toruń, Poland
[5] National Centre for Nuclear Research, Pasteura 7, 02-093 Warszawa, Poland
[6] Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK
[7] School of Physics & Astronomy, Cardiff University, Queen's building, The parade, Cardiff CF24 3AA, UK

Received 4 September 2025 / Accepted 8 October 2025

ABSTRACT

Molecular clouds (MCs) are active sites of star formation in galaxies, and their formation and evolution are largely affected by stellar feedback. This includes outflows and winds from newly formed stars, radiation from young clusters, and supernova explosions. High-resolution molecular line observations allow for the identification of individual star-forming regions and the study of their integrated properties. Moreover, state-of-the-art simulations are now capable of accurately replicating the evolution of MCs, including all key stellar feedback processes. We present ^{13}CO(2–1) synthetic observations of the STARFORGE simulations produced using the radiative transfer code RADMC-3D, matching the observational setup of the SEDIGISM survey. From these synthetic observations, we identified the population of MCs using hierarchical clustering and analysed them to provide insights into the interpretation of observed MCs as they evolve. The flux distributions of the post-processed synthetic observations and the properties of the MCs, namely, radius, mass, velocity dispersion, virial parameter, and surface density, are consistent with those of SEDIGISM . Both samples of MCs occupy the same regions in the scaling relation plots; however, the average distributions of MCs at different evolutionary stages do not overlap on the plots. This highlights the reliability of our approach in modelling SEDIGISM and suggests that MCs at different evolutionary stages contribute to the scatter in observed scaling relations. We study the trends in MC properties, morphologies, and fragmentation over time to analyse their physical structure as they form, evolve, and are destroyed. MCs appear as small diffuse cloudlets in early stages, and this is followed by their evolution to filamentary structures before being shaped by stellar feedback into 3D bubbles and getting dispersed. These trends in the observable properties of MCs are consistent with other realisations of simulations and provide strong evidence that clouds exhibit distinct morphologies over the course of their evolution.

Key words. stars: winds, outflows – ISM: bubbles – ISM: clouds – ISM: supernova remnants

1. Introduction

The molecular gas in the interstellar medium (ISM) is hierarchically clustered in the form of molecular clouds (MCs; (Dobbs et al. 2014; Chevance et al. 2023)). Molecular clouds are turbulent magnetically supercritical (Crutcher 2012) structures with sizes of ∼1–200 pc (Ballesteros-Paredes et al. 2020; Duarte-Cabral et al. 2021), mass M $\sim 10^2$–10^7 $M_\odot$ (Rebolledo et al. 2012), surface densities $\Sigma \sim 1$–1000 $M_\odot$ pc^{-2} (Barnes et al. 2018), and virial parameter (α_{vir}) around unity (Fukui et al. 2008; Roman-Duval et al. 2010). They show various morphologies (Neralwar et al. 2022a), with filaments being the most ubiquitous (André et al. 2010; Li et al. 2016; Duarte-Cabral & Dobbs 2017; Arzoumanian et al. 2019; Colombo et al. 2021; Priestley & Whitworth 2022).

The evolution of MCs is often presented as being regulated by either SN-driven turbulence (Padoan et al. 2016; Lu et al. 2020) or hierarchical gravitational collapse (Ballesteros-Paredes et al. 2011; Vázquez-Semadeni et al. 2019). These mechanisms lead to the fragmentation and formation of dense cores, which further lead to the formation of protostars. Protostars accrete gas from the surrounding ISM, leading to feedback in the form of bipolar outflows (Bally 2016). These outflows (jets) are relatively collimated structures that heat and compress the gas as they interact with the surrounding ISM up to parsec scales (e.g. Duarte-Cabral et al. 2012; Skretas et al. 2023; Karska et al. 2025). Outflows can vary significantly in energetics, with momentum rates between 10^{-5} and 10^{-2} $M_\odot$ kms^{-1} yr^{-1} for low-mass and O-type stars (e.g. Duarte-Cabral et al. 2013; Maud et al. 2015).

As the forming protostars move onto the main-sequence stage, they begin to drive isotropic stellar winds (Vink 2024), which help clear out the remaining gas in their envelope. Stellar winds form bubble-shaped cavities around stars (Weaver et al. 1977; Fierlinger et al. 2016; Ali et al. 2022) by releasing a significant amount of kinetic energy into the ISM ($\sim 10^{48}$ erg over the lifetime of the bubble, Luisi et al. 2021). In addition to stellar winds, high-energy radiation from massive stars ($M \geq 8$ $M_\odot$) ionises the surrounding gas, releasing thermal energy

---

[⋆] Corresponding author: `kneralwar@mpifr-bonn.mpg.de`

[⋆⋆] Member of the International Max Planck Research School (IMPRS) for Astronomy and Astrophysics at the Universities of Bonn and Cologne.

($\sim 10^{46}$ erg) and forming ionised (H II) regions (Simpson et al. 2012; Figueira et al. 2017; Santoro et al. 2022). Stellar winds and photoionising radiations are active during the main-sequence phase of stars (a few megayears) and play a major role in dispersing MCs (Rosen et al. 2020; Chevance et al. 2020). The last form of feedback, and perhaps one of the most important in setting the global conditions of the ISM in galaxies, are supernova explosions from massive stars (Geen et al. 2015; Lucas et al. 2020). Supernovae inject a large amount of energy ($\sim 10^{51}$ ergs) and drive the supersonic turbulence in the ISM on tens to hundreds of parsec scales. (Lu et al. 2020; Dubner & Giacani 2015).

Although various stellar feedback mechanisms have a significant impact on cloud formation, evolution, and dissolution (Chevance et al. 2023), observational studies of MCs rarely attempt to distinguish between clouds affected by different forms of feedback. Clouds are often generalised into a single population of quasi-static entities in near equilibrium (Colombo et al. 2019; Duarte-Cabral et al. 2021) and are collectively analysed using a scaling relation, such as those of Larson (1981) and Heyer et al. (2009). However, they are not necessarily hydrostatic structures, as the turbulence support (static clouds) and global hierarchical collapse (dynamical clouds) cloud formation models produce the same observational signatures (Vázquez-Semadeni et al. 2024). Studying the effects of local feedback events on cloud properties might provide some evidence in support of or in opposition to these models.

In the past decade, several high-resolution surveys of the Milky Way have revealed the intricate structure of MCs and allowed their properties to be resolved in detail (e.g. COHRS Dempsey et al. 2013, CHIMPS Rigby et al. 2016, SEDIGISM Schuller et al. 2021, and OGHReS Urquhart et al. 2024, 2025). Complementing such observations, state-of-the-art giant MC (GMC) simulations are now able to model the ISM at high resolution (e.g. SICLL Walch et al. 2015; Seifried et al. 2017, STARFORGE Grudić et al. 2021, and FIRE Hopkins et al. 2023) while including most of the physical phenomena related to star formation. Comparison of these observations and simulations requires mapping the simulations onto observational space using radiative transfer (e.g. RADMC-3D Dullemond et al. 2012) in order to model the emission that the simulated gas and stars would produce.

One goal of this paper is to use state-of-the-art simulations that incorporate all the relevant physics of star formation to produce synthetic observations that closely resemble observational data. We generated synthetic observations from the Star Formation in Gaseous Environments (STARFORGE) simulations following the observational setup of the SEDIGISM survey (Duarte-Cabral et al. 2021). Using RADMC-3D, we produced the ^{13}CO(2–1) line-emission cubes from the simulations and used a dendrogram-based cloud identification technique to extract MCs. We thus evaluated the extent to which current simulations can reproduce the observed MCs. Through this work, we also aim to analyse the observable properties of MCs as they evolve and determine whether their distributions vary across different evolutionary stages in order to provide some insight into the interpretation of observational MC surveys. As the cloud properties are comparable with SEDIGISM, we were able to provide a direct comparison of the cloud properties as they evolve and are affected by stellar feedback.

The structure of the paper is as follows. In Section 2 we describe the STARFORGE simulations used for our analysis and the SEDIGISM survey, which we used as a observational comparative benchmark. In Section 3, we describe the methodology to create the synthetic observations from STARFORGE (Sect. 3.1) along with the post-processing of the data (Sect. 3.2). We then describe the dendrogram algorithm used to extract structures from these synthetic observation cubes in Sect. 3.3, and this is followed by the definition of MC properties. We present our results of a comparison of the clouds from the simulations and those from SEDIGISM in Sect. 4.1. We investigate the evolution of the properties and morphologies of clouds over time and connect them to the formation of stars in Sect. 4.2. We analyse the scaling relations to compare the correlation between STARFORGE and SEDIGISM cloud properties and understand the time evolution of clouds on these plots in Sect. 5. In Sect. 6, we explain the trends in the properties, morphology, and substructures of MCs over time and connect them to their observed counterparts. We conclude by summarising our work in Sect. 7.

2. Data

2.1. STARFORGE

The STARFORGE[1] simulations are three-dimensional radiation magnetohydrodynamic (MHD) simulations that follow the evolution of GMCs, and they achieve spatial resolutions of a few tens of AU. These simulations model the formation, evolution, and dynamics of individual stars within a GMC, incorporating all forms of stellar feedback: jets, radiation, stellar winds, and supernovae. The simulation framework is built on the GIZMO code (Hopkins 2015), which uses a Lagrangian meshless finite mass method to solve MHD equations (Hopkins & Raives 2016). A comprehensive explanation of the numerical methods and validation tests is provided in Grudić et al. (2021).

We used the M2e4a2 suite of STARFORGE simulations (Table 1 of Guszejnov et al. 2022). It follows the evolution of a 20 000 $M_\odot$ GMC[2] to ~11 Myr and saves all the properties every 24.7 kyr, thus producing a total of 445 snapshots. Of the total number of snapshots, we analysed 410 that have significant ^{13}CO(2–1) emission in order to identify dendrogram structures. The GMC was initiated as a uniform surface density sphere with $R = 10$ pc, $\alpha_{vir} = 2$, and $T = 10$ K surrounded by a diffuse medium (density contrast of 1000) in a (100 pc)^3 box[3]. The gas was initiated as fully atomic, but the ionisation state rapidly converged to local equilibrium such that the interior of the cloud rapidly became fully molecular. The calculation we analysed includes the heating and cooling contributions from all key molecular, atomic, nebular, and continuum processes (see Hopkins et al. 2018, for more detail). Although the simulation does not explicitly follow the formation and destruction of H_2, it computes the molecular gas fraction in each cell using a fitting function that depends on the gas metalicity and surface density (Krumholz & Gnedin 2011). We then derived the number density of molecular hydrogen by assuming that the molecular gas mass in each cell is composed of molecular hydrogen.

We restricted our analysis to the gas within the inner 60 pc of the simulation box, as that is sufficient to capture the ^{13}CO(2–1) emission throughout the time evolution we studied. As the simulation progresses, the GMC collapses under self-gravity to form protostars (at 0.8 Myr) with outflows. Stars with photoionising radiation (at 2.7 Myr) and stellar winds (at 3.6 Myr) are then formed, and they disperse most of the GMC before the first supernova (at 9.8 Myr) occurs. This evolution follows a

---

[1] `https://www.starforge.space`

[2] Hereafter, 'GMC' refers to the simulated GMC defined by the STARFORGE collaboration.

[3] The equations to calculate these properties are described in Guszejnov et al. (2021).

global hierarchical collapse scenario with various stellar feedback mechanisms that can contribute to the injection of local turbulence and provide support against local collapse (Grudić et al. 2022; Guszejnov et al. 2022).

2.2. SEDIGISM

The Structure, Excitation, and Dynamics of the Inner Galactic InterStellar Medium (SEDIGISM; see Schuller et al. 2017, 2021 for an overview) survey spans an area of 84 deg^2 within the Galactic longitude range of $-60° \leq l \leq +18°$ and latitude $|b| \leq 0.5°$ (with some regional variations). It includes multiple molecular tracers specifically targeting the $J = 2$–1 transitions of ^{13}CO and C^{18}O. Observations were conducted from 2013 to 2017 using the 12 meter Atacama Pathfinder Experiment (APEX) telescope (Güsten et al. 2006). The survey provides a contiguous dataset divided into 77 data cubes, each covering approximately 2° × 1° and with a velocity coverage of −200 to 200 km s^{-1} and a pixel size of 9.5″. The first data release (DR1) features ^{13}CO observations with a full width at half maximum beam size of 28″ and a typical 1-σ sensitivity of 0.8–1.0 K per 0.25 km s^{-1}.

Using this dataset, Duarte-Cabral et al. (2021) constructed a catalogue of 10 663 MCs with their physical properties, further updated by Urquhart et al. (2021), Colombo et al. (2022), and Neralwar et al. (2022a). The MCs were identified using the Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES) algorithm (v.0.3.2, Colombo et al. 2015, 2019). Duarte-Cabral et al. (2021) also defined a science sample that consists of well-resolved clouds with reliable distance estimates. We selected MCs from the science sample that are at distances between 2.5 and 3.5 kpc away for our comparative analysis. This distance range provides a sufficiently large sample size (of 835 clouds) with a consistent physical resolution (~0.3–0.5 pc) and minimises the effects of different sensitivities at different distances. In the following, we refer to these clouds as SEDIGISM clouds. We use the deconvolved equivalent radius (radius_dec_pc[4]), cloud mass (Mass), velocity dispersion (sigv_kms), virial parameter (alpha_vir), and surface density (Surf_density_dec_Mpc2) to compare SEDIGISM clouds to synthetic MCs identified in this paper (Sect. 4.1).

3. Methods

3.1. Radiative transfer

We used the RADMC-3D (version 2.0, Dullemond et al. 2012) radiative transfer code on STARFORGE simulation to obtain the position-position-velocity (PPV) ^{13}CO(2–1) emission cubes that mimic the SEDIGISM survey. RADMC-3D uses the simulation data as input, including the distribution of the density, temperature, composition of the gas, and sources of radiation, and it generates a three-dimensional grid of points that are used to sample the environment and calculate the radiative transfer. We briefly outline the procedure in the following subsections.

3.1.1. Data preprocessing

We interpolated the STARFORGE data to a uniform grid with resolution 480 × 480 as input for RADMC-3D[5], matching the resolution in SEDIGISM at a distance of ~3 kpc. The RADMC-3D inputs are gas temperature, velocity, and the number densities of ^{13}CO and H$_2$, which acts as a collision partner. The collisional rate coefficients for ^{13}CO are provided by the Leiden Atomic and Molecular Database (LAMDA[6]). We obtained the number density of ^{13}CO from the H$_2$ number density estimated from the molecular gas fraction using an abundance ratio of H$_2$/CO = 10^4 (typical of the inner Milky Way, Dame et al. 2001; Bolatto et al. 2013) and a constant isotopic ratio of ^{12}CO/^{13}CO = 42.6 (Jacob et al. 2020, at the Galactocentric radius of 5 kpc). We also applied a freeze-out criterion by setting the ^{13}CO abundance ratio to 50% in regions with a gas temperature below 17 K and a gas density above 10^{5}cm^{-3} (Caselli et al. 1999; Lippok et al. 2013; Roueff et al. 2021). However, various temperature and density thresholds tested for the freeze-out criterion did not strongly affect the distribution of the ^{13}CO(2–1) emission studied in this work. RADMC-3D calculates line profiles based on thermal broadening by default, but it allows for the inclusion of turbulent widths (microturbulence[7]). We calculated the microturbulence as the product of the velocity gradient in a cell and the size of the cell, where the velocity gradient for the cells is obtained using the numerical differentiation functionality of the `MESHOID`[8] script.

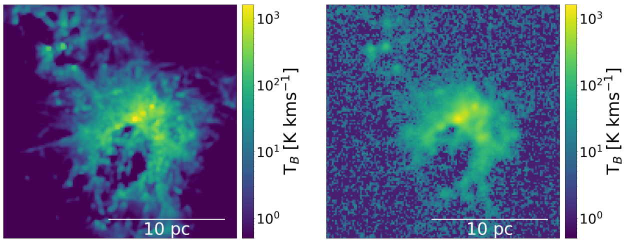


Fig. 1. Left: synthetic integrated ^{13}CO(2–1) emission map created with RADMC-3D. Right: same map but after convolution and with the noise of sigma = 0.2 ± 0.02 K.

3.1.2. RADMC-3D

The radiative transfer code RADMC-3D calculates line emission by obtaining the level population in each grid cell based on a line transfer model. We used the non-local thermodynamic equilibrium approximation together with the large velocity gradient mode and the escape probability method[9]. The ^{13}CO(2–1) transition emits at 220.398 GHz, or 1360.227 μm. We used this as the central wavelength to obtain PPV cubes with 65 channels separated by 0.25 kms^{-1}, leading to a velocity coverage of ± 8 kms^{-1}. We used the RADMC-3D `image` option with `loadlambda` to create PPV cubes of size 65 × 448 × 448 pixels at a distance of 3 kpc and a pixel size of 9.5″ (corresponding to 60 pc) in order to compare the SEDIGISM clouds directly (see Sect. 2.2) We also made use of second-order integrations while creating the image to avoid pixelation. For each snapshot, the ^{13}CO(2–1) cubes were constructed for three different lines of sight along the x, y, and z directions (Appendix A) using the phi={90, 0, 0} and incl={90, 90, 0} parameters.

3.2. PPV cube post-processing

We post-processed the output PPV cube from RADMC-3D (Fig. 1, left) in order to replicate the characteristics of the

---

[4] Column name in the catalogue in Duarte-Cabral et al. (2021).

[5] `https://github.com/Kartik-Neralwar/gizmo_carver`

[6] `https://home.strw.leidenuniv.nl/~moldata/datafiles/13co.dat`

[7] `https://www.ita.uni-heidelberg.de/~dullemond/software/radmc-3d/manual_radmc3d/lineradtrans.html#input-the-local-microturbulent-broadening-optional`

[8] `https://github.com/mikegrudic/meshoid`

[9] RADMC-3D lines_mode = 3.

SEDIGISM observations using the following procedure. First, the cube was convolved with a 2D Gaussian kernel with a full width at half maximum of 28″, corresponding to the APEX telescope beam at 220.398 GHz. The data from the noisy (first and last 50) channels of the SEDIGISM G305 cube were added to the cube as noise. To improve the signal-to-noise ratio, the data were spectrally smoothed at 0.5 kms^{-1} resolution and resampled to 0.25 kms^{-1} (Sect. 3.1.1 in Duarte-Cabral et al. 2021). An example of an original and post-processed cube is shown in Fig. 1.

To suppress the noise and prepare the data for the analysis of dendrograms, we applied the dilated masking method described in Grishunin et al. (2024), resulting in clean masked cubes (Appendix B). We performed two masking iterations as Grishunin et al. (2024) for robust data cleaning, with s2n_low = 2, s2n_high = 4, and s2n_vel = 3 for the first iteration and s2n_low = 2, s2n_high = 5, and s2n_vel = 3 for the second. The s2n_low follows the min_val parameter in Duarte-Cabral et al. (2021), setting the value of pixels whose emission is lower than 2σ of the local noise to zero. The higher value of s2n_high in the second iteration provides a stricter constraint to remove any spurious sources. These data processing steps allowed us to closely replicate the observed dataset of SEDIGISM (Sect. 4.1). Of the 445 snapshots, the first 35 (up to ~1 Myr) have no detectable ^{13}CO(2–1) above the s2n_low limit. Our analysis is therefore limited to 410 snapshots, resulting in a total of 1230 synthetic data cubes for the three projections. Excluding the early snapshots minimises potential biases in the reproduction of the synthetic line emission that could arise from the initial settling of the cloud. By ~1 Myr, the gas density distribution reaches a quasi-equilibrium state, satisfying a log-normal distribution, which is expected for supersonically turbulent gas (Padoan et al. 1997; Lane et al. 2022).

3.3. Dendrograms and cloud properties

Dendrograms describe the distribution and nesting of isosurfaces in data cubes and have long been used to discretise molecular gas emission at different scales in observations (Colombo et al. 2019; Duarte-Cabral et al. 2021) and simulations (Offner et al. 2022). We used dendrograms (Rosolowsky et al. 2008) to segment the molecular gas in each snapshot into leaves, branches, and trunks. Leaves are structures formed by single local maxima in the gas distribution and are nested in branches, which in turn are nested in trunks. The trunks can be isolated structures without substructures or hierarchical structures with multiple substructures (Appendix D). Dilated masking sets the value of all noisy voxels to zero, setting the lower threshold of the dendrograms. We built the dendrograms using min_val = 0 and n_delta = 2 σ_{rms}, where σ_{rms} = 0.87 represents the mean rms noise in our data (Sect. 3.2). We set the min_pix using n_area = 3 and n_vchan =2 such that all structures are both spatially (i.e. at least three beams) and spectrally (span at least two velocity channels) resolved. We found a total of 3710 hierarchical trunks and refer to them as MCs throughout this work. The entirety of detectable ^{13}CO(2–1) emission in a snapshot is referred to as the 'molecular gas complex' or the 'entire ^{13}CO(2–1) emission'. Additionally, the dendrograms store information about the nested structures (descendants) for every structure. We perform a quantitative analysis of these substructures (descendants) in each MC in Sect. 4.2.3.

The dendrogram analysis returned a catalogue of structures[10] with their properties (Rosolowsky & Leroy 2006), which we used to derive other MC properties while assuming a heliocentric distance of 3 kpc. The most useful directly measured properties are the projected footprint area (A), the velocity dispersion (σ_r), and the total brightness temperature (T$_B$). The effective radius is defined as $R_{eff} = \sqrt{A/\pi}$ and the deconvolved radius as $R = \sqrt{R^2_{eff} - R^2_{beam}}$, where R_{beam} = 0.2 pc is the physical size of the beam[11]. The luminosity mass was estimated from the luminosity (L) as M_{lum} [M$_\odot$] = $\alpha_{\rm CO}$ L [L$_\odot$], where $\alpha_{\rm CO}$ = 22.43 M$_\odot$ (K km s^{-1})$^{-1}$ pc $^{-2}$, estimated using $X_{^{13}{\rm CO(2-1)}} = 1^{+1}_{-0.5} \times 10^{21}$cm^{-2} (K km s^{-1})$^{-1}$ to be consistent with SEDIGISM . From this we derived the surface mass density $\Sigma = M_{lum}/A$ and the virial parameter $\alpha_{vir} = 5\sigma_v^2 R/GM_{lum}$, assuming a spherical and uniform cloud. The analysis presented in this paper uses deconvolved properties, although global trends are virtually unchanged by this choice.

We obtained the true molecular gas mass (M) directly from the simulation by projecting the H$_2$ masses of gas cells onto the RADMC-3D code and summing over the pixels associated with a given dendrogram structure. The true molecular gas mass is not subject to observational biases or limitations (e.g. $\alpha_{\rm CO}$ factor) and includes the CO-dark and freeze-out regions. We used this mass to study the time evolution of clouds in Sect. 4.2.1, but their derived properties were calculated using M_{lum} to be consistent with the observations. STARFORGE also tracks the positions, mass, ages, and evolutionary stages of individual stars and protostars (Grudić et al. 2021). We recorded the number of newborn stars (protostars and stars younger than 250 kyr) and the main-sequence stars (with $M > 2$ M$_\odot$) in each snapshot and used them to investigate the MC fragmentation in Sect. 4.2.3.

We analysed the cloud morphologies using the RJ plots (Clarke et al. 2022) algorithm, which is an extension of J plots (Jaffa et al. 2018). These algorithms classify pixelated structures into different morphologies based on the relationship between their mass distribution and moment of inertia. The J moments J_1 and J_2 are obtained by compairing the principal moments (I_1 & I_2) with those of a uniform surface density disc $\left(I_0 = \frac{\rm AM}{4\pi}\right)$ of the same area and mass: $J_i = \frac{I_0 - I_i}{I_0 + I_i}$ $\{i = 1, 2\}$. The RJ moments R_1 and R_2 are obtained by rotating the J plot 45 degrees in the anticlockwise direction, and R_2 was further normalised to remove the parameter space constraints given by $|J_i| \leq 1$, resulting in $R_1 = \frac{J_1 - J_2}{2}$ and $R_2 = \frac{J_1 + J_2}{\sqrt{2}(\sqrt{2} - R_1)}$. We note that $R_1 = 0$ corresponds to a perfectly circular structure, with increasing values indicating progressively higher degrees of elongation. The term R_2 measures the weight distribution of a structure compared to its centre of weight. The positive and negative values of R_2 represent centrally overdense and underdense structures, respectively.

4. Results

4.1. Validation of the synthetic dataset: Comparison with SEDIGISM

Before we could study the evolution of MCs using our sample of synthetic clouds, we first needed to ensure that our simulated sample was representative of the observed clouds in terms of properties and the probed parameter space. Therefore, we compared our synthetic MCs with those observed as part of the SEDIGISM survey (Duarte-Cabral et al. 2021). Figure 2 shows

---

[10] ppv_catalog: `https://dendrograms.readthedocs.io/en/latest/api/astrodendro.analysis.ppv_catalog.html`

[11] The SEDIGISM beam (28″) full width at half maximum at 3 kpc is 0.4 pc.

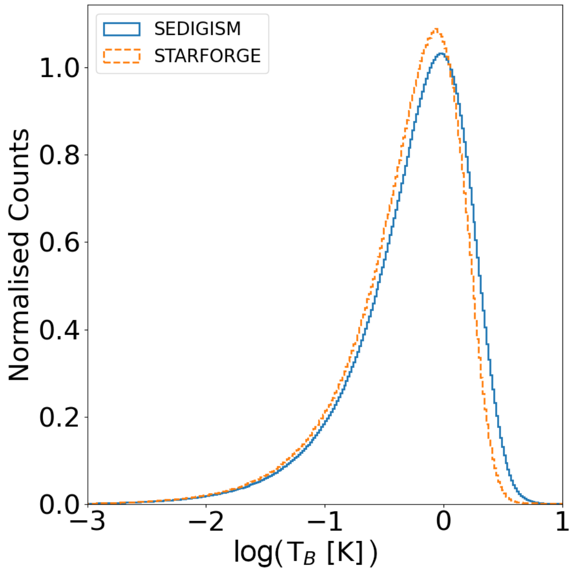


Fig. 2. Comparison between the flux values for the noisy pixels from all the STARFORGE snapshots along all three projections and the complete SEDIGISM survey (^{13}CO(2–1) emission).

the flux distribution in each pixel for the SEDIGISM ^{13}CO cubes and the emission of all STARFORGE snapshots along the three projections. The strong agreement between the two flux distributions serves as a validation that our synthetic emission maps replicate the SEDIGISM data to the first order.

Figure 3 shows the distribution of the integrated properties for our synthetic MCs and the SEDIGISM clouds. The good agreement between the two datasets for all properties indicates similar structural characteristics of the clouds in the two samples. This further highlights the relability of our approach in not only creating emission cubes similar to SEDIGISM but also identifying SEDIGISM-like clouds. However, this should not be confused as a single STARFORGE simulation replicating the entire diverse sample of clouds in SEDIGISM . Rather, due to robust data processing, our synthetic MCs occupy the same parameter space as SEDIGISM clouds.

The slight shift in the STARFORGE distributions towards higher values compared to SEDIGISM can be explained as follows. STARFORGE simulates an isolated GMC, and therefore the lowest level of dendrogram structures, i.e. hierarchical trunks, are called MCs. SEDIGISM , on the other hand, traces the larger gas structures in the Galaxy, and every dendrogram structure (e.g. non-overlapping trunks, branches, or leaves) is considered a cloud if it complies with the clustering criteria set by the segmentation algorithm. Using the same cloud identification criteria as SEDIGISM , i.e. the SCIMES algorithm (Colombo et al. 2019) is not suitable for this study. This is mainly due to the relatively small spatial coverage of each snapshot, which causes SCIMES to identify branches within the trunks as MCs, making the extracted clouds significantly smaller than those of SEDIGISM. The exclusion of leaves constrains the lower limit to the properties of our synthetic MCs. Despite these differences, our MCs are consistent with those in SEDIGISM (Figs. 8 and 9).

4.2. Time evolution of MCs

The aim of this section is to understand the changes in the structure and properties of MCs as they evolve and are affected by different stellar feedback mechanisms (Fig. 4). To highlight the broader evolutionary trend, we binned the MCs based on the evolutionary time. Each bin corresponds to ∼250 kyr and contains a total of 30 data cubes (three projections per snapshot). The evolution of properties of individual MCs is illustrated in the Appendix A. The following subsections focus on the general trends in integrated properties of the clouds (Sect. 4.2.1), their morphologies (Sect. 4.2.2), and substructures (Sect. 4.2.3). The corresponding figures (Figs. 5, 6 and 7) also show the onset of different stellar feedback mechanisms indicated on the time axis.

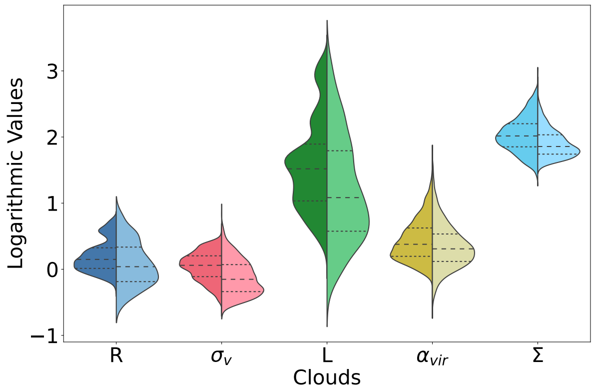


Fig. 3. Comparison of MC properties from STARFORGE (left side violins) and SEDIGISM (right side violins). The properties are radius (R), velocity dispersion (σ_v), luminosity (L), virial parameter (α_{vir}), and surface density (Σ). The horizontal dashed lines represent the quartiles of the distributions.

4.2.1. Integrated properties

In this section, we analyse the integrated properties of MCs as a function of time, over the ∼11 Myr covered by the simulations. Figure 5 shows that the radius, mass, luminosity, and surface density properties follow each other closely, which is expected because larger MCs are more massive on average (Larson 1981; Kauffmann et al. 2010). The similar trends of molecular gas mass and ^{13}CO luminosity distributions over time confirm that our results are not significantly affected by the choice of canonical ^{13}CO abundances (see Sect. 3.1.1). The four properties show a steady increase until ∼5, as progressively more ^{13}CO(2–1) is above the detection limit. The increase represents the active growth of the MCs as they transition from newly formed small diffuse cloudlets to large, massive, and dense MCs (illustrated in Fig. 4). The increase in mass is also influenced by the transition of the gas from the atomic to the molecular phase and is a key characteristic of the global hierarchical collapse model (Shimajiri et al. 2019; Vázquez-Semadeni et al. 2019). The 6 Myr transition marks the beginning of gas expulsion and cloud dispersal by stellar feedback, as noted by the decrease in the average properties. This is supported by the significant increase in the number of clouds at ∼7 Myr, which suggests that the gas is being removed and eroded from within and around clouds, resulting in a higher number of small cloud fragments (Sect. 7). These trends in MC properties are also seen in other simulation sets; however, their onset and duration vary depending on the initial conditions (Appendix E).

The velocity dispersion distribution remains relatively flat until ∼6 Myr and then decreases. The initial high values for some MCs are largely a result of the initial supersonic turbulence, with a possible contribution from the momentum injected by

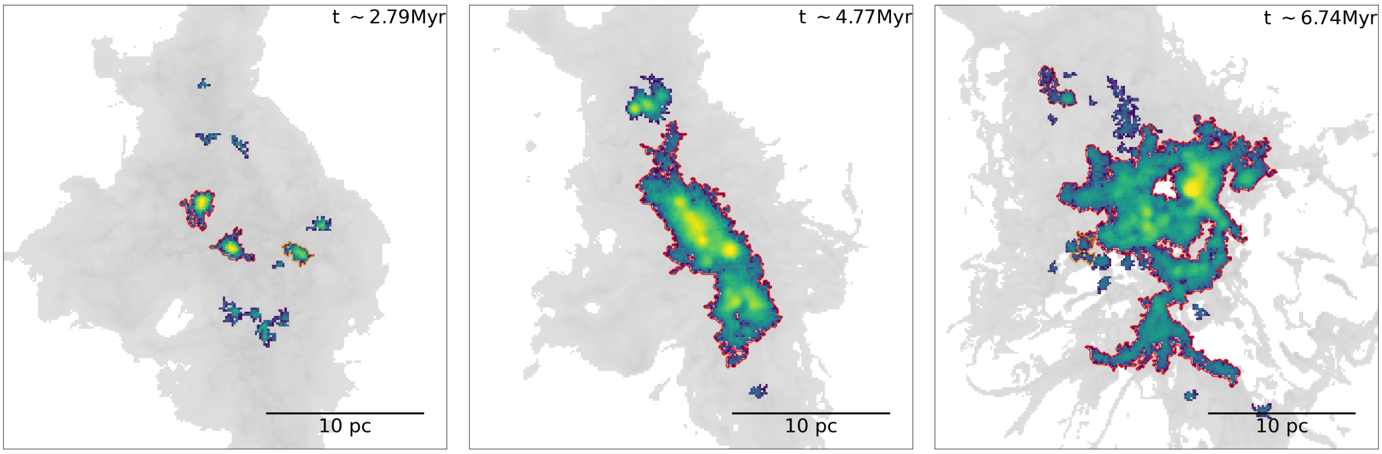


Fig. 4. Moment 0 maps of ^{13}CO(2–1) at different evolutionary times. The background grey scale represents the H_2 gas density with ^{13}CO(2–1) emission overlaid as viridis maps (molecular gas complex), and the coloured contours represent different MCs (dendrogram trunks), with red contours representing the largest MCs (R) in the cube. The ^{13}CO(2–1) maps for multiple snapshots along different projections are presented in Appendix B.

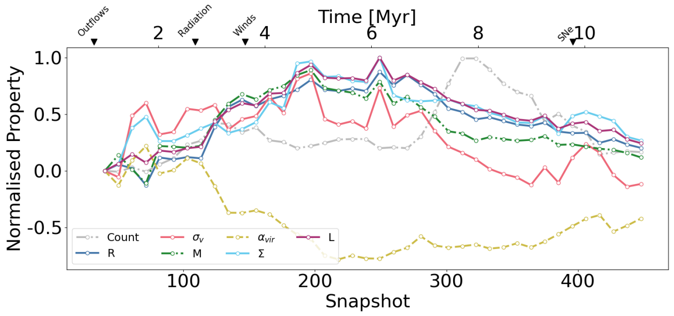


Fig. 5. Normalised medians of MC properties (log scales) as a function of time. The properties shown are radius (R), velocity dispersion (σ_v), molecular gas mass (M), virial parameter (α_{vir}), surface mass density (Σ), and luminosity (L). The count refers to the total number of clouds in a time bin. The normalisation process included subtracting the initial value (first bin) of the property from it, and this was followed by a min-max standardisation. The symbols on the top represent the times at which outflows, photoionisation radiation, stellar winds, and supernovae begin in the simulation.

the protostellar outflows. The average virial parameter decreases from ∼10 to ∼2 during the first 5 Myr and remains relatively constant throughout most of the evolution (Fig. 5). These distributions are not significantly affected by the growth and dispersal of MCs. However, they show peaks at ∼2 Myr and ∼10 Myr, which are also seen in the surface density distribution. The peaks correspond to the formation of turbulent gas structures and the onset of the first supernova, respectively. Supernovae inject turbulence into the ISM and produce relatively dense MCs in an environment that has mostly been dispersed by stellar winds and radiation. However, they do not significantly alter the general trends in the MC properties (Grudić et al. 2022).

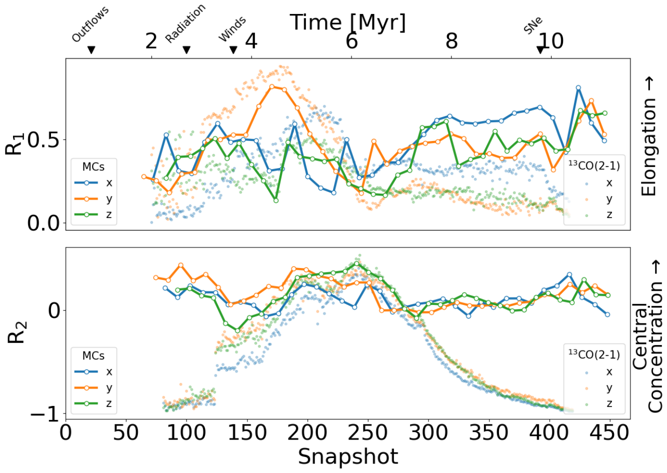


Fig. 6. Morphological analysis of the MCs (solid line) and molecular gas complexes (points) in different snapshots. R_1 represents the degree of elongation, and R_2 represents the degree of central concentration. The lines represent the R_1 and R_2 values for the MCs, and the scatter points represent the same for the molecular gas complexes, i.e. entire ^{13}CO(2–1) emission in a snapshot, obtained using the projections along the three axes. The colour scheme is the same as in Fig. 3.

4.2.2. Morphology

In this section, we study how the morphology of the MCs varies over time under the effect of various stellar feedback mechanisms. We used the *RJ* plots algorithm (Sect. 3.3) to analyse the morphology of the entire ^{13}CO(2–1) emission in the simulation box as well as the individual MCs as they evolve over time (Fig. 6). The MCs show a large scatter in R_1 and R_2 at all times, as cloud structures present a continuous spectrum rather than discrete classes. We therefore binned the MCs following the same criteria as used in Sect. 4.2.1 in order to analyse the overall morphological changes and present the scatter plots in Appendix A. We also analysed the R_1 and R_2 moments for molecular gas complexes showing the structure of the entire emission in a snapshot.

Figure 6 shows that, on average, the clouds (both the individual MCs and the molecular gas complexes) have different elongations (R_1) along the different projections. This is expected, as they evolve in a unisotropic environment, which causes them to evolve asymmetrically. However, this difference is not very significant for individual MCs, as shown by the large scatter in Fig. A.1 (bottom left). The trends in R_2 are similar along the three projections, which highlights that the internal structure of the clouds appears the same regardless of the viewing angle.

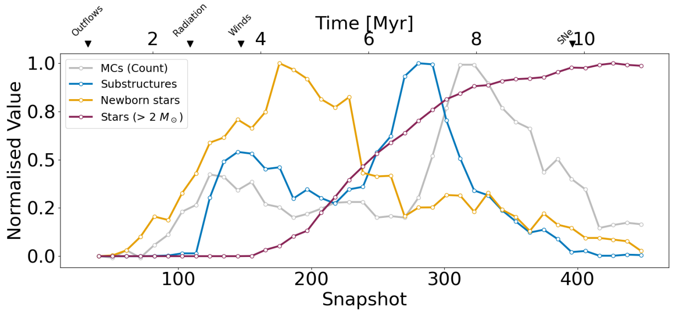


Fig. 7. Quantitative representation of MCs (counts), their substructures (dendrogram descendants), the newborn stars (age < 250 kyr), and the main-sequence stars more massive than $2M_\odot$ as a function of time. The values are normalised using a min–max standardisation.

The large average values of R_1 for the MCs throughout most of the simulation highlight the ubiquitousness of filamentary structures (solid lines in Fig. 6, top). A notable trend in the distribution of R_1 is a peak occurring in the 3–6 Myr range, suggesting an elongated shape for CO emission in snapshots (Fig. 4, centre). The increase in elongation is more dominant in one projection direction, suggesting that the molecular gas complex is collapsing faster along one axis. Although the average value of R_1 for MCs increases after 6 Myr, this is not the case for molecular gas complexes. It points towards the emergence of filamentary MCs within feedback-affected spherical bubble-like regions and potentially indicates the formation of intra-cloud filaments. However, this increase should be interpreted with caution due to the large scatter in R_1 among individual MCs. The distribution of R_2 for the molecular gas complexes shows a constant increase until ~6 Myr, which is then followed by a constant decrease. The rise in central concentration (R_2) is in agreement with the gravitational collapse of the simulated GMC. The decrease signifies the dispersion of the molecular gas due to stellar feedback, leading to centrally underdense bubble-like structures (Fig. 4, right). The increase in R_2 along all three projections suggests that these MCs represent 3D bubbles rather than 2D rings (illustrated in Appendix B).

The different morphologies of the clouds for different projections reveal the anisotropy of the MCs. The anisotropy is driven by a combination of the initial turbulence, magnetic fields, and various stellar feedback events. A caveat of STARFORGE is that it assumes an isolated cloud and does not account for galactic-scale processes, such as gas accretion and the galactic potential. However, we expect the galactic potential to have a minor effect on the cloud scale over a timescale of 10 Myr. Although we see the trends in the morphology for the entire ^{13}CO emission in snapshots, the individual MCs on average remain elongated centrally concentrated structures throughout the simulation. Visual analysis further confirmed their filamentary and clumpy nature (Appendix B) during most of their lifetime (~3–7) Myr.

4.2.3. MC substructures

The fragmentation of MCs is often attributed to various stellar feedback mechanisms (Mazumdar et al. 2021; Grishunin et al. 2024); however, fundamental questions about the physics responsible for fragmenting MCs are still open. Analysing the number of MCs, their substructures, and stars provides information about how the MCs fragment to form dense substructures that lead to star formation, which in turn produces stellar feedback and disperses these gas structures. Here, we present a quantitative analysis of the substructures within our MCs. The substructures are stored by the dendrogram algorithm as descendants and represent subparsec-scale compact structures, which are typically referred to as clumps. Figure 7 shows an increase in the number of MCs and their substructures around (~3 Myr), representing progressively more emission that is above the noise level. The second peaks (~7–8 Myr) in both distributions are the result of gas dispersion by stellar feedback.

We also analysed the number of stars and protostars with ages less than 250 kyr and collectively refer to these as newborn stars. These reflect the instantaneous star formation rate of the clouds, with the 3–7 Myr period representing the peak of star formation activity. The newborn stars evolve to the main sequence, producing stellar winds and photoionising radiation. The stellar evolution is a strong function of their mass (Hosokawa et al. 2011), and a large number of low-mass stars remain in the main-sequence phase throughout the lifetime of the simulated GMC, seen as a constant increase in the number of stars over time (Fig. 7).

The significant increase in the number of stars at ~5 Myr is followed by peaks in the number of substructures (7 Myr) and MCs (8 Myr). Stellar winds and radiation from individual stars disperse and expel gas in their neighbourhood, fragmenting dense gas structures (6–7 Myr; Fig. 4). Over time, the feedback becomes stronger and erodes these fragmented clumps, causing them to decrease in number (>7 Myr). This strong gas dispersal affects the entire MC and leads to the removal of gas between dense MCs, i.e. the entire ^{1}3CO emission is identified as multiple small MCs instead of a single continuous structure, or trunk (7–8 Myr). These smaller MCs are dispersed over time as a result of continuous feedback events (>8 Myr). The lack of a dense molecular gas further decreases the number of embedded stars.

5. Scaling relations

The scaling relations (Larson's and Heyer's) show the correlations between the physical properties of clouds. Larson's first relation, originally derived from the analysis of numerous MCs by Larson (1981), was later refined by Solomon et al. (1987), resulting in the relation $\sigma_v = 0.74\,L^{0.5}$ (discussed in Colombo et al. 2019). The spatial and velocity structures of MCs following power laws are often considered a proof of universal cloud turbulence (Padoan et al. 2016). It is a simplification of Kolmogorov's law for turbulence, indicating that larger clouds exhibit broader linewidths. Larson's laws individually do not provide information about the virial state of a cloud. To take this into account, Heyer et al. (2009) combined Larson's second and third laws, creating Heyer's relation, which compares the surface density of a cloud with its scaling parameter ($\sigma_v^2/R \propto \Sigma$).

We present the two scaling relations for our MCs and compare them with the SEDIGISM clouds in Figs. 8 and 9. The synthetic MC distribution shows an almost complete overlap with the 3-σ kernel density estimator for SEDIGISM clouds. This provides strong evidence that the MCs from our synthetic observations have global properties similar to those of real clouds. Moreover, this shows that the correlation of properties is consistent across both samples. That is, similar-sized MCs have similar velocity dispersions, leading to their overlap on the scaling relation plots.

Figure 8 shows an increase in the average size and linewidth of the clouds up to ~6 Myr. This is largely a result of the formation of dense gas structures that merge and result in progressively more ^{13}CO(2–1) emission being detectable. In addition, stellar

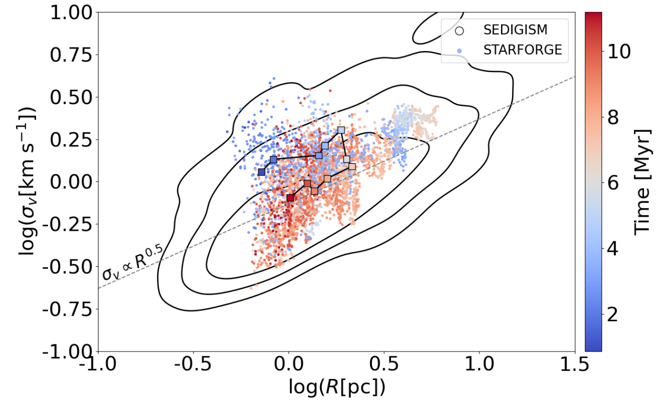


Fig. 8. Size-linewidth relation (σ_v versus R) for our MCs (scatter points), colour-coded with respect to the time elapsed (in Myr) since the start of the simulation. The squares represent medians of distributions in ~1 Myr bins. The black contours represent the 1σ, 2σ, 3σ levels for the SEDIGISM clouds. The dashed line represents Larson's first relation (Larson 1981; Solomon et al. 1987).

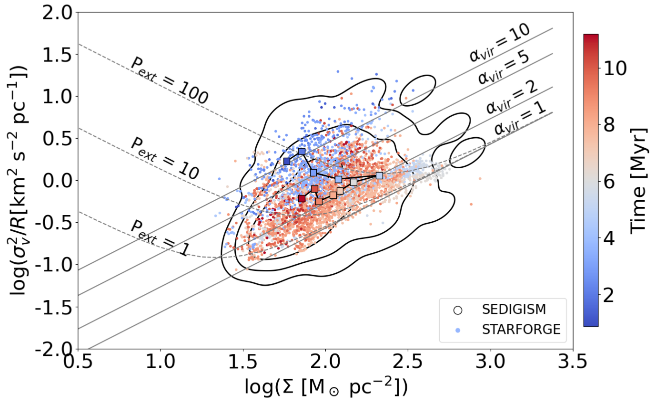


Fig. 9. Scaling relation between σ_v^2/R and surface mass density (Σ). The symbols and notations follow Fig. 8. The solid grey lines represent isocontours of virial parameters. The dashed lines represent $\alpha_{\rm vir} = 1$ when including an external pressure of $P_{\rm ext}$ = 1, 10, and 100 $M_\odot$ pc^{-3} km^2 s^{-2}.

feedback mechanisms drive the velocities in the MCs and expand them, resulting in larger sizes and linewidths over time. At evolutionary times beyond ~6 Myr, stellar feedback mechanisms begin to disperse the gas significantly, resulting in the identification of smaller MCs. This appears as a sharp drop in the average velocity dispersion followed by a gradual decrease in their size. The higher values in the average velocity dispersion at early times (<6 Myr) for similar-sized structures are most likely due to the initial supersonic turbulence injected in the simulation.

Figure 9 highlights an initial trend of MCs as they transform from underdense and highly supervirial structures to denser virialised structures. The decrease in the average scaling ratio (σ^2/R) is due to the significant increase in the MC radius compared to the velocity dispersion. The erosion of MCs due to feedback after 6 Myr causes a horizontal shift in Fig. 9 towards lower surface densities and higher virial parameters.

Molecular clouds are often analysed collectively in scaling relation plots, which typically show a large scatter (Colombo et al. 2019; Duarte-Cabral et al. 2021). Neralwar et al. (2022b) have shown that the cloud morphology and internal substructures influence their distribution in these relations and hypothesised that the different morphologies might correspond to different evolutionary stages. Figures 8 and 9 show that MC populations at different evolutionary times occupy different positions in the scaling relation plots. This suggests that the large scatter in the scaling relations could be due to the ensemble of observed MCs being at different stages of their evolution. SEDIGISM clouds could have gone through a diversity of physical conditions, as they are influenced by the larger Galactic environment and feedback events and follow various evolutionary paths. The gas flows and the effects of external factors are not simulated in STARFORGE . However, our MCs lie in the same parameter space as SEDIGISM , and the simulation traces all relevant physics of star formation at parsec scales. Therefore, at least some of the SEDIGISM clouds follow an evolutionary path similar to that of our MCs. When combined with the fact that MC morpholgies evolve over time (Sect. 4.2.2), our results support the hypothesis proposed by Neralwar et al. (2022b).

6. Discussion

6.1. The life cycle of synthetic MCs and their observed counterparts

The distribution of integrated properties, morphology, and fragmentation show that the MCs evolve from small, diffuse structures to dense filamentary structures before being dispersed by stellar feedback. They appear as filamentary and clumpy structures throughout most of their lifetimes, which is consistent with other simulations (e.g. Clarke et al. 2017). The smaller structures they host collapse and form stars even though the parent MC appears unbound ($\alpha_{vir} \gg 1$). The stars produce stellar feedback that disperses the molecular gas, resulting in smaller, less massive diffuse structures. The presence of MCs as small clumps, filamentary structures, and bubble-like structures is also supported by observations (Neralwar et al. 2022a).

The initial turbulence in the simulations produces overdensities that become denser over time because of gravitational collapse. These MCs (<3 Myr) appear as small, diffuse, low mass, gravitationally unbound ($\alpha_{vir} \sim 10$) and approximately spherical structures. We refer to them as MCs following their definition as hierarchical trunks to be consistent throughout the paper; however, they are closer to starless molecular gas clumps in observations (e.g. starless clumps in Traficante et al. 2018 and quiescent clumps in Urquhart et al. 2022).

Molecular clouds at 3–7 Myr appear as large filamentary structures with dense clumps. The entire ^{13}CO(2–1) emission in the simulation box appears as a single (or a few) large MC(s) since most of the gas in the simulation domain is molecular[12]. These represent a majority of the MCs detected in observational surveys (Molinari et al. 2010; Arzoumanian et al. 2011; Colombo et al. 2021; Neralwar et al. 2022a; Ge et al. 2023). The long lives of filamentary MCs are often attributed to continuous gas flows from the larger environment onto small-scale clumps through the filaments (Gómez & Vázquez-Semadeni 2014). Peretto et al. (2023) have shown that the substructures within the MCs produce a deep gravitational potential and accrete the gas from the filament and thus dynamically decouple from the MCs in order to grow faster. The formation of these dense clumps[13] leads to a central infall of gas along the filament, which feeds the clumps, forms new small clumps, and causes turbulent movements (previously discussed in Gong et al. 2018; Lu et al. 2018;

---

[12] This is shown by the molecular gas fraction https://starforge-tools.readthedocs.io/en/latest/data.html#gas-data-fields values stored in STARFORGE for each snapshot.

[13] These central overdensities in MCs are visible in Fig. B.2 and evident from the large values of R_1 (Sect. 4.2.2).

Williams et al. 2018; Krumholz & McKee 2020). The higher number of dense clumps results in an accelerated formation of protostars and stars (Fig. 7; 4–6 Myr).

Stellar winds and radiation become more effective throughout the simulation domain after ∼6 Myr, resulting in gas expulsion and dispersion. These phenomena result in the ^{13}CO(2–1) emission appearing as centrally underdense structures with a shell-like morphology. These 3D bubble-like clouds are widely studied as wind- and radiation-driven bubbles (Churchwell et al. 2004; Palmeirim et al. 2017; Tiwari et al. 2021) associated with H II regions (Neupane et al. 2024) and classified as the last evolutionary stage of clouds (Kawamura et al. 2009). Feedback disperses most of the ^{13}CO emission by ∼8 Myr, causing the broken shells to be identified as individual MCs. The formation of massive stars at ∼3 Myr and most of the ^{13}CO(2–1) emission being dispersed by ∼8 Myr agrees with the fast dispersal of MCs by feedback (up to ∼5 Myr, Kruijssen et al. 2019; Chevance et al. 2020; Figueira et al. 2020; Knutas et al. 2025).

6.2. Caveats and outlooks

STARFORGE simulates an isolated GMC within a closed box with a fixed total gas mass ($2 \times 10^4 M_\odot$) that restricts the upper mass limit of the MCs. Moreover, the simulation does not track the real-time abundance of CO, so the use of canonical abundance values and ad hoc freeze-out prescriptions is a simplification. This results in under- or overestimation of the real abundances, thus introducing an uncertainty on the measured M_{lum} from the synthetic observations and potentially skewing these distributions. To minimise this error, we set the CO abundance to zero in regimes where it can freeze out. We also set strict constraints while performing dilated masking (Sect. 3.2) and chose only the hierarchical trunks (Sect. 3.3) as MCs to avoid spurious sources. Inclusion of a chemical network in the simulations or radiative transfer could improve the accuracy of property estimates, but this goes beyond our current scope and does not significantly affect our overall analysis (Appendix C).

Predicting the evolutionary stages of SEDIGISM MCs based on their properties and morphology might be possible since our MCs share the same parameter space as SEDIGISM (Sect. 5). However, the degeneracy in these distributions on either side of the 6 Myr peak, visible as the large scatter in the scaling relation plots, makes this task extremely challenging when using only ^{13}CO(2–1) observations. The early MCs show an H_2 envelope (Fig. B.2) that could be traced with diffuse gas tracers such as ^{12}CO, thus separating them from the feedback-affected MCs (Fig. B.3). Moreover, an analysis of the dense gas structures within MCs using tracers such as N_2H^+ and NH_3 could reveal the fragmentation trends and aid in the evolutionary classification of the MCs. However, such a multi-wavelength study is beyond the scope of this work.

Observational works often perform multi-wavelength studies using tracers of dense gas, young stellar objects, and HII regions in order to classify MCs and clumps into various evolutionary stages (Kawamura et al. 2009; Traficante et al. 2018; Urquhart et al. 2022; Watkins et al. 2025). In a follow-up paper, we will study how clumps (dendrogram branches) and cores (dendrogram leaves) within our MCs are affected by various stellar feedback mechanisms (similar to Neralwar et al. 2024). This will improve our understanding of the evolution of the molecular gas structures from clouds to core scales. We will also compare these gas structures with their observed counterparts in observations at various evolutionary stages (Urquhart et al. 2022) to understand the degree to which such multi-wavelength analysis are able to predict the evolutionary stages of gas structures.

7. Summary and conclusions

In this paper, we have created synthetic observations from a 20 000 $M_\odot$ STARFORGE simulation modelled after the SEDIGISM survey. We used the RADMC-3D radiative transfer code to convert the gas density cubes into ^{13}CO(2–1) emission maps and performed a dendrogram analysis to identify MCs. We analysed this sample of synthetic MCs and investigated the trends in properties, morphology, and substructures to understand how MCs evolve under the effects of different stellar feedback mechanisms.

The flux distributions of the SEDIGISM ppv cubes and our synthetic data cubes are in strong agreement, validating the replication of the SEDIGISM data to the first order. The properties of the synthetic MCs show good agreement with the SEDIGISM clouds, and the two samples fill the same parameter space in the scaling relation plots, which further confirms the robustness of our approach. Although the two cloud samples show an overall good agreement, the sythetic MCs reproduce only a subset of the diversity seen in the observations. Moreover, synthetic MCs at different evolutionary stages occupy distinct regions of the scaling relation plots, suggesting that evolutionary time plays a significant role in driving the observed scatter.

We studied the formation, evolution, and destruction of MCs through variations in their observable properties. The initial turbulence in the simulations creates gas overdensities that collapse under self-gravity and are detected as ^{13}CO(2–1) emission. These reflect the early cloudlets in observations that accrete gas from the larger environment and appear as moderately dense gas structures. Gas flows from large to small scales shape MCs into elongated filamentary structures with multiple substructures. The fractal substructures in MCs form stars, which eject matter and radiation into the surrounding environment, driving the formation of gas bubbles. These 3D bubble-like MCs are often associated with stellar winds, radiation, and H II regions. Our analysis presents MCs as evolving from small, diffuse structures to dense filamentary MCs followed by 3D gas bubbles, and these evolutionary trends are consistent with simulations initialised differently. This confirms the key hypothesis from our previous observational work that MCs evolve from concentrated to elongated to ring-like structures.

In conclusion, we have produced ^{13}CO(2–1) synthetic observations modelling the SEDIGISM survey using the STARFORGE simulations that include all the relevant physics for star formation. By analysing the properties, morphologies, and fragmentation trends of MCs, we have shown that they evolve from small, diffuse structures to dense filamentary structures to bubble-like structures. The distributions of MCs occupy different parameter spaces in the scaling relation plots, suggesting that they drive the scatter in the observed scaling relations. In an upcoming paper, we will study the effect of individual feedback mechanisms – outflows, stellar winds, radiation, supernovae – on these MCs and their substructures. We will also explore the possibility of compairing the structures at different evolutionary stages in simulations and observations.

Data availability

Movies associated to Appendix B are available at `https://www. aanda.org`

Acknowledgements. The authors thank the anonymous referee for a constructive report, which has significantly improved the quality of the manuscript. KN thanks Prof. Stefanie Walch-Gassner, Dr. Daniel Seifried, and Dr. Piyush Sharda for helpful discussions. AK acknowledges support from the Polish National Science Center SONATA BIS grant No. 2024/54/E/ST9/00314. MF acknowledges support from the Polish National Agency for Academic Exchange grant No. BPN/BEK/2023/1/00036/DEC/01 and from the Polish National Science Centre SONATA grant No. 2022/47/D/ST9/00419. S.N. gratefully acknowledges the Collaborative Research Center 1601 (SFB 1601 sub-project B1) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 500700252.

References

Ali, A. A., Bending, T. J. R., & Dobbs, C. L. 2022, MNRAS, 510, 5592
André, P., Men'shchikov, A., Bontemps, S., et al. 2010, A&A, 518, L102
Arzoumanian, D., André, P., Didelon, P., et al. 2011, A&A, 529, L6
Arzoumanian, D., André, P., Könyves, V., et al. 2019, A&A, 621, A42
Ballesteros-Paredes, J., Hartmann, L. W., Vázquez-Semadeni, E., Heitsch, F., & Zamora-Avilés, M. A. 2011, MNRAS, 411, 65
Ballesteros-Paredes, J., André, P., Hennebelle, P., et al. 2020, Space Sci. Rev., 216, 76
Bally, J. 2016, ARA&A, 54, 491
Barnes, P. J., Hernandez, A. K., Muller, E., & Pitts, R. L. 2018, ApJ, 866, 19
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Caselli, P., Walmsley, C. M., Tafalla, M., Dore, L., & Myers, P. C. 1999, ApJ, 523, L165
Chevance, M., Kruijssen, J. M. D., Hygate, A. P. S., et al. 2020, MNRAS, 493, 2872
Chevance, M., Krumholz, M. R., McLeod, A. F., et al. 2023, in Astronomical Society of the Pacific Conference Series, 534, Protostars and Planets VII, eds. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 1
Churchwell, E., Whitney, B. A., Babler, B. L., et al. 2004, ApJS, 154, 322
Clarke, S. D., Whitworth, A. P., Duarte-Cabral, A., & Hubber, D. A. 2017, MNRAS, 468, 2489
Clarke, S. D., Jaffa, S. E., & Whitworth, A. P. 2022, MNRAS, 516, 2782
Colombo, D., Rosolowsky, E., Ginsburg, A., Duarte-Cabral, A., & Hughes, A. 2015, MNRAS, 454, 2067
Colombo, D., Rosolowsky, E., Duarte-Cabral, A., et al. 2019, MNRAS, 483, 4291
Colombo, D., König, C., Urquhart, J. S., et al. 2021, A&A, 655, L2
Colombo, D., Duarte-Cabral, A., Pettitt, A. R., et al. 2022, A&A, 658, A54
Crutcher, R. M. 2012, ARA&A, 50, 29
Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792
Dempsey, J. T., Thomas, H. S., & Currie, M. J. 2013, ApJS, 209, 8
Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2014, in Protostars and Planets VI, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 3
Duarte-Cabral, A., & Dobbs, C. L. 2017, MNRAS, 470, 4261
Duarte-Cabral, A., Chrysostomou, A., Peretto, N., et al. 2012, A&A, 543, A140
Duarte-Cabral, A., Bontemps, S., Motte, F., et al. 2013, A&A, 558, A125
Duarte-Cabral, A., Colombo, D., Urquhart, J. S., et al. 2021, MNRAS, 500, 3027
Dubner, G., & Giacani, E. 2015, A&A Rev., 23, 3
Dullemond, C. P., Juhasz, A., Pohl, A., et al. 2012, RADMC-3D: a multi-purpose radiative transfer tool, Astrophysics Source Code Library [`record ascl:1202.015`]
Fierlinger, K. M., Burkert, A., Ntormousi, E., et al. 2016, MNRAS, 456, 710
Figueira, M., Zavagno, A., Deharveng, L., et al. 2017, A&A, 600, A93
Figueira, M., Zavagno, A., Bronfman, L., et al. 2020, A&A, 639, A93
Fukui, Y., Kawamura, A., Minamidani, T., et al. 2008, ApJS, 178, 56
Ge, Y., Wang, K., Duarte-Cabral, A., et al. 2023, A&A, 675, A119
Geen, S., Rosdahl, J., Blaizot, J., Devriendt, J., & Slyz, A. 2015, MNRAS, 448, 3248
Gómez, G. C., & Vázquez-Semadeni, E. 2014, ApJ, 791, 124
Gong, Y., Li, G. X., Mao, R. Q., et al. 2018, A&A, 620, A62
Grishunin, K., Weiss, A., Colombo, D., et al. 2024, A&A, 682, A137
Grudić, M. Y., Guszejnov, D., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2021, MNRAS, 506, 2199
Grudić, M. Y., Guszejnov, D., Offner, S. S. R., et al. 2022, MNRAS, 512, 216
Güsten, R., Nyman, L. Å., Schilke, P., et al. 2006, A&A, 454, L13
Guszejnov, D., Grudić, M. Y., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2021, MNRAS, 502, 3646
Guszejnov, D., Grudić, M. Y., Offner, S. S. R., et al. 2022, MNRAS, 515, 4929
Heyer, M., Krawczyk, C., Duval, J., & Jackson, J. M. 2009, ApJ, 699, 1092
Holdship, J., Viti, S., Jiménez-Serra, I., Makrymallis, A., & Priestley, F. 2017, AJ, 154, 38
Hopkins, P. F. 2015, MNRAS, 450, 53
Hopkins, P. F., & Raives, M. J. 2016, MNRAS, 455, 51
Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800
Hopkins, P. F., Wetzel, A., Wheeler, C., et al. 2023, MNRAS, 519, 3154
Hosokawa, T., Offner, S. S. R., & Krumholz, M. R. 2011, ApJ, 738, 140
Jacob, A. M., Menten, K. M., Wiesemeyer, H., et al. 2020, A&A, 640, A125
Jaffa, S. E., Whitworth, A. P., Clarke, S. D., & Howard, A. D. P. 2018, MNRAS, 477, 1940
Karska, A., Figueira, M., Mirocha, A., et al. 2025, A&A, 697, A186
Kauffmann, J., Pillai, T., Shetty, R., Myers, P. C., & Goodman, A. A. 2010, ApJ, 712, 1137
Kawamura, A., Mizuno, Y., Minamidani, T., et al. 2009, ApJS, 184, 1
Knutas, A., Adamo, A., Pedrini, A., et al. 2025, ApJ, 993, 13
Kruijssen, J. M. D., Schruba, A., Chevance, M., et al. 2019, Nature, 569, 519
Krumholz, M. R., & Gnedin, N. Y. 2011, ApJ, 729, 36
Krumholz, M. R., & McKee, C. F. 2020, MNRAS, 494, 624
Lane, H. B., Grudić, M. Y., Guszejnov, D., et al. 2022, MNRAS, 510, 4767
Larson, R. B. 1981, MNRAS, 194, 809
Li, G.-X., Urquhart, J. S., Leurini, S., et al. 2016, A&A, 591, A5
Lippok, N., Launhardt, R., Semenov, D., et al. 2013, A&A, 560, A41
Lu, X., Zhang, Q., Liu, H. B., et al. 2018, ApJ, 855, 9
Lu, Z.-J., Pelkonen, V.-M., Padoan, P., et al. 2020, ApJ, 904, 58
Lucas, W. E., Bonnell, I. A., & Dale, J. E. 2020, MNRAS, 493, 4700
Luisi, M., Anderson, L. D., Schneider, N., et al. 2021, Sci. Adv., 7, eabe9511
Maud, L. T., Moore, T. J. T., Lumsden, S. L., et al. 2015, MNRAS, 453, 645
Mazumdar, P., Wyrowski, F., Urquhart, J. S., et al. 2021, A&A, 656, A101
Molinari, S., Swinyard, B., Bally, J., et al. 2010, A&A, 518, L100
Neralwar, K. R., Colombo, D., Duarte-Cabral, A., et al. 2022a, A&A, 663, A56
Neralwar, K. R., Colombo, D., Duarte-Cabral, A., et al. 2022b, A&A, 664, A84
Neralwar, K. R., Colombo, D., Offner, S., et al. 2024, A&A, 690, A345
Neupane, S., Wyrowski, F., Menten, K. M., et al. 2024, A&A, 692, A114
Offner, S. S. R., Taylor, J., Markey, C., et al. 2022, MNRAS, 517, 885
Padoan, P., Jones, B. J. T., & Nordlund, Å. P. 1997, ApJ, 474, 730
Padoan, P., Pan, L., Haugbølle, T., & Nordlund, Å. 2016, ApJ, 822, 11
Palmeirim, P., Zavagno, A., Elia, D., et al. 2017, A&A, 605, A35
Peretto, N., Rigby, A. J., Louvet, F., et al. 2023, MNRAS, 525, 2935
Priestley, F. D., & Whitworth, A. P. 2022, MNRAS, 509, 1494
Priestley, F. D., Clark, P. C., & Whitworth, A. P. 2023, MNRAS, 519, 6392
Rebolledo, D., Wong, T., Leroy, A., Koda, J., & Donovan Meyer, J. 2012, ApJ, 757, 155
Rigby, A. J., Moore, T. J. T., Plume, R., et al. 2016, MNRAS, 456, 2885
Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, ApJ, 723, 492
Rosen, A. L., Offner, S. S. R., Sadavoy, S. I., et al. 2020, Space Sci. Rev., 216, 62
Rosolowsky, E., & Leroy, A. 2006, PASP, 118, 590
Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, ApJ, 679, 1338
Roueff, A., Gerin, M., Gratier, P., et al. 2021, A&A, 645, A26
Santoro, F., Kreckel, K., Belfiore, F., et al. 2022, A&A, 658, A188
Schuller, F., Csengeri, T., Urquhart, J. S., et al. 2017, A&A, 601, A124
Schuller, F., Urquhart, J. S., Csengeri, T., et al. 2021, MNRAS, 500, 3064
Seifried, D., Walch, S., Girichidis, P., et al. 2017, MNRAS, 472, 4797
Shimajiri, Y., André, P., Palmeirim, P., et al. 2019, A&A, 623, A16
Simpson, R. J., Povich, M. S., Kendrew, S., et al. 2012, MNRAS, 424, 2442
Skretas, I. M., Karska, A., Wyrowski, F., et al. 2023, A&A, 679, A66
Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730
Tiwari, M., Karim, R., Pound, M. W., et al. 2021, ApJ, 914, 117
Traficante, A., Duarte-Cabral, A., Elia, D., et al. 2018, MNRAS, 477, 2220
Urquhart, J. S., Figura, C., Cross, J. R., et al. 2021, MNRAS, 500, 3050
Urquhart, J. S., Wells, M. R. A., Pillai, T., et al. 2022, MNRAS, 510, 3389
Urquhart, J. S., König, C., Colombo, D., et al. 2024, MNRAS, 528, 4746
Urquhart, J. S., König, C., Colombo, D., et al. 2025, MNRAS, 539, 3105
Vázquez-Semadeni, E., Palau, A., Ballesteros-Paredes, J., Gómez, G. C., & Zamora-Avilés, M. 2019, MNRAS, 490, 3061
Vázquez-Semadeni, E., Palau, A., Gómez, G. C., et al. 2024, arXiv e-prints [arXiv:`2408.10406`]
Vink, J. S. 2024, arXiv e-prints [arXiv:`2406.16517`]
Walch, S., Girichidis, P., Naab, T., et al. 2015, MNRAS, 454, 238
Watkins, E. J., Peretto, N., Rigby, A. J., et al. 2025, MNRAS, 536, 2805
Weaver, R., McCray, R., Castor, J., Shapiro, P., & Moore, R. 1977, ApJ, 218, 377
Williams, G. M., Peretto, N., Avison, A., Duarte-Cabral, A., & Fuller, G. A. 2018, A&A, 613, A11

Appendix A: Effects of projection angles

The ppv cubes for this study were produced using RADMC-3D by projecting along three orthagonal axes. This is achieved using three combinations of incl-phi: 0-0, 90-0 and 90-90 in the RADMC-3D script for spectral line imaging (`radmc3d image`)[14]. The simulation box is thus projected along the z, y, and x axes, respectively. MCs identified in different projections have similar properties, which is consistent with previous similar works (e.g. Priestley et al. 2023). This is largely due to the fact that ^{13}CO(2-1) emission is optically thin and thus the entire MC is traced along all projections. We conclude that using a specific projection does not alter the MC properties and provides a sanity check that the simulations and RADMC-3D produce model clouds reasonably well.

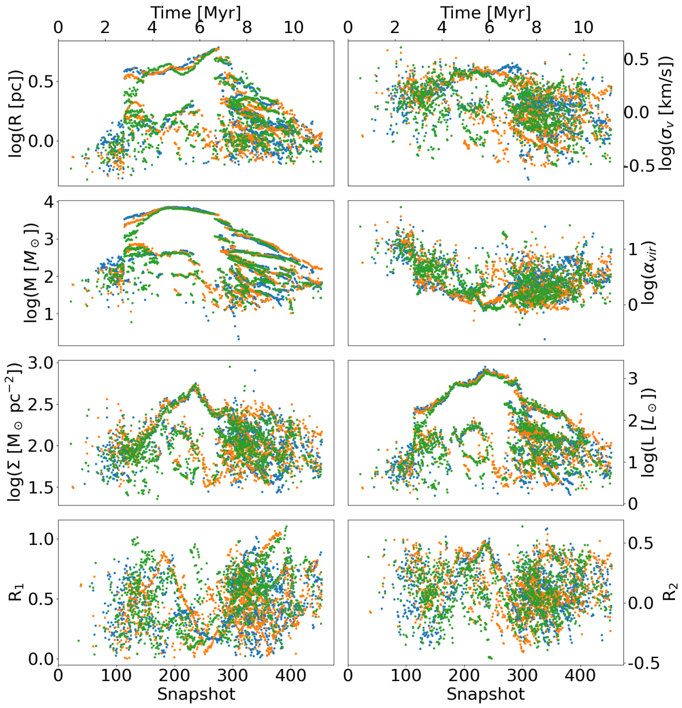


Fig. A.1. The three colours represent the axes along which the cube is projected. The colour scheme follows fig. 6.

Appendix B: ^{13}CO(2-1) emission maps

In this section, we show a sequence of the ^{13}CO(2-1) moment 0 maps for the GMC (i.e. the molecular gas complexes) as it evolves along with the dendrogram trunks (MCs). The MC obtained by projecting along three orthagonal axes (Sect. 3.1.2) are shown in Figures B.1–B.3. We also provide videos that represent all snapshots along the three projections as ancillary materials. This helps us to visualise the clouds and understand it's structure at different evolutionary stages. Fig. B.1 shows the formation of MCs as small diffuse structures. Fig. B.2 shows a single (or a few) contour(s) that cover the entire emission in the viridis, representing most of the observed MCs. The filamentary, fractal, and complex nature of these structures is also visible in the emission maps. Fig. B.3 shows the MCs that are significantly impacted by stellar feedback processes. These lead to gas expulsion and dispersion, thus presenting the ^{13}CO(2-1) as a 3D bubble. As more and more gas are dispersed, the number of MCs decreases. Some of these late MCs represent the early structures (Fig. B.1), with the difference that the early MCs have accompanying H_2 gas. The absence of a molecular gas prevents the formation of MCs and ends the simulation.

---

[14] `https://www.ita.uni-heidelberg.de/~dullemond/software/radmc-3d/manual_radmc3d/imagesspectra.html`

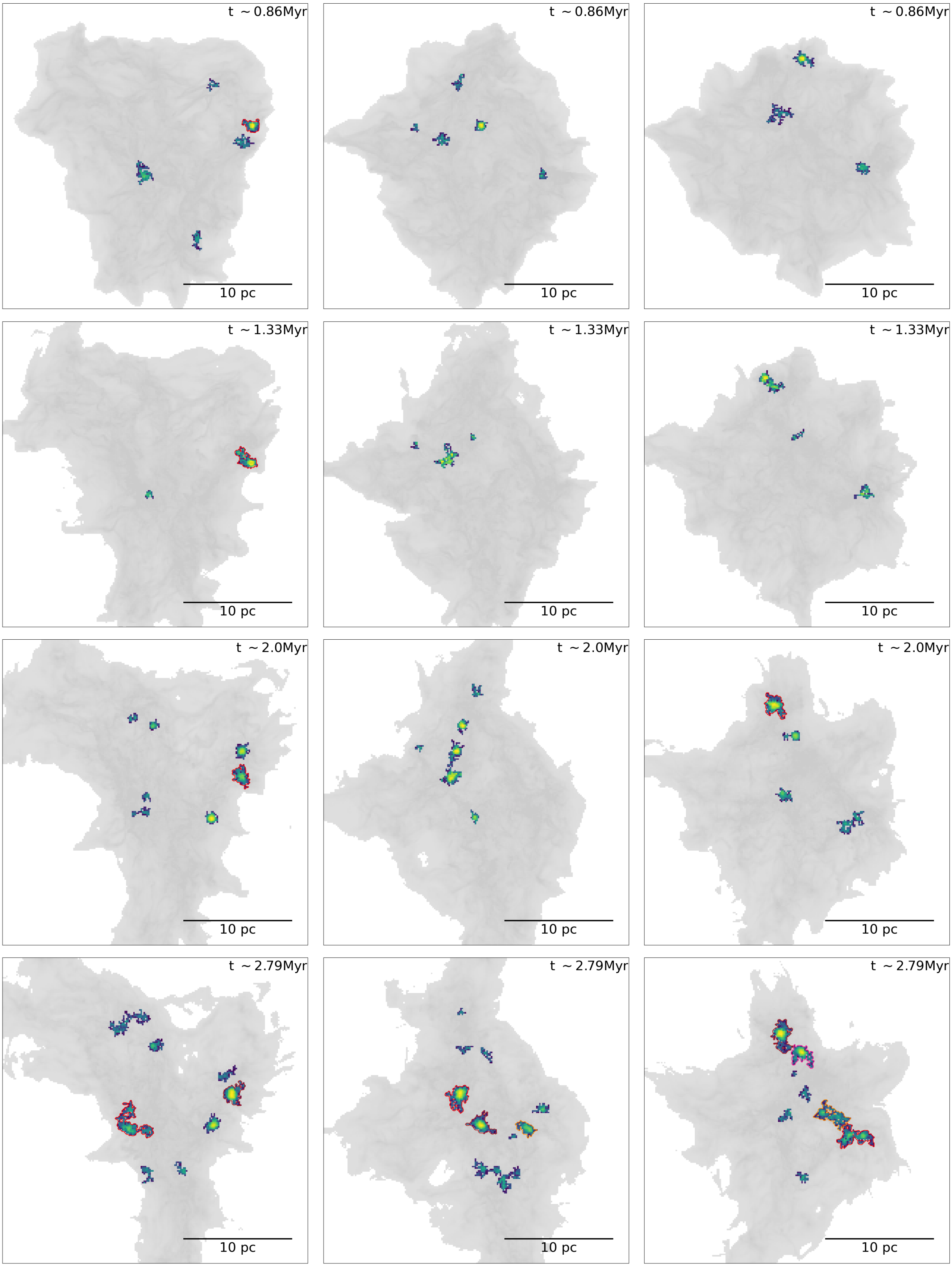


Fig. B.1. Moment 0 maps of ^{13}CO(2-1) for different snapshots. The three columns represent the clouds projected along the x, y and z axes, respectively. The background grey scale represents H_2 gas density with ^{13}CO(2-1) emission overlaid as viridis maps, and coloured contours represent different MCs (dendrogram trunks), with red contours representing the largest MCs (R) in the cube.

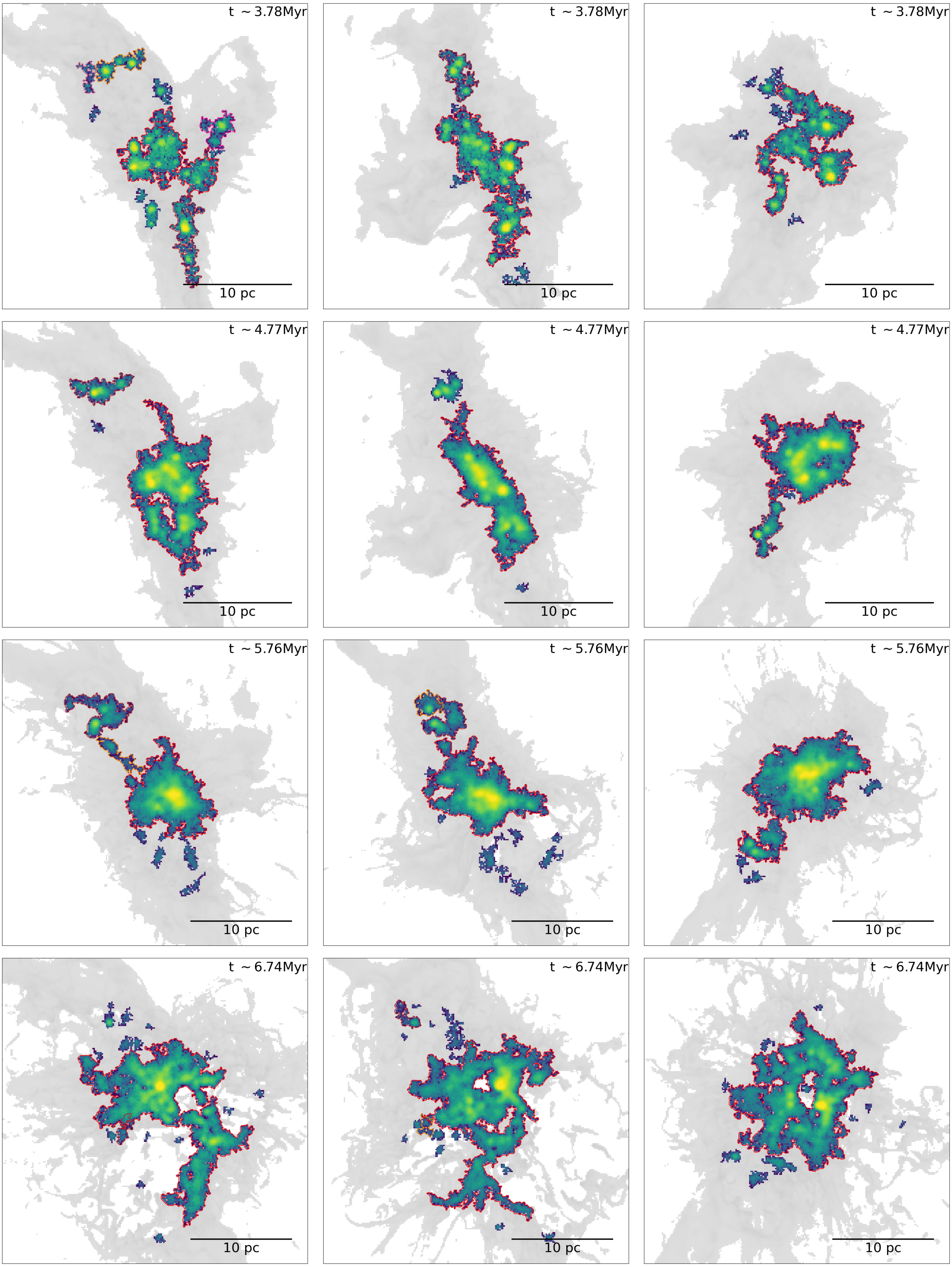


Fig. B.2. Moment 0 maps of ^{13}CO(2-1) for different snapshots. The colours and symbols follow Fig. B.1.

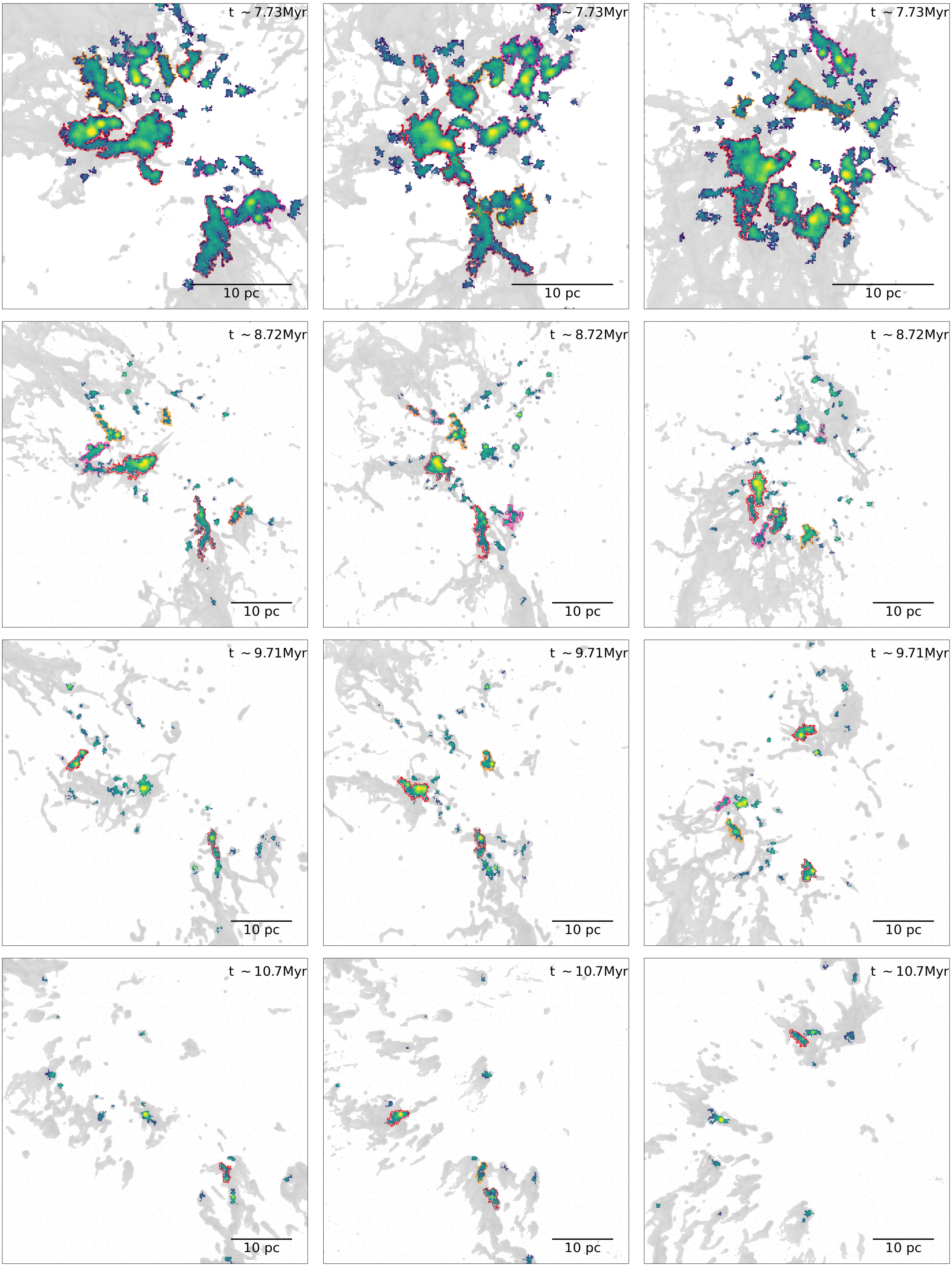


Fig. B.3. Moment 0 maps of ^{13}CO(2-1) for different snapshots. The colours and symbols follow fig. B.1.

Appendix C: Inclusion of CO chemistry

We post-processed the STARFORGE simulations with UCLCHEM (Holdship et al. 2017) chemical code[15] to estimate the abundance of CO (Sharda et al. in prep). However, due to computational cost, this has only been possible for three snapshots. Figures C.1 – C.3 show that although our fiducial approach slightly overestimates the ^{13}CO emission, the data processing steps produce MCs of comparable size and morphology in both cases. This is further highlighted in Fig. C.4 & C.5, which show that both sets of MCs have similar properties in the same snapshots. The inclusion of chemistry might cause a small difference in the properties of our MCs and result in smaller MCs not being detectable. However, since we study the trends in the distribution of properties over time, we expect that these are not significantly influenced by the absence of CO chemistry.

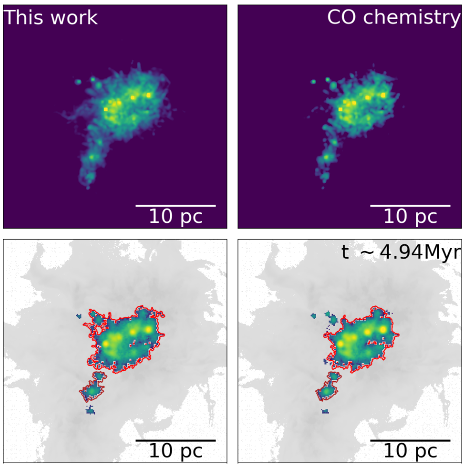


Fig. C.1. Comparison between the ^{13}CO(2-1) ppv cube used in this work and created by including CO chemistry for snapshot 200 (4.94 Myr). The top rows show the RADMC-3D output and the bottom show the masked cubes with MCs (contours) overlayed on the ^{13}CO(2-1) emission (viridis) and projected H_2 density (grey scale) maps.

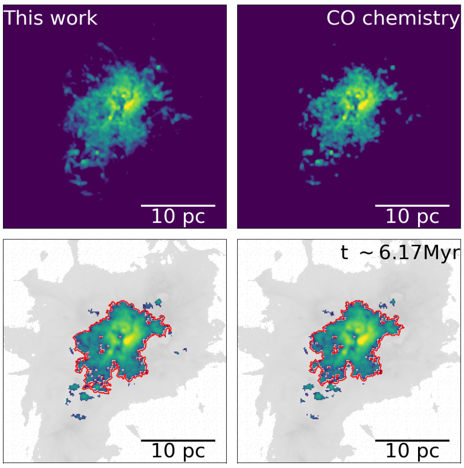


Fig. C.2. Comparison between the ^{13}CO(2-1) ppv cube used in this work and created by including CO chemistry for snapshot 250 (6.17 Myr). The symbols and notations follow Fig. C.1.

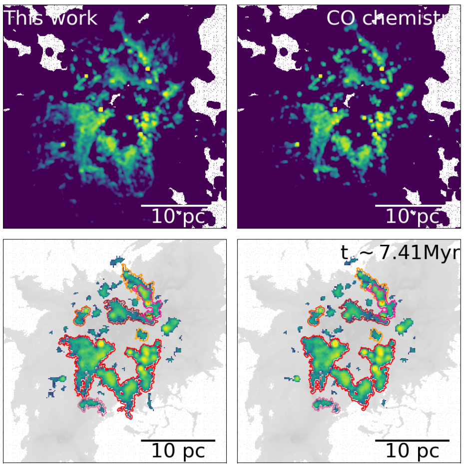


Fig. C.3. Comparison between the ^{13}CO(2-1) ppv cube used in this work and created by including CO chemistry for snapshot 300 (7.41Myr). The symbols and notations follow Fig. C.1.

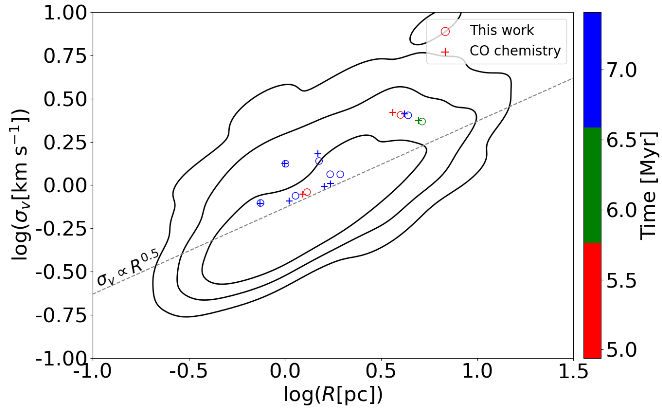


Fig. C.4. Size-linewidth relation (σ_v versus R) for our MCs (circles) and those created by including CO chemistry (plus). The MCs have been extracted from three ppv cubes in Figs. C.1–C.3. The symbols and notations are consistent with Fig. 8.

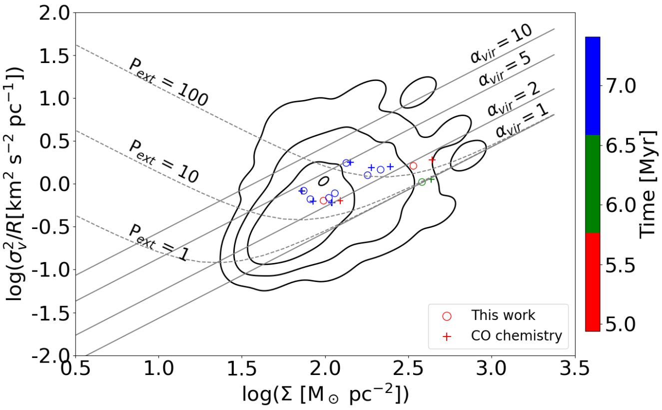


Fig. C.5. Scaling relation between σ_v^2/R and surface mass density (Σ). The symbols and notations follow Fig. C.4. The solid grey lines represent isocontours of virial parameters. The dashed lines represent α_{vir} = 1 when including an external pressure P_{ext} = 1, 10, 100 $M_\odot$ pc^{-3} km^2 s^{-2}.

---

[15] The pipeline is provided here: https://github.com/psharda/gizmo_carver/tree/pschanges

Appendix D: Hierarchical and isolated trunks

This section explains the reason for selecting only trunks that are branches as MCs, rather than including all trunks. In figure D.1, we show the distribution of properties for all dendrogram trunks. The results clearly demonstrate that only hierarchical trunks (branches) exhibit consistent trends in their properties as they evolve. In contrast, isolated trunks (leaves) have scattered distributions with no clear trends. This is likely because most of these isolated structures represent transient gas features that do not correspond to the fractal MCs seen in observations (Fig. B.1). Furthermore, the large sample of isolated trunks results in the average properties of the trunks (Fig. D.1, central line) that show no significant trends over time.

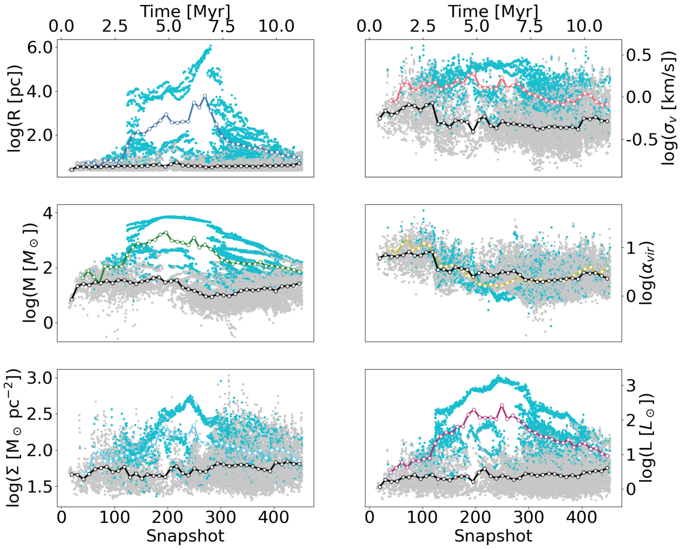


Fig. D.1. The cyan points represents the hierarchical trunks (branches) and grey points represents the isolated trunks (leaves). The coloured lines represents the average distribution of hierarchical trunks, with the colours following Fig. 5. The black line represents the average distribuion for isolated trunks.

Appendix E: Simulations with different initial turbulence

We perform our analysis on two other simulation sets. These have the same initial conditions as our fiducial runs, with the exception of initial turbulence. These are M2e4a1 and M2e4a4 with $\alpha_{vir} = 1$ and $\alpha_{vir} = 4$, respectively. Figs. E.1 & E.2 show the evolution of the MC properties for these two simulations, respectively. Although the GMC lifetime and the onset of different feedback mechanisms are different in the two simulations, they show a trend of increase in the MC properties representing actively growing MCs followed by a decrease in the properties due to MC dispersal by feedback. The morphology and fragementation trends of these MCs are consistent with the fiducial simulations.

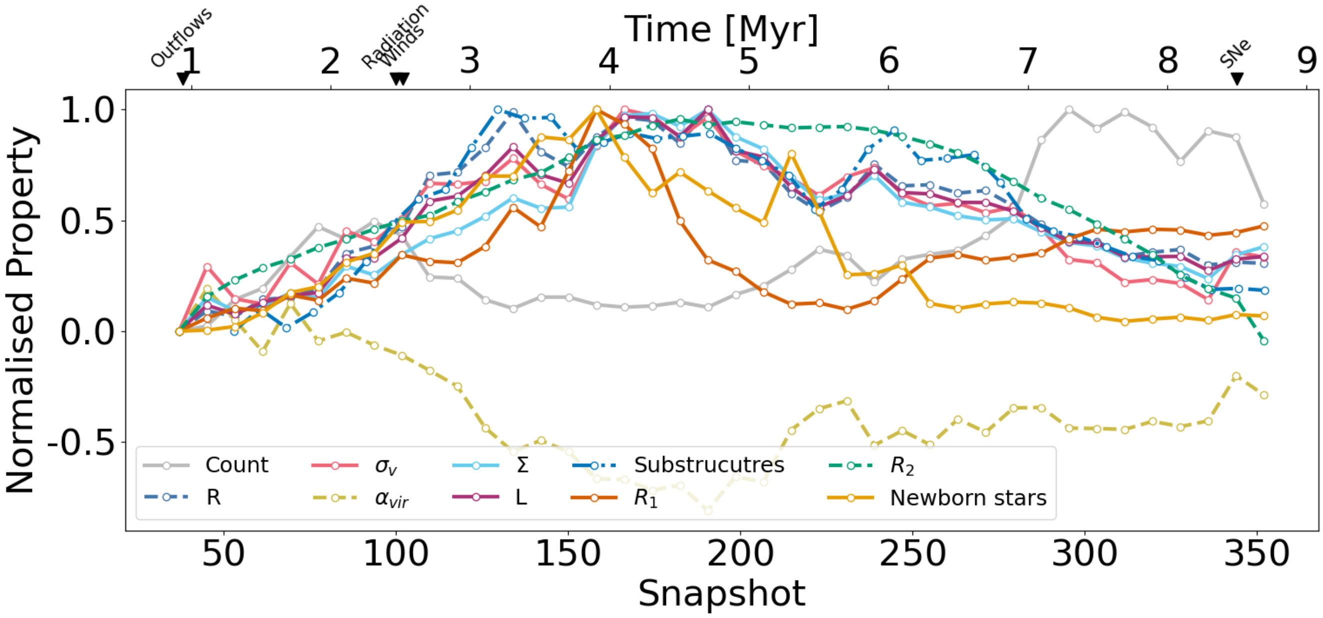


Fig. E.1. Simulation with α_{vir} = 1. Properties of the clouds at various evolutionary stages. The R_1 and R_2 values represent the morphologies of the molecular gas complexes. The symbols and colours follow fig. 5 & 7.

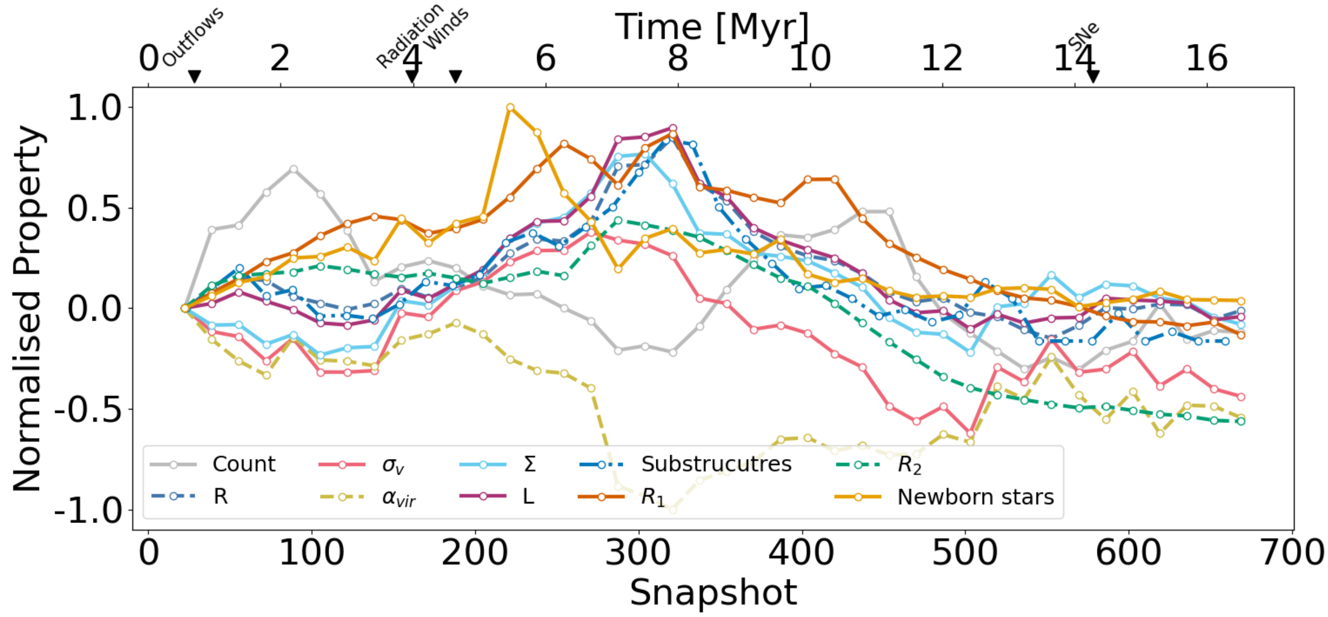


Fig. E.2. Simulation with α_{vir} = 4. Properties of the clouds at various evolutionary stages. The symbols and colours follow fig. E.1.

APPENDIX C

Evolution of individual clumps in STARFORGE

C.1 Moment maps for clump 27526

In this section, we present the moment maps for the hyperclump 27526, which illustrate its complete evolution. Figures C.1–C.8 show a 21 pc×21 pc region of the simulation box from t $\sim$ 6.8 Myr – 10.9 Myr after each 0.5 Myr. These maps enable a visual interpretation of the evolution of the clump 27526 as it fragments and merges into multiple time substructures.

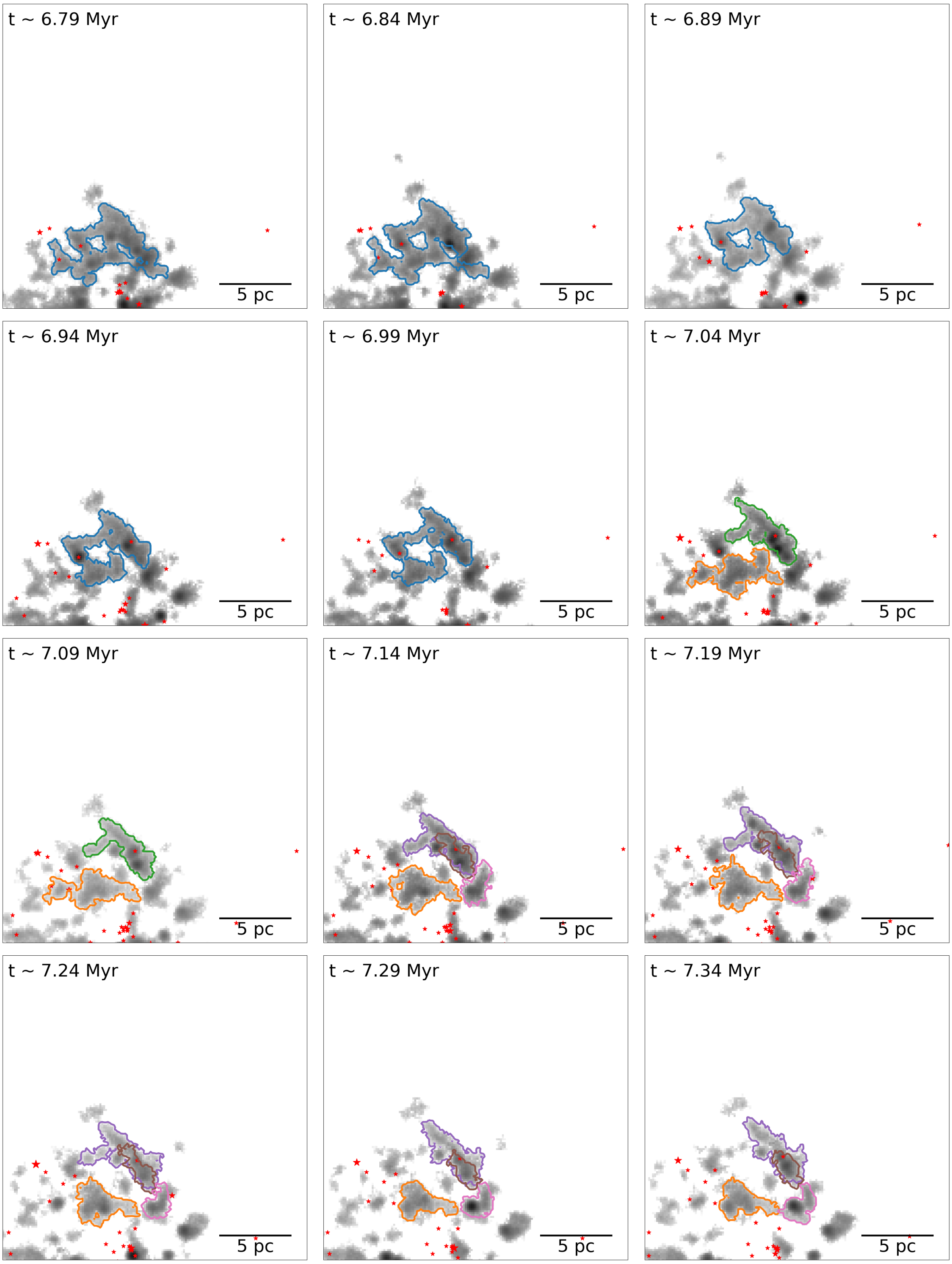


Figure C.1: Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors and notation follow Fig. 7.6.

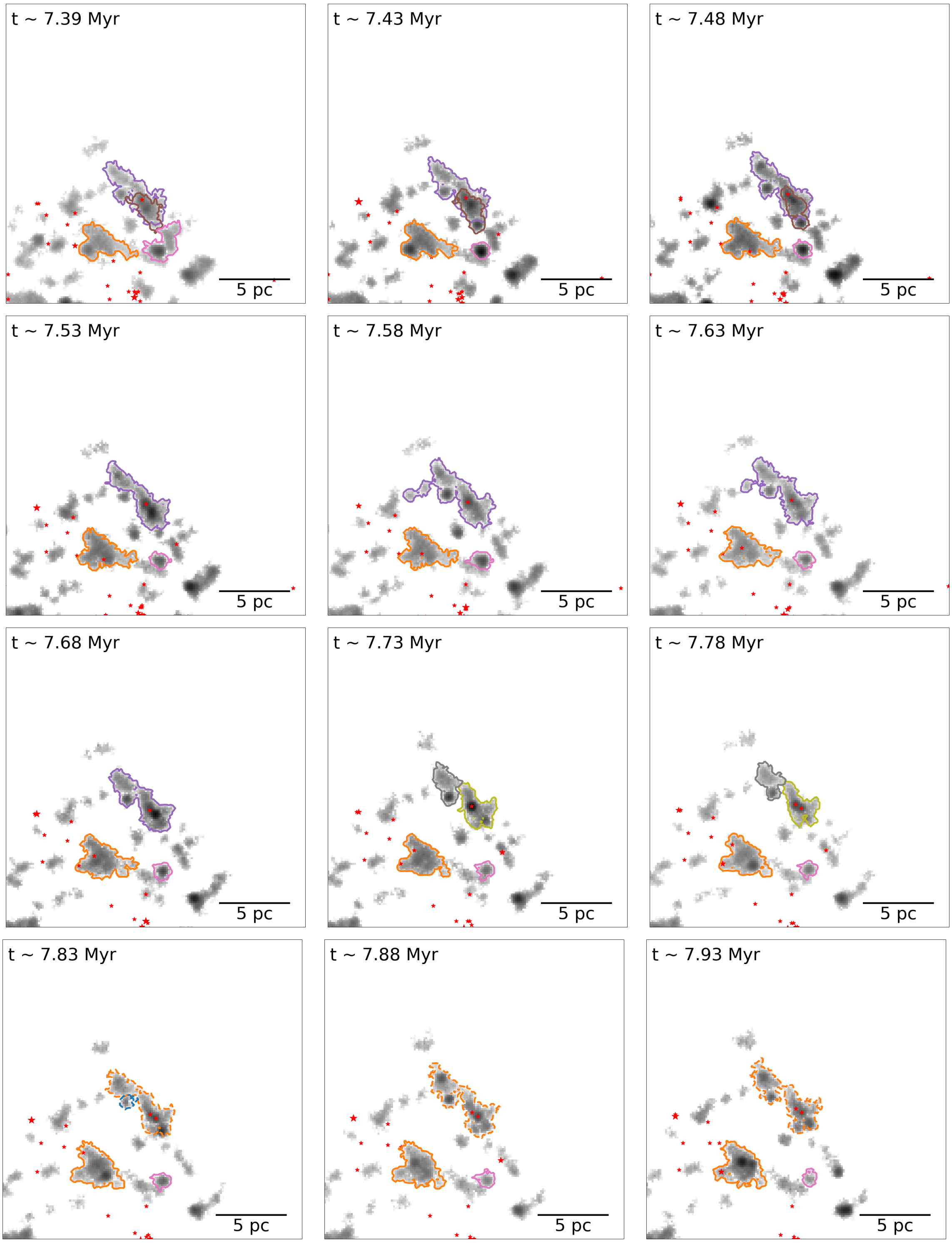


Figure C.2: Continuation of Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors, and notation follow Fig. 7.6.

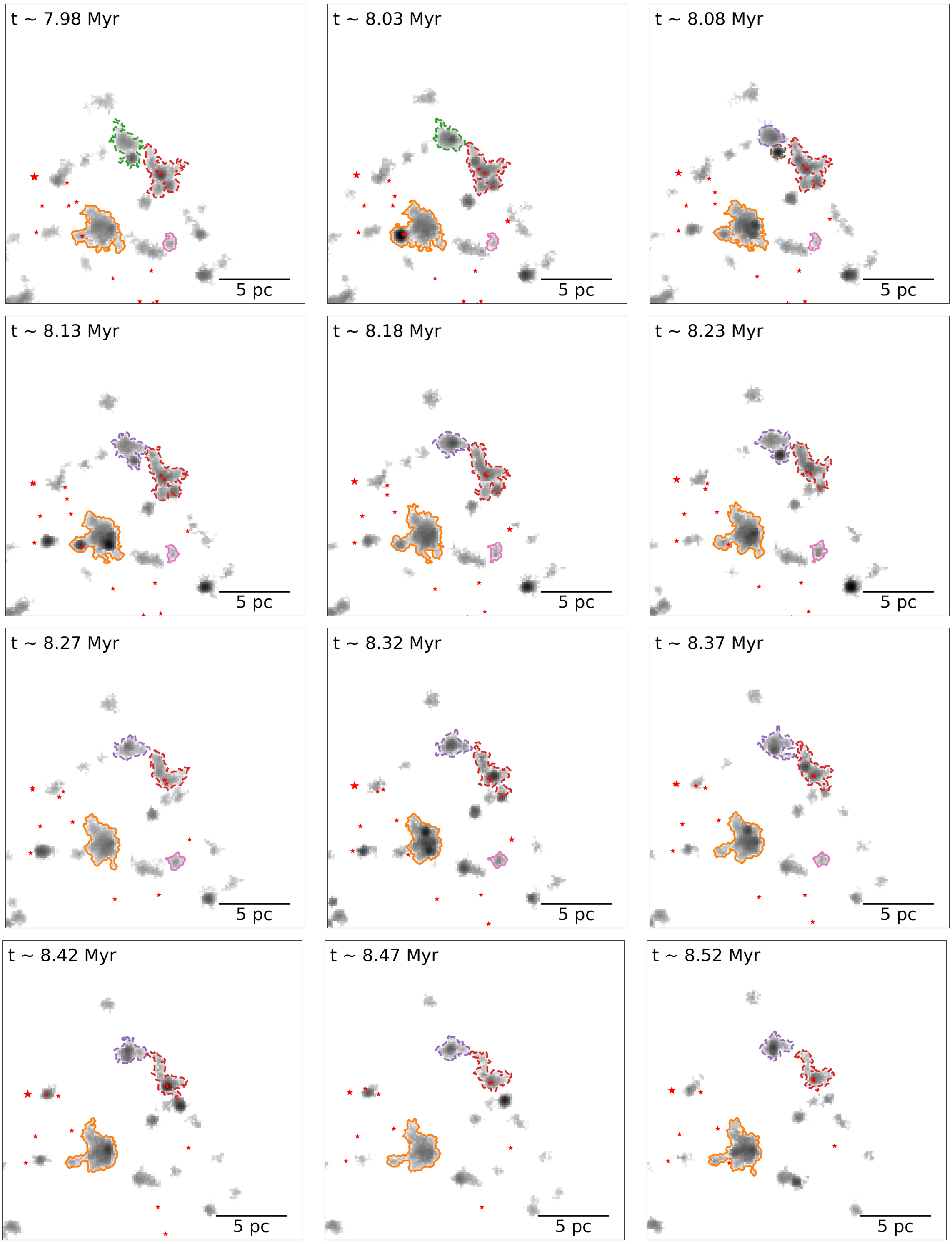


Figure C.3: Continuation of Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors, and notation follow Fig. 7.6.

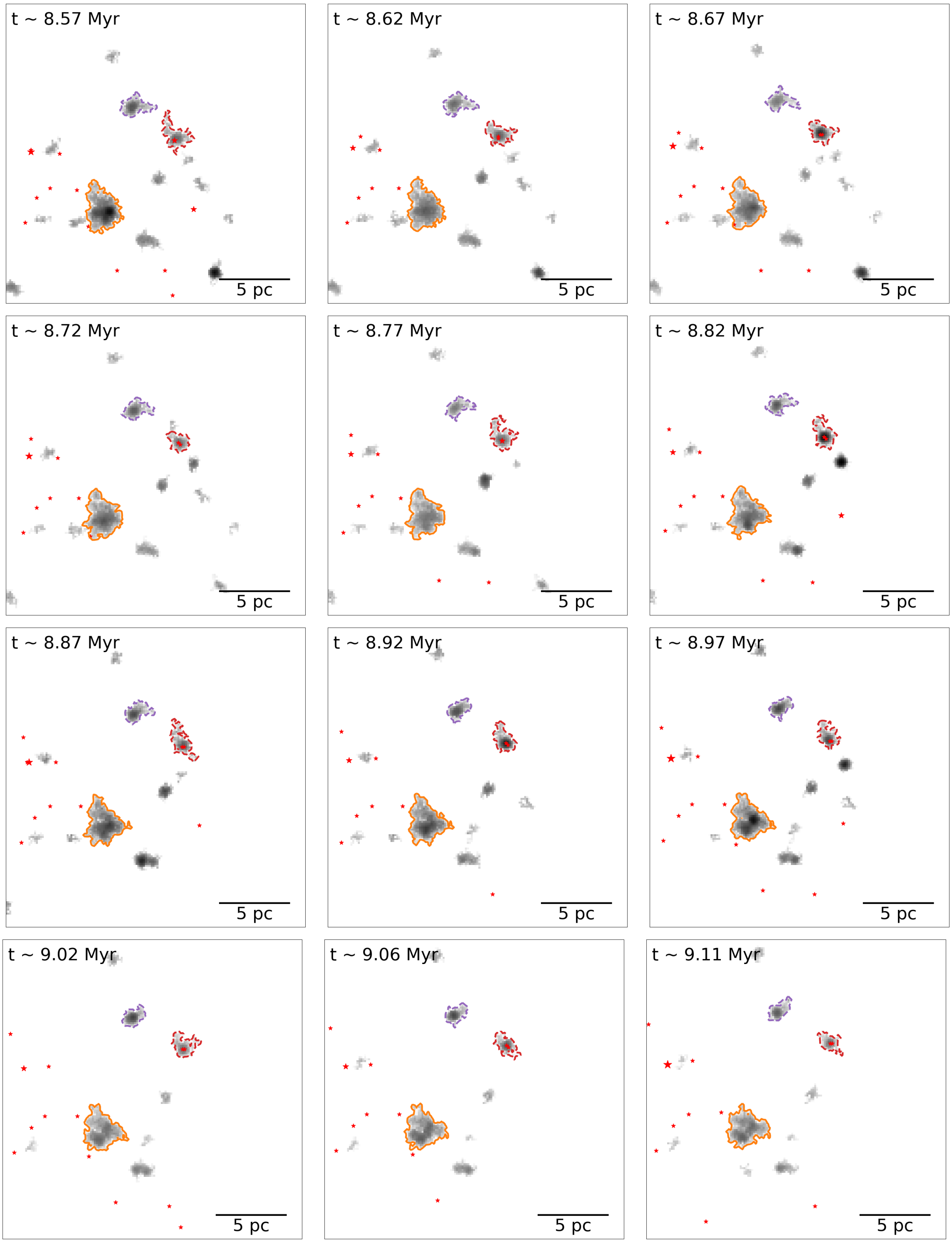


Figure C.4: Continuation of Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors, and notation follow Fig. 7.6.

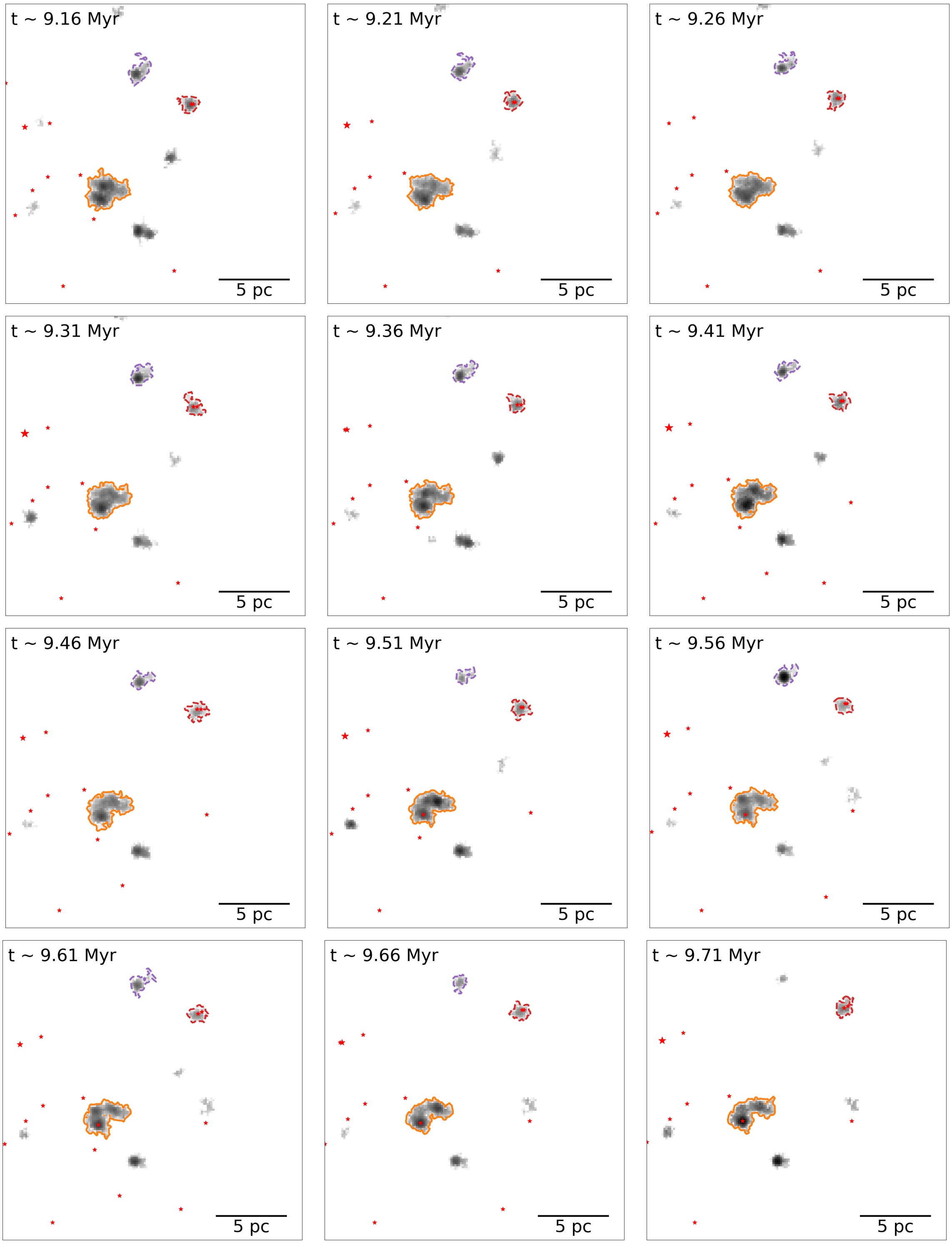


Figure C.5: Continuation of Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors, and notation follow Fig. 7.6.

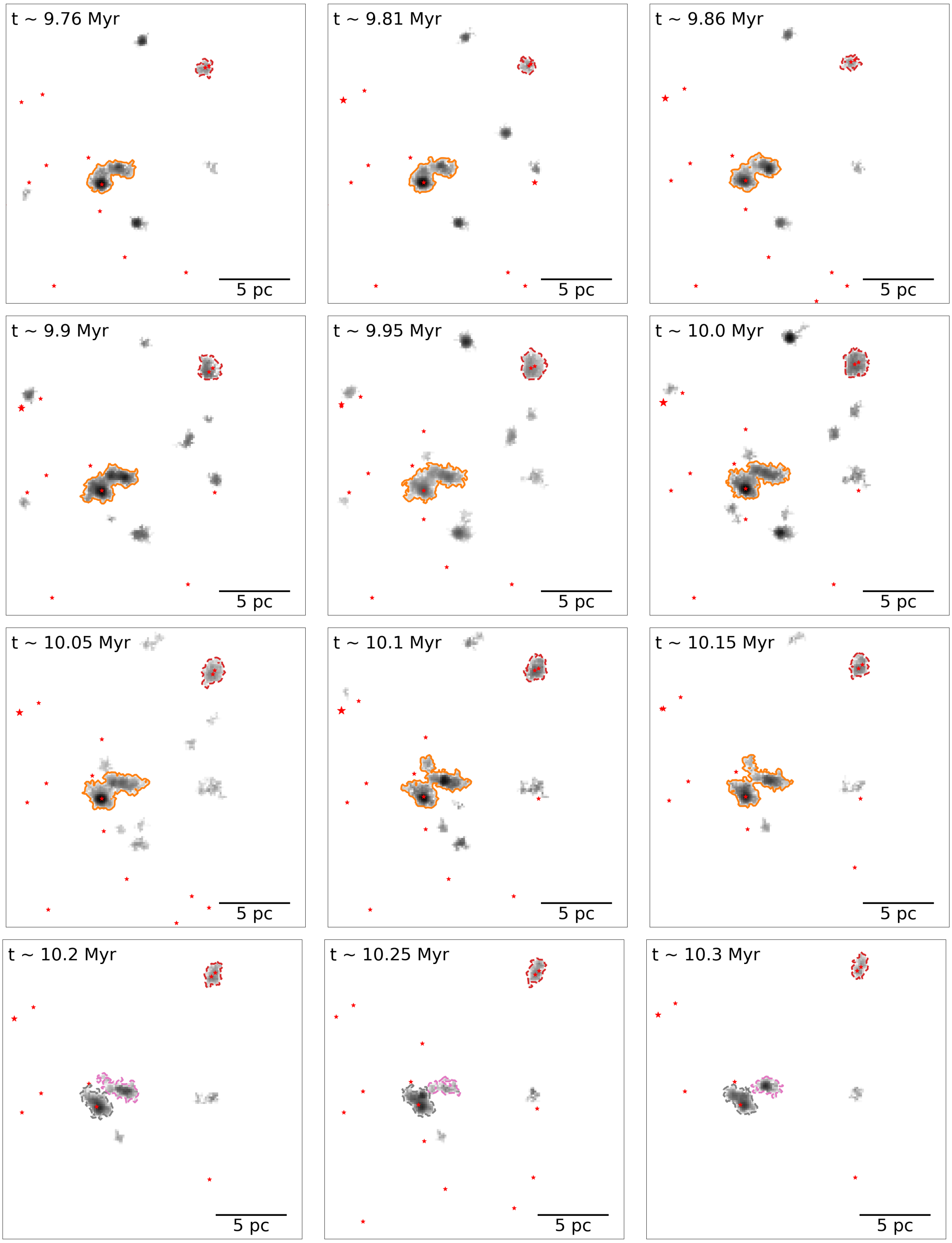


Figure C.6: Continuation of Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors, and notation follow Fig. 7.6.

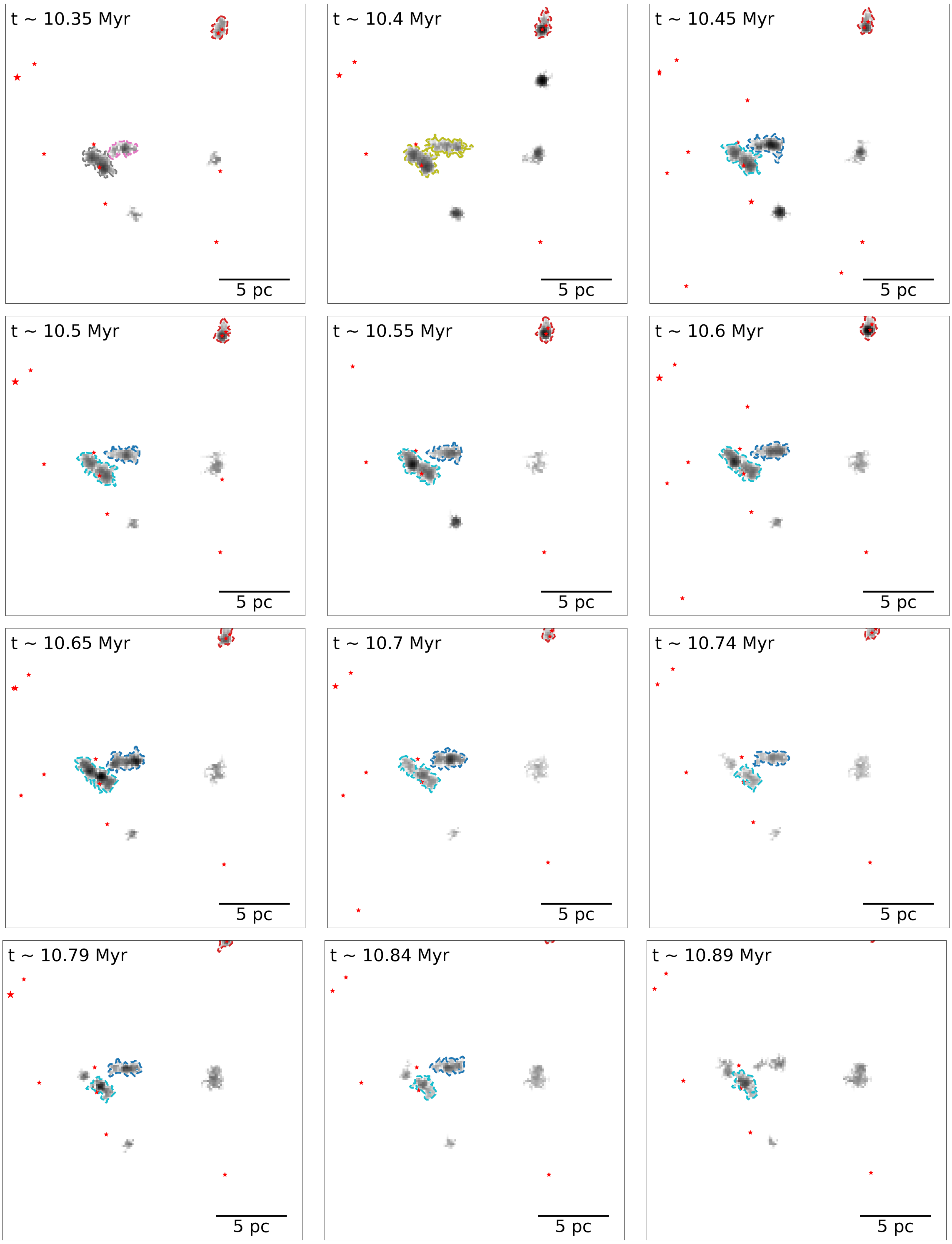


Figure C.7: Continuation of Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors, and notation follow Fig. 7.6.

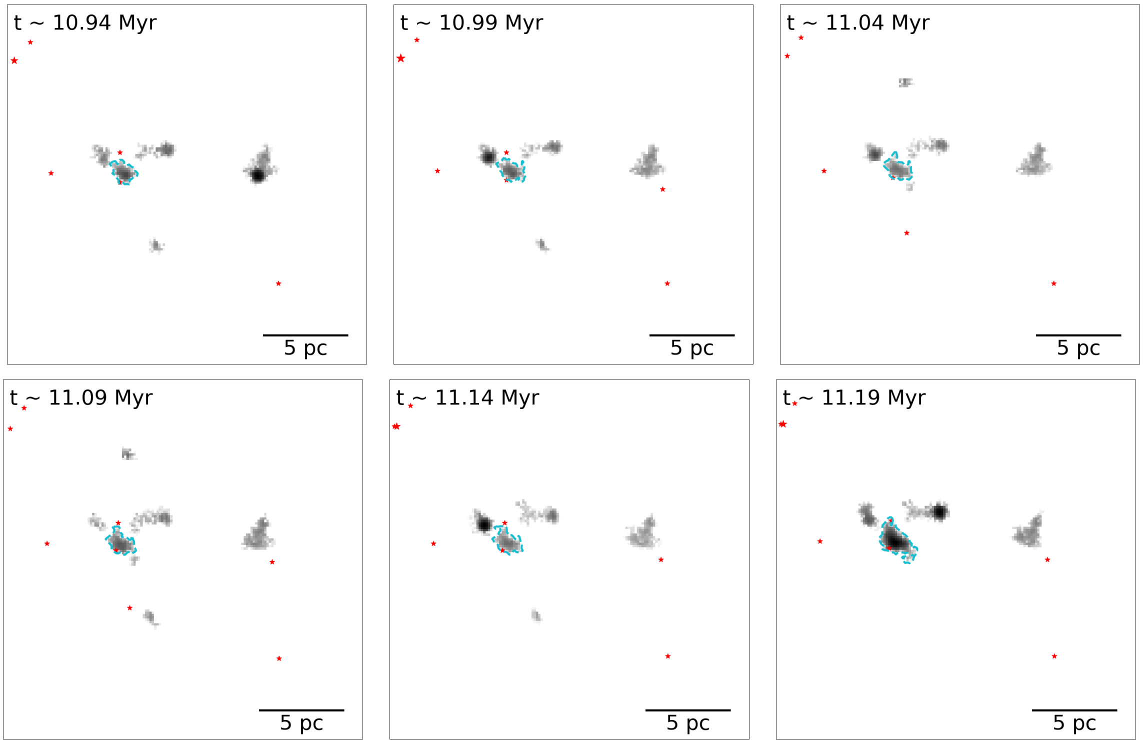


Figure C.8: Continuation of Moment 0 maps of the clump 27526 and its time-substructures (greyscale). The symbols, colors, and notation follow Fig. 7.6.

C.2 Evolution of properties of individual clumps

In this section, we show the evolution of the radius, mass, and velocity dispersion of individual clumps and their time substructures (Figs C.9 & C.10). The clumps are selected such that their lifetime is closest to the average lifetime of the clumps identified in the nearby snapshots.

Clumps identified in the actively-growing phase of the GMC (≤ 6 Myr; Sect 7.5.1) show an increase in size and mass over time (till ~ 7 Myr) before fragmenting into multiple smaller structures, which subsequently decrease in size. The clumps identified in the feedback-dominated phase (≥ 6 Myr; Sect 7.5.1) show diverse behaviours. They are either rapidly dispersed, evolve largely as a single coherent structure, or undergo many fragmentations. These are generally small structures embedded within a GMC, which is being dispersed by stellar feedback. As a result, they match small, fragmented clumps in subsequent snapshots. However, once these clumps are dispersed, CHANDRAST can match them to a larger overlapping structure, causing them to appear as large clumps in Fig. C.10. This reflects a caveat of CHANDRAST (Sect. 7.6), i.e., the absence of an upper size limit when matching clumps across snapshots.

Overall, the properties of long-lived clumps (Fig. C.9) are consistent with our findings in Neralwar, Colombo, S. Offner, Karska et al. (2025). This presents a scenario where molecular gas structures grow from small cloudlets into large, massive clouds, before being fragmented and eventually dispersed by stellar feedback.

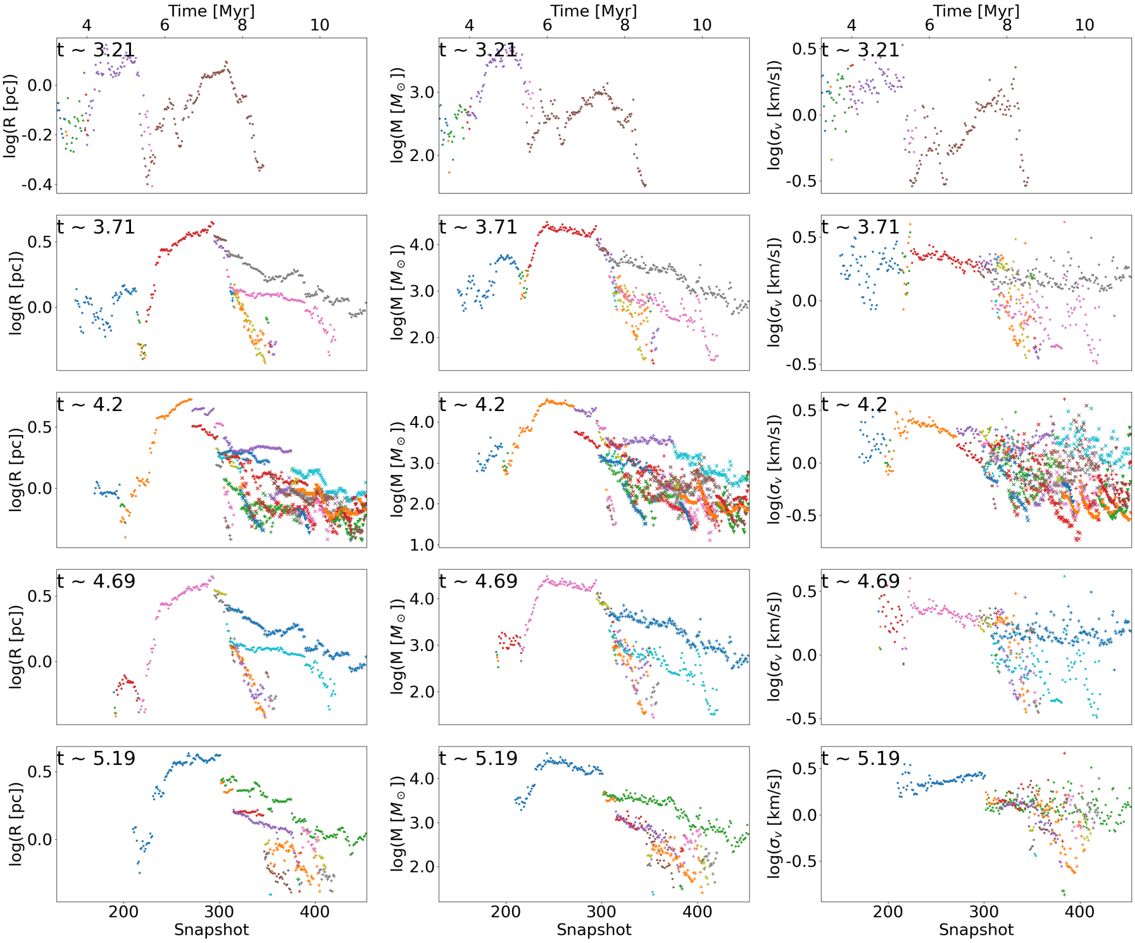


Figure C.9: Evolution of radius, surface density and velocity dispersion of clumps. Each row corresponds to a single hyperclump, with 't' denoting the time after the start of the simulation at which the hyperclump is first detected.

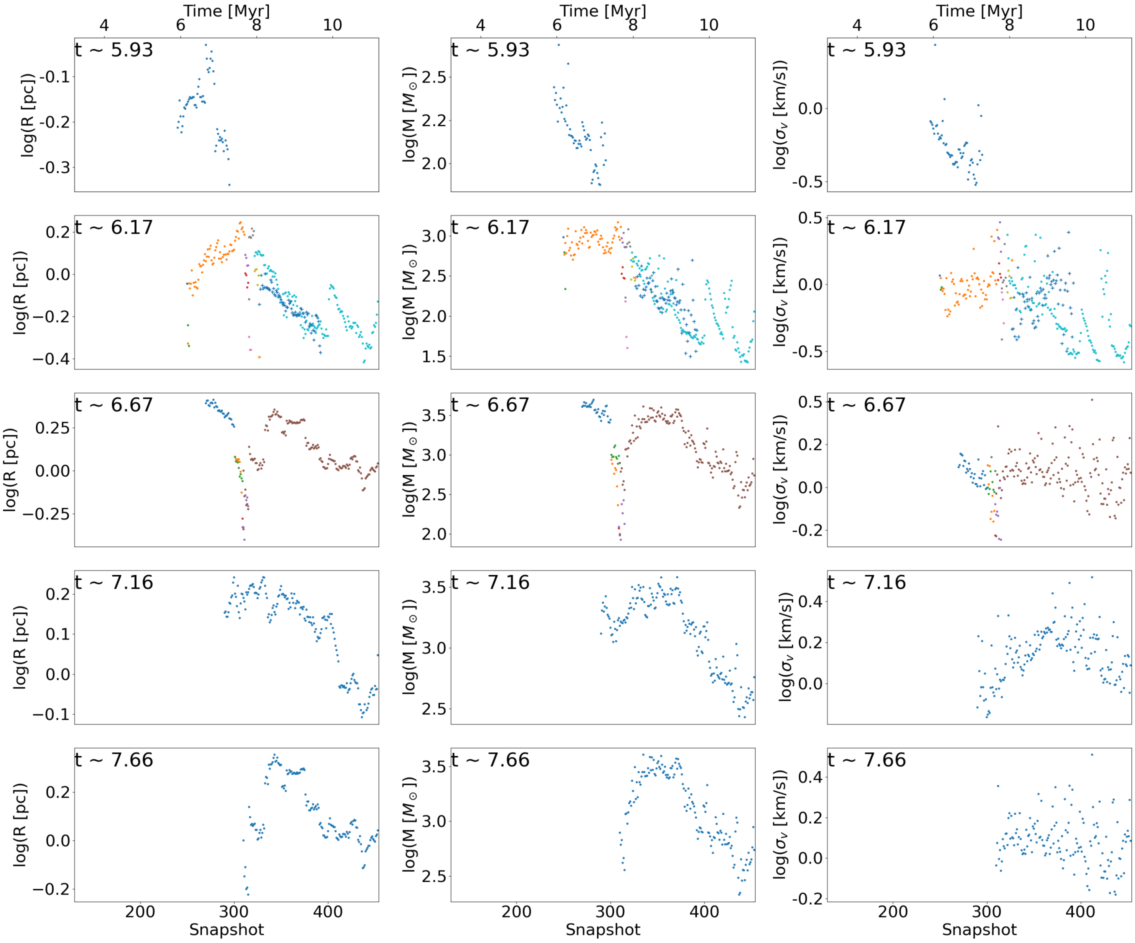


Figure C.10: Evolution of radius, surface density and velocity dispersion of clumps. Each row corresponds to a single hyperclump, with 't' denoting the time after the start of the simulation at which the hyperclump is first detected.

C.3 Evolution of properties of feedback-affected clumps

In this section, we present the evolution of the radius, mass, and velocity dispersion of individual clumps within hyperclumps that are either marginally or significantly affected by stellar feedback (Figs C.12 – C.13).

Figure C.12 shows the evolution of clump properties for in five hyperclumps that are not significantly affected by protostellar outflows (Sect. 7.5.3). These clumps are small, less massive, and short-lived structures and likely correspond to starless clumps that are rapidly dispersed shortly after their formation.

Figure C.11 presents the evolution of clump properties in five hyperclumps that are significantly affected by protostellar outflows. These are large and massive clumps that survive long enough to form protostars. The outflows driven by the embedded protostars contribute to the fragmentation of the clumps and increase their overall outflow feedback fractions. Compared to the short-lived ones, these clumps have a slower decrease in radius and mass. This indicates that they require longer timescales

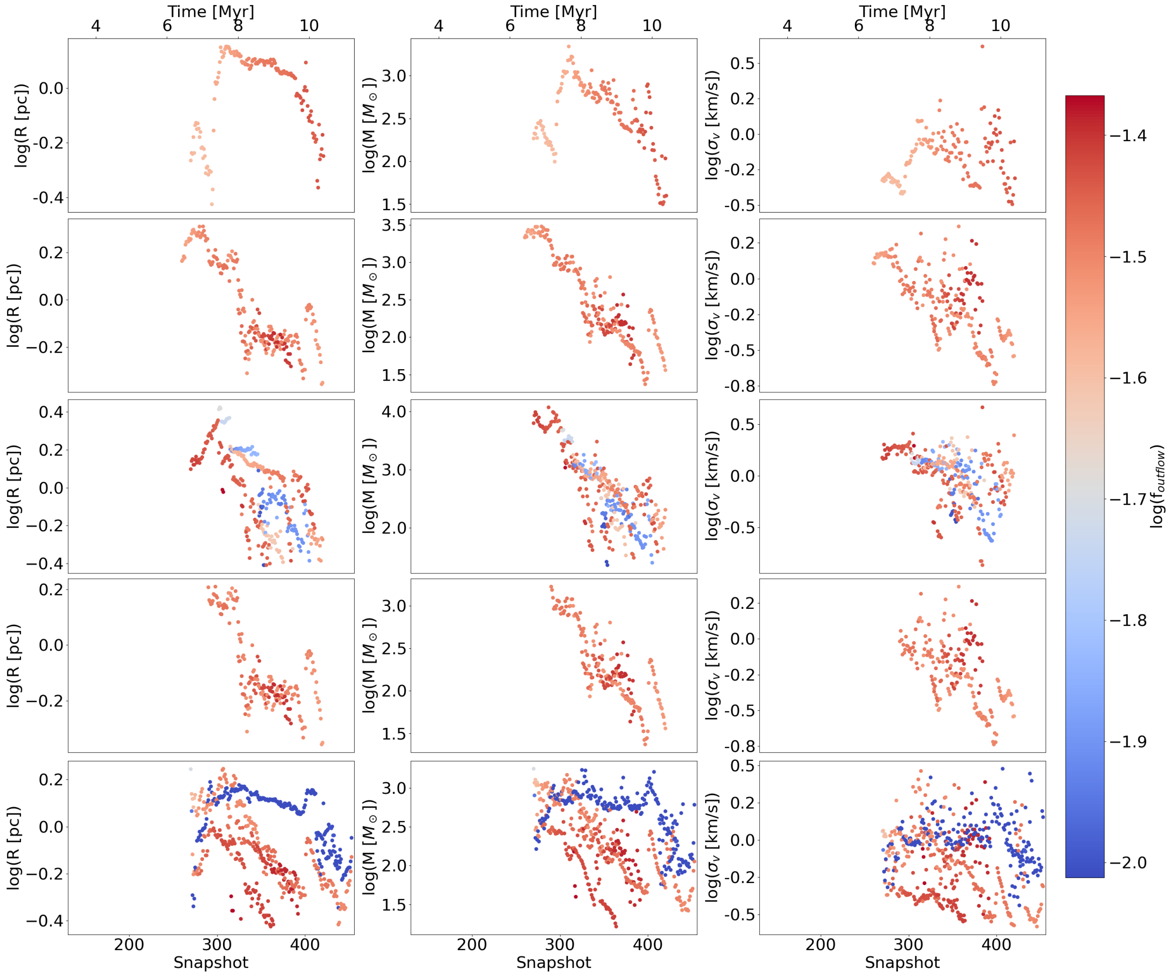


Figure C.11: Evolution of radius, mass and velocity dispersion of the clumps in short living hyperclumps that are significantly affected by outflows. The colors represent the outflow feedback fractions for the clumps.

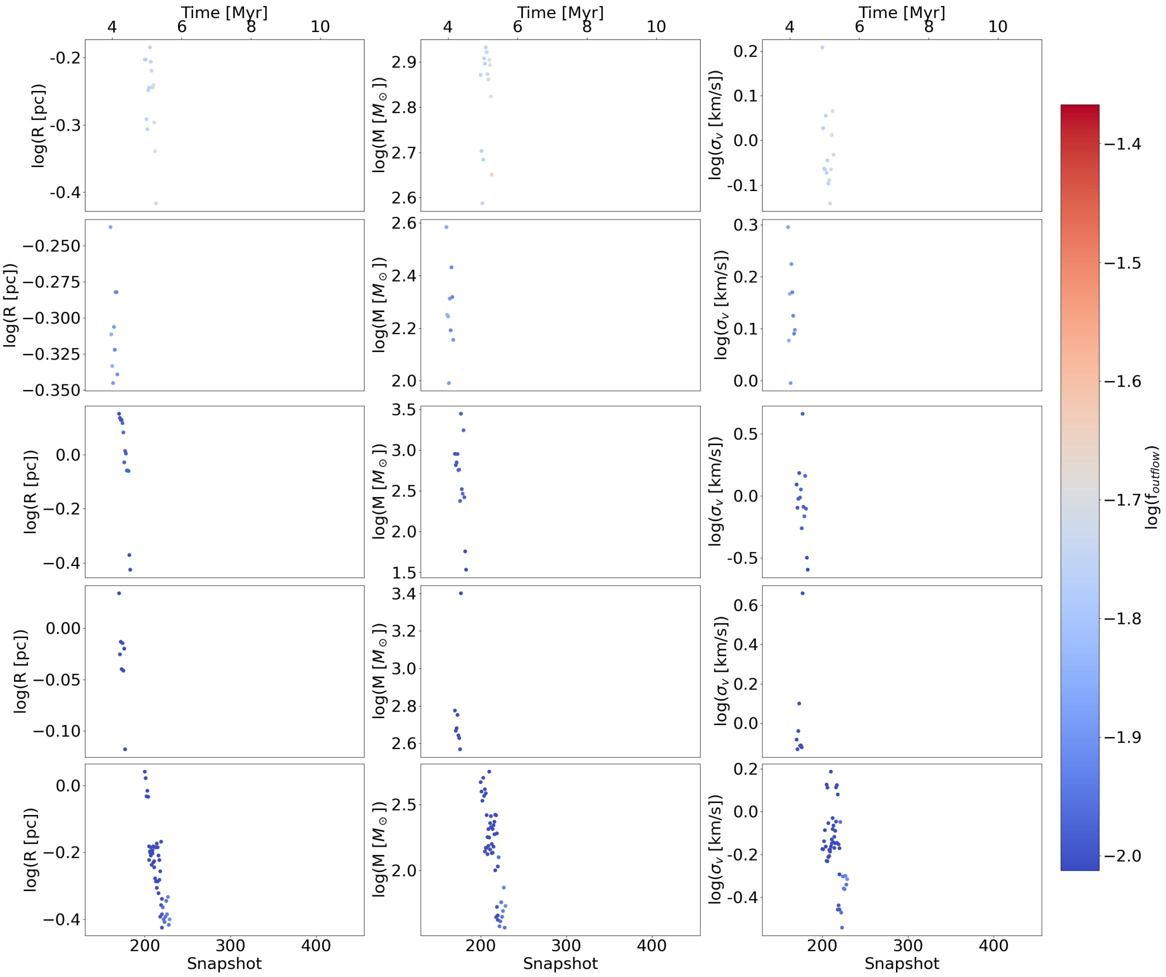


Figure C.12: Evolution of radius, mass and velocity dispersion of the clumps in short living hyperclumps that are marginally affected by outflows. The colors represent the outflow feedback fractions for the clumps.

to disperse or undergo gravitational collapse.

Figure C.13 illustrates the evolution of the properties of the clumps in five hyperclumps that show significant far infrared emission. The FIR emission traces the reprocessed UV radiation absorbed by dust. Therefore, these FIR-bright hyperclumps represent regions that have been strongly affected by radiative feedback in the past. As a result, they contain small and low-mass leftover clumps that are rapidly dispersed by feedback.

Figure C.14 presents the evolution of clump properties in five hyperclumps that show marginal FIR emission. Their evolution can be broadly divided into two categories. In the first three rows of the figure, the clumps do not exhibit FIR emission through the evolution of the hyperclumps. These hyperclumps likely correspond to regions exposed to weak radiation fields, where gas is not significantly dispersed. This enables the clumps to survive longer. The two bottom rows of Figure C.13 show the hyperclumps whose initial clumps have large FIR emission. However, after fragmentation, the resulting fragments do not show a strong association with FIR. This likely reflects the fragments drifting away from a radiation-dominated region and evolving gradually in regions less affected

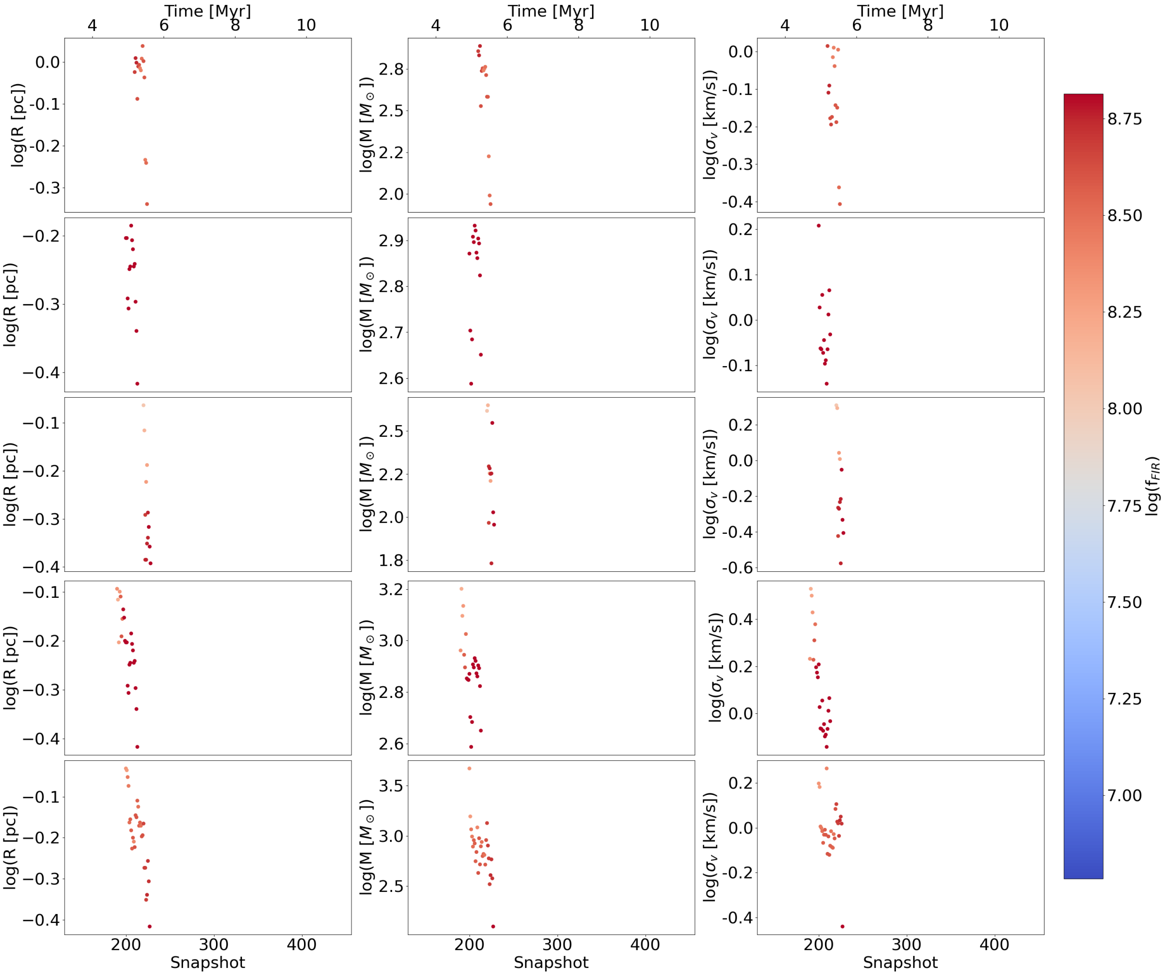


Figure C.13: Evolution of radius, mass and velocity dispersion of the clumps in short living hyperclumps that are associated with significant FIR radiation. The colors represent the FIR feedback fractions for the clumps.

by radiation. As a result, these hyperclumps are long-lived and show lower average FIR feedback fractions.

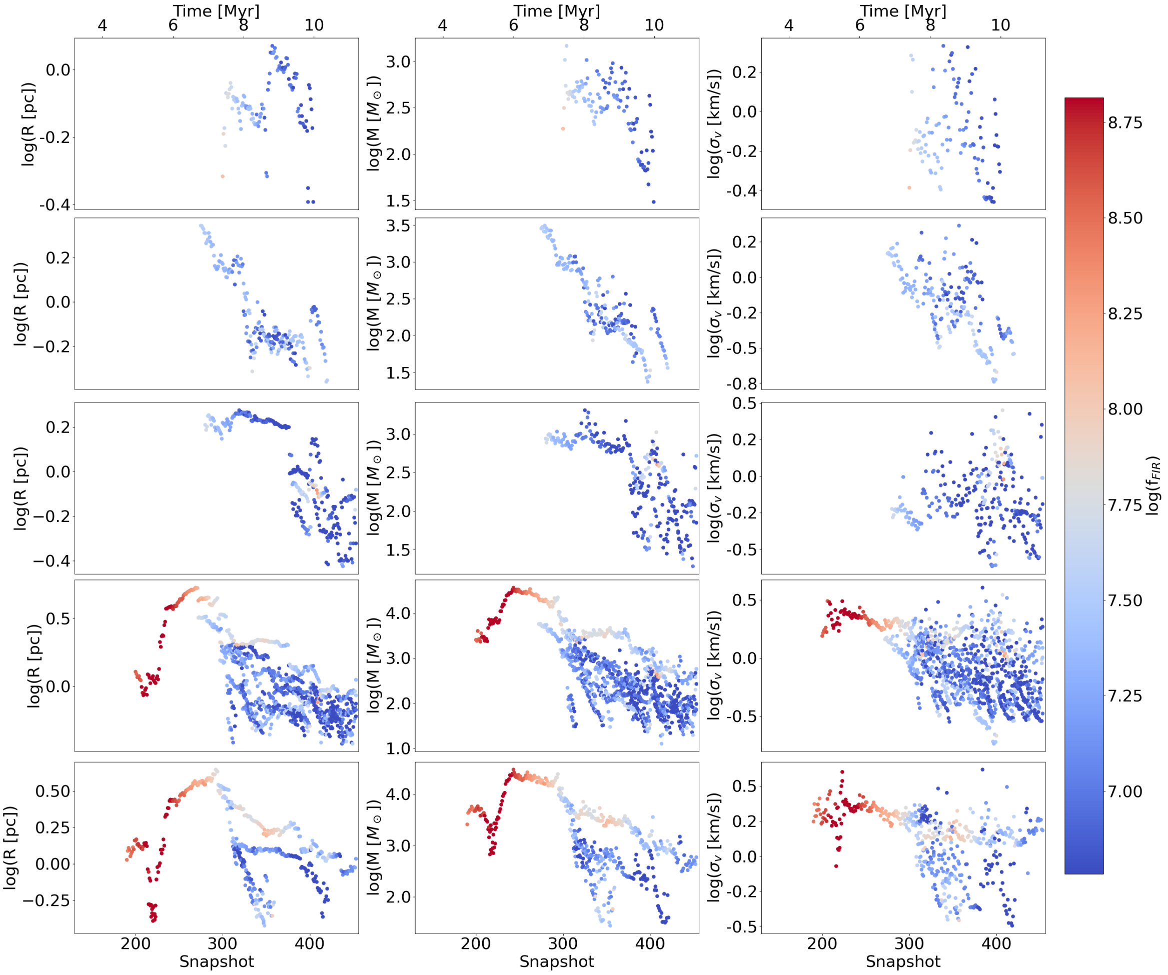


Figure C.14: Evolution of radius, mass and velocity dispersion of the clumps in short living hyperclumps that are associated with marginal FIR emission. The colors represent the FIR feedback fractions for the clumps.

List of Figures

List of Tables

Acknowledgements

While it is always best to believe in oneself, a little help from others can be a great blessing. – Uncle Iroh

A Ph.D. is never a solitary journey, and this has been especially true for me. As I write these acknowledgements in the last days before submission, I reflect on a long academic journey that began not with my PhD, but with my master's degree. Over the years, I have been fortunate to receive support, guidance, and encouragement from many people, without whom this work would not have been possible.

I am deeply grateful to the late Prof. Dr. Karl M. Menten, who gave me the opportunity to pursue my master's thesis and PhD in his group. His passing is a profound loss, and he will always remain in the hearts of his students. I sincerely thank my thesis committee members, Dr. Daniel Seifried, Prof. Pavel Kroupa, Prof. Akaki Rusetsky, and Prof. Matthias Hullin, for their time, valuable feedback, and guidance throughout this process. In particular, I thank Dr. Daniel Seifried for his help in restructuring the thesis to its final form.

I am especially grateful to my supervisor, Dr. Dario Colombo, who supervised me during both my master's and PhD research. I cannot imagine completing my theses or publishing my papers without his unwavering support. His constant guidance and availability throughout the research, his encouragement towards my professional development, and his help during the writing of proposals and applications have been invaluable.

My deepest gratitude goes to my family. Without their unconditional love and support, I would not be where I am today. Thank you, Aai and Baba, for always believing in me; I know you are probably even more excited about my PhD than I am. Thank you, Dada, for being my strength pillar and for giving me the confidence to pursue my goals without fear of failure. Thank you, Vahini, for your fun perspective on life and for always lifting my spirits, and thank you, Prahi, for sharing your wonderful daily stories with me. Thank you Gurudev for your support and guidance.

I would like to thank the OG foosball gang, Mandar, Amith, Aparna, and Raphael, for being there through the ups and downs of my life. Mandar, thank you for being a great roommate and for introducing me to Jeeravan. Amith, thank you for getting me through the labs; I could not have asked for a better lab partner. Aparna, thank you for the interesting and long chatter. Raphael forced me to thank him. Just kidding, thank you for always taking our jokes in such good humour. I thank my badminton partners, Rohit, Hans, Sarwar, and Anahat, for all the fun and for providing a much-needed outlet to de-stress. Iason, Ben, Effi, and Dat, thank you for hosting wonderful parties and board game sessions. Arshia, thank you for helping me with countless things, from sending me the links for IMPRS forms to refining my applications. Although I have not been in touch as closely as I would have liked, I remain grateful to my friends from India; Omkar, Shreya, Samarth, Saurabh, Chandrashekhar, Tapish, and Ziad.

I am thankful to my fellow PhD students and postdocs in the submillimeter group — Miguel, Konstantin, Camilla, Survi, Veselina, Nikolaus, Zsófi, Ivalu, and Agata — for creating a fun work environment. Agata and Miguel, thank you so much for all your help with the papers, thesis, and applications. Konstantin, thank you for all the entertaining discussions and for making the German classes enjoyable. I also thank Dr. Friedrich Wyrowski, Dr. Sudeep Neupane, Dr. Veena V.S., Dr. Wonju Kim, Dr. Carsten König, and Dr. Arnaud Belloche for many helpful discussions throughout my PhD. Sudeep, thank you additionally for your help with Spanish at APEX and for joining me on bike rides. I am grateful to Tuyet-Le Tran, Eva Schmelmer, and Barbara Menten for making all the administrative processes so easy and smooth. Dirk Murder, thank you very much for your support with computing matters, for helping me install various software, for exploring interesting IT solutions at my request, and for the many engaging stories and discussions.

Finally, I thank all my collaborators for making the publication process for multiple papers both smooth and enjoyable. It would take several pages to acknowledge everyone individually, so I mention here those whose contributions were particularly significant. I am especially grateful to Prof. Stella Offner for her support throughout my PhD journey and for consistently going above and beyond to help me. I sincerely thank her for her extensive guidance with STARFORGE and machine learning. I would also like to thank Dr. Ana Duarte-Cabral and Prof. James Urquhart for their valuable contributions to my publications and for many insightful discussions. In particular, I am grateful to Ana for her detailed and thoughtful comments on the publications, which really helped me a lot.

Cover page background image: Projected gas density of the GMC at ~ 6 Myr from STARFORGE M2e4a2 simulation set, representing the primary dataset of this thesis. The image spans a ~ 20 pc region and was generated in Python using matplotlib.