

Incorporating Observational Gas Data into Simulations of Embedded Star
Clusters

INCORPORATING OBSERVATIONAL GAS DATA INTO
SIMULATIONS OF EMBEDDED STAR CLUSTERS

By Anita MATHEWS, B.Sc,

*A Thesis Submitted to the School of Graduate Studies in the Partial
Fulfillment of the Requirements for the Degree Master of Science*

McMaster University © Copyright by Anita MATHEWS, B.Sc
October 13, 2020

McMaster University

Master of Science, (2020)

Hamilton, Ontario (Department of Physics & Astronomy)

TITLE: Incorporating Observational Gas Data into Simulations of Embedded Star
Clusters

AUTHOR: Anita MATHEWS, B.Sc (McMaster University)

SUPERVISOR: Dr. Alison SILLS

NUMBER OF PAGES: xx, 165

Abstract

Realistic star cluster simulations that dynamically evolve embedded star clusters require an accurate treatment of the stars, gas and their interaction. We present and validate a novel technique which creates an initial gas density distribution based on observational gas column density data. We consider two approaches to this technique where the first is based on randomly sampling from the original gas density distribution and the second assigns one particle to represent the gas in each observed image pixel. To create a three-dimensional distribution, we consider two estimates of cloud depth where one is a constant value and the second involves variable depths calculated using image processing techniques based on density features seen in the plane of the sky. We apply these methods to evolve the Carina region using initial stellar positions derived from the MYSTiX catalogue and gas data from the Herschel Hi-GAL survey. We evolve the stars using an N-body code and the gas using a smoothed-particle hydrodynamics code which are coupled through the AMUSE framework. We analyzed our results using dendrograms to describe the gas distribution over time and the DBSCAN clustering algorithm to track the clustering of stars over time. We model the gas using an adiabatic ideal gas equation of state and find that increasing the initial gas velocity dispersion prevents gas from accumulating and therefore could hinder future star formation. We also find that the stars, initially in subclusters spatially (not necessarily bound), tend to merge together to form one large cluster regardless of the initial conditions of the gas. It is only after the subclusters have merged that the initial conditions of the gas start to have a noticeable effect on the structure of the star cluster. Of the two approaches to our novel technique, the second approach leads to more accurate

and realistic results. The second approach also has a significant effect on the stars as the subclusters merge together approximately 1 Myr earlier compared to the first approach. Therefore the choice of initial gas conditions affects the dynamical evolution of star cluster systems and being able to incorporate observational gas data leads to the increasingly accurate dynamical evolution of such systems.

Acknowledgements

I would like to thank my supervisor Dr. Alison Sills for her support and guidance over the last two years. Thank you for your immense patience as I delved into astrophysics research. It has been quite an experience and I am truly grateful for all of your help. Thank you to my parents for all of the support through the years. I would not be where or who I am today if not for you. Thank you to Claude and Jeremy for being wonderful group members and to my officemates for providing a friendly environment. I would also like to thank Jacqueline, Joey, Hector, Dennis, James, Wyatt and Sachin for all of their moral support. Through the laughter and tears, you have been an incredible support to me. I feel extremely lucky to be part of such an amazing group and my experience at McMaster was made all the better because of you. Thank you.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Overview of Star Formation	3
1.2 Modelling Cluster Evolution	6
1.3 The MYSTiX Project	7
1.4 Sub-cluster Identification	9
1.5 The Herschel Hi-GAL Survey	13
1.5.1 The Carina Region	17
1.6 Other Gas Observations	24
1.6.1 Gas Velocity Dispersion Measurements	24
1.7 Thesis Overview	27
2 Simulation Methods	28
2.1 Smoothed-particle Hydrodynamics	28
2.1.1 GADGET-2	31
2.2 N-body Code	34
2.3 AMUSE	36

2.4	Analysis Tools	41
2.5	Astrodendro	41
2.6	DBSCAN	46
2.7	Mass-Segregation Ratios	50
3	Gas Particle Creation Method	54
3.1	Carina Gas Data	56
3.2	Simulated Observational Image	59
3.3	Deriving XY particle positions	60
3.3.1	Method I	64
3.3.2	Method II	67
3.4	Z-depth assignment	67
3.4.1	Constant z depth	68
3.4.2	Thresholding with use of Distance Transform	68
3.5	Internal Energies	72
3.6	Velocities	72
3.7	Validity of Our Methods	73
3.7.1	χ^2	74
3.7.2	1D KS Test	75
3.7.3	Results	76
3.7.4	Histogram Intersection Method	79
3.7.5	Histograms used for the Statistical Tests and Intersection Method	80
3.8	Summary	83
4	Results	86

4.1	Summary of Simulations	87
4.2	Method I Run I	89
4.3	Method I Run II	98
4.4	Method I Run III	103
4.5	Method I Run IV	104
4.6	Method I Run V	105
4.7	Method I Run VI	108
4.8	Method II Run I	112
4.9	Method II Run II	117
4.10	Stellar Distribution Comparisons	126
4.11	Further Comparisons	137
5	Discussion and Conclusion	149
5.1	Incorporation of Observational Gas Data	149
5.2	Effect of Initial Gas Properties on Star Cluster Evolution	150
5.3	Future Work	151

List of Figures

1.1	Ellipse fits to Carina region detailed in Kuhn et al. 2014 with colored stars indicating sub-cluster assignments (20 different clusters indicated with ellipses). Non-clustered stars are light blue and unassigned stars are gray. (© AAS. Reproduced with permission.)	12
1.2	Field of view of Carina region as observed with VISTA (Preibisch et al. 2014), HAWK-I (Preibisch et al. 2011) and Chandra (Townesley et al. 2011a) overlaid on optical image of Carina where the solid line corresponds to the Galactic plane. (Credit: Preibisch et al. 2014 reproduced with permission © ESO).	20
1.3	Carina H_2 number density obtained from processed (Marsh et al. 2017) Herschel Hi-GAL data (Molinari et al. 2010) assuming a line of sight depth of 20 parsecs (value from our collaborators Dr. Klaassen & Dr. Reiter). This follows from dividing the column densities by 20 parsecs resulting in a number of gas molecules per cm^{-3} for each pixel. Logged values are shown.	21
1.4	Carina line of sight mean temperature obtained from processed (Marsh et al. 2017) Herschel Hi-GAL data (Molinari et al. 2010)	22

1.5	Measured stellar positions from MYSTiX project (dark blue) (Feigelson et al. 2013) overlaid on Carina H_2 number density described in Figure 1.3. Positions of Tr 14, 15, 16 and η Carinae are in orange, cyan, yellow and red, respectively.	23
1.6	Total $C_{18}O$ ($J = 1 - 0$) intensity map used to obtain velocity dispersion of massive cores in the Carina region. Contours begin at 0.5 K km/s and increase in steps of 0.5 K km/s. Reproduced from Figure 4 in Yonekura et al. 2005 with permission of the AAS.	25
1.7	Positions of cores overlaid on $C^{18}O$ integrated intensity map. Contours are the same as those in Figure 1.6. Contours begin at 0.5 K km/s and increase in steps of 0.5 K km/s. Reproduced from Figure 5 in Yonekura et al. 2005 with permission of the AAS.	26
2.1	The MUSE framework as described in Portegies-Zwart et al. 2013 and reprinted with permission from Elsevier.	40
2.2	Example of dendrogram structures used to describe a 2D hierarchical distribution. Credit: Astrodendro online documentation (Robitaille et al. 2013).	42
2.3	Example of constructing a dendrogram from a 1D dataset (Steps are from right to left, top to bottom). Credit: Astrodendro online documentation (Robitaille et al. 2013).	44

2.4	2D histogram of the initial gas distribution for the Carina region (specifically for Method I run I defined in Chapter 3). Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$	45
2.5	XY view of identification of clusters in the Carina region using DBSCAN. The initial distribution of stars is seen on the left which consists of 20 smaller clusters in a uniform background distribution of stars. The results from DBSCAN are seen on the right which shows that 16 clusters are identified where each cluster corresponds to a different colour.	48
2.6	XZ view of identification of clusters in the Carina region using DBSCAN. The initial distribution of stars is seen on the left which consists of 20 smaller clusters in a uniform background distribution of stars. The results from DBSCAN are seen on the right which shows that 16 clusters are identified where each cluster corresponds to a different colour.	49
2.7	YZ view of identification of clusters in the Carina region using DBSCAN. The initial distribution of stars is seen on the left which consists of 20 smaller clusters in a uniform background distribution of stars. The results from DBSCAN are seen on the right which shows that 16 clusters are identified where each cluster corresponds to a different colour.	50

3.1	Carina Number Density Map in pixels. Orientation is due to the conversion between the FITS image array to a <i>numpy</i> array we use in Python.	57
3.2	Carina Mean Line of Sight Temperature Map in pixels. Orientation is due to the conversion between the FITS image array to a <i>numpy</i> array in Python.	58
3.3	Simulated Observational Image of a Gas Cloud in pixels where the number of gas molecules per cm^{-3} is shown for each pixel. Data taken from Ward 2015 simulation data.	60
3.4	Distribution of 100,000 particles used to approximate the simulated observational image where the color bar corresponds to the number of particles representing each pixel (simulated image from Ward 2015).	65
3.5	Total mass in each pixel of the simulated image (simulated image from Ward 2015).	66
3.6	Segmentation of the fake data in terms of the non-zero pixel region (LEFT) and the high density region (RIGHT). Regions of interest are in white while the background is in black.	71
3.7	Distance transform of the non-zero region (LEFT) and dense region (RIGHT). Color bar corresponds to distances to the background measured in pixels.	71
3.8	Comparison of XZ mass-weighted distribution (in M_{\odot}) from using the dense transform (LEFT), both non-zero and dense transforms (MIDDLE) and the simulated data set (RIGHT).	71
3.9	1D mass-weighted histograms from simulation data	81

3.10	Examples of histogram intersections in X which are calculated to be the percent overlap between the simulation data (BLUE), and the results from our methods (RED) where a larger overlap is desirable. The comparisons of the mass-weighted X positions derived from Method I (left) and from Method II (right) are shown.	81
3.11	Mass-weighted Z histogram intersections which are calculated to be the percent overlap between the simulation data (BLUE), and the results from our methods (RED) where a larger overlap is desirable. The results from a constant 20 pc depth (left), the use of the dense distance transform (middle) and the use of the non-zero and dense distance transforms (right) are shown.	82
3.12	Mass-weighted (in M_{\odot}) Cartesian XY (left) and XZ (right) of particles created using Method I to represent gas densities in the observed Carina region. Gas particles range in mass from roughly 0.07 to 15 M_{\odot} with most having masses of roughly 0.5 M_{\odot} . (Note: 1000 bins in each of the dimensions were used.)	84
3.13	Mass-weighted (in M_{\odot}) Cartesian XY (left) and XZ (right) of particles created using Method II to represent gas densities in the observed Carina region. Maximum mass of a gas particle is 0.5 M_{\odot} . (Note: 1000 bins in each of the dimensions were used.)	85
3.14	Distance transform of the non-zero region (LEFT) and dense region (greater than 100 cm^{-3}) (RIGHT). Color bar corresponds to distances to the background measured in pixels.	85

4.1	Evolution of Carina Region in XY plane with gas velocity dispersion of 1.4 km/s.	94
4.2	Evolution of Carina Region in XZ plane with gas velocity dispersion of 1.4 km/s.	95
4.3	Gas distribution in Carina region over time with initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities between 100 - 1000 M_{\odot}/pc^2	96
4.4	Clusters in Carina picked out by DBSCAN over time with gas velocity dispersion of 1.4 km/s. Each cluster is denoted by a different colour. The colours do not necessarily correspond to the same stars across the images as they are only used to highlight the different clusters identified.	97
4.5	Evolution of Carina Region in XY plane with gas initially in a uniform sphere with a gas velocity dispersion of 1.4 km/s	100
4.6	Evolution of Carina Region in XZ plane with gas initially distributed in a uniform sphere and with a gas velocity dispersion of 1.4 km/s.	101
4.7	Gas distribution in Carina region over time with gas initially in a uniform sphere with initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities between 100 - 1000 M_{\odot}/pc^2	102

4.8	Comparison of gas distribution in Carina region between runs IV (left) and V (right) at $t = 0$ Myr which highlights the effect of random sampling in our particle creation technique (Method I). Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$	106
4.9	Comparison of gas distribution in Carina region over time between runs I (left), IV (middle) and V (right) between $t = 5$ Myr to $t = 7.7$ Myr. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$	107
4.10	Gas distribution in Carina region over time with initial gas velocity dispersion of 5 km/s and new sampling. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$	110
4.11	Comparison of gas distribution in Carina region over time between runs I, V and VI between $t = 2$ Myr to $t = 7.7$ Myr. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$	111

4.12	Evolution of Carina Region in XY plane using Method II with a varied Z depth and an initial gas velocity dispersion of 1.4 km/s. . .	114
4.13	Evolution of Carina Region in XZ plane using Method II with a varied Z depth and an initial gas velocity dispersion of 1.4 km/s. . .	115
4.14	Gas distribution in Carina region over time using Method II with a varied Z depth, and an initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities greater than 100 M_{\odot}/pc^2	115
4.15	Gas distribution in Carina region over time using Method II with a varied Z depth, and an initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram. Cyan corresponds to densities between 100 - 1000 M_{\odot}/pc^2 , yellow corresponds to densities between 1000 - 5000 M_{\odot}/pc^2 and red corresponds to densities between 5000 - 10000 M_{\odot}/pc^2	116
4.16	Evolution of Carina Region in XY plane using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep.	120
4.17	Evolution of Carina Region in XZ plane using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep.	121

4.18 Gas distribution in Carina region over time using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep. Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities greater than 100 M_{\odot}/pc^2	122
4.19 Comparison of gas distribution in Carina region over time between Method I (MI) runs I, II and Method II (MII) run II between $t = 0$ Myr to $t = 4$ Myr. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities greater than 100 M_{\odot}/pc^2	123
4.20 Gas distribution in Carina region over time using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep. Overplotted are contours from branches and leaves in the dendrogram. Cyan corresponds to densities between 100 - 1000 M_{\odot}/pc^2 , yellow corresponds to densities between 1000 - 5000 M_{\odot}/pc^2 and red corresponds to densities between 5000 - 10000 M_{\odot}/pc^2	124

4.21	Clusters in Carina picked out by DBSCAN over time using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep. Each cluster is denoted by a different colour. The colours do not necessarily correspond to the same stars across the images as they are only used to highlight the different clusters identified.	125
4.22	$N_{clusters}$ over time as identified by DBSCAN for all runs. Run details can be seen in Table 4.1.	127
4.23	DBSCAN results for Method I runs I-VI at $t = 6$ Myr. Each colour corresponds to a different cluster. Colours do not necessarily correspond to the same stars between the images.	128
4.24	DBSCAN results for Method I runs I-VI at $t = 7$ Myr. Each colour corresponds to a different cluster. Colours do not necessarily correspond to the same stars between the images.	129
4.25	DBSCAN results for Method I runs I-VI at $t = 7.7$ Myr. Each colour corresponds to a different cluster. Colours do not necessarily correspond to the same stars between the images.	130
4.26	Clusters picked out by DBSCAN for Method I Run I with most mass-segregated clusters in black.	134
4.27	Clusters picked out by DBSCAN for Method II Run II with most mass-segregated clusters in black.	135
4.28	Max Λ_{msr} over time for clusters identified by DBSCAN for all runs. Run details can be seen in Table 4.1.	136
4.29	Kinetic and potential energies of the stars and gas over time for Method I runs I-VI.	142

4.30	Kinetic and potential energies of the stars and gas over time for Method I and Method II runs.	143
4.31	LEFT: Bound gas mass over time for Method I runs I-VI. RIGHT: Bound stellar mass over time for Method I runs I-VI.	144
4.32	LEFT: Gas Lagrange radii for Method I and Method II runs. RIGHT: Stellar Lagrange radii for Method I and Method II runs.	148
4.33	LEFT: Average gas temperature for Method I runs. RIGHT: Average gas temperature for Method I and Method II runs.	148

List of Tables

3.1	Summary of tests with different particle techniques along different dimensions.	77
3.2	Results of χ^2 and KS tests.	77
3.3	Results of Histogram Intersection Method.	80
4.1	Summary of runs.	88
4.2	Λ_{msr} values for clusters picked out by DBSCAN.	133

Chapter 1

Introduction

The oldest star clusters are known as globular clusters which are spherical, devoid of gas and invaluable tools for tracing the formation history of galaxies. These clusters have evolved from young gas-rich star clusters, the studies of which may provide insight as to the early stages of globular clusters. Stars are formed in dense regions embedded in molecular clouds with only approximately 10% of stars remaining in bound clusters after 10 Myr (Lada and Lada 2003). Several questions remain about the evolution of these clusters in terms of the loss of stellar substructure (clumps of stars) and gas over time. It is expected that relatively young clusters would show stellar substructure as opposed to those that are relatively old and have had time to relax. Young clusters (1 Myr) have been observed with centrally concentrated stellar distributions (e.g the Orion Nebula Complex, Hillenbrand and Hartmann 1998) suggesting the erasure of substructure on short timescales and simulations have confirmed this can occur on timescales of 1-2 Myr (Sclally and Clarke 2002; Goodwin and Whitworth 2003; Allison et al. 2010a). However, observations show that substructure is present in some clusters that are

roughly 30 Myr (Cartwright and Whitworth 2004). Therefore, there does not seem to be a clear relationship between the ages of star clusters and the loss of stellar substructure. As well, the interaction between the stars and gas in these clusters is an important factor to consider in the evolution of these systems. It is thought that star clusters end up being destroyed (or unbound) within tens of millions of years due to gas expulsion where the gas is lost due to various stellar feedback processes and that the degree to which clusters end up unbound is dependent on the star formation efficiency which is the fraction of gas that is converted into stars (Hills 1980; De Grijs and Goodwin 2008; Baumgardt and Kroupa 2007; Dinnbier and Kroupa 2020 and references therein). Therefore, the evolution of young clusters is complex and we wish to study this process in more detail by simulating the dynamics of the stars, gas and their interactions. To this end, it is imperative to incorporate both stars and gas into simulations in a meaningful and accurate manner. We aim to simulate realistic star-forming regions using initial conditions based on observational data with a focus on specifically incorporating observational gas data. Our approach allows for a wide range in initial gas density distributions which can include complex morphologies like clumps or filamentary structures depending on the observational data.

This is an improvement over other star cluster simulations (e.g Federrath et al. 2010; Sills et al. 2018) that use simplifying assumptions to model the gas distribution which includes using spheres of gas or gas that closely follows the stellar distribution. This is unrealistic and does not take the specific gas morphology of a region into account which may include filamentary structures or clumps of gas. Our work is based on that done in Sills et al. 2018 where they simulated different

star-forming regions using measured stellar positions and a gas distribution similar to that of the stars. In this thesis, we present and validate a novel approach to incorporate observed gas column density data into a star cluster simulation thereby creating a realistic simulation informed by observations of both stars and gas. In this way, we open the door to modeling more complex gas morphologies that are grounded in observations which has not been done before. Although a method exists where gray-scale images of gas clouds can be processed into initial conditions for such simulations, it uses the brightness of the pixels as a proxy for the density (Arth et al. 2019) while our approach makes direct use of the gas column density data. Our current focus is on simulating the Carina region although our technique to include observational gas data is general enough to model other embedded star-forming regions as well.

In this chapter, we begin with an overview of the star formation process, followed by an explanation of the MYSTiX project and the Herschel Hi-GAL survey from which we obtain our initial conditions to model the Carina region.

1.1 Overview of Star Formation

Stars are formed in dense regions embedded in molecular clouds. These clouds are $10^3 - 10^7 M_{\odot}$ in mass and predominantly contain molecular hydrogen (roughly 70% H_2 and 30% He) (Elmegreen and Falgarone 1996). Although H_2 is abundant, it is extremely difficult to observe directly in star-forming regions which are relatively cool ($T = 10K$). The lowest radiative transition in H_2 is a weak quadrupole transition that occurs at a temperature of 510K which is much warmer than the cold

star-forming regions (Krumholz 2015). As a result, different proxies are used to trace the H_2 distribution in these clouds. One tracer is CO which is used because it is the most abundant molecule in the interstellar medium after H_2 and it can undergo transitions at low temperatures making it observable at the temperatures in the star-forming regions (Krumholz 2015). Another tracer used is the thermal radiation emitted from dust grains in the clouds. The intensity of this emission along with the temperature of the gas can be used to infer column densities of H_2 in the observed region which are often recorded as the number of H_2 molecules per cm^2 along the line of sight. An example of how the dust properties can be used to determine column density values is described in section 1.5 where we discuss the observational gas data that we use in our simulations.

These clouds have fractal morphologies (Elmegreen and Falgarone 1996) and undergo hierarchical collapse when they become gravitationally unstable. In order for collapse to occur, the gravitational potential energy of the cloud must be greater than the kinetic energy which supports the cloud from collapse. The virial parameter characterizes the collapse or expansion of a cloud assuming it is homogeneous and spherical and this is given by:

$$\alpha = \frac{5\sigma_v^2 R}{GM} \tag{1.1}$$

where σ_v is the 1D velocity dispersion, R is the cloud radius and M is the cloud gas mass. α values of less than 1 indicate that a cloud is collapsing while values greater than 1 indicate expansion of the cloud. α of exactly 1 indicates that the internal pressure or turbulence of the cloud sufficiently balances out gravity.

Bertoldi and McKee 1992 show that the virial parameter is also equal to:

$$\alpha = a2 \frac{E_{kin}}{|E_{pot}|} \quad (1.2)$$

where a is a parameter that takes deviations from spherical symmetry and different density gradients into account. However, Kauffmann et al. 2013 mention that this parameter may not be applicable to describe regions that lack symmetry and are not centrally condensed.

If the gravitational potential energy of the cloud is greater than its kinetic energy, collapse occurs which amplifies anisotropies present in the cloud creating filamentary structures (Lin et al. 1965). Dense cores form at the intersections of these filaments which ultimately form stars (Schneider et al. 2010; Chen et al. 2019). Due to the fragmentation of the initial cloud, collapse can occur in several regions at once giving rise to star formation that occurs in groups which includes clusters and associations (Lada and Lada 2003; Kruijssen 2012). The fraction of gas that is converted to stars, or star-formation efficiency, per free-fall time is observed to be roughly 1% in giant molecular clouds (Krumholz et al. 2019b).

Stars in the early stages of formation, referred to as protostars, are embedded in gas and dust which is optically thick in the visible to near IR wavelengths. As a result, direct observations of the stellar emissions are not possible as they are absorbed by the gas and dust in the surrounding area. Instead, the energy from the protostars is re-radiated by the surrounding dust at longer IR wavelengths. As well, young stars without surrounding dust or disks can be observed in the

X-ray wavelengths. Combining observations of young embedded clusters in different wavelengths allows for a more complete census to be obtained which is the motivation behind the MYSTiX project (Feigelson et al. 2013) and we use stellar positions derived from this (specifically of the Carina region) in our simulations to create a realistic initial stellar distribution.

1.2 Modelling Cluster Evolution

Star formation models usually begin with the collapse of a sphere of gas that creates filaments where stars are formed in dense regions (e.g Federrath et al. 2010). To track the dynamical evolution of the stars, N-body codes are often used which take the gravitational effects between the stars into account (e.g Fujii and Portegies Zwart 2016). However, in order to understand cluster evolution, it is imperative to model both the stars and gas in a self-consistent and accurate manner. This means incorporating gas dynamics, N-body dynamics and the gravitational effects between the stars and gas. Our simulations are constructed with this type of setup and are based on the work done in Sills et al. 2018. Our current focus is modelling the evolution of the Carina region and we use realistic initial conditions for both the stars and gas. As mentioned, similar models usually employ a spherical distribution of gas which is not realistic to model a young embedded stellar system considering the many feedback processes that shape the gas distribution. Examples of feedback processes include proto-stellar outflows, photoionization and supernovae explosions. Proto-stellar outflows eject material into the surrounding region and have been observed in several star-forming regions (Frank et al. 2014).

Photoionization feedback refers to UV photons emitted from young massive stars that ionize the surrounding neutral gas. This creates bubbles of ionized hydrogen (HII) which can reach temperatures of approximately 10^4 K (Sales et al. 2014, Krumholz et al. 2019a). The temperature difference between the HII regions and the surrounding regions creates a pressure difference which results in the movement of the gas towards the lower pressure areas (Shu 1992). Supernovae explosions occur when a massive star undergoes core-collapse during the late stages of its life as iron forms in the core from fusion processes or when a white dwarf in a binary system accretes enough material to pass the Chandrasekhar limit (which defines the mass of a stable white dwarf) and can no longer resist gravitational collapse. These feedback processes shape the gas in star-forming regions and our aim is to model this complex gas distribution. We present a novel approach to include observational gas data which can be seen in Chapter 3. In our simulations, we do not include stellar feedback processes as our stars are low-mass and we consider the evolution of the cluster on short timescales (up to 10 Myr) before supernovae feedback begins to be significant. To model the Carina region, we use stellar positions derived from the MYSTiX project and gas column density data from the Herschel Hi-GAL survey which are detailed in the following sections.

1.3 The MYSTiX Project

The MYSTiX project (Feigelson et al. 2013) aims to further our understanding of star formation in massive star-forming regions (MSFRs). Particular questions for this survey include those that pertain to the conditions required in molecular

clouds for rich cluster formation, massive OB star-formation as well as how mass-segregation (massive stars tending to be found near the center) develops in clusters. This project was created to advance our knowledge in this area by forming a high-quality dataset of stellar populations in massive star-forming regions. It consists of 31,754 MYSTiX probable complex members in 20 nearby (distances less than approximately 4 kpc) MSFRs that have been compiled based on archival data from the Chandra X-ray Observatory, the United Kingdom Infrared Telescope (UKIRT), the Two Micron All Sky Survey (2MASS) and NASA's Spitzer Space Telescope. Examples of these massive star-forming regions include the Orion Nebula, DR21 and the Carina Nebula.

The combination of observations in the different wavelengths allows for a more complete census of cluster members. Determining stellar membership in massive clusters is difficult as issues can arise if observations are limited to the optical or IR wavelengths. Specifically, variable nebular line emission from heated gas or dust in HII regions, obscuration of molecular clouds and contamination from Galactic field stars can limit the resulting observations to OB stars that are not heavily obscured and young stars with proto-planetary disks. Datasets based solely on optical or IR observations therefore do not include the young, disk-free, lower mass stars known to be predominant in young clusters (Feigelson et al. 2013). X-rays observations of stars, specifically with NASA's Chandra X-ray Observatory, can be obtained despite heavy obscuration (Feigelson 2010), can identify the source of emissions (e.g pre-main sequence stars producing flares) and can reduce field contamination. As well, X-rays observations are well suited for identifying disk-free stars (stars with disks can be identified as well but to a lesser degree) (Getman et al. 2009;

Stelzer et al. 2012). Thus, combining the archival data is advantageous to creating a more complete census of young stellar populations which includes OB stars, pre-main sequence stars and young stars with and without proto-planetary disks. The MYSTiX catalogue provides positions for these stars that have masses greater than approximately $0.83 M_{\odot}$.

To identify sub-clusters within the observed regions, Kuhn et al. 2014 processed the MYSTiX data using isothermal ellipsoid fits. They then grouped the stellar positions based on the cluster identification. Our current focus is on simulating the Carina region and we use the stellar data that has been processed in this way for our initial conditions.

1.4 Sub-cluster Identification

Kuhn et al. 2014 identified sub-clusters of stars using finite mixture models (McLachlan and Peel 2000) and isothermal ellipsoids. Their model includes an approximately constant core density that then drops off as a power law which they find to be generally representative of young stellar clusters. They begin with an isothermal sphere surface density profile given by:

$$\Sigma(r) = \frac{\Sigma_0}{1 + (r/r_c)^2} \quad (1.3)$$

where Σ is the stellar surface density at a distance r from the cluster center, Σ_0 is the central stellar surface density and r_c is the core radius.

This density profile was then stretched to form an elliptical model which is more general and can represent young clusters which have been observed to have elliptical surface density profiles (e.g the Orion Nebula Complex described in Hillenbrand and Hartmann 1998).

The stretched profile is:

$$\Sigma_{elliptical} = \Sigma_0 \left[1 + \frac{\left| \begin{pmatrix} (1-\epsilon)^{-1/2} & 0 \\ 0 & (1-\epsilon)^{-1/2} \end{pmatrix} \hat{R}(\phi)(\mathbf{r} - \mathbf{r}_0) \right|^2}{r_c^2} \right]^{-1} \quad (1.4)$$

where $\mathbf{r} - \mathbf{r}_0 = (x - x_0, y - y_0)$, x_0 and y_0 are the central right ascension and declination, respectively, ϕ is the ellipse orientation, ϵ is the ellipticity and $\hat{R}(\phi)$ rotates $\mathbf{r} - \mathbf{r}_0$ by ϕ .

The finite mixture model includes an ellipsoid model for each cluster along with a flat density profile to characterize the unclustered stars. The best-fit model is determined using a maximum-likelihood estimation process to determine the best number of clusters to characterize a region, the details of which can be found in Kuhn et al. 2014.

The sub-clusters identified were further classified into one of four morphologies: long chains, clumpy structures, isolated clusters with a core-halo structure and isolated clusters well fit with a single isothermal ellipsoid. The Carina region in

particular contains 20 sub-clusters most of which have a chain structure while two have a core-halo structure. It should be noted that the sub-clusters that are found do not necessarily correspond precisely to known clusters. For example, the Trumpler 14 cluster is classified into two sub-clusters (B and C) (Kuhn et al. 2015). The ellipse fits can be seen in Figure 1.1.

In addition to the measured stellar positions, the ellipse parameters that characterize each of the sub-clusters are used to generate the initial stellar positions in our simulations. The right ascension and declination of the observed stars in each subcluster belonging to the Carina region in the MYSTiX catalogue are transformed into Cartesian X and Y coordinates. The Z coordinates (along the line of sight) of the stars are then drawn from a range equivalent to the length of the minor axis of the ellipse fits. We follow the same process outlined in Sills et al. 2018 to create the initial stellar conditions which includes a population of low mass stars (estimated from the number of observed stars) and a background population for which the Z coordinate range is the extent of all the sub-clusters in the XY plane.

In Sills et al. 2018, the DR21, M17 and NGC 6357 star-forming regions were simulated using stellar positions derived from the MYSTiX catalogue and gas distributions that were made to closely follow the stellar distribution. Here, we aim to include increasingly realistic initial gas conditions by developing a new method to incorporate observational gas specifically applying this to observations of the Carina region obtained with the Herschel Hi-GAL Survey.

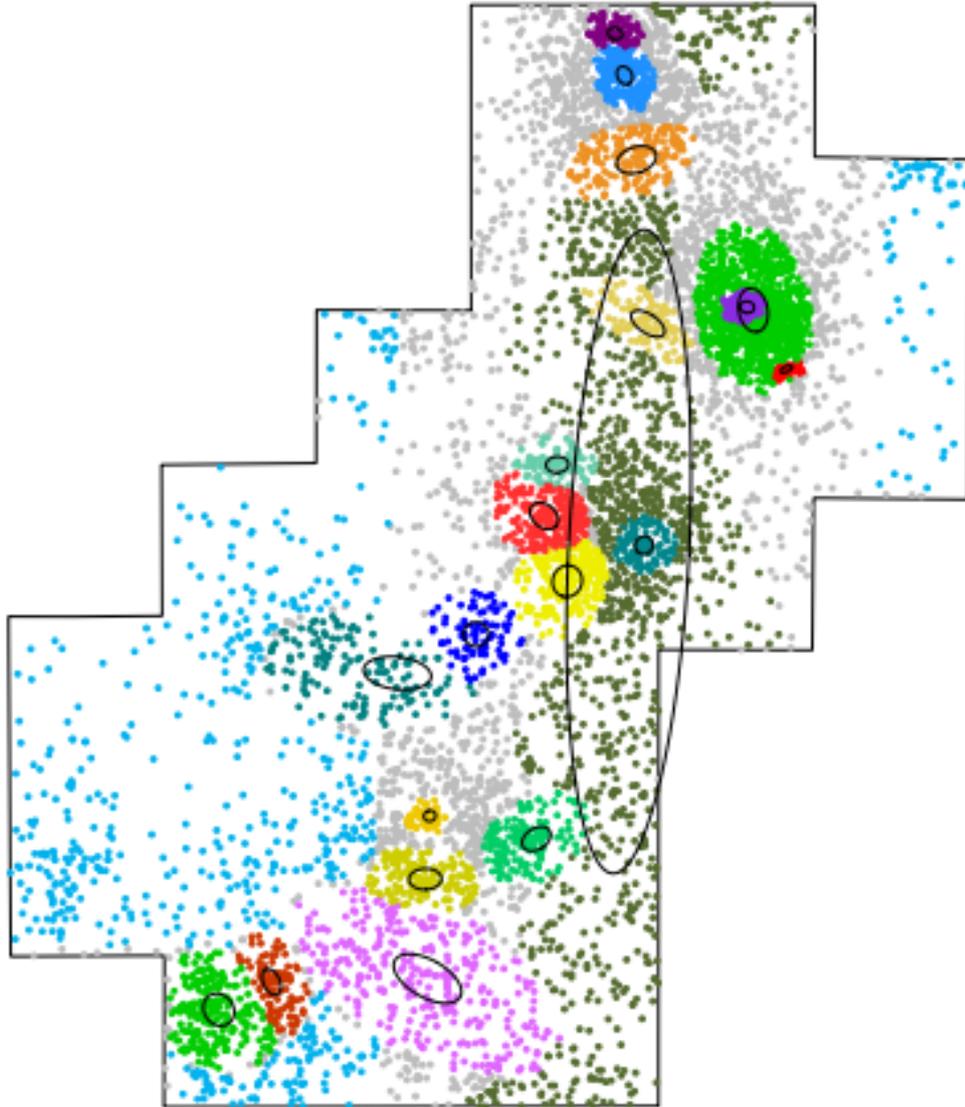


FIGURE 1.1: Ellipse fits to Carina region detailed in Kuhn et al. 2014 with colored stars indicating sub-cluster assignments (20 different clusters indicated with ellipses). Non-clustered stars are light blue and unassigned stars are gray. (© AAS. Reproduced with permission.)

1.5 The Herschel Hi-GAL Survey

The Herschel Hi-GAL Survey (Molinari et al. 2010) was initially conducted to detect the earliest phases of molecular cloud formation and the formation of high mass stars. Far infrared radiation, produced by large grains (15 - 110 μm), that traces the interstellar medium was observed in order to derive dust temperature measurements and create dust continuum maps of the entire Galactic plane in the wavelength range of 70-500 μm .

These maps are useful to describe and characterize molecular clouds containing young clusters which are often still embedded. Stellar feedback processes including stellar winds and supernovae act to disperse gas from the cluster and disrupt the surrounding cloud. For this reason, it is thought that molecular clouds themselves are transient structures which are likely dispersed after approximately 10^7 years (Larson 2003). As molecular clouds undergo gravitational collapse, filaments form which are elongated structures of gas in the interstellar medium. Initially observed in the far-IR with the IRAS survey (Low et al. 1984), filaments have also been observed in HI (McClure-Griffiths et al. 2006) and CO (Bally et al. 1987; Goldsmith et al. 2008).

The Herschel Space Observatory provided increased sensitivity and spatial resolution in galactic imaging surveys which not only showed more observational evidence for filaments but that they are ubiquitous in cold regions of the interstellar medium and are present in star forming regions (André et al. 2010). As will be seen in Chapter 3, we explore ways of capturing this filamentary nature of molecular clouds as we incorporate the observational gas data.

Our simulations use the Herschel Hi-Gal maps of the Carina region that were processed using the PPMAP algorithm (Marsh et al. 2017) which was designed to create accurate column density maps with a high spatial resolution of 12 arcsec. The standard method for obtaining molecular gas column densities and dust temperatures involves fitting spectral energy distributions to the raw IR emission data which has been done with Herschel Hi-Gal images before. For example, Rebolledo et al. 2016 first degraded Herschel images of the Carina Nebula to the 500 μm image which corresponds to a 36.5 arcsec resolution. They then fit a grey-body function (Burton et al. 2004) given by equation 1.5 to each pixel with the dust temperature (T_{dust}) and mass column density n_{dust} (units of g cm^{-2} , may be referred to as Σ_{dust} elsewhere) as free parameters where:

$$F_\nu = \Omega B_\nu(T_{dust}) \kappa_0 \left(\frac{\nu}{\nu_0}\right)^\beta n_{dust} \quad (1.5)$$

where F is the flux, B_ν is Planck's law of black-body radiation, Ω is the angular size of the observed region, β is the dust emissivity index and κ_0 is the dust opacity at a reference frequency of ν_0 and the gas column density is given by:

$$N_{H_2}(dust) = \frac{n_{dust}}{\mu_{H_2} m_H R_{dg}} \quad (1.6)$$

where μ_{H_2} is the mean molecular weight in units of the hydrogen atom mass (2.72), m_H is the mass of hydrogen and R_{dg} is the dust-to-gas ratio (0.01) assuming that the all the gas traced by the dust is in molecular form (see Rebolledo et al. 2016 for more details).

The restrictions of this fitting method include the loss of resolution as all the images are degraded to the 500 μm image and the assumption of a uniform dust temperature along the line of sight which may not be accurate.

In contrast, the PPMAP (point process mapping) algorithm does not require the images to be smoothed as it is able to take the point spread functions (describes the intensity of a point-like object that will appear extended due to diffraction by the telescope aperture) of the telescopes into account. After processing, a higher spatial resolution of 12 arcsec is obtained which can resolve structures in the gas between 0.06 and 0.6 pc in size at distances between 1 and 10 kpc in the Herschel Hi-GAL survey. In addition, this method does not assume a uniform dust temperature along the line of sight which allows temperature-differential column densities to be obtained.

The PPMAP algorithm first represents an astrophysical structure (e.g filament or molecular cloud) as a set of components each of unit column density and parametrized by two spatial coordinates (x, y) and a dust temperature which can be represented as a point in a three-dimensional state space. This state space is divided into a grid of N_{st} cells corresponding to the number of states and the column densities (as a function of (x, y) and T_{dust}) end up being the occupation numbers of these cells. The column densities (or state of the system) is defined as $\mathbf{\Gamma}$ which is estimated given the data \mathbf{d} . The relationship between the two is as follows:

$$\mathbf{d} = \mathbf{A}\mathbf{\Gamma} + \mu \tag{1.7}$$

where μ is the measurement noise and \mathbf{A} is the system response matrix. \mathbf{A} is a function of the point spread functions, the Planck function, the solid angle of the observed area and a dust opacity law defined as:

$$\kappa(\lambda) = 0.1 \text{cm}^2 \text{g}^{-1} \left(\frac{\lambda}{300 \mu\text{m}} \right)^{-2} \quad (1.8)$$

To estimate the state of the system, the average Γ is obtained which is given by:

$$\rho(\mathbf{z}_n | \mathbf{d}) = \sum_{\Gamma} \Gamma_n P(\Gamma | \mathbf{d}) \quad (1.9)$$

where z_n is a vector of the (x, y, T_{dust}) coordinates of the n^{th} cell. ρ is essentially the differential column density which is solved for (details in Marsh et al. 2015) and then mapped back into the (x, y, T_{dust}) coordinates.

PPMAP takes as input a set of dust continuum images, their associated point spread functions, an assumed dust opacity law and a grid of temperatures at which the differential column density will be estimated. 12 temperatures are used which are equally spaced between logged values ranging from 8-50 K which is the expected range based on the spectral energy distributions. The output from this algorithm is an image cube of the temperature differential column density of gas and dust in units of $10^{20} H_2$ molecules per cm^2 per Kelvin which can then be used to obtain a 2D mean line of sight dust temperature map and a column density map integrated in temperature. The temperature integrated map was then divided by a value of 20 pc in order to obtain a number density map in units of H_2 per cm^3 . This value

of 20 pc was selected by our collaborators Dr. Klaassen and Dr. Reiter as an appropriate value in between a filament width of 0.1 pc and a low density cloud of 100 pc. Depths of clouds can be determined using stellar reddening relations along the line of sight (Lallement et al. 2014). However at this time, we do not have such measurements. As a result, the Z depth of the cloud is an estimate.

The PPMAP technique was applied to all 163 tiles (each tile corresponds to $2.4^\circ \times 2.4^\circ$ field of view) of the Herschel Hi-Gal survey (Marsh et al. 2017) and the resulting maps can be found on the publicly available ViaLactea (an EU funded project) database (*ViaLactea n.d.*). We use maps of the Carina region (specifically the number density map and the 2D mean line of sight dust temperature map) that have been processed with PPMAP as input in our simulations to create a realistic initial gas distribution.

1.5.1 The Carina Region

Our current focus is on the Carina region which is a large star-forming complex approximately 2.3 kiloparsecs away (Smith et al. 2006). It is one of the largest HII regions observed having been first described in 1847 (Herschel 1847). Smith et al. 2000 showed that there are large dense clouds extending over 50 parsecs around Trumpler 14, 15 and 16, which are notable star clusters in this region. Pillar structures in the gas and dust have been observed (Cortes-Rangel et al. 2020, Klaassen et al. 2020) while estimates of the total gas and dust mass in this region is on the order of $650,000 M_\odot$. Star-formation is thought to have started about 10 Myr ago with Trumpler 15 forming 6.5 Myr ago and Trumpler 14 and 16 forming

2-3 Myr ago (DeGioia-Eastwood et al. 2001) allowing a significant amount of time for the clouds to become disrupted due to massive stellar feedback. For example, this region is known for the energetic eruptions of the η Carinae binary system which consists of a luminous blue variable star and an evolved main sequence star (Humphreys and Martin 2012). In addition, most of the massive stars in Trumpler 15 have undergone supernovae explosions (Wang et al. 2011). Such events shape the surrounding material giving rise to different structures like the Homunculus nebula (Kashi and Soker 2010) which surrounds the η Carinae binary system.

Although the Carina region is a somewhat evolved star-forming region in this regard, there is evidence of ongoing star-formation as well. This is seen in the high levels of UV radiation on the clouds in the central part of the region which is similar to the radiation observed in extragalactic star-burst regions (Roccatagliata et al. 2013). Furthermore, the feedback from the massive stars contributes to the compression of the surrounding dense clouds which appears to be resulting in new star-formation (Gaczkowski et al. 2013). Therefore this region is complex in terms of the gas distribution and the star-formation history.

The Carina Nebula has been observed in several surveys. The stellar distributions in the Carina region have been extensively observed in the x-ray, near-infrared and mid-infrared wavelengths. X-ray observations obtained with the Chandra telescope show over 14,000 point sources along with diffuse emission (Townsley et al. 2011b). Similar observations in the near-infrared have been obtained using the HAWK-I instrument on the VLT (Preibisch et al. 2011) and in the mid-infrared using the IRAC instrument on the Spitzer telescope (Povich et al. 2011). The field of view of the VISTA (Preibisch et al. 2014), HAWK-I and Chandra observations

overlaid on an optical image of the Carina region can be seen in Figure 1.2.

The MYSTiX project combines the HAWK-I and Chandra observations based on the overlap in the observed regions. This overlap therefore corresponds to the region covered by HAWK-I which can be seen in Figure 1.2. As well, the gas observations we use in our simulations from the Herschel Hi-GAL survey overlap with the stellar distribution observed in the HAWK-I region. Figure 1.3 shows the observed H_2 number density and Figure 1.4 shows the mean line of sight temperature distribution. Figure 1.5 shows the observed H_2 number density along with the observed stars from the MYSTiX project. The location of η Carinae and the centers of Trumpler 14, 15 and 16 are indicated in this image as well.

We are interested in simulating the dynamical evolution of this region specifically where the stars and gas overlap (Note: Our simulations model only the overlapping region in a 40 parsec radius centered on the stellar distribution and we assume this region to be isolated when in reality the surrounding gas will have an effect). As seen in Figure 1.5, the gas distribution around the stars appears to be somewhat elongated and filamentary and does not envelop the stars in a symmetric way. Therefore to accurately simulate the complex gas and dust distribution of such regions we must go beyond simplifying assumptions and develop a new method to incorporate observational gas data.

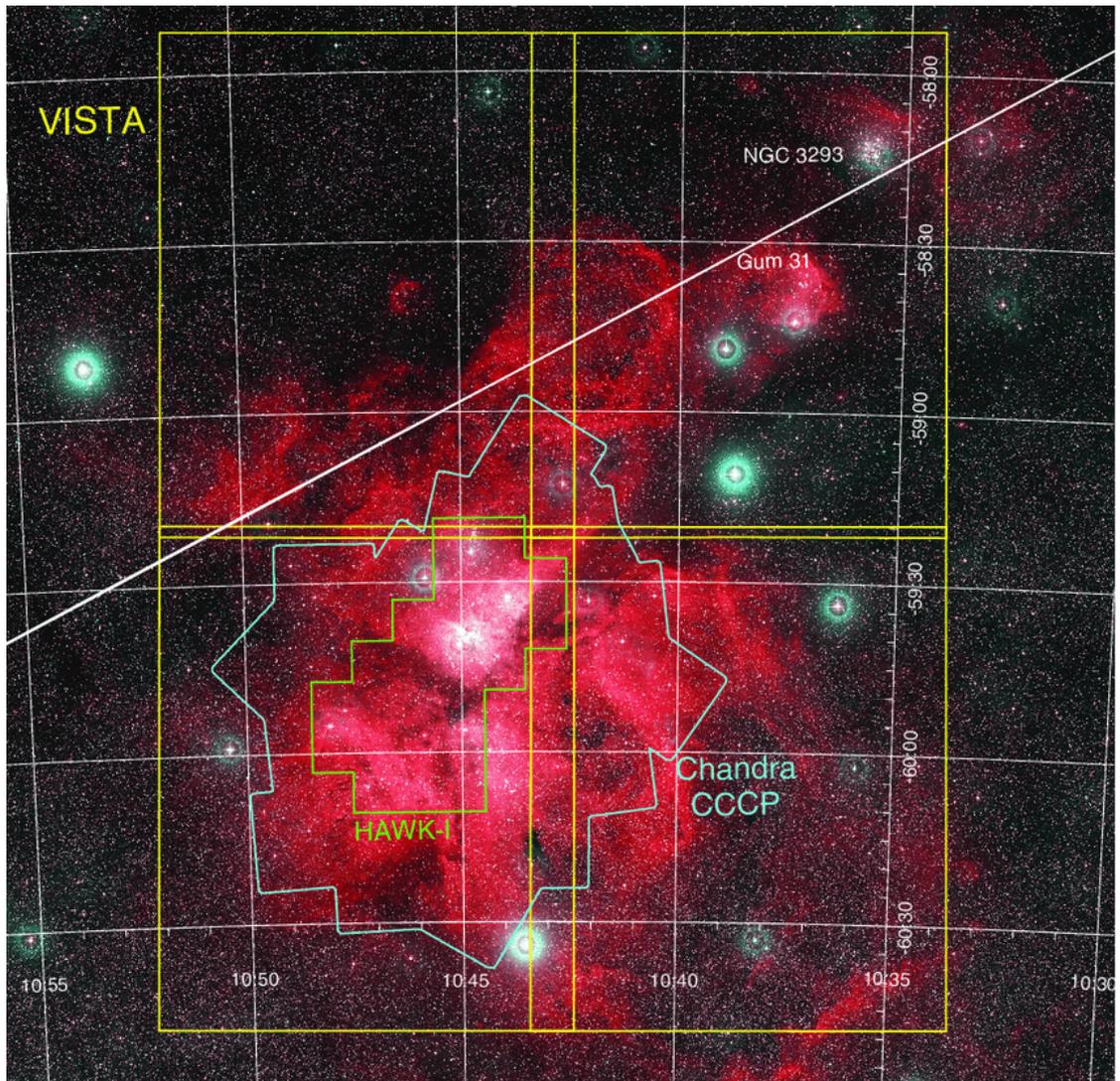


FIGURE 1.2: Field of view of Carina region as observed with VISTA (Preibisch et al. 2014), HAWK-I (Preibisch et al. 2011) and Chandra (Townsend et al. 2011a) overlaid on optical image of Carina where the solid line corresponds to the Galactic plane. (Credit: Preibisch et al. 2014 reproduced with permission © ESO).

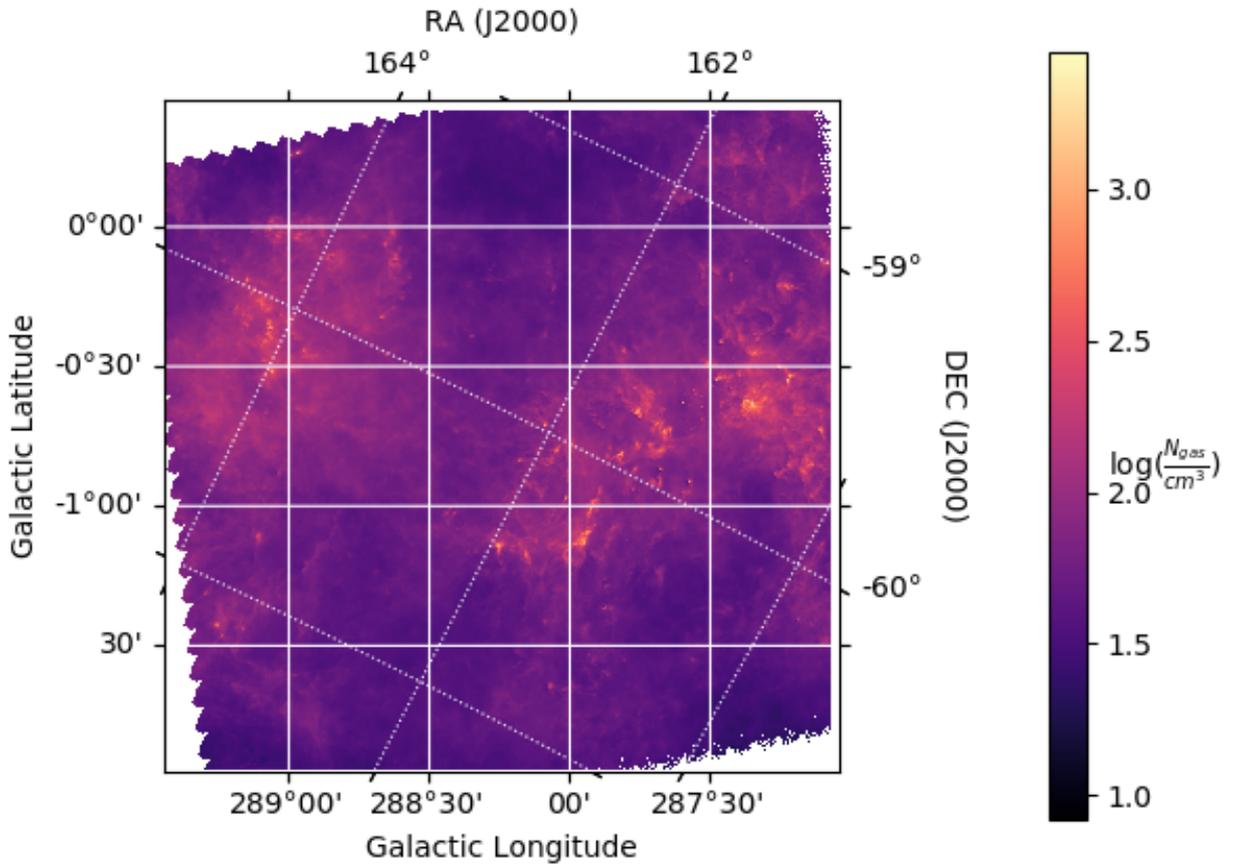


FIGURE 1.3: Carina H_2 number density obtained from processed (Marsh et al. 2017) Herschel Hi-GAL data (Molinari et al. 2010) assuming a line of sight depth of 20 parsecs (value from our collaborators Dr. Klaassen & Dr. Reiter). This follows from dividing the column densities by 20 parsecs resulting in a number of gas molecules per cm^{-3} for each pixel. Logged values are shown.

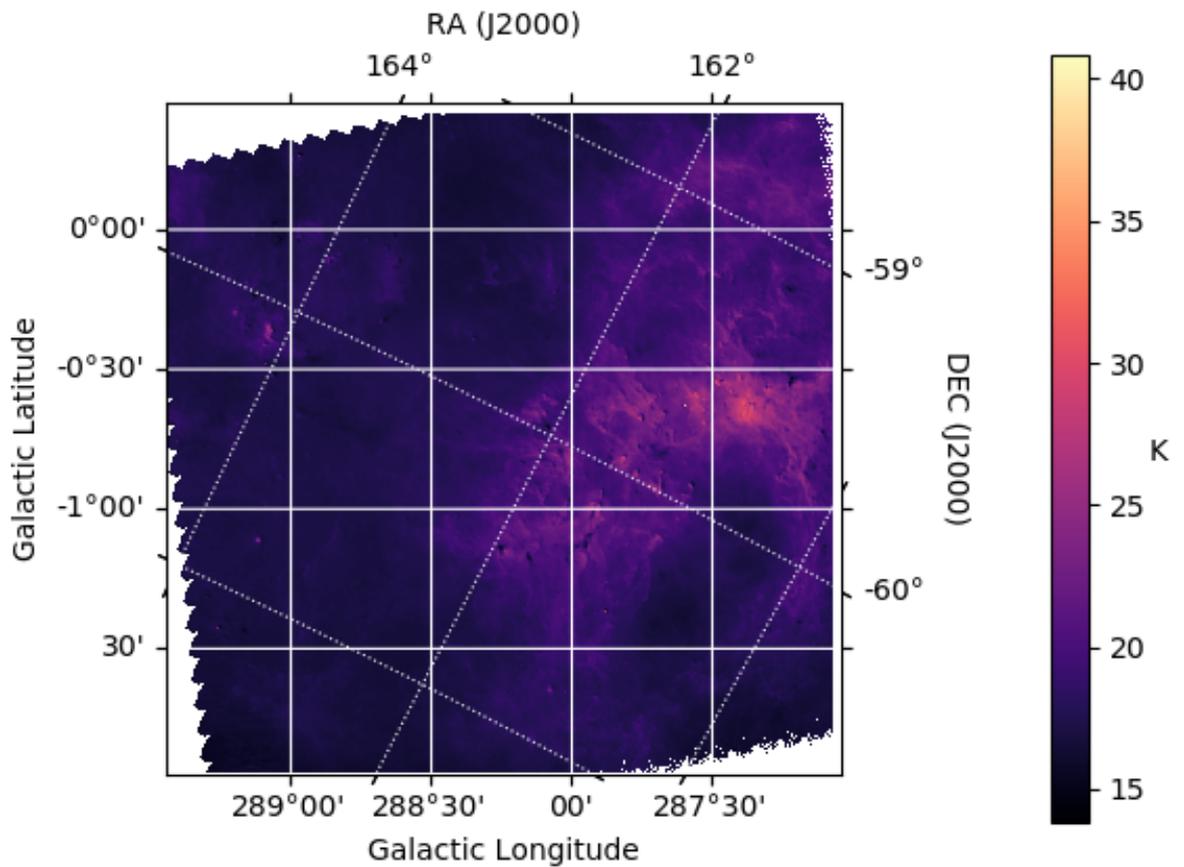


FIGURE 1.4: Carina line of sight mean temperature obtained from processed (Marsh et al. 2017) Herschel Hi-GAL data (Molinari et al. 2010)

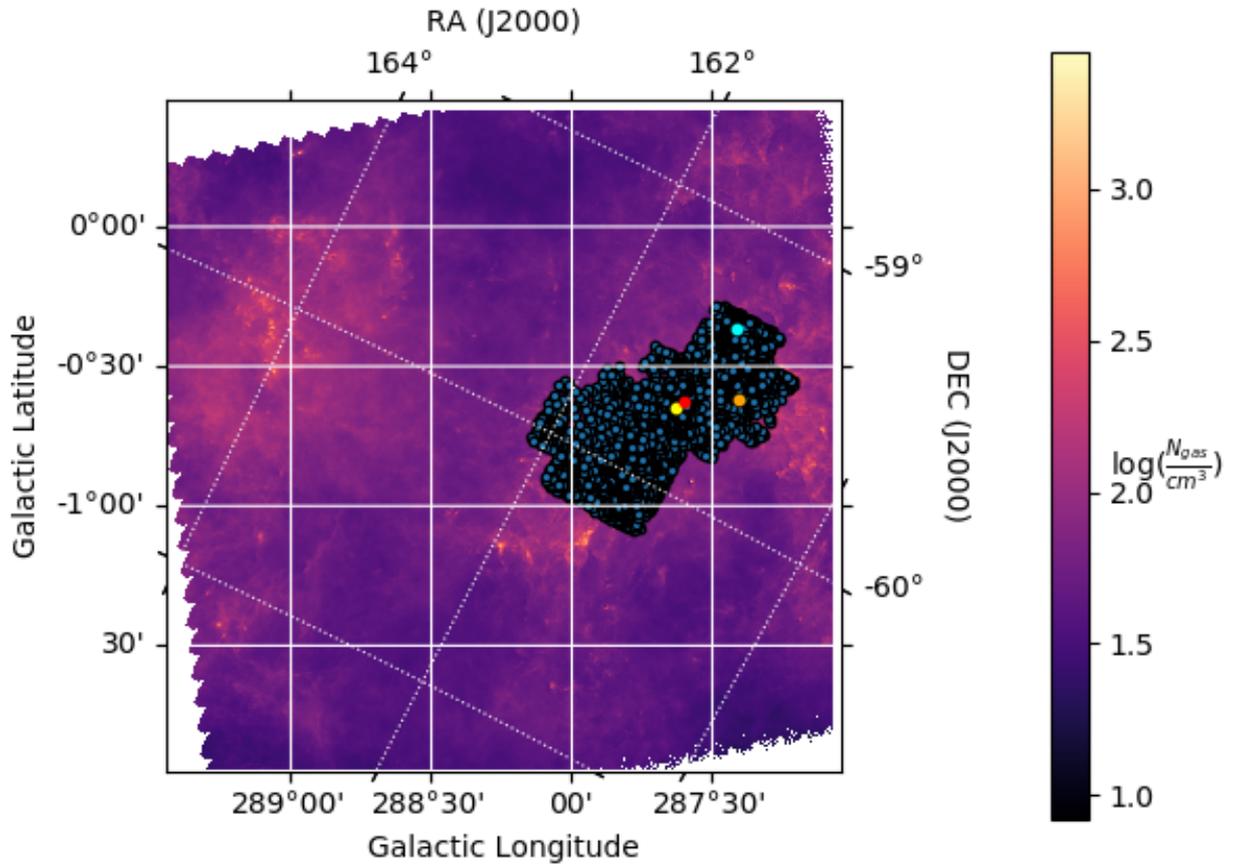


FIGURE 1.5: Measured stellar positions from MYSTiX project (dark blue) (Feigelson et al. 2013) overlaid on Carina H_2 number density described in Figure 1.3. Positions of Tr 14, 15, 16 and η Carinae are in orange, cyan, yellow and red, respectively.

1.6 Other Gas Observations

The previous sections described the stellar and gas observations we incorporate in our simulations. Specifically, these observations give spatial information in terms of stellar positions and the gas density distribution. To further increase the accuracy of the input gas, we can allow the gas in our simulations to have a velocity dispersion based on observations.

1.6.1 Gas Velocity Dispersion Measurements

Larson's relation states that the internal velocity dispersion is correlated with the size and mass of the molecular cloud which was determined through observations of molecular clouds (Larson 1981). Using this relation would provide an estimate for the dispersion, however it would depend on the size of the observed region which may not reflect the size of the entire molecular cloud. As well, the estimate depends on a single characteristic size of the cloud which may not accurately describe an elongated cloud.

Velocity dispersion values can also be derived from observations of ^{12}CO , ^{13}CO and C^{18}O which trace different cloud densities of roughly 10^2 cm^{-3} , 10^3 cm^{-3} and 10^4 cm^{-3} , respectively (Mizuno et al. 1995; Kim et al. 2004; Yonekura et al. 2005). These density differences are due to the corresponding optical depth of the transitions. For example, the ^{12}CO emission is optically thick which is why it is only able to trace the lower density layer of a molecular cloud.

In our simulations, we use a velocity dispersion value derived from observations of massive cores obtained with the NANTEN telescope (Yonekura et al. 2005) which are traced with $C^{18}O$ emission lines. Figure 1.6 shows the total intensity map of $C^{18}O$ while Figure 1.7 shows the positions of the cores on the $C^{18}O$ integrated intensity map. Cores numbered 8-15 overlay roughly with the Herschel Hi-GAL column density data.

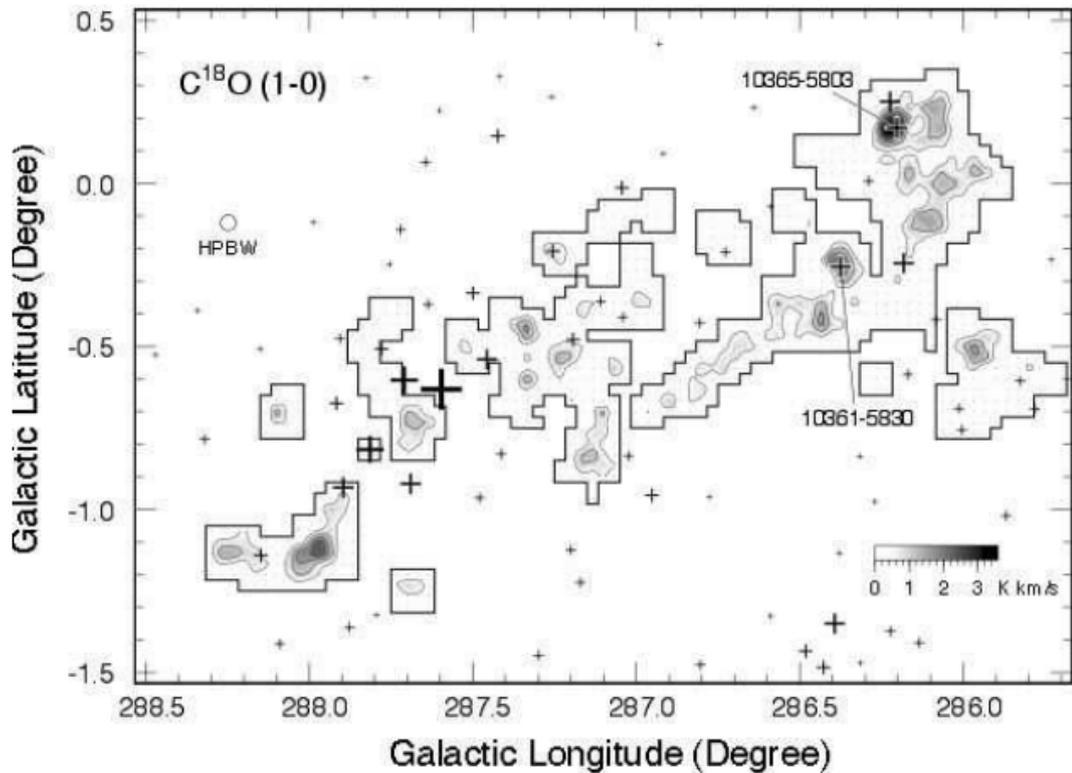


FIGURE 1.6: Total $C_{18}O$ ($J = 1 - 0$) intensity map used to obtain velocity dispersion of massive cores in the Carina region. Contours begin at 0.5 K km/s and increase in steps of 0.5 K km/s. Reproduced from Figure 4 in Yonekura et al. 2005 with permission of the AAS.

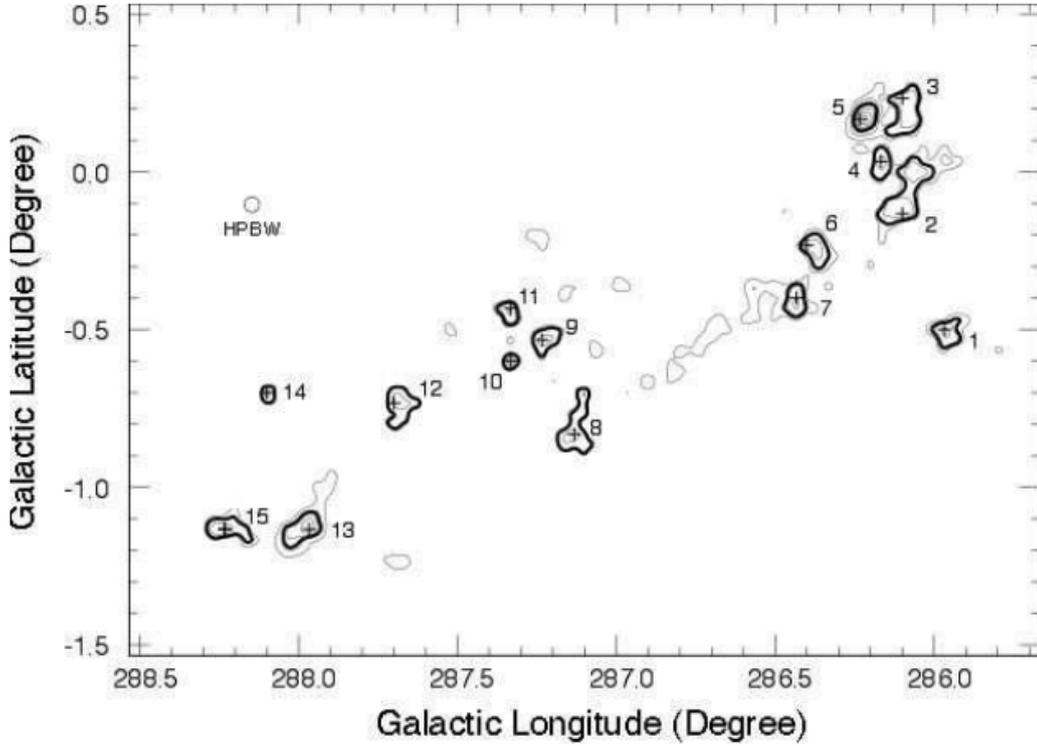


FIGURE 1.7: Positions of cores overlaid on $C^{18}O$ integrated intensity map. Contours are the same as those in Figure 1.6. Contours begin at 0.5 K km/s and increase in steps of 0.5 K km/s. Reproduced from Figure 5 in Yonekura et al. 2005 with permission of the AAS.

These observations have a typical line width or full-width half-maximum (FWHM) of 3.3 km/s. The velocity dispersion σ is calculated using:

$$FWHM = 2\sqrt{2 \ln 2} \sigma \quad (1.10)$$

which corresponds to a velocity dispersion of 1.4 km/s.

Although these line width measurements are precise and localized, they may not

be an indicative measure of the velocity dispersion over the entire area observed with the Herschel Hi-GAL survey due to the fact that $C^{18}O$ only traces the high density regions. Consequently, we allow the velocity dispersion to vary in our simulations.

1.7 Thesis Overview

The purpose of this thesis project is to develop a new technique to incorporate observational gas data into a star cluster simulation. We present and validate a new method that is able to do this and we evolve the Carina region as an example. The numerical methods are outlined in Chapter 2 and the new method to incorporate observational gas data is described in Chapter 3. Results are presented in Chapter 4 and conclusions in Chapter 5.

Chapter 2

Simulation Methods

In this chapter, we describe the numerical techniques used to model the stars and gas in our simulations. Specifically, we use an N-body code to evolve the stars and a smoothed-particle hydrodynamics code to evolve the gas which are coupled through the AMUSE (a multi-physics software environment implementation for astrophysics) framework. We present the tools used to analyze our simulations as well. These include dendrograms which are used to describe the hierarchical structure in the gas distribution, a clustering algorithm (DBSCAN) to track the merging or separation of sub-clusters, and the calculation of the mass-segregation ratio of the clusters picked out by DBSCAN.

2.1 Smoothed-particle Hydrodynamics

Smoothed particle hydrodynamics (SPH) is a mesh-free particle method initially developed to model non-spherical stars (e.g the formation of protostars from clouds of gas) and is based on fluid dynamics (Gingold and Monaghan [1977](#)). This method

involves representing fluids as discrete particles and employs a Lagrangian description of fluid flow where co-moving spatial coordinates of fluid elements are used and the particle properties are followed over time. The equations of motion that describe the fluid are therefore a set of ordinary differential equations discretized with respect to time. Smoothed particle properties are obtained with interpolation through the use of a kernel approximation which relies on an associated smoothing length value, h , that defines a local region of influence around each particle. Kernel functions that drop to zero at a distance h (or $2h$ depending on notation), are therefore ideal (Monaghan 1992). By smoothing the particle properties in this way, continuous fields can be constructed from the set of discrete particles. Following Monaghan 1992, the integral interpolant of a function $A(\mathbf{r})$ over all space is:

$$A(\mathbf{r}) = \int A(\mathbf{r}')W(\mathbf{r} - \mathbf{r}', h)\mathbf{d}\mathbf{r}' \quad (2.1)$$

where W is a kernel function that has the following two properties:

$$\int W(\mathbf{r} - \mathbf{r}', h)\mathbf{d}\mathbf{r}' = 1 \quad (2.2)$$

$$\lim_{h \rightarrow 0} W(\mathbf{r} - \mathbf{r}', h) = \delta(\mathbf{r} - \mathbf{r}') \quad (2.3)$$

The interpolant can be approximated using a summation over all the particles:

$$A_s(\mathbf{r}) = \sum_b m_b \frac{A_b}{\rho_b} W(\mathbf{r} - \mathbf{r}_b, h) \quad (2.4)$$

where b is a particle that has mass m_b , position r_b , density ρ_b and any other quantity A at r_b defined as A_b .

For example, the density everywhere can be estimated using:

$$\rho(\mathbf{r}) = \sum_b m_b W(\mathbf{r} - \mathbf{r}_b, h) \quad (2.5)$$

which depends on the mass of the particles and the kernel function which itself depends on the particle position and smoothing length.

An example of a kernel is a spline function seen in Monaghan and Lattanzio (1985). The function used in GADGET-2 (the specific SPH code we use, Springel 2005) is based on this and is given as:

$$W(\mathbf{r}, h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3 & \text{if } 0 \leq \frac{r}{h} \leq \frac{1}{2} \\ \left(2\left(1 - \frac{r}{h}\right)\right)^3 & \text{if } \frac{1}{2} \leq \frac{r}{h} \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.6)$$

Using this spline function as a kernel is advantageous because it has compact support and so particles have zero interactions with those that are $r > h$ away. This means that each particle is only influenced by its immediate neighbours. As

well, the second derivative of this function is continuous which implies that it is not sensitive to disorder in the particles. If the particle disorder is low, the errors in approximating the integral interpolants by summation interpolants remains low ($O(h^2)$) as well which is another advantage of using this particular kernel.

As seen, the SPH method is closely tied to smoothing length values. If particles are greater than $r > h$ away, they cease to influence each other. Early use of SPH (Monaghan 1992) used a smoothing length value that was proportional to the average density of all the particles and was constant in space but varied with time. In this way, all the particles would have the same smoothing length value at each step in time. Hernquist & Katz (1989), Evrard (1988) and Benz et al. (1990) improved on this and explored increased resolution by using smoothing lengths dependent on the density of each individual particle where particles in less dense regions would be assigned larger values of h while particles in high density regions would be assigned smaller h values. Adaptive smoothing lengths are used in the SPH code we use called GADGET-2, the main details of which are summarized in the next section.

2.1.1 GADGET-2

GADGET-2 (Springel 2005) is a cosmological simulation code structured as a TreeSPH code (Hernquist and Katz 1989) to compute gravitational interactions and gas dynamics using smoothed-particle hydrodynamics. The tree code uses a hierarchical multipole expansion to calculate the gravitational forces. It involves particles being grouped into cells with distant particles being grouped into larger

cells. The gravitational forces are then calculated with respect to the cells which results in the total number of computations being $O(N \log(N))$, an improvement from $O(N^2)$ in traditional N-body codes where forces between pairs of particles are calculated. GADGET-2 is capable of modelling collisionless dynamics (e.g dark matter or stars in a galaxy) as well as collisional gas dynamics. This code can be used to simulate systems with greatly differing scales including the large-scale structure of the universe, colliding galaxies and isolated star clusters, our focus being the latter. We use GADGET-2 to specifically model the self-gravity and hydrodynamics of the gas in our simulations. The density for each particle is estimated according to the previously described SPH method:

$$\rho_i = \sum_{j=1}^N m_j W(|\mathbf{r}_i - \mathbf{r}_j|, h_i) \quad (2.7)$$

where the kernel function is given by equation 2.6. GADGET-2 makes use of adaptive smoothing lengths and is able to conserve both energy and entropy. The smoothing lengths are adaptive in order for the kernel volumes to contain a constant mass and therefore obey the implicit equation:

$$\frac{4\pi}{3} h_i^3 \rho_i = N_{sph} \bar{m} \quad (2.8)$$

where N_{sph} is the typical number of neighbours and \bar{m} is the average particle mass.

Springel and Hernquist 2002 show that the equations of motion of the particles are:

$$\frac{dv_i}{dt} = - \sum_{j=1}^N m_j [f_i \frac{P_i}{\rho_i^2} \nabla_i W_{ij}(h_i) + f_j \frac{P_j}{\rho_j^2} \nabla_i W_{ij}(h_j)] \quad (2.9)$$

where P_i are the particle pressures, $W_{ij}(h_i) = W(|\mathbf{r}_i - \mathbf{r}_j|, h)$ and each f_i is:

$$f_i = (1 + \frac{h_i}{3\rho_i} \frac{\partial \rho_i}{\partial h_i})^{-1} \quad (2.10)$$

Shocks are implemented using an artificial viscosity with a viscous force given by:

$$\frac{d\mathbf{v}_i}{dt} = - \sum_{j=1}^N m_j \Pi_{ij} \nabla_i \bar{W}_{ij} \quad (2.11)$$

where Π_{ij} describes the artificial viscosity which was originally introduced (for numerical reasons) to resolve shocks in fluids. $\Pi_{ij} \geq 0$ and is non-zero when particles approach each other. The parametrization of the viscosity by Monaghan 1997 is given by:

$$\Pi_{ij} = - \frac{\alpha [c_i + c_j - 3w_{ij}] w_{ij}}{2 \rho_{ij}} \quad (2.12)$$

where $\alpha \approx 0.5 - 1.0$ (our simulations use $\alpha = 0.5$), $w_{ij} = \frac{\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{|\mathbf{r}_{ij}|}$ and c_i, c_j are the sound speeds of particles i and j .

This summarizes certain key aspects of the SPH formalism used in the GADGET-2 code. Further details can be seen in Springel 2005. We use SPH to model the gas which is a fluid. To model the stars, we use an N-body code.

2.2 N-body Code

The N-body code we use is a fourth-order Hermite integration scheme (Makino and Aarseth 1992) which takes into account the gravitational forces between the stars. This integration scheme is a predictor-evaluator-corrector (PEC) scheme.

The first step involves position (\mathbf{r}) and velocity (\mathbf{v}) predictions given by:

$$\mathbf{r}_{i,p} = \mathbf{r}_{i,0} + \mathbf{v}_{i,0}\Delta t_i + \frac{1}{2}\mathbf{a}_{i,0}\Delta t_i^2 + \frac{1}{6}\dot{\mathbf{a}}_{i,0}\Delta t_i^3 \quad (2.13)$$

$$\mathbf{v}_{i,p} = \mathbf{v}_{i,0} + \mathbf{a}_{i,0}\Delta t_i + \frac{1}{2}\dot{\mathbf{a}}_{i,0}\Delta t_i^2 \quad (2.14)$$

Next, the acceleration and jerk are calculated using:

$$\mathbf{a}_{i,1} = \sum_{j=0, j \neq i}^N Gm_j \frac{\mathbf{r}_{ij}}{(r_{ij}^2 + \epsilon^2)^{\frac{3}{2}}} \quad (2.15)$$

$$\dot{\mathbf{a}}_{i,1} = \sum_{j=0, j \neq i}^N Gm_j \left(\frac{\mathbf{v}_{ij}}{(r_{ij}^2 + \epsilon^2)^{\frac{5}{2}}} \right) \quad (2.16)$$

where ϵ is a softening parameter, $\mathbf{r}_{ij} = \mathbf{r}_{j,p} - \mathbf{r}_{i,p}$, $\mathbf{v}_{ij} = \mathbf{v}_{j,p} - \mathbf{v}_{i,p}$ and $r_{ij} = |\mathbf{r}_{ij}|$. We use gravitational softening to prevent the gravitational force from going to infinity if particles approach very close to each other (r tending to zero). Our softening length is chosen to be 200 AU.

The correction step is as follows:

$$\mathbf{r}_{i,1} = \mathbf{r}_{i,p} + \frac{1}{24}\Delta t_i^4 \ddot{\mathbf{a}}_{i,0} + \frac{1}{120}\Delta t_i^5 \dddot{\mathbf{a}}_{i,0} \quad (2.17)$$

$$\mathbf{v}_{i,1} = \mathbf{v}_{i,p} + \frac{1}{4}\Delta t_i^3 \ddot{\mathbf{a}}_{i,0} + \frac{1}{24}\Delta t_i^4 \dddot{\mathbf{a}}_{i,0} \quad (2.18)$$

The second and third derivatives of acceleration are calculated using the third-order interpolation polynomial:

$$\mathbf{a}_{i,1} = \mathbf{a}_{i,0} + \dot{\mathbf{a}}_{i,0}\Delta t_{i,0} + \frac{1}{2}\Delta t_{i,0}^2 \ddot{\mathbf{a}}_{i,0} + \frac{1}{6}\Delta t_{i,0}^3 \dddot{\mathbf{a}}_{i,0} \quad (2.19)$$

where:

$$\ddot{\mathbf{a}}_{i,0} = \frac{-6(\mathbf{a}_{i,0} - \mathbf{a}_{i,1}) - \Delta t_{i,0}(4\ddot{\mathbf{a}}_{i,0} + 2\ddot{\mathbf{a}}_{i,1})}{\Delta t_{i,0}^2} \quad (2.20)$$

$$\dddot{\mathbf{a}}_{i,0} = \frac{-12(\mathbf{a}_{i,0} - \mathbf{a}_{i,1}) - 6\Delta t_{i,0}(\dot{\mathbf{a}}_{i,0} + \dot{\mathbf{a}}_{i,1})}{\Delta t_{i,0}^3} \quad (2.21)$$

The time step is updated according to:

$$t_{i,1} = t_{i,0} + \Delta t_{i,0} \quad (2.22)$$

$$\Delta t_{i,1} = \sqrt{\eta \frac{|\mathbf{a}_{i,1}| |\ddot{\mathbf{a}}_{i,1}| + (|\dot{\mathbf{a}}_{i,1}|)^2}{|\dot{\mathbf{a}}_{i,1}| |\ddot{\mathbf{a}}_{i,1}| + (|\ddot{\mathbf{a}}_{i,1}|)^2}} \quad (2.23)$$

where η is a control parameter that controls accuracy, $\ddot{\mathbf{a}}_{i,1} = \ddot{\mathbf{a}}_{i,0}$ due to the third order interpolation and $\dot{\mathbf{a}}_{i,1} = \dot{\mathbf{a}}_{i,0} + \Delta t_i \ddot{\mathbf{a}}_{i,0}$.

We couple the SPH code GADGET-2 to the fourth-order Hermite integration scheme (specific code is ph4) within the AMUSE framework to simultaneously evolve both the stars and gas. It should be noted that although GADGET-2 contains an N-body solver which could be used, it is less precise than the ph4 code we use.

2.3 AMUSE

AMUSE is based on the MUSE (Multi-physics Software Environment) framework with specific applications in astrophysics (Portegies-Zwart et al. 2013; Portegies-Zwart et al. 2018; Pelupessy et al. 2013; Pelupessy et al. 2009). The motivation for creating the MUSE framework stemmed from a desire to be able to combine the many independent physics community codes available. These codes span different areas of physics including stellar evolution and hydrodynamics although each code

is developed and maintained separately by those in the scientific community. By integrating these existing codes, simulations would not always need to be built from the ground up. As well, there would be more flexibility as switching between different physics modules could be done with ease. MUSE is implemented in Python and its approach depends on a user script and a community module which itself consists of a manager, communication layer and community code layer. This can be seen in Figure 2.1. Details about the specific components of the architecture can be seen in Portegies-Zwart et al. 2013. The user script is the interface onto the MUSE framework and is where a model is generated with the appropriate initial conditions and the different community codes to be used. Once the simulation is run, any results are returned to this layer for analysis. The manager is the interface onto the communication layer. The communication layer is bi-directional and facilitates communication between the manager and the community code layer. The communication layer consists of a proxy and an associated partner. The proxy translates Python commands which are then sent to the partner to be decoded and executed to run the community codes. This step allows the community codes to be executed which may not be written in Python (e.g they could be written in C, C++ or FORTRAN). AMUSE (a MUSE implementation for astrophysics) is the implementation of MUSE used to simulate astrophysical phenomena. Different community codes that could be coupled together include N-body, smoothed-particle hydrodynamics and stellar evolution codes. In this way, increasingly complex simulations can be realized with relative ease through the use of this framework.

Within AMUSE, we also couple the stars and gas using the Bridge method

(Fujii et al. 2007). The main idea is that the gravitational effects between the stars and gas are computed by giving the star and gas particles periodic velocity kicks. As shown in Fujii et al. 2007, the integration for the star particles is as follows:

$$\mathbf{v}'_{\text{star},0} = \mathbf{v}_{\text{star},0} + \frac{1}{2}\Delta t \mathbf{a}_{\text{gas} \rightarrow \text{star},0} \quad (2.24)$$

$$\mathbf{x}_{\text{star},0} \rightarrow (\textit{HermiteScheme}) \rightarrow \mathbf{x}_{\text{star},1} \quad (2.25)$$

$$\mathbf{v}'_{\text{star},0} \rightarrow (\textit{HermiteScheme}) \rightarrow \mathbf{v}'_{\text{star},1} \quad (2.26)$$

$$\mathbf{v}_{\text{star},1} = \mathbf{v}'_{\text{star},1} + \frac{1}{2}\Delta t \mathbf{a}_{\text{gas} \rightarrow \text{star},1} \quad (2.27)$$

The integration for the gas particles is as follows which is a leapfrog method:

$$\mathbf{v}_{\text{gas},\frac{1}{2}} = \mathbf{v}_{\text{gas},0} + \frac{1}{2}\Delta t \mathbf{a}_{\text{all} \rightarrow \text{gas},0} \quad (2.28)$$

$$\mathbf{x}_{\text{gas},1} = \mathbf{x}_{\text{gas},0} + \Delta t \mathbf{v}_{\text{gas},\frac{1}{2}} \quad (2.29)$$

$$\mathbf{v}_{\text{gas},1} = \mathbf{v}_{\text{gas},\frac{1}{2}} + \frac{1}{2}\Delta t \mathbf{a}_{\text{all}\rightarrow\text{gas},1} \quad (2.30)$$

First the accelerations of all particles acting on the gas and the gas acting on the star particles are calculated. The stars are given a velocity kick while the gas velocity is updated. The stellar positions and velocities are integrated forward using the Hermite scheme while the gas positions are updated using the leapfrog method. Next, the accelerations are calculated for $t + \Delta t$ which are then used to give the stars another kick and update the gas velocities.

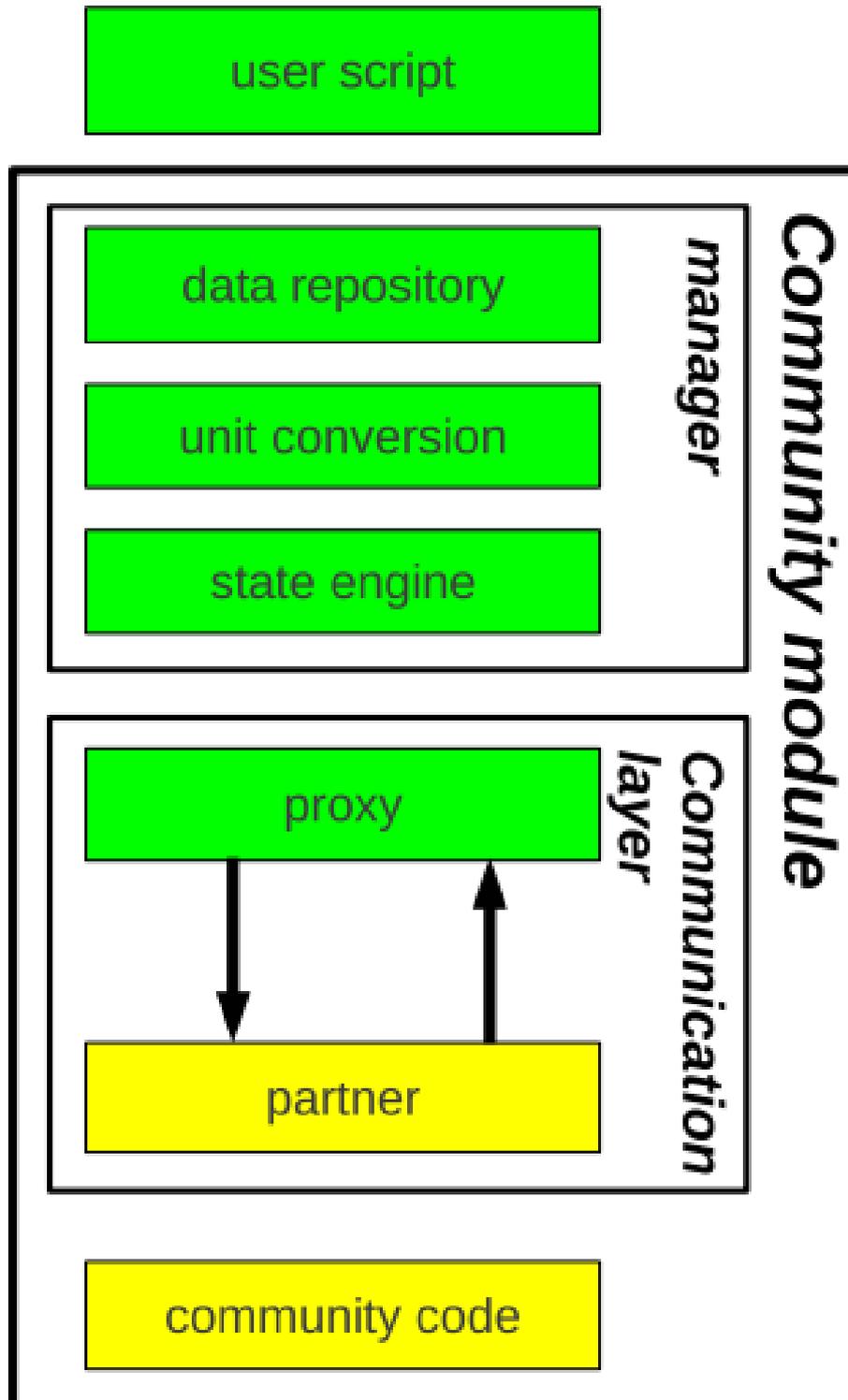


FIGURE 2.1: The MUSE framework as described in Portegies-Zwart et al. 2013 and reprinted with permission from Elsevier.

2.4 Analysis Tools

This section summarizes the tools used to analyze the stars and gas in our simulations. We compute dendrograms of the gas distribution in order to track clumps and filamentary structures over time using the *astrodendro* package in Python (Robitaille et al. 2013). As well, we use a clustering algorithm called DBSCAN (Density-Based Spatial Clustering of Applications with Noise) also implemented in Python (Ester et al. 1996; Pedregosa et al. 2011b) to characterize the degree of sub-clustering in the stars over time. Furthermore, we track the mass segregation of the subclusters identified with DBSCAN by calculating Λ_{MSR} (Allison et al. 2009) which is implemented in AMUSE.

2.5 Astrodendro

Dendrograms are a way to describe the hierarchical structure of objects or observations with a trunk, leaves and branches (Robitaille et al. 2013). It is a general technique used to visualize the results of hierarchical clustering and has applications in classification problems. Leaves contain no substructure while branches contain multiple leaves. The trunk refers to the main structure that has no parent structure. The maxima of a distribution are represented by leaves that are significantly different than their surroundings. *astrodendro* is a Python package (Robitaille et al. 2013) that specifically computes dendrograms from astronomical data. We can use this tool to compute dendrograms to describe the gas density distribution in 2D planes over time where the maxima would correspond to dense

regions of gas. We can then track the evolution of structures in the gas like clumps or fragments over time. We will also be able to compare how the regions of dense gas evolve over time compared to the stellar distribution. We chose to use dendrograms in this way because gas in molecular clouds is hierarchical where high density regions are surrounded by lower density gas. We model our gas using particles which could be used as input to a clustering algorithm like DBSCAN, however such algorithms will only pick out clusters of particles without keeping track of any hierarchical relationships as seen in the next section. Therefore, dendrograms are a useful tool to characterize gas distributions as regions of high and low densities can both be traced. This tool has been previously used to characterize molecular cloud structures from ^{12}CO and ^{13}CO observations (Nayak et al. 2016) and to characterize the fragmentation of molecular clouds in simulations (Chira et al. 2018). Raw gas observations of molecular clouds are in the form of 2D column densities, therefore we chose to analyze our 3D simulations in 2D to determine what would be seen from an observational perspective. An example of a dendrogram used to describe a hierarchical structure can be seen in Figure 2.2.

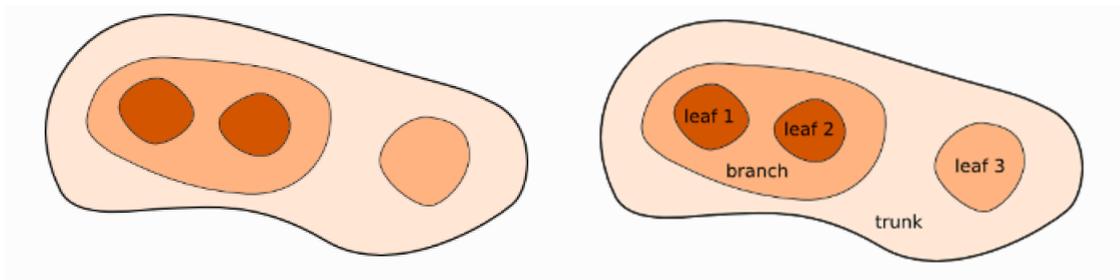


FIGURE 2.2: Example of dendrogram structures used to describe a 2D hierarchical distribution. Credit: Astrodendro online documentation (Robitaille et al. 2013).

Similarly, any dense clumps or filamentary structures in the gas distribution in our simulations can be described by leaves that are connected by branches that

are then encompassed by the surrounding gas or the trunk. In order to construct a dendrogram of a distribution, the tree is built starting from the pixels or bins with the highest value. The first structure or leaf is created from this peak value and the pixel with the next highest value is considered. This next pixel may be added to the first structure or it may form a new structure if it is a local maximum. As subsequent pixels are considered, branches are formed which do not correspond to local maxima but are adjacent to the leaves that are formed. This process continues creating more leaves and branches that describe the overall distribution. This process can be seen in Figure 2.3.

The dendrogram structures in Figures 2.2 and 2.3 assumes that all pixel values are significant and therefore will consider all local maxima as leaves regardless of the size of the maxima in pixels. As well, this means that any noise in the data could potentially be added to the tree. The method used to compute the dendrogram takes a *min_value*, *min_delta* and *min_npix* as parameters in order to construct structures that are significant. The *min_value* represents the noise level below which the tree will not be constructed, *min_delta* represents the significance required for a new structure to form and *min_npix* is the minimum number of pixels for a leaf to be its own entity. The initial gas particle distribution from one of our Carina simulations (Method I run I, see 4.1 in Chapter 4) was binned in 2D and weighted by the particle masses per square parsec such that each bin is 0.5 parsecs in each dimension. This results in a range in densities from 0 to roughly $1000 M_{\odot}/pc^2$. Based on this range, we set *min_value* to be $5 M_{\odot}/pc^2$ to remove the relatively small fluctuations in the density. We then set *min_delta* to be $10 M_{\odot}/pc^2$ which is roughly 100 times smaller than the maximum density.

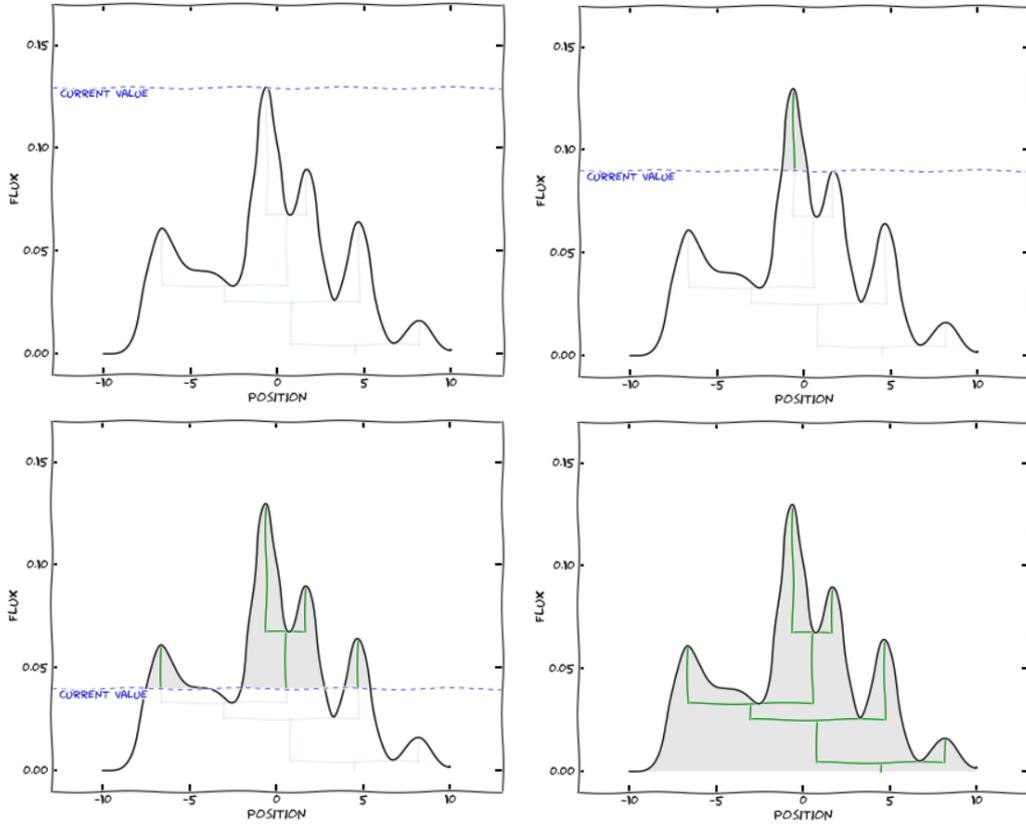


FIGURE 2.3: Example of constructing a dendrogram from a 1D dataset (Steps are from right to left, top to bottom). Credit: Astrodendro online documentation (Robitaille et al. 2013).

This will create several structures and allows flexibility in how they are grouped into different density ranges. We set min_npx to be 10 which corresponds to a square of area $\approx 2.5 pc^2$ which is an order of magnitude greater than a thin filament (0.1 parsecs) and is reasonable for picking out clumps of gas that span more than one bin reducing bin to bin fluctuations. Figure 2.4 shows the dendrogram that characterizes the initial gas distribution for the Carina region.

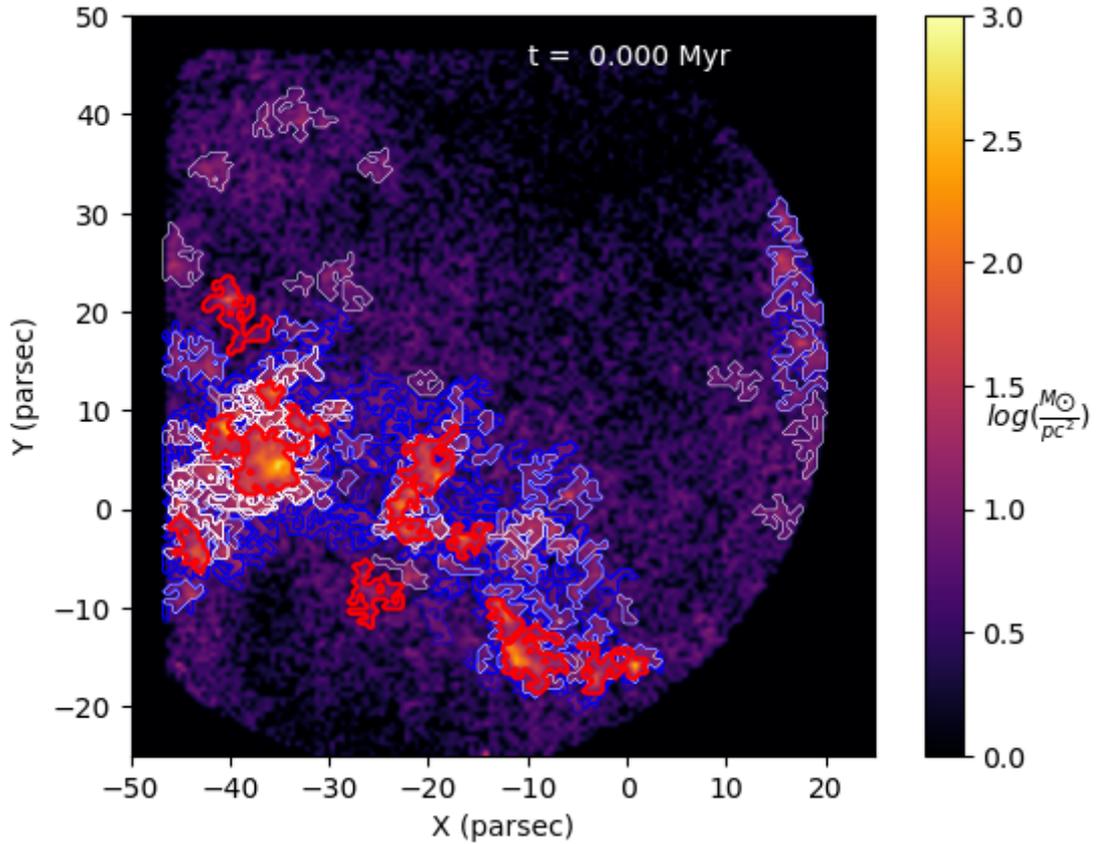


FIGURE 2.4: 2D histogram of the initial gas distribution for the Carina region (specifically for Method I run I defined in Chapter 3). Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$.

2.6 DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is an algorithm used to identify clusters of arbitrary shape from a set of data points (Ester et al. 1996). Its dependence on the density of data points allows clusters to be discovered which correspond to areas of relatively high densities and noise which corresponds to lower densities outside the clusters. Although it was initially developed for identifying clusters in large spatial databases with noise, the data points themselves can be spatial coordinates, velocities or other quantities allowing clusters to be determined with respect to various different parameters. Since its development, this algorithm has been implemented in Python (Pedregosa et al. 2011a) and has been applied to observational stellar data resulting in the detection of stars that are clustered in space, dynamically correlated or clusters which are chemically homogeneous (Beccari et al. 2020; Price-Jones and Bovy 2019).

The Python implementation relies on two main input parameters which are a search radius (*eps*) and a minimum number of samples in a cluster (*min_samples*). The *eps* value is the maximum distance between two samples in order for them to be in the neighborhood of each other while the *min_samples* is the number of samples in the neighborhood for that point to be considered a ‘core’ point where the number of cores corresponds to the number of clusters found. It is possible for some samples to be border points where they could be grouped into more than one cluster. If this occurs, the border points get assigned to clusters depending on the order in which the clusters are processed which may result in some clusters having fewer samples than the *min_samples* value. As a result, the *min_samples*

parameter does not always equate to the minimum number of samples in a cluster and is a caveat of the DBSCAN algorithm.

We apply DBSCAN to our simulation of the Carina region where the *eps* value is set to 0.5 parsecs (similar to the value used in Beccari et al. 2020) and the *min_samples* value is set to 20 in order to track the spatial evolution of the clusters in 1 Myr timesteps. The clustering is performed in three dimensions and the *eps* and *min_samples* values were chosen in order for clusters containing at least 20 stars (in general considering the aforementioned caveat) to be identified which are more likely to be persistent clusters rather than noise. DBSCAN is applied in 3D to estimate the realistic clustering of stars. Figures 2.5, 2.6 and 2.7 depict the results of DBSCAN on the initial stellar distribution of the Carina region as examples. In our analysis, we present 2D views (in the plane of the sky) to mimic the perspective of an observer which is seen in ???. The initial stellar distribution is derived from observations in the MYSTiX project and consists of 20 clusters (Kuhn et al. 2014) for which the 2D stellar positions are known along with a uniform background distribution of stars. The Z coordinates of the stars in subclusters are drawn from a range corresponding to the minor axis of the ellipse from the fits performed by Kuhn et al. 2014 while the Z coordinates of the background stars are drawn from a range set by the distance that encompasses the entire region. Further details can be seen in Sills et al. 2018.

From Figures 2.5, 2.6 and 2.7, we can see that DBSCAN picks out 16 clusters (all are visible in the XY plane) from the initial stellar distribution where the background stars are interpreted as noise. This is different than the number of clusters originally included from the isothermal ellipsoid fitting procedure in Kuhn

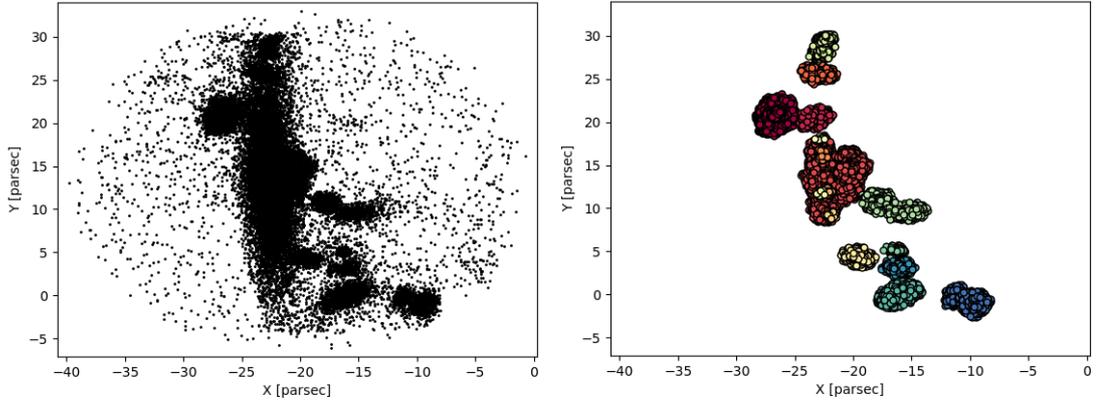


FIGURE 2.5: XY view of identification of clusters in the Carina region using DBSCAN. The initial distribution of stars is seen on the left which consists of 20 smaller clusters in a uniform background distribution of stars. The results from DBSCAN are seen on the right which shows that 16 clusters are identified where each cluster corresponds to a different colour.

et al. 2014 which identified 20 clusters. This arises because the clustering methods focus on different criteria. The ellipsoid model consists of an approximately flat core surface density which then drops off as a power law away from the core. The model corresponding to the “best” number of clusters (a parameter which is varied) is then chosen to represent the sub-cluster (see Kuhn et al. 2014 for further details). This fitting procedure focuses on identifying clusters based on their density distribution while DBSCAN depends on a search radius and a minimum number of samples. Therefore, these algorithms are inherently different resulting in the differing number of clusters that are picked out. We also attempted to vary the DBSCAN parameters which would recover the 20 subclusters identified with the ellipsoid fits but were unable to do so exactly. We chose to use DBSCAN due to its ease of use and we pick physically reasonable parameters to track the degree of sub-clustering in three dimensions over the course of our simulations in

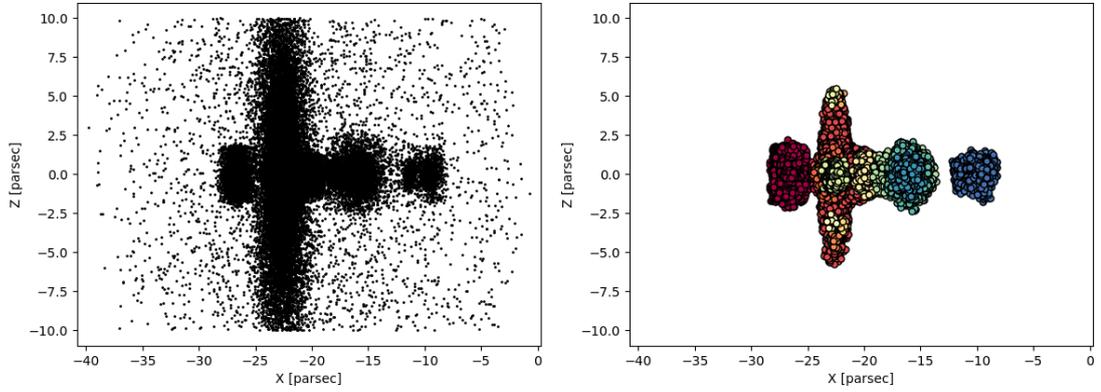


FIGURE 2.6: XZ view of identification of clusters in the Carina region using DBSCAN. The initial distribution of stars is seen on the left which consists of 20 smaller clusters in a uniform background distribution of stars. The results from DBSCAN are seen on the right which shows that 16 clusters are identified where each cluster corresponds to a different colour.

a consistent manner.

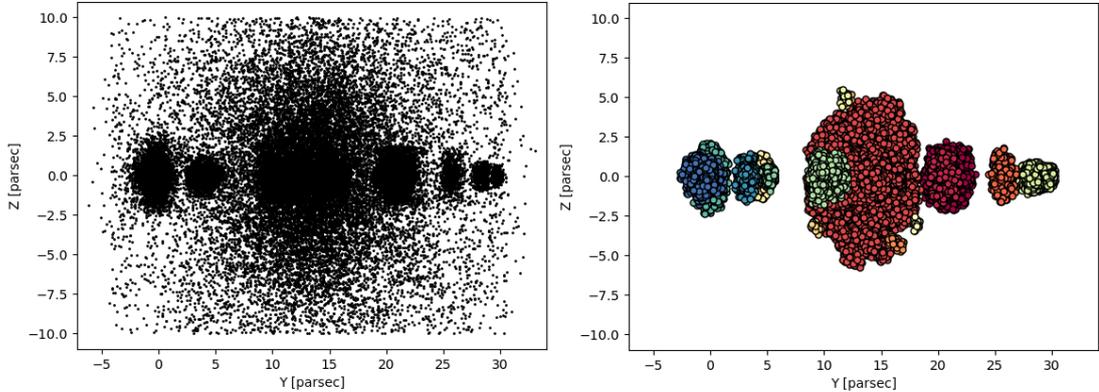


FIGURE 2.7: YZ view of identification of clusters in the Carina region using DBSCAN. The initial distribution of stars is seen on the left which consists of 20 smaller clusters in a uniform background distribution of stars. The results from DBSCAN are seen on the right which shows that 16 clusters are identified where each cluster corresponds to a different colour.

2.7 Mass-Segregation Ratios

Mass segregation refers to massive stars being more concentrated towards the center of a star cluster while low mass stars reside further away from the center. This has been observed in several clusters including young clusters (e.g Littlefair et al. 2003; Pang et al. 2013; Habibi et al. 2013) however it is not clear whether mass segregation occurs as a result of dynamical interactions or if it is a primordial effect. The dynamical interactions between the stars leads to two-body relaxation effects where the massive stars sink towards the center of the cluster (McMillan et al. 2007; Allison et al. 2009) and this theory is supported by numerical simulations where it was found that mass segregation occurs on timescales similar to the cluster ages in young clusters (Allison et al. 2010b; Parker et al. 2015). We track changes in the initial stellar distribution (we do not form new stars) which allows us to explore the effects of dynamical interactions related to mass segregation as the

system evolves.

In our simulations we are interested in the dynamical evolution of young embedded star clusters and our initial stellar distribution is informed from observations. The stellar masses for stars within a cluster are randomly drawn from an initial mass function within a mass range of 0.08 to 10 solar masses (details can be seen in Sills et al. 2018) and are not correlated with stellar positions. Thus, our initial distribution of clusters is not mass segregated. We can investigate whether mass segregation occurs in each individual cluster before merging or if it takes place after the clusters merge. To quantify the degree of mass segregation, we can calculate the mass segregation ratio (Allison et al. 2009) which relies on constructing a minimum spanning tree (MST). An MST is the unique set of edges that connect a set of points without closed loops where the sum of the lengths of the edges is minimized (Boruvka (1926); Kruskal (1956); Prim (1957); Gower & Ross (1969)).

Although different methods exist to calculate the MST of a set of points, the mass segregation ratio calculation presented in Allison et al. 2009 is based on the algorithm described in Prim (1957). To create an MST that connects all vertices of a set, the algorithm first creates a tree with a single vertex randomly chosen from the set. The tree grows one edge at a time by selecting the vertex that results in the smallest separation from the tree at each step.

Applying this to a cluster, Allison et al. 2009 first create an ordered list of the distances between all pairs of stars in a subset of interest and they then connect the stars in order of increasing separation. The degree of mass segregation can be quantified by comparing the length of the MST of the most massive stars in a

cluster to the average length of the MST of k sets of N stars chosen at random. If the length of the MST corresponding to the most massive stars is shorter than the average length, it demonstrates that the massive stars are more concentrated and therefore the cluster is mass segregated.

The mass segregation value and uncertainty is then given by (Allison et al. 2009):

$$\Lambda_{MSR} = \frac{\langle l_{random} \rangle}{l_{massive}} \pm \frac{\sigma_{random}}{l_{massive}} \quad (2.31)$$

where $\langle l_{random} \rangle$ is the average MST length of k sets of N stars chosen at random, $l_{massive}$ is the MST length of the most massive stars and σ_{random} is the standard deviation obtained from the k sets.

From this, we can see that $\Lambda_{MSR} \approx 1$ indicates no mass segregation while $\Lambda_{MSR} \gg 1$ is indicative of mass segregation.

This calculation is implemented in AMUSE allowing both the Λ_{MSR} and uncertainty values to be obtained. Such values are only meaningful if the cluster contains a reasonable number of stars. We follow Dib et al. 2018 where they calculate mass segregation ratios of clusters in the Milky Way from observations and restrict their analysis to clusters with at least 40 members. As well, they state that if a cluster is indeed mass segregated, it can be seen in the MST that contains at most 20 of the most massive stars regardless of the size of the cluster. In this way, we calculate Λ_{MSR} values for clusters picked out by DBSCAN that have at

least 40 members and consider MSTs for 20 of the most massive stars. To obtain the average MST length, 100 sets of 20 stars each are used.

By tracking the Λ_{MSR} values of the clusters picked out by DBSCAN over time, we can investigate how the mass segregation changes as the system evolves. We can also determine the effect of the initial gas properties (e.g varying the initial gas velocity dispersion) on the resulting stellar distributions by considering these mass segregation ratio values.

Chapter 3

Gas Particle Creation Method

As mentioned in Chapter 1, star cluster simulations often rely on simplifying assumptions when modelling the gas distribution which includes using spheres of gas (e.g Federrath et al. 2010) or gas that closely follows the stellar distribution (e.g Sills et al. 2018). Such assumptions do not lend themselves to a realistic depiction of gas in clusters which is observed to contain filaments and clumps due to various stellar feedback processes. To improve this, we developed a new technique that is able to create a gas distribution that captures the unique characteristics of the molecular gas from observational data and we apply this to the Carina region as an example. In this way, we simulate the Carina region using realistic initial conditions for the stars (details about the stellar distribution setup can be seen in Sills et al. 2018) and the gas.

This new method of modelling the gas based on observations is beneficial for exploring the accurate and realistic evolution of an embedded star cluster. By taking the observed gravitational potential and temperature into account, we can make inferences about future star formation that may occur depending on how the

gas collects over time. Furthermore, incorporating accurate initial gas conditions will be useful in later versions of these simulations as stellar feedback processes are included.

In our simulations of the Carina region, we use processed Herschel Hi-GAL maps (Molinari et al. 2010; Marsh et al. 2017) (specifically the number density map and the 2D mean line of sight dust temperature map) to create the initial gas conditions. Our technique involves sampling from these 2D distributions to create SPH particles (see Chapter 2 for details about SPH) with masses, 3D Cartesian coordinates, 3D velocities and internal energy values. The masses and positions of the particles in the XY plane (the plane of the sky) are made to match the observed number density distribution while the internal energies are made to match the observed mean temperatures. We developed two such sampling methods with slight variations and compare their relative performance. Since we do not have cloud depth measurements, we consider different approaches of estimating the depth based on features seen in the 2D distributions. As well, we use the measured velocity dispersion of the cores in the Carina region (mentioned in Chapter 1) to assign particle velocities and we also consider other velocity dispersion values to examine their effect. To determine the validity of the techniques used to assign the gas particle attributes, we apply our method and their variations to a simulated observational image that is created from a known particle data set. In this chapter, we discuss specific features of the Carina gas data that we take advantage of in our sampling techniques, we describe the simulated image data used for testing and we provide detailed descriptions of the methods used to assign the gas particle attributes.

3.1 Carina Gas Data

The gas data we use comes from the Herschel Hi-GAL survey (Molinari et al. 2010) that has been processed using the PPMAP algorithm (Marsh et al. 2017). Details regarding the survey and processing algorithm can be seen in Chapter 1. The distributions previously shown in Chapter 1 are in galactic coordinates that have been assigned from the raw pixels in the images. The number density map and the 2D mean line of sight dust temperature map can be seen in Figures 3.1 and 3.2 with the axes now in pixels. Note that the orientation of these images in pixel coordinates is different than that shown previously in galactic coordinates due to the conversion between the FITS image array to a Python *numpy* array (e.g NAXIS1 and NAXIS2 in the FITS image corresponds to the second and first index in the Python array, respectively). This was necessary in order to correctly match the other matrices required (all created in Python) to create the initial conditions (seen in section 3.3) and to ensure that the orientations of the stars and gas were correct in Cartesian space.

These maps can be loaded as matrices in Python where each pixel is specified by a set of two pixel coordinates (which can easily be converted to galactic coordinates using *astropy* (Astropy-Collaboration et al. 2018, Astropy-Collaboration, Robitaille, et al. 2013) and FITS handling in Python) and contains the observed value. For example, the number density of H_2 at a pixel location of (i, j) can be accessed by the value in the pixel number density matrix at row i and column j.

The observational gas data gives number densities in the plane of the sky or the XY plane. We do not have accurate cloud depth measurements, therefore we

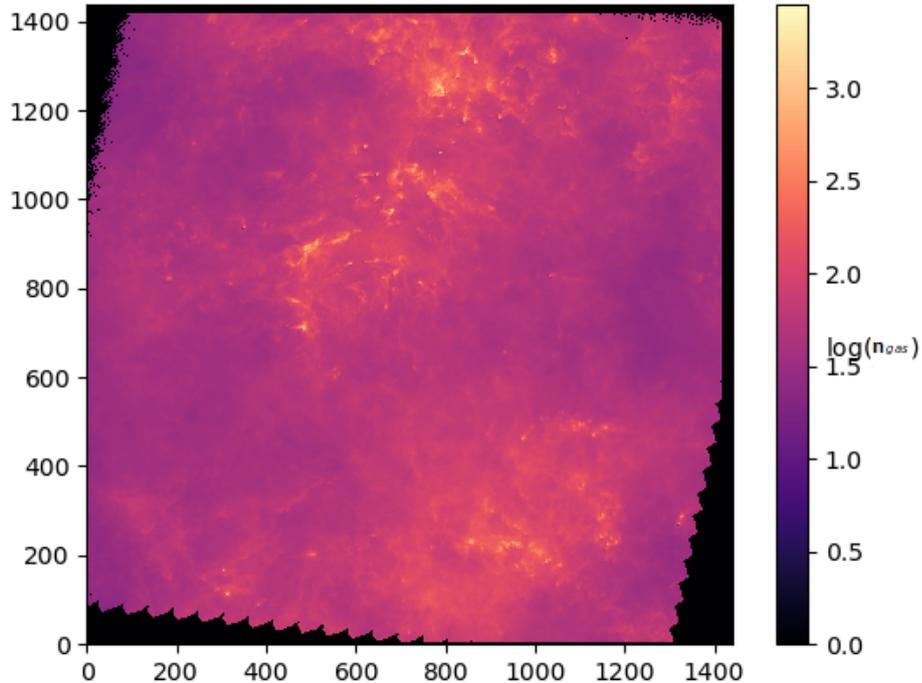


FIGURE 3.1: Carina Number Density Map in pixels. Orientation is due to the conversion between the FITS image array to a *numpy* array we use in Python.

consider different depth values (Z depths) and explore these estimates separately.

Our goal is to create SPH particles (relatively large, at least order $0.001 M_{\odot}$, compared to the size of an H_2 molecule) with 3D positions and masses which resemble the original number density distribution in the plane of the sky (XY plane) using the observations and an estimate for the cloud depth (Z coordinates).

We consider two methods to obtain the SPH particle distribution in the XY plane. The main idea behind Method I involves creating SPH particles by normalizing the number density matrix and then randomly sampling from this normalized

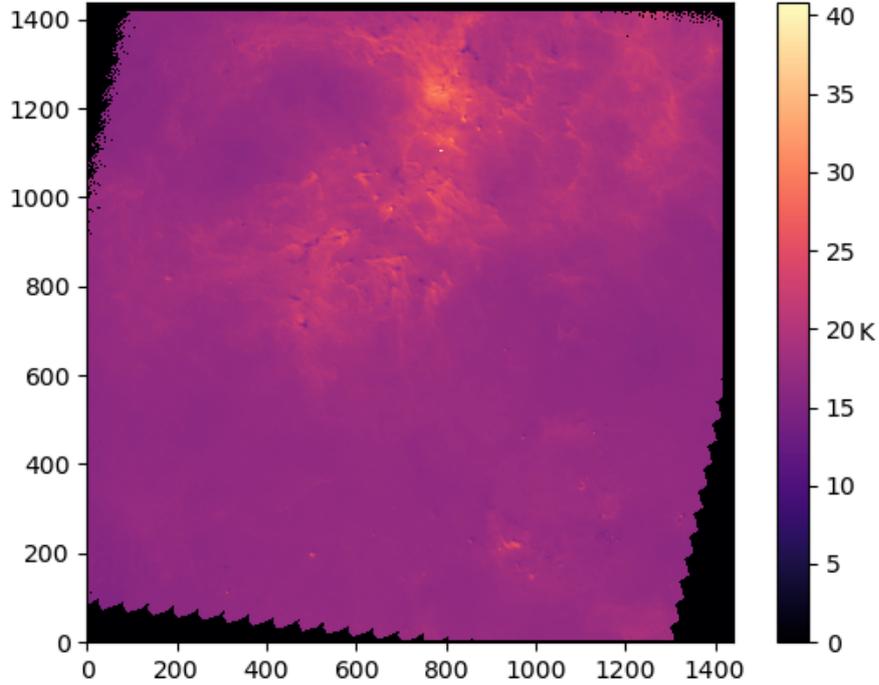


FIGURE 3.2: Carina Mean Line of Sight Temperature Map in pixels. Orientation is due to the conversion between the FITS image array to a *numpy* array in Python.

distribution to give selected pixels. The number of particles created for each pixel would correspond to the number of times the pixel was selected from this sampling. The particle XY positions (requiring a conversion of pixel coordinates to Cartesian coordinates) and masses (related to the number density of the represented pixel) would then be set. The Z coordinates would be assigned based on an estimate. As well, the temperature of each particle (and hence the internal energy) would then be assigned based on the corresponding pixel value in the temperature matrix.

Method II is similar except it involves creating one particle to represent each

pixel where the variations in the particle masses will depend on the original gas number density distribution. As well, our simulations operate in Cartesian coordinates therefore a transformation from galactic coordinates is a necessary step. The details pertaining to these methods are discussed in section [3.3](#).

3.2 Simulated Observational Image

To assess the validity of our techniques to accurately describe the image data, we apply our methods on a simulated observational image where the true particles (and their attributes) that comprise the image are known. We would be comparing the particle set generated from our techniques to the known particle set to determine the accuracy of our methods. We use a mock 2D image of a gas cloud obtained from particles in a Gasoline (Wadsley et al. [2004](#)) simulation (Ward [2015](#)). This is seen in Figure [3.3](#).

The particles have masses, 3D Cartesian coordinates and 3D velocities while the image of the cloud seen in Figure [3.3](#) is given in the XY plane (plane of the sky). In this way, the cloud depth and velocity information is lost when converting the particle distribution to the 2D image. The goal is to apply our sampling techniques on the mock observations to create a particle set that describes the image data in terms of the number density in the XY plane. In addition, we can investigate the extent to which the cloud depth (Z) distribution can be estimated from features seen in the XY plane. We can compare our particle data set that we obtain from processing the image to the original initial particle distribution from which the image was created in order to assess the advantages and limitations of

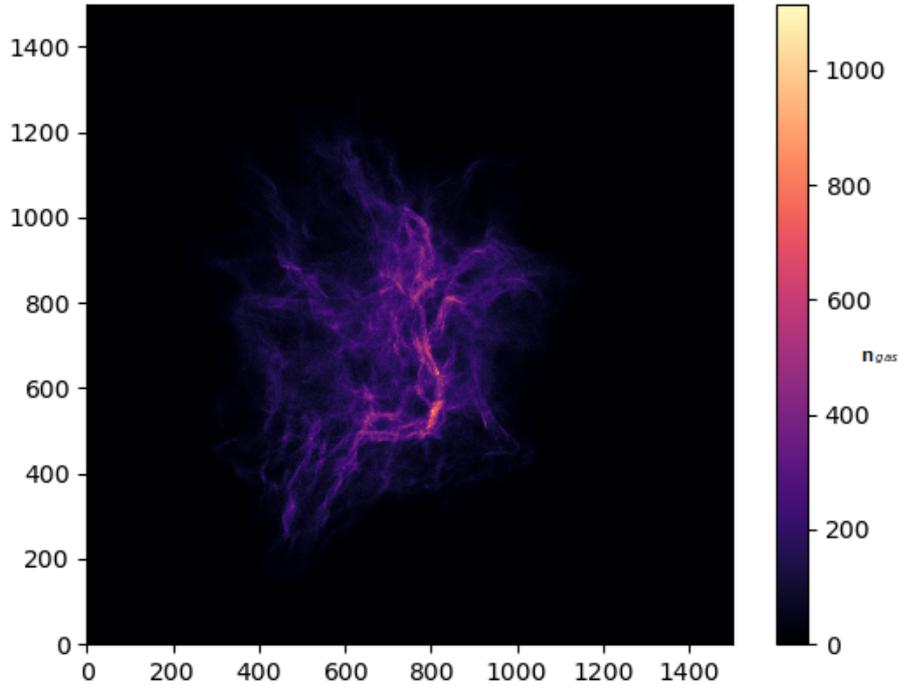


FIGURE 3.3: Simulated Observational Image of a Gas Cloud in pixels where the number of gas molecules per cm^{-3} is shown for each pixel. Data taken from Ward 2015 simulation data.

our techniques. Moreover, we can employ statistical methods to quantify these comparisons, the details of which can be seen in section 3.7.

3.3 Deriving XY particle positions

In this section, we discuss how the particle positions were determined to create a density distribution in the XY plane similar to the original data. Our approach is to match the original density distribution by assigning varying masses to the

SPH particles. There are two versions of our approach. Method I is based on randomly sampling the distribution a set number of times while Method II creates a particle for every pixel with a non-zero observed number density. Both versions rely on the same set of initialization steps. This involves creating matrices from the raw data that link the pixel coordinates to the world coordinates (galactic or equatorial) and converting the pixels to Cartesian coordinates. The conversion between the world coordinates and the pixels is handled using the *astropy* Python package (Astropy-Collaboration et al. 2018, Astropy-Collaboration, Robitaille, et al. 2013) while we explicitly code the conversion to Cartesian coordinates. The list of matrices required and common to both versions are as follows:

- 2D Number density matrix in pixel coordinates
- 2D Temperature matrix in pixel coordinates
- GLON/GLAT (or RA, DEC) matrix for the **centers** of the pixels
- GLON/GLAT (or RA, DEC) matrix for the **edges** of the pixels
- Cartesian matrix for the **edges** of the pixels
- Mass matrix for the pixels (total mass contained within each pixel)

The number density and temperature matrices are loaded from the raw image data where the (i, j) (row i and column j) pixel corresponds to the observed value. We create a matrix linking the world coordinates to the centers of the pixels in order to be able to retrieve $glon_{central}$ (or $RA_{central}$) and $glat_{central}$ (or

$DEC_{central}$) values. These correspond to the world coordinates of the central pixel of the observed image which is used in the conversion to Cartesian coordinates. The central values can usually be found in the FITS header, however initializing this matrix serves as a check and ensures that the subsequent matrices are created properly. The conversion from equatorial coordinates to Cartesian XY coordinates is as follows:

$$X = \frac{\cos(\delta) \sin(\alpha - \alpha_{central})}{\cos(\delta_{central})\cos(\delta) \cos(\alpha - \alpha_{central}) + \sin(\delta_{central}) \sin(\delta)} D \quad (3.1)$$

$$Y = \frac{\cos(\delta_{central})\sin(\delta) - \sin(\delta_{central})\cos(\delta)\cos(\alpha - \alpha_{central})}{\sin(\delta_{central})\sin(\delta) + \cos(\delta_{central})\cos(\delta)\cos(\alpha - \alpha_{central})} D \quad (3.2)$$

where α and δ are the RA and DEC equatorial coordinates, $\alpha_{central}$ and $\delta_{central}$ are the RA and DEC coordinates of the central pixel of the observed image and D is the distance to the observed region (1.5 kpc for the simulated image, 2.3 kpc for the Carina region). Equations 3.1 and 3.2 can also be used to convert from galactic coordinates to Cartesian coordinates by using $glon$, $glat$, $glon_{central}$ and $glat_{central}$ values. The next matrix we create links the world coordinates to the edges of the pixels which serves as an intermediate step to ultimately create a matrix linking the edges of the pixels to Cartesian coordinates. Again, the conversion from the world coordinates to Cartesian coordinates is performed using equations 3.1 and 3.2. We require the matrix linking the edges of the pixels to Cartesian coordinates in order to create the mass matrix. The mass matrix allows the total mass at pixel (i, j) to be retrieved which is then divided among the number of particles that end

up representing that pixel. The total mass in a pixel can be calculated by:

$$M_{pixel} = \rho_{pixel} A_{pixel} z_{pixel} \quad (3.3)$$

where ρ is the number density, A is the area in the XY plane and z is the depth of each pixel. The number density data is the 2D column density divided by a value of 20 parsecs (chosen by our collaborators Dr. Klaassen and Dr. Reiter), therefore z for all pixels is equal to 20 parsecs. The value of 20 parsecs was chosen by our collaborators Dr. Klaassen and Dr. Reiter as a characteristic depth for the Carina region specifically and we use the same depth for the simulated data. We always take z to be 20 parsecs when calculating the mass matrix in order for the total mass to be constant. We can then distribute the mass along the line of sight in different ways (e.g with varying z-depth) while the 2D density distribution in the plane of the sky (what is observed) remains the same. The conversion from pixels to Cartesian coordinates introduces distortion which changes the shape of the pixels from squares to rhombuses. Therefore the area of each pixel is calculated using Gauss's area formula (also known as the shoelace method) which can be used to compute the area of a quadrilateral (or other polygons in general) from the Cartesian coordinates of its corners. This is given by:

$$A_{pixel} = \frac{1}{2} \left| \sum_{i=1}^4 \det \left(\begin{pmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{pmatrix} \right) \right| \quad (3.4)$$

where i is a corner and the Cartesian coordinates for the edges of each pixel are

obtained from the matrix linking the two. Once the total mass in each pixel is calculated and the mass matrix created, particle masses are then assigned by dividing the total mass in a pixel by the number of particles that end up representing that pixel.

3.3.1 Method I

This version of our method is based on a weighted random sampling of the original number density distribution to create SPH particles (that are significantly larger than an H_2 molecule) that spatially resemble the original density distribution in the XY plane. To do this, another matrix describing the normalized number density distribution is required in addition to the aforementioned matrices. This normalized matrix represents weighted pixels which are then randomly sampled 100,000 times to create this many SPH particles. We chose to use this number of particles in order to make efficient use of computational power. The number of particles that end up in each pixel due to this weighted sampling process are then saved as a matrix. For each particle, the Cartesian XY coordinates are determined using the world coordinates of the corresponding pixel edges. The world coordinates for the pixels are rectangular with regular increments in the two dimensions which allows values to be randomly drawn along each pixel edge with ease. The Cartesian XY coordinates for a particle are calculated by first randomly picking a value in world coordinates along each edge of the corresponding pixel. This creates the world coordinates of the particle which is then transformed into Cartesian coordinates using equations 3.1 and 3.2. Assigning XY positions in this way allows for particles to be distributed randomly within a pixel should multiple

fall into the same one. Figure 3.3 shows the original simulated image and Figure 3.4 shows the original density distribution sampled 100,000 times.

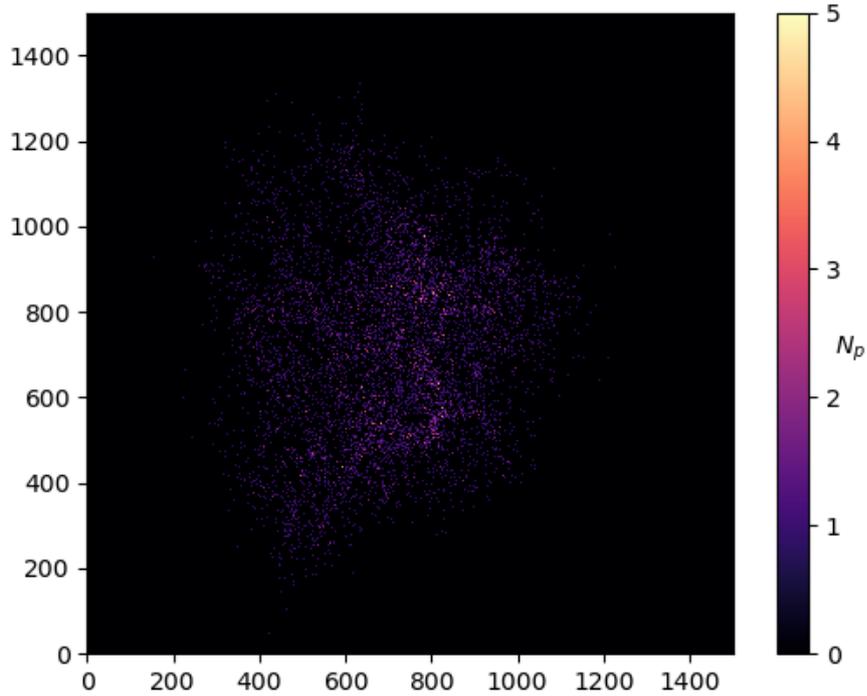


FIGURE 3.4: Distribution of 100,000 particles used to approximate the simulated observational image where the color bar corresponds to the number of particles representing each pixel (simulated image from Ward 2015).

The mass matrix is then calculated assuming a constant depth of 20 parsecs which can be seen in Figure 3.5. The mass of the particles is determined by dividing the corresponding mass in the matrix by the number of particles in each pixel from Figure 3.4. The particle masses range from 0.001 to roughly $0.07 M_{\odot}$ with most being around $0.006 M_{\odot}$.

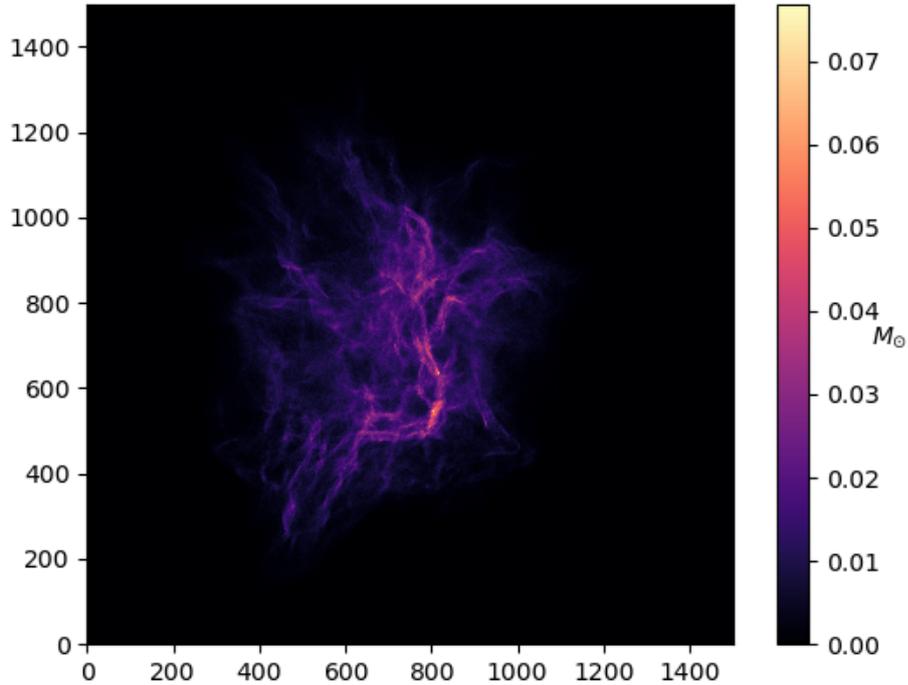


FIGURE 3.5: Total mass in each pixel of the simulated image (simulated image from Ward 2015).

As seen in Figure 3.4, randomly sampling 100,000 times is able to capture some of the large-scale features of the original distributions. However, a limitation of this is that there is no guarantee that all of the non-zero pixels will be sampled after sampling 100,000 times. As a result, we are not making use of all the available data. To improve this, we assign one particle to represent each pixel which is the basis of Method II of this technique.

3.3.2 Method II

To make use of all the available data, we represent each non-zero pixel with a particle. The total number of particles would then be on the order of 10^6 (an image is roughly 1000 x 1000 pixels) particles. By assigning particles in this way, we will be able to capture the finer features present in the original distribution. Initially, one particle represents each pixel. The particles are then assigned masses that are equivalent to the total calculated mass in each pixel from the mass matrix. All pixels are initially represented by one particle including pixels for which density data (and hence mass data) does not exist or is zero. The particles that do not have corresponding observational data are removed when the particle positions are assigned. If the mass of a particle exceeds $0.5 M_{\odot}$, more particles are introduced to represent that pixel in order to lower the mass of each particle. For example, if a particle is initially assigned a mass of $1.5 M_{\odot}$ then 2 more particles (for a total of 3) are put in to represent that pixel each with a mass of $0.5 M_{\odot}$. This is done to prevent a wide spread in particle masses. Once the particle masses are assigned, the mass-weighted distribution looks identical to that seen in Figure 3.5.

3.4 Z-depth assignment

We investigate different cloud depth (Z-depth) estimates and consider two cases. The first case is a constant depth and the second is a varied depth based on features seen in the 2D distribution.

3.4.1 Constant z depth

Since we do not have information about the Z depth, our collaborators Dr. Klaassen and Dr. Reiter assumed a depth of 20 parsecs centered at $Z = 0$ which is a value between a filament width of 0.1 pc and a low density cloud of 100 pc. In this way, we gave each particle a Z coordinate by uniformly sampling from -10 to 10 parsecs. Therefore the cloud density is uniform along the Z axis.

3.4.2 Thresholding with use of Distance Transform

As mentioned in Chapter 1, filaments have been observed in star-forming regions with the Herschel Space Observatory (André et al. 2010). This motivated our efforts in using the size of dense features in the XY plane as an estimate for the Z-depth of these structures to create filamentary structures. We began with a watershed segmentation implemented in Python (Pedregosa et al. 2011b). This type of segmentation takes user-defined “marker” values which represents basins. The markers are pixel values which in our case correspond to the 2D number densities. The segmentation algorithm begins flooding from the specified basins until basins corresponding to different markers meet along watershed lines. We then apply closing functions to remove any noise or “holes” in the resulting image. We perform two such segmentations. The first separates the non-zero regions from the background (marker value of zero corresponds to the background while marker values greater than zero are of interest). This serves to select the entire cloud region. The second segmentation involves selecting for the dense regions where the marker values less 200 cm^{-3} were considered to be the background while those greater than

200 cm^{-3} were of interest. After obtaining the segmentations and smoothing the results using the previously mentioned closing functions, we then obtained the distance transforms of these images (calculated using Python’s scikit-image, Walt et al. 2014). The distance transform gives the distance to the background (in number of pixels) for each pixel. These pixel distances are converted to physical distances using the Cartesian matrices previously mentioned (specifically taking the average of the distance in X and Y). Using these converted distances as z-depth values for each pixel introduces variations which mimic filamentary structures observed in clouds. The segmentations and distance transforms can be seen in Figures 3.6 and 3.7.

First, only the distance transform of the dense region (second segmentation) was used to set the z-depth values for the corresponding pixels. These z-depths were assigned by drawing from a range equal in size to the distance transform value and centered on $Z = 0$. For example, if a pixel had a distance transform value of 50 pixels which corresponded to a physical distance of 10 parsecs, then the z-depth would be assigned by drawing from -5 to 5 parsecs. Pixels that did not belong to the dense regions were given a z-depth that was approximately the size of the cloud in the XY plane. The size is approximately 8 parsecs and the z-depths were drawn from (-4, 4) parsecs. We then explored using both distance transforms. If a pixel only belonged to the non-zero region and not the dense region (selected in the first segmentation but not the second), the z-depth value was drawn from a range equal to the distance transform of the non-zero region (value from the first segmentation) and it was centered on $Z = 0$. If a pixel belonged to both, the z-depth was first **offset** by the value from the distance transform of the non-zero region which

is equally likely to lie above or below $Z = 0$. The z -depth value from the dense distance transform was then added on to this. This was done to prevent variations in Z concentrated about $Z = 0$ and to create slightly more realistic filamentary structures that are embedded in the entire cloud. To illustrate, if a pixel belonged to both and the distance transform of the non-zero region was 50 pixels (which for example corresponds to a distance of 10 pc) and the distance transform of the dense region was 10 pixels (for example corresponding to a distance of 2 pc), then the z -coordinate would either be $5 + (-1, 1)$ or $-5 + (-1, 1)$ due to the offset of 5 pc being equally likely to be above or below $Z = 0$. In our segmentations, there would not be any pixels that belong only to the dense region and not the non-zero region. The resulting distribution of these two methods compared to the true distribution of the simulated image data can be seen in Figure 3.8. The left panel shows that using the distance transform of the dense region results in a thin filamentary structure of gas elongated along the X axis embedded in a rectangular distribution of lower density gas. The filamentary structure has small variations in the Z -depth that are from the distance transform. It is an improvement on using a constant Z -depth as a filament is introduced, however more are needed to match the simulated data set seen in the right panel of the figure. The middle panel shows the use of two distance transforms which gives rise to filamentary structures that are symmetric about $Z = 0$ embedded in a cloud of less dense gas. The symmetry in the filaments is not realistic, however it does appear to better resemble the simulated data.

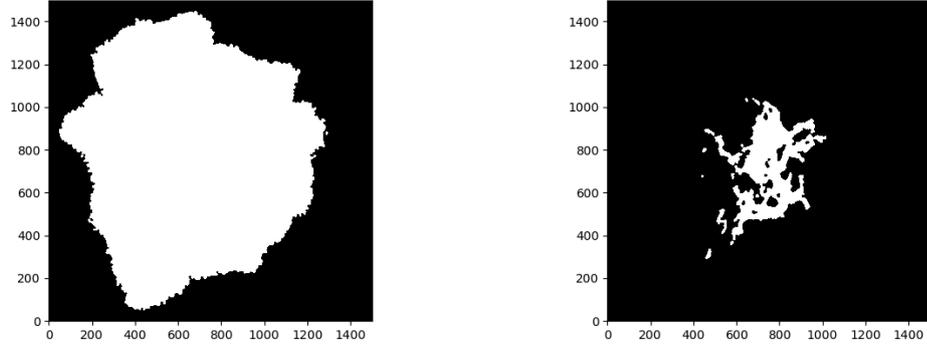


FIGURE 3.6: Segmentation of the fake data in terms of the non-zero pixel region (LEFT) and the high density region (RIGHT). Regions of interest are in white while the background is in black.

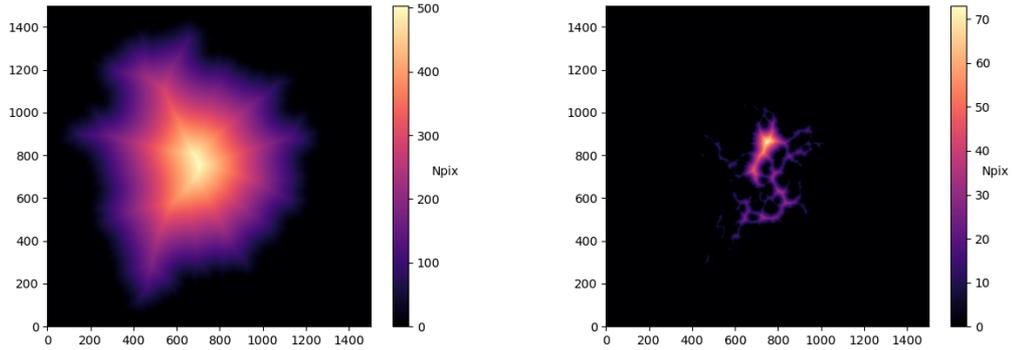


FIGURE 3.7: Distance transform of the non-zero region (LEFT) and dense region (RIGHT). Color bar corresponds to distances to the background measured in pixels.

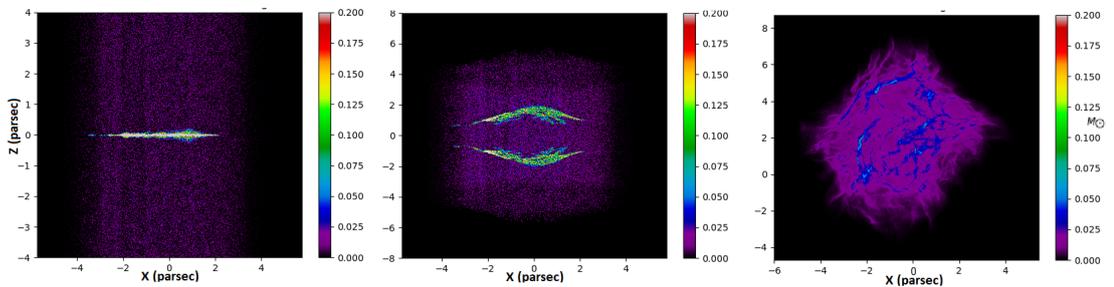


FIGURE 3.8: Comparison of XZ mass-weighted distribution (in M_{\odot}) from using the dense transform (LEFT), both non-zero and dense transforms (MIDDLE) and the simulated data set (RIGHT).

3.5 Internal Energies

The internal energies of the gas particles (all have radius of 200 AU \approx 0.001 pc) are calculated according to:

$$u = \frac{3k_B T}{2\mu} \tag{3.5}$$

where k_B is the Boltzmann constant, T is the temperature and μ is the mean molecular weight of the gas defined as $4.0 m_{proton}/(1.0 + 3.0 * 0.76)$ assuming the gas to be neutral and composed of 76% hydrogen, 23% helium and 1% dust. T is set to a constant value of 10 K (representative of a star-forming region) for the particles representing the simulated observed image. When this method is applied to the Carina region, the temperature of the particles is varied according to the corresponding pixel temperature seen in Figure 3.2. We model our gas using an adiabatic ideal gas equation of state.

3.6 Velocities

As previously mentioned, the velocity data is lost when the particle set from the Gasoline simulation was converted to an observational image. We cannot infer the velocity distribution from the 2D spatial information alone, therefore we assume the velocity dispersion to be Gaussian where we estimate a value for σ . When we model the Carina region, we use observations of dense gas cores to estimate a realistic σ . Since we do not have similar observations (or other information) to

use in combination with the simulated data, we take σ to be 1.4 km/s (the same as the value for the Carina region from observations) and we do not compare the velocities when testing the validity of our methods as we cannot make an informed estimate.

3.7 Validity of Our Methods

In this section, we test the validity of the methods we developed to generate particles from the simulated image data. We are interested in matching the spatial density distribution, therefore we compare the mass-weighted positions (in each of the X, Y and Z dimensions separately) of the particles created using our two techniques (Methods I and II) and Z estimates to the known particle set from which the image was made. We quantify these comparisons using two statistical techniques and a histogram comparison method.

The two statistical techniques used are the χ^2 test and the Kolmogorov-Smirnov (KS) test which can compare two 1D distributions. There are some that are strongly opposed to using the KS test in higher dimensions stating issues with defining a unique way to order the data points of the distribution functions to be compared (see Feigelson and Babu [n.d.](#) and references therein), which is why we limit our analysis to the 1D distributions. The two sample form of the χ^2 and Kolmogorov-Smirnov (KS) tests was conducted to determine if the output from the particle creation methods and the simulated data come from a common distribution. In addition to these statistical techniques, a histogram intersection

method was used which quantifies the overlap between two histograms. We first discuss the statistical techniques with the results summarized after.

3.7.1 χ^2

The known particle set of the simulated data is large (roughly 30 million particles). Therefore, it is less computationally expensive to compare mass-weighted positions that are binned. The χ^2 technique was selected as it is commonly used to compare histograms. The two sample formula is as follows (Press et al. 1992):

$$\chi^2 = \sum_i^k \frac{(k_1 R_i - k_2 S_i)^2}{R_i + S_i} \quad (3.6)$$

with scaling factors:

$$K_1 = \frac{\sqrt{\sum_i^k S_i}}{\sqrt{\sum_i^k R_i}} \quad (3.7)$$

$$K_2 = \frac{\sqrt{\sum_i^k R_i}}{\sqrt{\sum_i^k S_i}} \quad (3.8)$$

The cumulative distribution function (CDF) is a regularized gamma function given by:

$$CDF = \frac{\gamma(\frac{N}{2}, \frac{\chi^2}{2})}{\Gamma(\frac{N}{2})} = P(\frac{N}{2}, \frac{\chi^2}{2}) \quad (3.9)$$

The p value is then given by $1 - CDF$. This χ^2 test compares two samples R and S (may be unequal in size) and tests the null hypothesis where the two samples come from a common distribution. A high p value indicates that the null hypothesis cannot be rejected. In our case, we are testing whether our particle sets and the known set comes from a common distribution and a high p value is therefore desirable. The sums are over k bins with $k - 1$ (and minus any zero count bins in R and S) degrees of freedom. One limitation with this technique is that the χ^2 value is heavily dependent on how the data are binned.

3.7.2 1D KS Test

The 1D KS test was selected as it can be used to compare continuous data where the null hypothesis is that the two samples come from a common distribution. As a result, this test is not dependent on how the data are binned, however it is more computationally expensive than the χ^2 test. The test involves calculating the cumulative distribution function (CDF) of each sample and the D statistic according to:

$$D = \max_{-\infty < x < \infty} |F_1(x) - F_2(x)| \quad (3.10)$$

where $F_1(x)$ and $F_2(x)$ are two different CDFs to be compared.

The CDF is given by:

$$CDF = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} \exp(-2i^2 z^2) \quad (3.11)$$

The p value is then calculated using:

$$p(D > observed) = 1 - CDF(\sqrt{N_e}D) \quad (3.12)$$

where N_e is the effective sample size given by:

$$N_e = \frac{N_1 N_2}{N_1 + N_2} \quad (3.13)$$

N_1 and N_2 are the sizes of samples 1 and 2, respectively. The sum converges quickly as the sample size increases, therefore the sum was taken to be from $i = 1$ to $i = 1000$. The data being compared are mass weighted. Therefore, the CDF is the sum over the total mass. As well, N_1 and N_2 are each changed to be of the form (Monahan 2011):

$$N = \frac{(\sum_i W_i)^2}{\sum_i W_i^2} \quad (3.14)$$

where W_i 's are the weights.

3.7.3 Results

Table 3.1 summarizes the mass-weighted dimension being compared. In each case, the null hypothesis is that the two samples (from our method and the simulated data) are drawn from the same distribution. Each case will be examined using the χ^2 and 1D KS test. The critical p value at which the results reject the null

hypothesis is taken to be 0.05 where we desire the p value to be higher than this critical value in order to fail to reject the null hypothesis.

Test No.	Particle Technique	Dimension
1	Method I	X
2	Method I	Y
3	Method II	X
4	Method II	Y
5	Method II Z=20pc	Z
6	Method II mostly at Z=0	Z
7	Method II around Z=0	Z

TABLE 3.1: Summary of tests with different particle techniques along different dimensions.

Table 3.2 summarizes the results of the χ^2 and 1D KS tests.

Test No.	χ^2	bins	DOF	χ^2 pvalue	D statistic	z	KS pvalue
1	68.33	1000	972	1.0	0.0594	37.42	3.24E-210
2	122.28	1000	980	1.0	0.1029	64.82	0
3	1.26	1000	979	1.0	0.0025	1.57	0.014
4	1.62	1000	991	1.0	0.0053	3.34	4.16E-10
5	3920.81	1000	999	0.0	0.2867	180.59	0
6	5091.11	1000	994	0.0	0.3365	211.96	0
7	934.27	1000	999	0.93	0.0992	62.49	0

TABLE 3.2: Results of χ^2 and KS tests.

To compare the same bins for the χ^2 test, the maximum and minimum of the histograms were chosen to be the maximum and minimum values of that in the corresponding simulated data set (with the exception of the Z=20pc case where the max and min were taken to be -10 and 10 pc). The simulated image data contained 1501 x 1501 pixels. The number of bins was chosen to be 1000 to reduce the number of bins with small counts while still maintaining structure similar to the original image. The variation in the degrees of freedom (DOF) stems from the varying number of zero count bins in each case. The results from this test show that the null hypothesis cannot be rejected for tests 1-4. This is expected considering how Method I samples from the image data and how Method II purposely samples every pixel. In terms of variations in z-depth, only test 7 cannot reject the null hypothesis while tests 5-6 reject the null with p values of zero. The p value of test 7 is relatively high which is surprising considering that the z-depth was estimated from the size of features in X and Y. The extreme p values obtained could be due to how the data were binned. Although performing this test with 1000 bins is meaningful because it allows comparisons to be made on similar structure scales as the original image, it does not result in a range of p values across the tests.

The results from the 1D KS tests show that tests 1-7 all reject the null hypothesis which states that the compared distributions have a common distribution (i.e that the two different particle sets come from a common gas density distribution). This is true even for tests 3 and 4 that result in relatively low D statistic values. This is most likely due to the extremely large sample sizes considered. As sample size increases, the required D statistic to fail to reject the null hypothesis becomes increasingly small. To discern how well the distributions compare, we can instead

focus on the D statistic values themselves noting how tests 3 and 4 have the lowest values. This suggests that Method II is the optimal method for determining the mass-weighted positions of the particles along the X and Y axes. In addition, test 7 resulted in the lowest D statistic value among tests 5-7 which consider variations in the z-depth which supports the result from the χ^2 test.

3.7.4 Histogram Intersection Method

Another way of comparing the particle sets can be done using the histogram intersection method. Histogram comparisons are useful in image processing applications. Specifically, the intersection of two color histograms corresponding to two images can be compared to determine if they are the same object (Swain and Ballard 1991). This idea can be extended to the histograms used here.

The formula to calculate the intersection is:

$$Intersection = \frac{\sum_{i=1}^k \min(I_i, M_i)}{\sum_{i=1}^k M_i} \quad (3.15)$$

where I_i and M_i is the value of histogram I , M (model) at bin i and the intersection value ranges from 0 to 1.

Table 3.3 summarizes the results of applying this method to the test cases. The intersection method results in a greater spread of values compared to the other statistical tests. From this we see that tests 3-4 result in mass-weighted distributions that capture 99% of the true distribution (from the simulated data) in the X and Y directions. We also see that test 7 outperforms tests 5-6 and

captures 72% of the mass-weighted distribution along the Z direction. Thus, the combination of the two statistical tests and the histogram intersection method shows that Method II of our particle generation technique is best for creating particles in the X and Y dimensions and that Method II with particles placed symmetrically about Z=0 using the distance transforms is best for estimating the distribution in Z.

Test No.	Bins	Intersection
1	1000	0.21
2	1000	0.21
3	1000	0.99
4	1000	0.99
5	1000	0.44
6	1000	0.37
7	1000	0.72

TABLE 3.3: Results of Histogram Intersection Method.

3.7.5 Histograms used for the Statistical Tests and Intersection Method

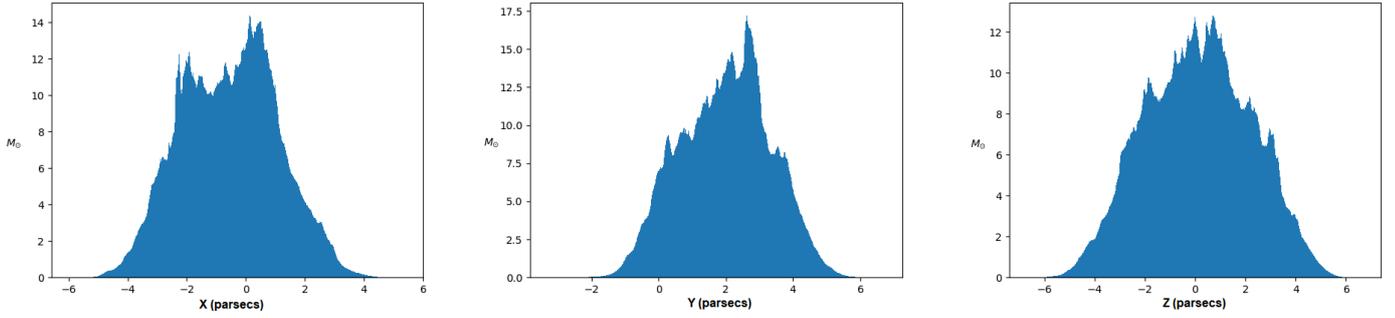


FIGURE 3.9: 1D mass-weighted histograms from simulation data

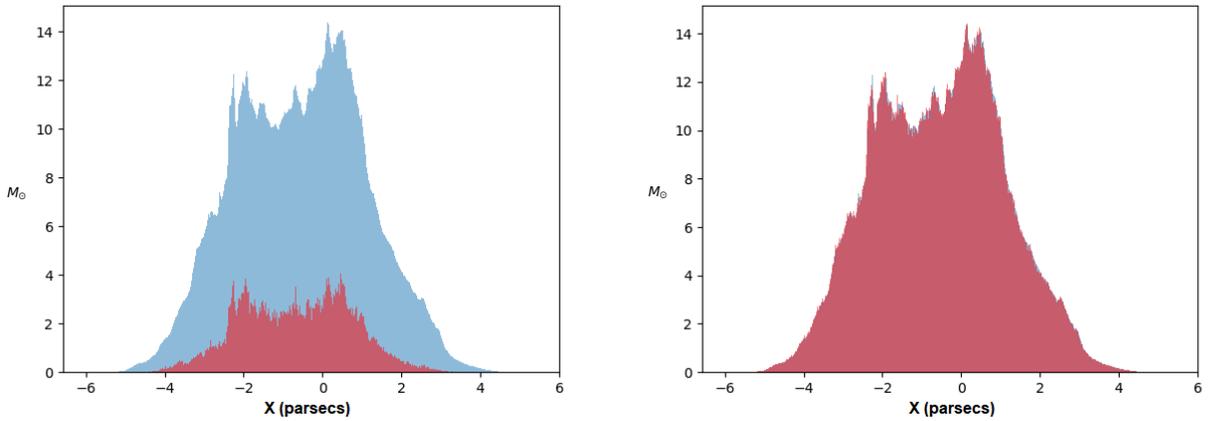


FIGURE 3.10: Examples of histogram intersections in X which are calculated to be the percent overlap between the simulation data (BLUE), and the results from our methods (RED) where a larger overlap is desirable. The comparisons of the mass-weighted X positions derived from Method I (left) and from Method II (right) are shown.

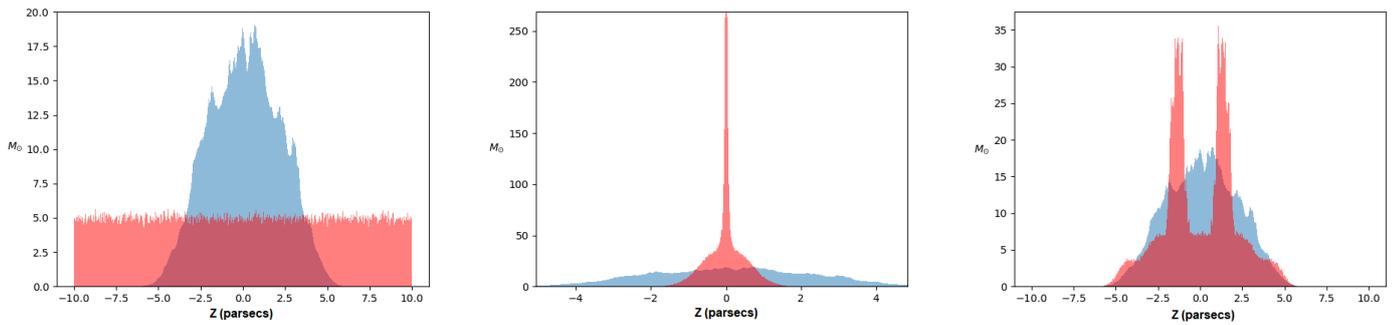


FIGURE 3.11: Mass-weighted Z histogram intersections which are calculated to be the percent overlap between the simulation data (BLUE), and the results from our methods (RED) where a larger overlap is desirable. The results from a constant 20 pc depth (left), the use of the dense distance transform (middle) and the use of the non-zero and dense distance transforms (right) are shown.

3.8 Summary

We developed particle generation techniques able to create SPH particles from observational gas data. Two methods were developed to assign particle positions and masses in the XY plane (plane of the sky). The first is a random sampling technique while the second involves each non-zero pixel being represented by one particle. In addition to this, we explored different estimates of the Z-depth of the cloud (this information is not included in the gas observations) including a constant depth of 20 parsecs and a varying depth based on dense features seen in the plane of the sky. We applied these techniques to a simulated observed image. Based on the validity tests (two statistical tests and a histogram comparison method), we find that the optimal procedure for creating particles that resemble the original density distribution involves each non-zero pixel being represented by one particle and the Z coordinates of the particles being symmetric about $Z=0$ and estimated from the size of dense features seen in the XY plane. This corresponds to Method II of our method and the Z coordinates being determined by calculating the distance transforms of the image in the XY plane. We use both versions of our method in our simulations of the Carina region to investigate further differences. We pair Method I with a constant z-depth of 20 parsecs and the resulting spatial distribution can be seen in Figure 3.12. We pair Method II with the Z coordinates calculated using two distance transforms which can be seen in Figure 3.14 where the dense regions correspond to pixels with values greater than 100 cm^{-3} . The resulting spatial distribution from Method II can be seen in Figure 3.13. The central region in the left panel of Figure 3.14 corresponds to the greatest distance and therefore the greatest range in Z which leads to the overall diamond shape

seen in XZ. The triangular features correspond to areas of relatively higher mass (seen in XY) with lower ranges in Z while the filaments are primarily due to the dense segmentation. The two segmentations were combined in a way to give rise to a symmetric distribution about $Z = 0$ as was done with the simulated Gasline data. Note that Figure 3.14 corresponds to the orientation of the gas observations seen in Figure 3.1.

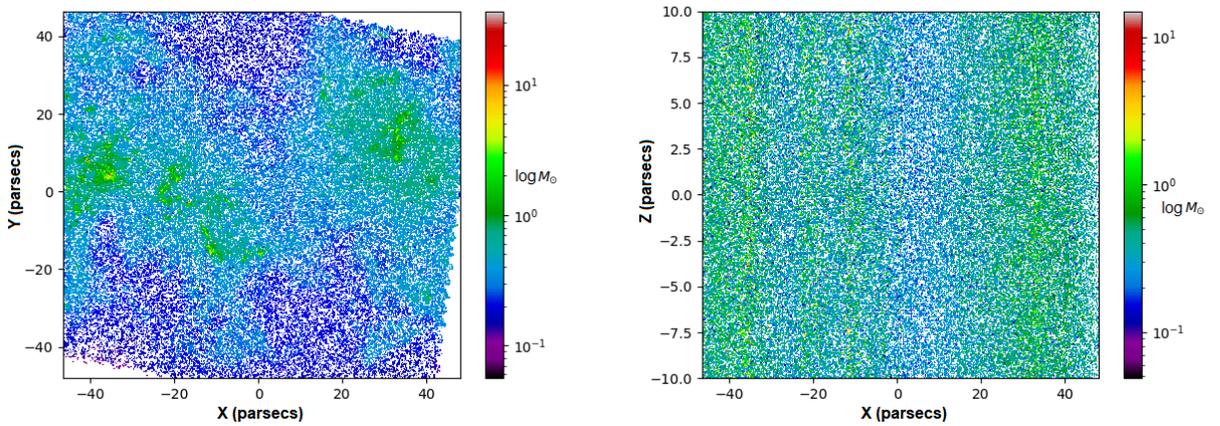


FIGURE 3.12: Mass-weighted (in M_{\odot}) Cartesian XY (left) and XZ (right) of particles created using Method I to represent gas densities in the observed Carina region. Gas particles range in mass from roughly 0.07 to $15 M_{\odot}$ with most having masses of roughly $0.5 M_{\odot}$. (Note: 1000 bins in each of the dimensions were used.)

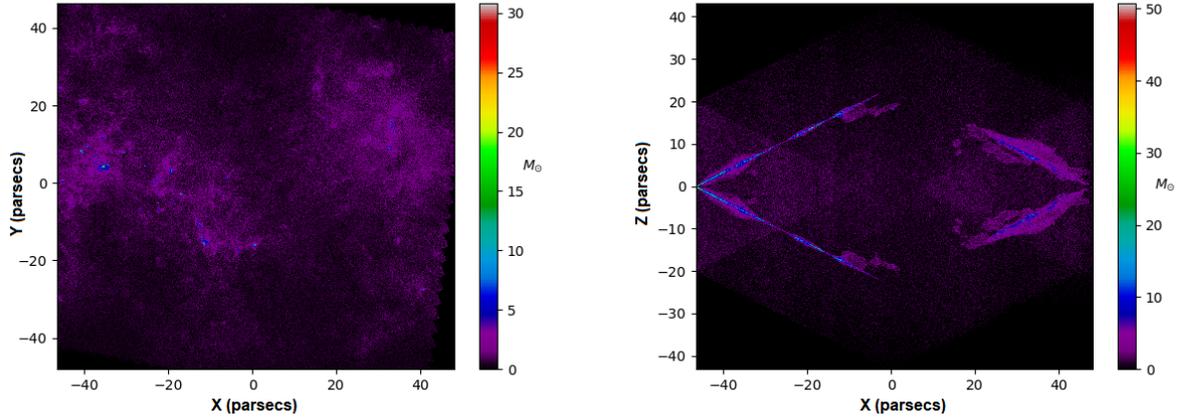


FIGURE 3.13: Mass-weighted (in M_{\odot}) Cartesian XY (left) and XZ (right) of particles created using Method II to represent gas densities in the observed Carina region. Maximum mass of a gas particle is $0.5 M_{\odot}$. (Note: 1000 bins in each of the dimensions were used.)

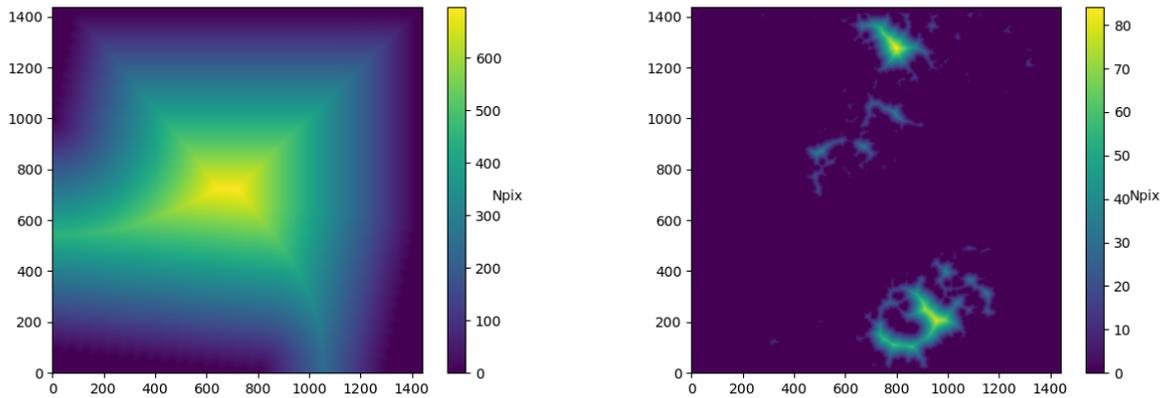


FIGURE 3.14: Distance transform of the non-zero region (LEFT) and dense region (greater than 100 cm^{-3}) (RIGHT). Color bar corresponds to distances to the background measured in pixels.

Chapter 4

Results

In this chapter, we summarize the results from our simulations. We investigated two methods of obtaining the positions of the gas particles in the XY plane (in the plane of the sky based on observational data) in conjunction with two methods of estimating the Z coordinate (depth along the line of sight). Method I (methods described in Chapter 3) obtains the X and Y coordinates of the gas particles using a random sampling method that samples from the original gas column density distribution 100000 times. Method II obtains these coordinates by purposefully representing all non-zero density pixels by at least one particle which makes use of all the available information in the plane of the sky. In our simulations, we pair Method I with a constant Z depth of 20 parsecs and we pair Method II with a varied Z depth created from calculating the distance transforms of the non-zero and dense regions. We pair Method I with a constant Z depth since this method was developed first and we wanted to explore the assumption of a constant cloud depth. We pair Method II with a varied Z depth which are techniques later developed found likely to be more representative of the Carina region (see Validity section

in 3). Other pairings are indeed possible, however our analysis is limited to these cases at this time. Using Method I, we also investigate the effect of varying the initial gas velocity dispersion. We compare the sets of simulations to determine the effect of different initial gas distributions and conditions on the resulting evolution of the star cluster system. Specifically, we consider how the stars cluster together over time using DBSCAN (see analysis tools in Chapter 2), how and if significant mass segregation occurs within the clusters as they merge (by calculating Λ_{msr}), how the gas clumps together and whether high enough densities are reached to fuel new star-formation (using a threshold of roughly $129 \pm 14 M_{\odot}/pc^2$ (Heiderman et al. 2010)) by analyzing the gas distribution using dendrograms created with *astrodendro*.

4.1 Summary of Simulations

We are interested in how the interaction between the stars and gas affects star cluster structure, therefore we focused on the overlapping regions between the stars and the gas. For this reason, we apply cuts to both the stellar and gas distributions which also contributes to efficient use of the computational resources. To this end, the gas distribution in a 40 parsec radius (in XY) from the cluster center was used. The stellar distribution was also modified to only include those within the Z depth of 20 parsecs in the gas. The set of simulations can be seen in Table 4.1. The same stellar distribution was used for all runs in order to determine the effects of the varying gas distribution. In addition to runs where the initial conditions of the gas are derived from observations of the Carina region, we also ran a simulation

where the gas is initially distributed in a uniform sphere with a constant initial temperature of 19K (the average temperature from the Carina observations). In doing so, we can compare the effect of varying the initial gas distribution spatially and determine how or if that affects the resulting evolution of the cluster.

Run	XY Method	Z-depth	NStars	NGas	Gas σ	Gas Mass (M_{\odot})	Stellar Mass (M_{\odot})
1	Method I	20pc	44101	53249	1.4	25846	27113
2	Method I	19K Uniform Sphere	44101	53249	1.4	25846	27113
3	Method I	20pc and 19K	44101	53249	1.4	25846	27113
4	Method I	20pc	44101	53249	2 scaled	25846	27113
5	Method I	20pc	44101	53268	2	25895	27113
6	Method I	20pc	44101	53125	5	25957	27113
7	Method II	Varied	44101	1106330	1.4	351016	27113
8	Method II	Varied	44101	1106330	1.4	351016	27113

TABLE 4.1: Summary of runs.

We chose to use three gas velocity dispersion values of 1.4, 2 and 5 km/s. The 1.4 km/s value comes from observations of dense cores in the Carina region (see Chapter 1). The other two values were chosen to explore both a small and large

increase in the dispersion. In Table 4.1, note that run 4 is a scaled version of run 1 in terms of the gas velocity distribution. The gas particle velocities in run 1 were multiplied by 2/1.4 km/s to create initial conditions with a gas velocity dispersion of 2 km/s used in run 4 (with all other conditions equivalent to run 1). Run 5 also has a gas velocity dispersion of 2 km/s, however this is with a new set of gas particles obtained from the random sampling method. Run 8 differs from run 7 in terms of the interaction timestep, the details of which are explained later. All Method I runs were evolved up to 7.7 Myr while the Method II runs were evolved up to a maximum of 4.6 Myr (due to the increased computation time from the large increase in gas particles). The substantial increase in gas particles from the Method I runs to the Method II runs stems from assigning at least one particle to represent each pixel with more introduced if the particle mass exceeded $0.5 M_{\odot}$. The substantial increase in the gas mass from the Method I runs to the Method II is again due to each observed pixel being used.

4.2 Method I Run I

This simulation used the random sampling method, a constant Z depth of 20 parsecs, varied gas temperature from observations and a gas velocity dispersion of 1.4 km/s. The dynamical evolution of the stars and gas can be seen in Figures 4.1 and 4.2. The stars are depicted as points while the gas distribution is a 2D histogram (surface density plot) that shows the density of the gas in units of $\log M_{\odot}/pc^2$. The surface densities were obtained by binning the gas particles using two bins per parsec (each bin increments by 0.5 pc in X and 0.5 pc in Y)

along each axis with weights corresponding to the gas mass per square parsec. A Gaussian interpolation was used to plot the gas distribution which smooths out the distribution between bins and allows large scale differences in density to be more clearly seen. Between $t = 0$ Myr and $t = 2$ Myr, slight contraction and expansion of the stellar sub-clusters can be seen in Figure 4.1 as the clusters attract each other and begin to merge. They begin to group together somewhat along a line in the Y direction (seen at $t = 3$ Myr) which resembles the initial stellar distribution. By $t = 5$ Myr, the clusters at the ends of the linear distribution have begun to merge with those that are near the center. By $t = 7.7$ Myr, the stars are centrally concentrated but not quite spherical as there still appears to be some elongation in the Y direction. In the XZ plane, (Figure 4.2) the stars are initially in a cross-like distribution with the gas between $Z = -10$ and $Z = 10$ parsecs. By $t = 7.7$ Myr, the stellar distribution is centrally concentrated and the gas has spread out reaching roughly $Z = -20$ to $Z = 20$ parsecs.

We analyze the evolution of the stellar and gas distributions separately. To characterize and track the regions of relatively dense gas, we use dendrograms which are a tool to describe hierarchical structures. This can be seen in Figure 4.3 where the contours are the structures (branches or leaves) of the dendrogram that represent different densities.

Initially the gas is clumpy and fragmented. The high density regions (in red which correspond to densities of $100 - 1000 M_{\odot}/pc^2$) lie approximately between $X = -50$ to 0 and from $Y = -20$ to 20 parsecs. The densest region is located roughly at $(-35, 5)$ parsecs and is clumpy while regions located at roughly $(-20, 5)$ and $(-10, -15)$ appear to be slightly more elongated and filamentary. In

general, the dense regions appear to be surrounded by less dense gas (seen as blue and white contours representing densities of 6 - 10 M_{\odot}/pc^2 and 10 - 100 M_{\odot}/pc^2 , respectively). Together, this forms the main filamentary structure of gas approximately 20 parsecs in width which appears to curve downwards (negative Y direction) and to the right (positive X direction). There are also clumps of low density gas (6 - 10 M_{\odot}/pc^2) located between X = -50 to -30 parsecs and Y = 20 to 40 parsecs that are relatively separated. As well, there are regions of less dense gas (blue and white contours) located at X = 15 between Y = 0 and 30 parsecs which are clumped together and borders the 40 parsec radius cutoff we used. Compared to Figure 4.1 image (A), we can see that the lower end of the stellar distribution overlaps with the dense region at (-20, 5) parsecs (the middle of the filamentary gas structure) but in general, the stellar distribution does not follow the gas distribution. In fact, a significant portion of the stars are located in regions with gas densities lower than 6 M_{\odot}/pc^2 .

Over time, the main filamentary gas structure persists until $t = 2$ Myr. At this point, the dense feature in the middle of the main filament is no longer present. From $t = 3$ Myr to $t = 4$ Myr, we can see the main filament fragmenting into two structures. One structure appears oval in shape with the center corresponding to the most dense region while the second structure appears more filamentary and extended in the Y direction. By $t = 5$ Myr, each of these structures ends up with high central densities of 100 - 1000 M_{\odot}/pc^2 . At this point, the less dense regions of gas that were separated above the main filament and clumped along the 40 parsec border at $t = 0$ Myr have more or less dispersed to form regions that are less than

$6 M_{\odot}/pc^2$. The two structures eventually merge to form one structure that is Y-shaped at $t = 7$ Myr where the high density central region (in red) is surrounded by the lower density gas (in white followed by blue) and by $t = 7.7$ Myr the high density region becomes more centrally concentrated. What was initially a clumpy distribution with high density regions that were spaced out results in a centrally concentrated distribution by $t = 7.7$ Myr. The distribution does not appear to be spherical as there are two regions extending from the main body of gas. Compared to the stellar distribution in Figure 4.1, we can see that the dense region of gas coincides with the central region of the stellar distribution which has also become more concentrated.

It is interesting to note that at all times, there is at least one dense region of gas. Within these dense regions, there are areas where the density exceeds $100 M_{\odot}/pc^2$ which can be seen in the orange and yellow colors. This implies that star formation could be present throughout the evolution of the system according to the density threshold of $129 \pm 14 M_{\odot}/pc^2$ (Heiderman et al. 2010). More checks would be required (e.g determining if the dense gas remains bound) similar to those used for simulations with sink particles (e.g Federrath et al. 2010) which use a 3D density criterion to determine with more certainty if a star is formed. At this time, we analyze our results in 2D (in the plane of the sky) to mimic an observational view and the 2D density threshold does appear to have been met at all times in this run.

We used DBSCAN in 3D (clustering algorithm described in Chapter 2) to characterize the changing sub-structure in the star clusters over time. This is seen in Figure 4.4 with 1 Myr increments where each identified cluster is represented by a

different color. Initially, there are 16 clusters identified with DBSCAN. Over time, the sub-clusters merge together and the number of clusters identified generally decreases with time until $t = 5$ Myr. By $t = 2$ Myr, several of the clusters from roughly $X = -28$ to $X = -14$ and $Y = 7$ to $Y = 30$ parsecs have merged together into one cluster. The clusters originally at $X = -20$ to $X = -15$ and $Y = -5$ to 5 parsecs at $t = 0$ Myr have also merged together into a cluster by $t = 2$ Myr. Further loss of substructure occurs as the stars merge together in the X direction. By $t = 5$ Myr, the clusters have merged together into one large main cluster and one small cluster. The main cluster is elongated in the Y direction which is similar to the orientation of the clusters at $t = 0$ Myr. From this point onwards, the number of clusters identified by DBSCAN increases, however there remains one large cluster of stars while the other clusters are relatively small. The smaller clusters are due to noise as they are not persistent. The large main cluster of stars does not form a spherical distribution by $t = 7.7$ Myr.

The other runs include variations in the initial spatial distribution of the gas and in the initial gas velocity dispersion. We will compare our results by focusing on the significant differences between the evolution of the gas and stellar distributions.

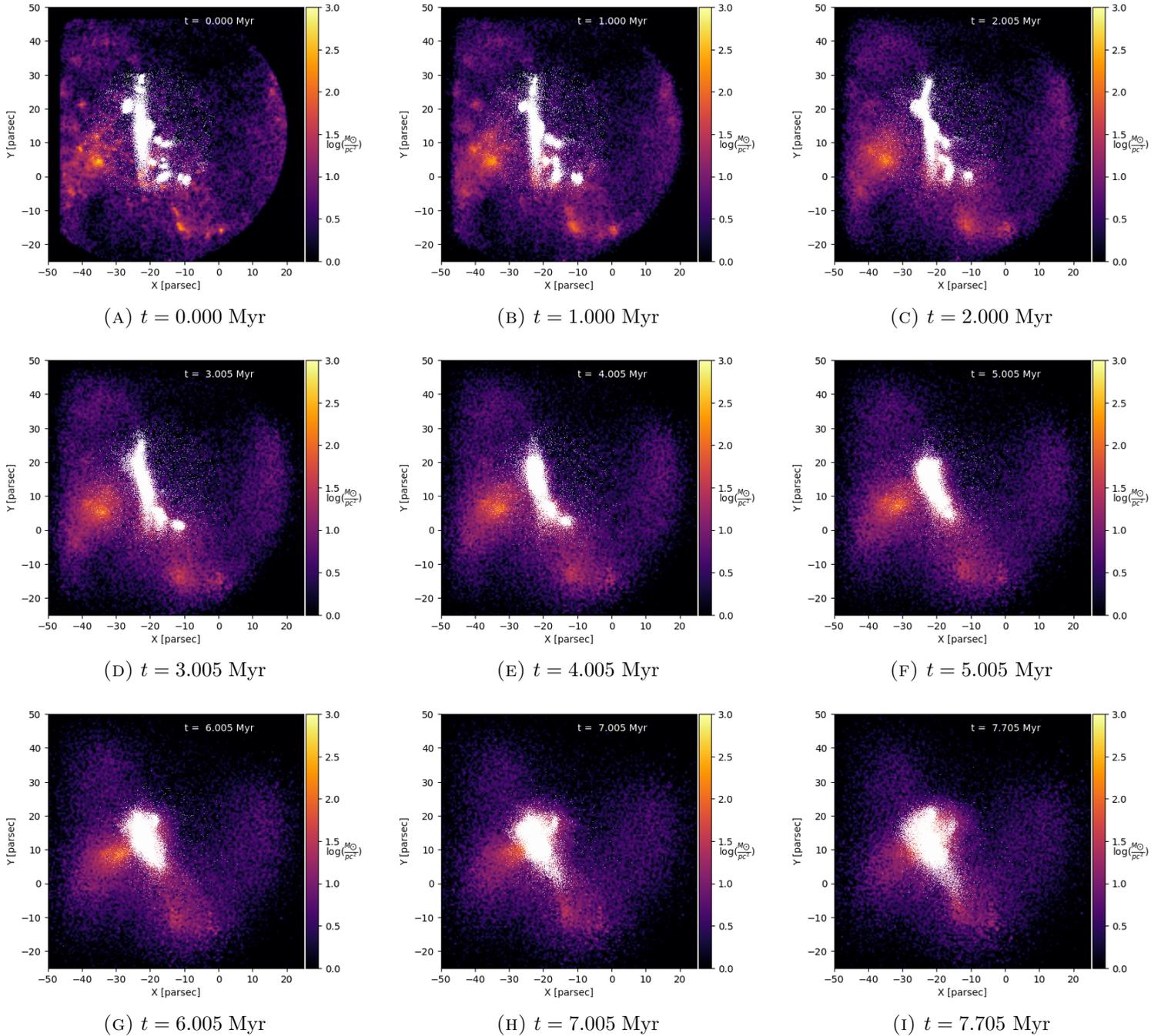


FIGURE 4.1: Evolution of Carina Region in XY plane with gas velocity dispersion of 1.4 km/s.

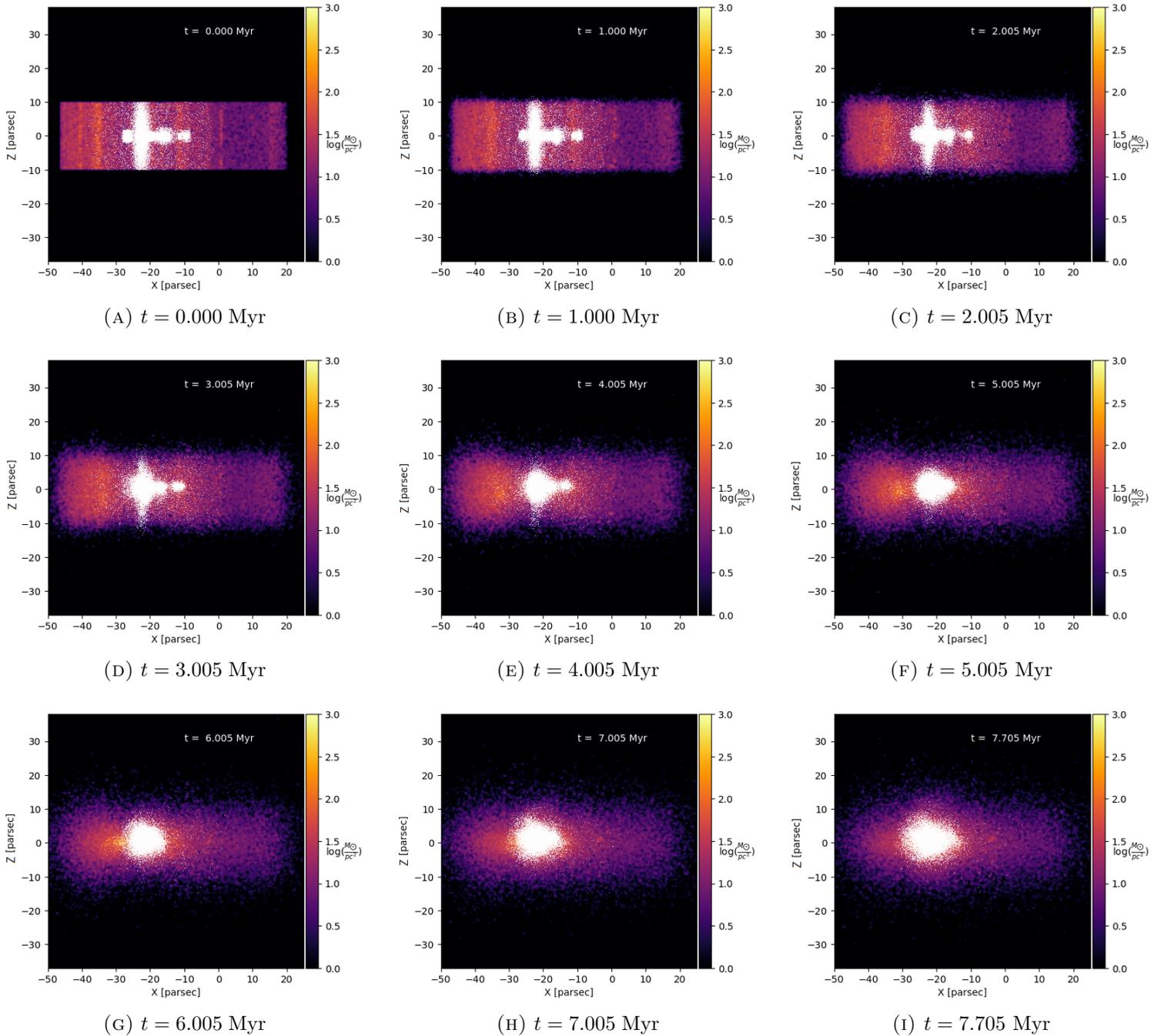


FIGURE 4.2: Evolution of Carina Region in XZ plane with gas velocity dispersion of 1.4 km/s.

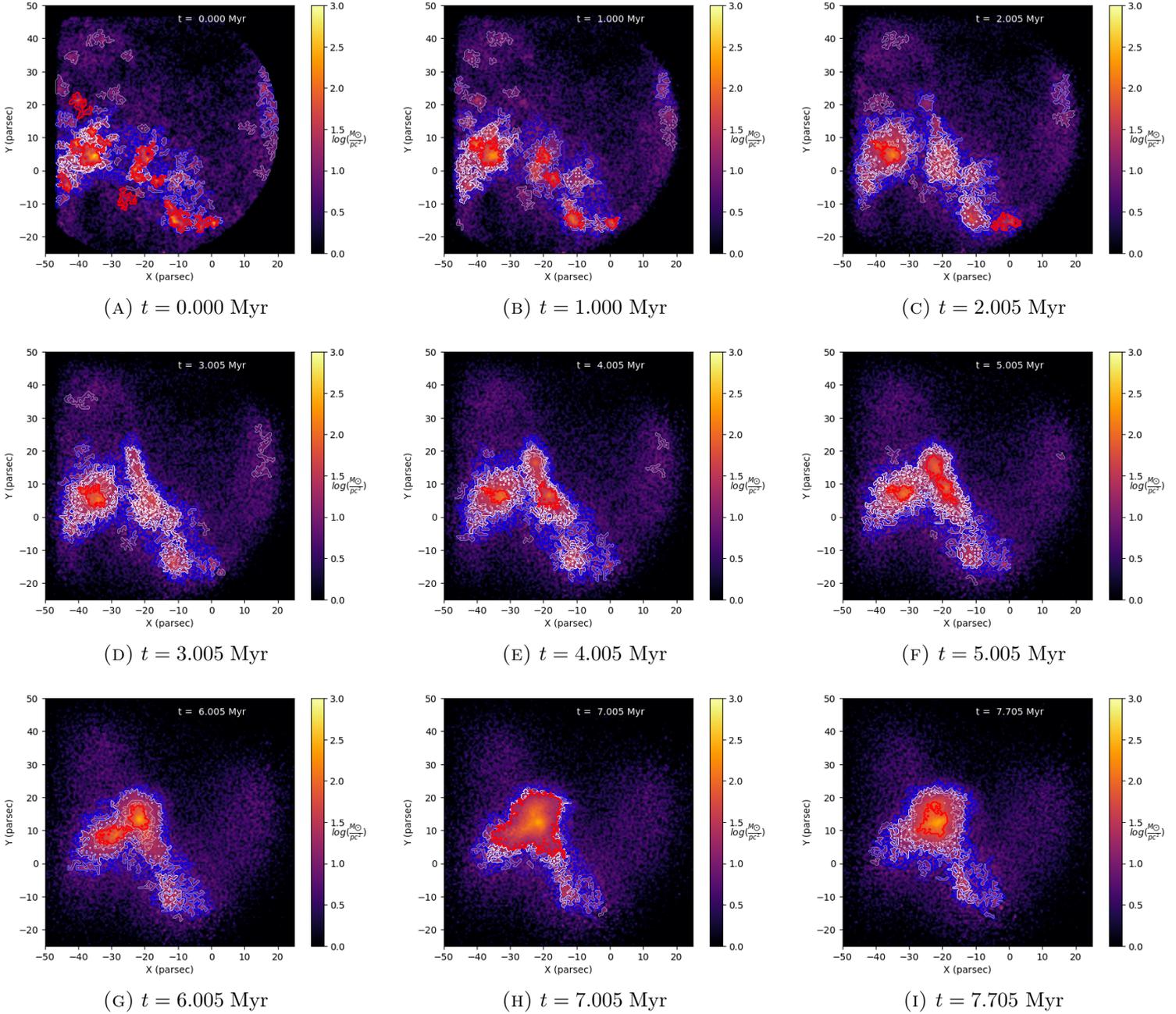


FIGURE 4.3: Gas distribution in Carina region over time with initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$.

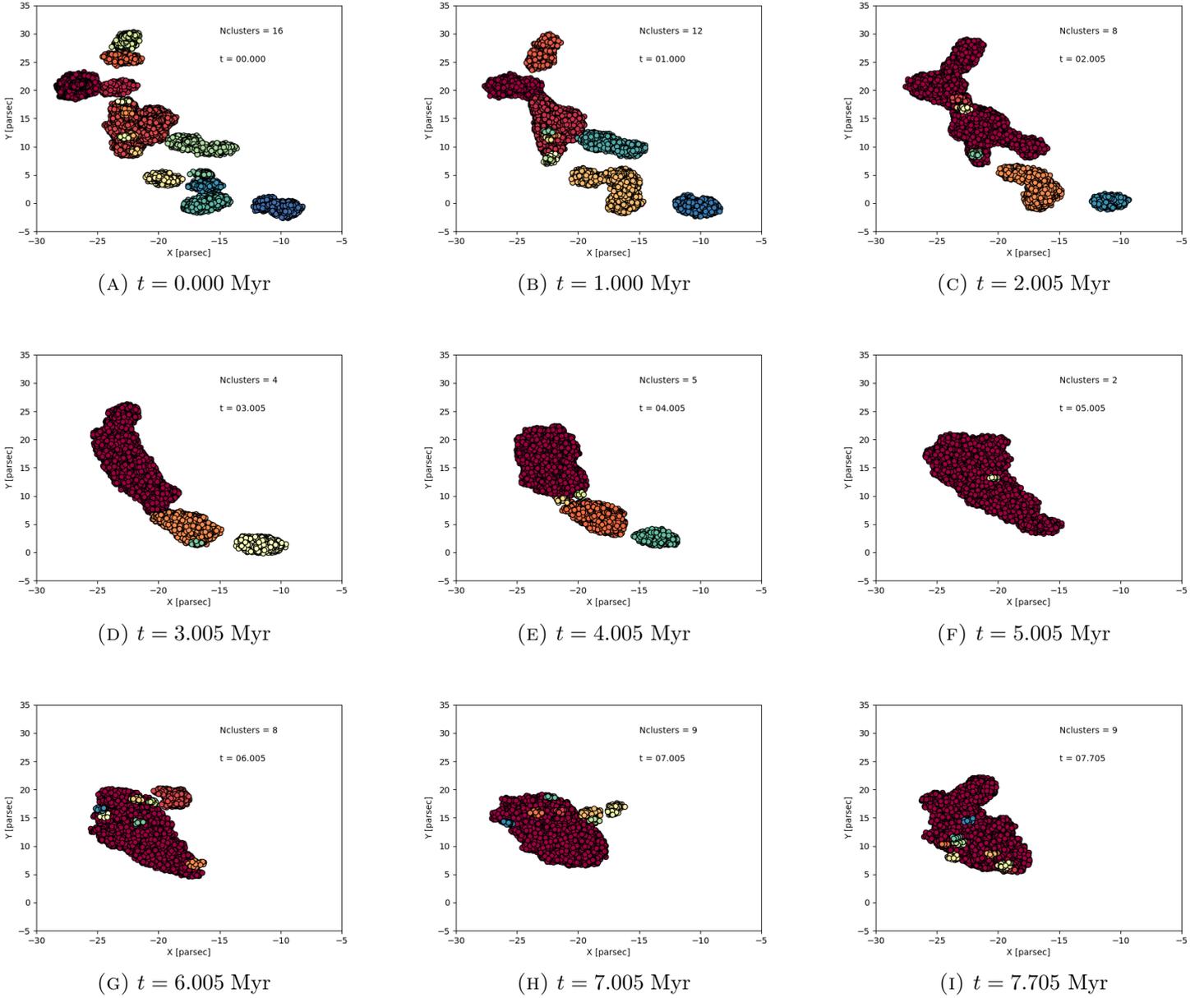


FIGURE 4.4: Clusters in Carina picked out by DBSCAN over time with gas velocity dispersion of 1.4 km/s. Each cluster is denoted by a different colour. The colours do not necessarily correspond to the same stars across the images as they are only used to highlight the different clusters identified.

4.3 Method I Run II

In this run, the gas is initially distributed in a uniform sphere of radius 40 parsecs, the gas temperature is set to 19K (which is the average temperature for the observed Carina region) and the initial gas velocity dispersion is 1.4 km/s. We chose to use a constant gas temperature of 19K since the sphere extends slightly into the regions where we do not have corresponding temperature measurements. The total gas mass and number of gas particles are the same as Run I. While varied masses were used in Run I (as per the Method I prescription outlined in Chapter 3), here we use a constant gas particle mass of approximately $0.49 M_{\odot}$. The purpose of this run is to investigate the effect of a simplified initial spatial distribution of the gas with an initial constant gas temperature. As previously mentioned, star formation simulations often begin with a spherical distribution of gas. This does not realistically describe molecular clouds which have been observed to contain filamentary structures. We want to compare the results of our methods which incorporate observational gas data to an initial spherical distribution of gas to see if any significant differences arise as the cluster evolves over time. The dynamical evolution of the stars and gas can be seen in Figures 4.5 and 4.6. We can see that overall, the change to an initial uniform sphere of gas did not have a significant impact on the evolution of the stellar distribution in the XY plane or the XZ plane compared to run I. They merge in a similar fashion at the same times as run I. Although the gas distribution was initially uniform, we can see that it becomes centrally concentrated by $t = 7.7$ Myr.

If we consider the dendrogram analysis seen in Figure 4.7, we can see how

initially there are regions of lower density gas mixed together (densities of 6 - 10 M_{\odot}/pc^2 and 10 - 100 M_{\odot}/pc^2 seen as blue and white contours). Even though the distribution of gas was set to be a uniform sphere, the variations arise due to the random distribution of particles. Between $t = 0$ Myr and $t = 2$ Myr, the gas distribution does not change significantly. Between $t = 3$ Myr and $t = 5$ Myr, we can start to see a region of relatively dense gas (still less than 100 M_{\odot}/pc^2) forming in the center. By $t = 7.7$ Myr, the gas distribution becomes centrally concentrated with the inner regions having a density between 10 - 100 M_{\odot}/pc^2 which is surrounded by lower density gas of 6 - 10 M_{\odot}/pc^2 (seen as white contours surrounded by blue). It is interesting to note that even though the gas becomes centrally concentrated over time, high density regions of gas (densities between 100 - 1000 M_{\odot}/pc^2) do not form. Hence, star formation is not a possibility (according to our density threshold) in this run.

We analyzed the stellar distribution over time using DBSCAN but found no large differences in the number of clusters identified compared to run I. We compare the number of identified clusters over time for all of our runs in section [4.10](#).

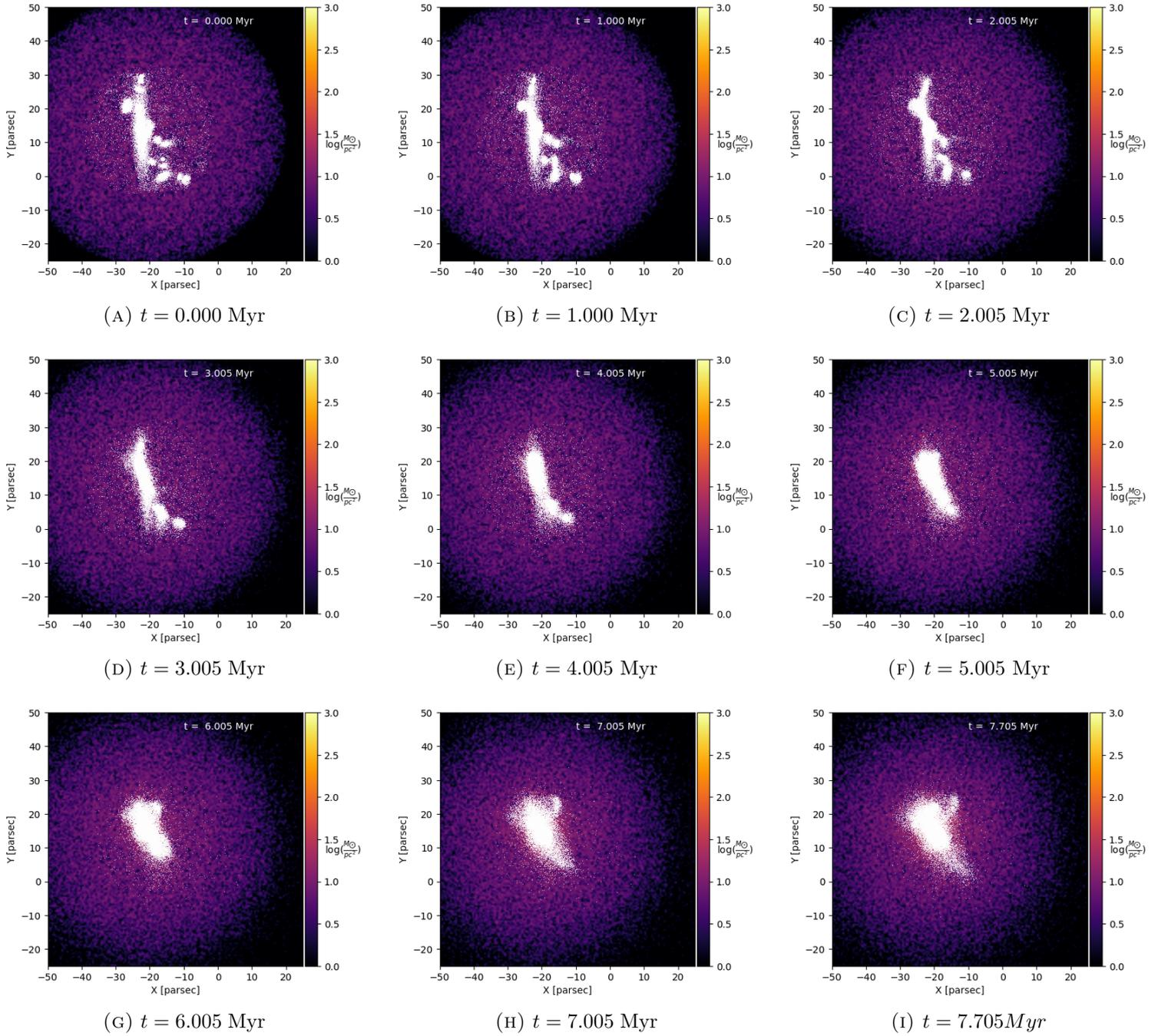


FIGURE 4.5: Evolution of Carina Region in XY plane with gas initially in a uniform sphere with a gas velocity dispersion of 1.4 km/s

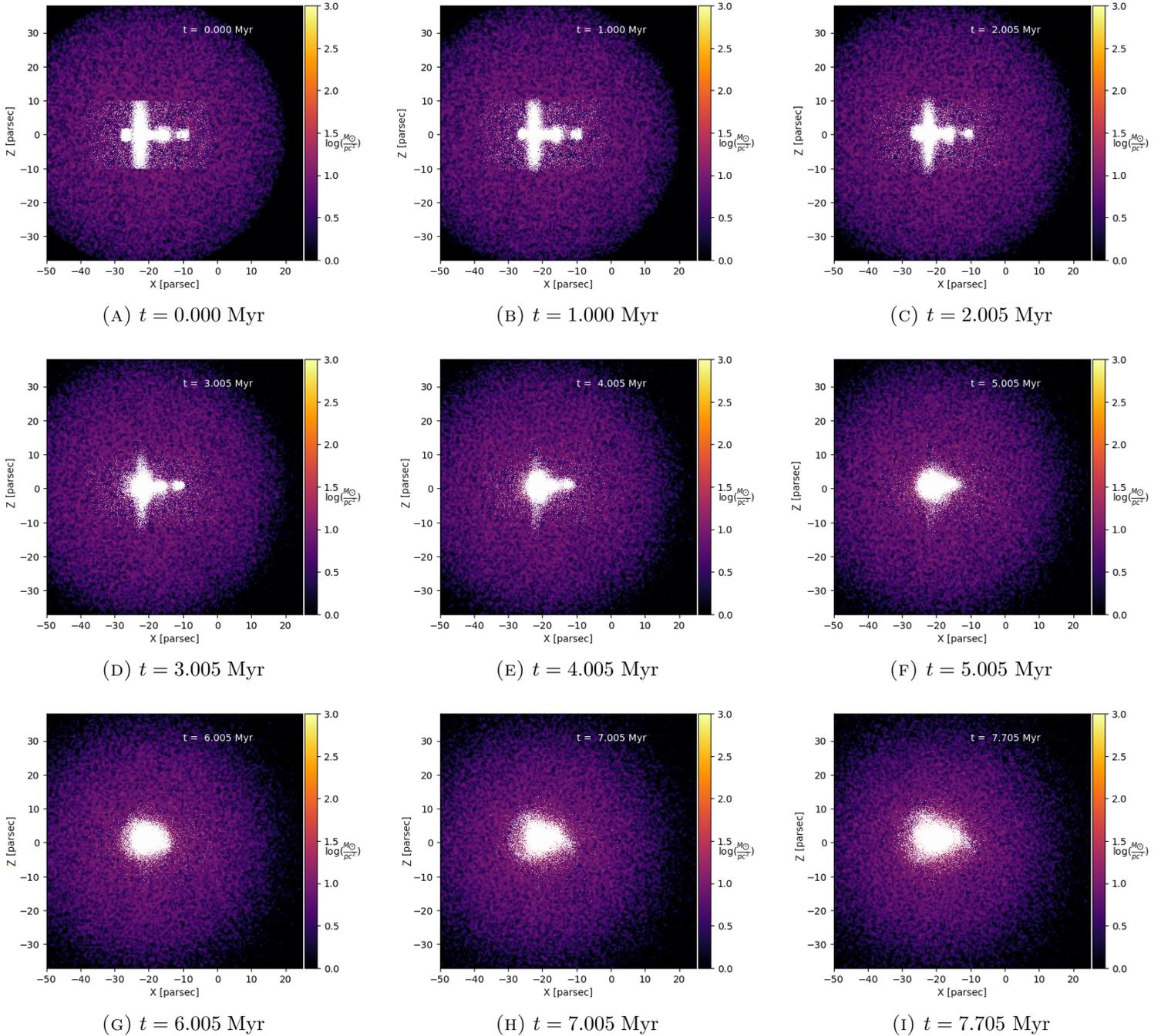


FIGURE 4.6: Evolution of Carina Region in XZ plane with gas initially distributed in a uniform sphere and with a gas velocity dispersion of 1.4 km/s.

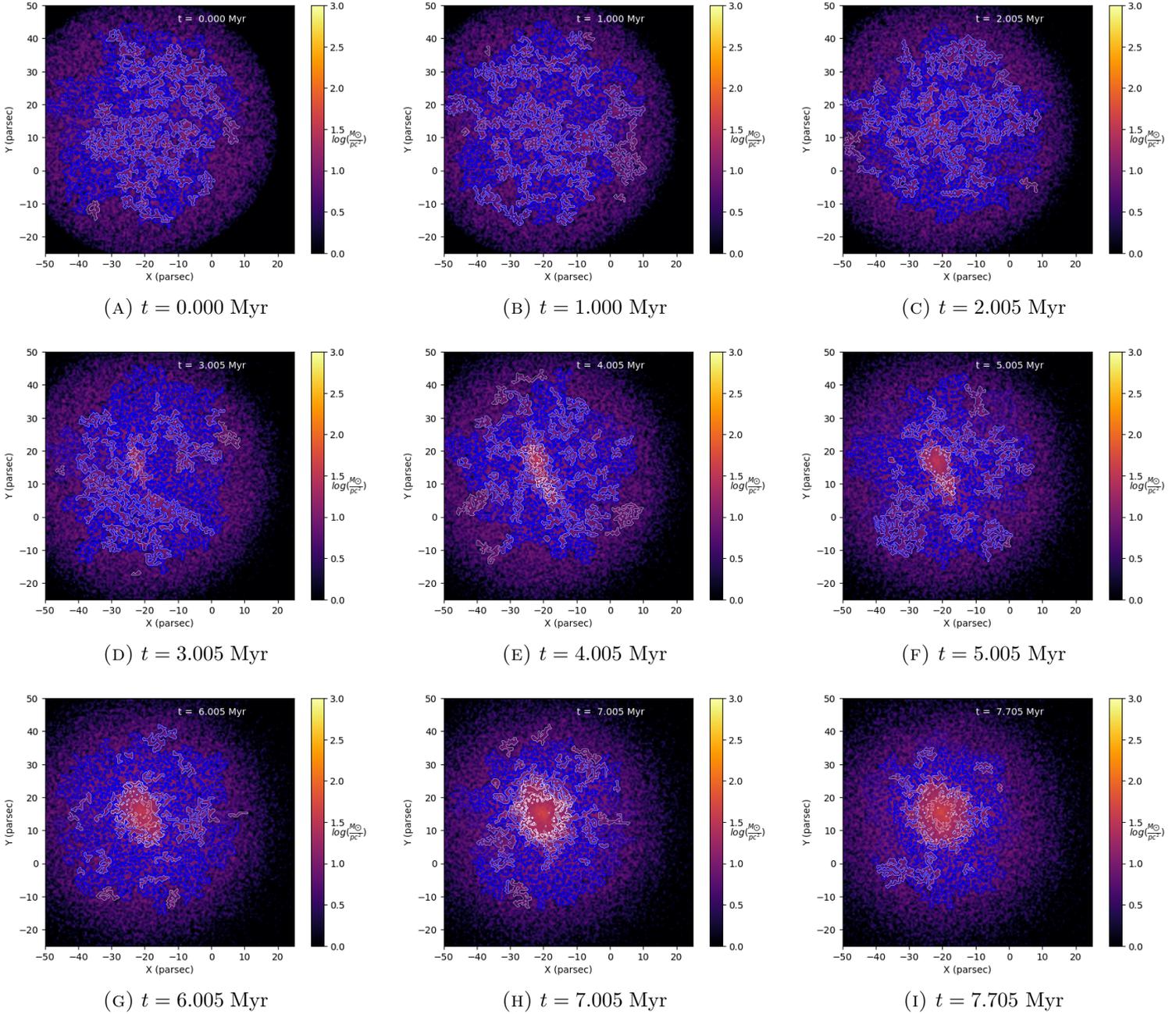


FIGURE 4.7: Gas distribution in Carina region over time with gas initially in a uniform sphere with initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities between 100 - 1000 M_{\odot}/pc^2

4.4 Method I Run III

This run used the random sampling method, a constant Z depth of 20 parsecs, a constant gas temperature of 19K (average temperature over the observed Carina region) and an initial gas velocity dispersion of 1.4 km/s. The initial conditions are similar to those in run II with the only difference being the initial density distribution of the gas. While run II considered simplifying assumptions about the initial gas density distribution and gas temperature, the purpose of this run is to consider the effect of only a different initial gas density distribution. We found that the evolution of the stellar distribution over time was very similar to run II in both the XY and XZ planes. We also found the evolution of the gas distribution over time to be similar to run I where two main structures formed from the gas which eventually merged together. As well, we found no large differences in the number of clusters identified by DBSCAN compared to run I. The results from these first three runs show that for an initial gas velocity dispersion of 1.4 km/s, the evolution of the stellar distribution over time is not significantly sensitive to a change in the initial gas density distribution specifically considering a distribution derived from observations compared to gas particles distributed in a uniform sphere. We find that incorporating a realistic initial gas distribution is advantageous for determining possible star-formation sites as high enough densities are reached. This was not the case when we considered the gas distributed in a uniform sphere. The next set of runs focuses on varying the initial gas velocity dispersion while using the initial gas density distribution derived from the Carina observations.

4.5 Method I Run IV

This run used the random sampling method, a constant Z depth of 20 parsecs, varied gas temperature from observations and an initial gas velocity dispersion of 2 km/s. The initial positions and masses of the gas particles are identical to those in run I. The evolution of the stellar distribution over time in the XY and XZ planes is similar to that seen in run I. If we consider the evolution of the gas distribution over time, the dendrograms appear to differ from run I at $t = 5$ Myr. This can be seen in Figure 4.9 (where we compare with run V as well). We can see that there exists two main high density regions (represented by the red contours) in runs I and IV at $t = 5$ Myr. However, the high density region at around $X = -20$ and $Y = 10$ parsecs appears to be smaller in this run compared to run I. At $t = 6$ Myr, the high density region around $X = -35$ and $Y = 10$ parsecs also appears to be smaller in this run compared to run I. At $t = 6$ Myr, there also appears to be more gas with densities between $10 - 100 M_{\odot}/pc^2$ (white contours) surrounding the dense gas in this run compared to run I. By $t = 7.7$ Myr, we can see that the gas distribution becomes centrally concentrated in both cases. However the Y-shaped gas structure is not as prominent in this run as the gas appears to be more tightly collected together. The results from the DBSCAN analysis did not show any large differences in the way the clusters merged together over time. Therefore, the results from this run show that an increase in the gas velocity dispersion from 1.4 km/s to 2 km/s decreases the size of some high gas density regions and results in a gas distribution that appears more spherical after $t = 7.7$ Myr.

4.6 Method I Run V

This run used the random sampling method, a constant Z depth of 20 parsecs, varied gas temperature from observations and an initial gas velocity dispersion of 2 km/s. It differs from run IV as a new particle set was generated using the sampling method which introduces a slight variation in the number of gas particles, the spatial distribution of the gas and the gas mass as seen in Table 4.1. Again, the overall evolution of the stellar distribution over time in the XY and XZ planes does not appear to be significantly different from the other runs. If we look at the dendrograms seen in Figure 4.8 describing the gas distribution, we can see that they differ at $t = 0$ Myr. There appears to be more clumps of the high density gas (red contours) here initially compared to run IV even though the main filamentary structure is still apparent. This is due to the sampling method we use which samples the observational gas distribution 100000 times. We expect some slight variation in the random sampling from trial to trial which is what we see here. The dendrograms comparing runs I, IV and V from $t = 5$ Myr onwards (this is where run IV started to significantly differ from run I) can be seen in Figure 4.9. We see that the number and size of the dense regions (in red contours) differs between this run and run IV which again is due to the slightly different initial conditions resulting from the sampling technique. However, the overall structure of the gas appears to be similar between these two runs where the high density regions are surrounded by the less dense gas (white and then blue contours) which form a Y-shape present at $t = 5$ Myr and $t = 6$ Myr. By $t = 7.7$ Myr, the Y-shaped gas structure is not as prominent as the gas becomes more centrally concentrated. Comparing these two runs to run I (where the initial gas velocity

dispersion was set to 1.4 km/s), we can see that increasing the velocity dispersion to 2 km/s results in smaller dense regions at $t = 5$ Myr, one main dense region at $t = 6$ Myr and a centrally concentrated distribution that is less Y-shaped at $t = 7.7$ Myr. We analyzed the stellar distribution over time using DBSCAN but found no significant differences in the clustering of the stars over time.

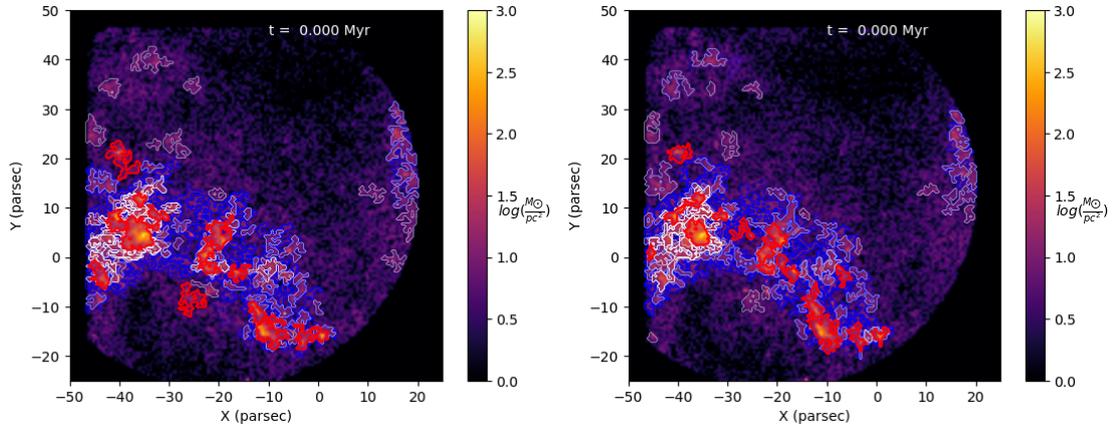


FIGURE 4.8: Comparison of gas distribution in Carina region between runs IV (left) and V (right) at $t = 0$ Myr which highlights the effect of random sampling in our particle creation technique (Method I). Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities between 100 - 1000 M_{\odot}/pc^2 .

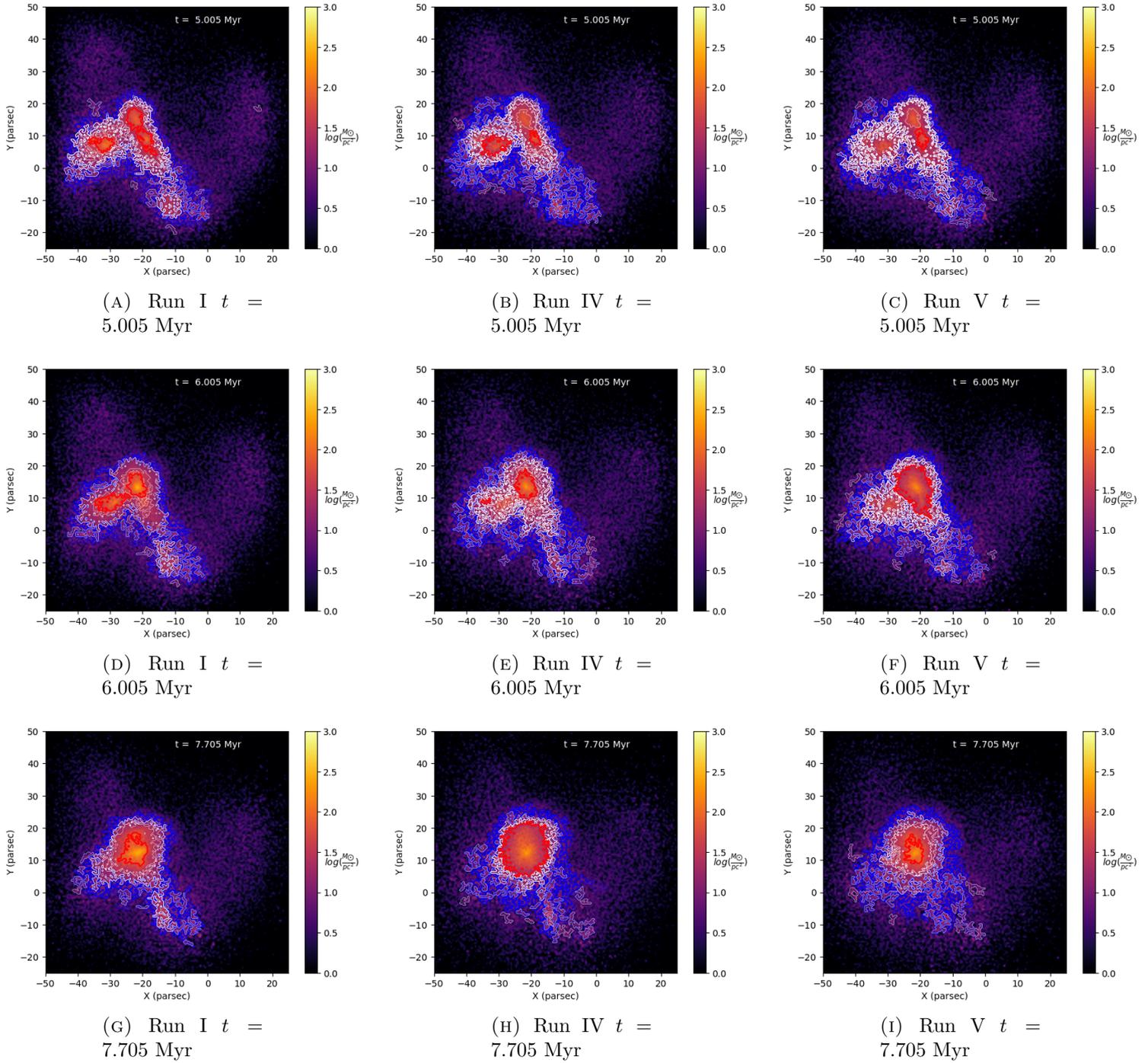


FIGURE 4.9: Comparison of gas distribution in Carina region over time between runs I (left), IV (middle) and V (right) between $t = 5$ Myr to $t = 7.7$ Myr. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$

4.7 Method I Run VI

This run used the random sampling method, a constant Z depth of 20 parsecs, varied gas temperature from observations and an initial gas velocity dispersion of 5 km/s. This run also used a new particle set generated using the sampling method which introduces a slight variation in the number of gas particles, the spatial distribution of the gas and the gas mass as seen in Table 4.1. The purpose of this run is to examine the effect of a large increase in the initial gas velocity dispersion on the cluster evolution. Despite the relatively large increase, we find that the overall evolution of the stellar distribution over time in the XY and XZ planes does not appear to be significantly different from the other runs. As well, we analyzed the stellar distribution over time using DBSCAN but found no significant differences in the clustering of the stars over time. If we consider the dendrogram analysis seen in Figure 4.10, we see that initially at $t = 0$ Myr, the main filamentary structure is present with dense regions (red contours) embedded in lower density gas (white and blue contours). The exact contours differ from those in run I due to the new particle set generated for this run. The gas distribution appears to significantly differ from runs I, IV and V starting at $t = 2$ Myr. This can be seen in Figure 4.11 which compares runs I, V and VI (runs IV and V were found to be quite similar in terms of the overall evolution of the gas distribution). At $t = 2$ Myr, runs I, IV and V still have regions of dense gas (red contours) while this run does not. No dense regions of gas develop over time with densities of $100 - 1000 M_{\odot}/pc^2$ which implies that star formation is not a possibility (according to our density threshold) beyond $t = 2$ Myr. In addition, some curvature of the main filamentary structure has been lost by $t = 2$ Myr in this run where the structure

looks slightly more compact along the X axis. At $t = 5$ Myr the gas distribution starts to appear more spherical. By $t = 7.7$ Myr, the distribution has become quite spherical and centrally concentrated with gas densities of $10 - 100 M_{\odot}/pc^2$ (white contours) surrounded by gas with densities between $6 - 10 M_{\odot}/pc^2$ (blue contours). The results from this run show that increasing the initial gas velocity dispersion to 5 km/s prevents dense gas from accumulating (densities of $100 - 1000 M_{\odot}/pc^2$) and causes initial structures present in the gas to be lost relatively quickly.

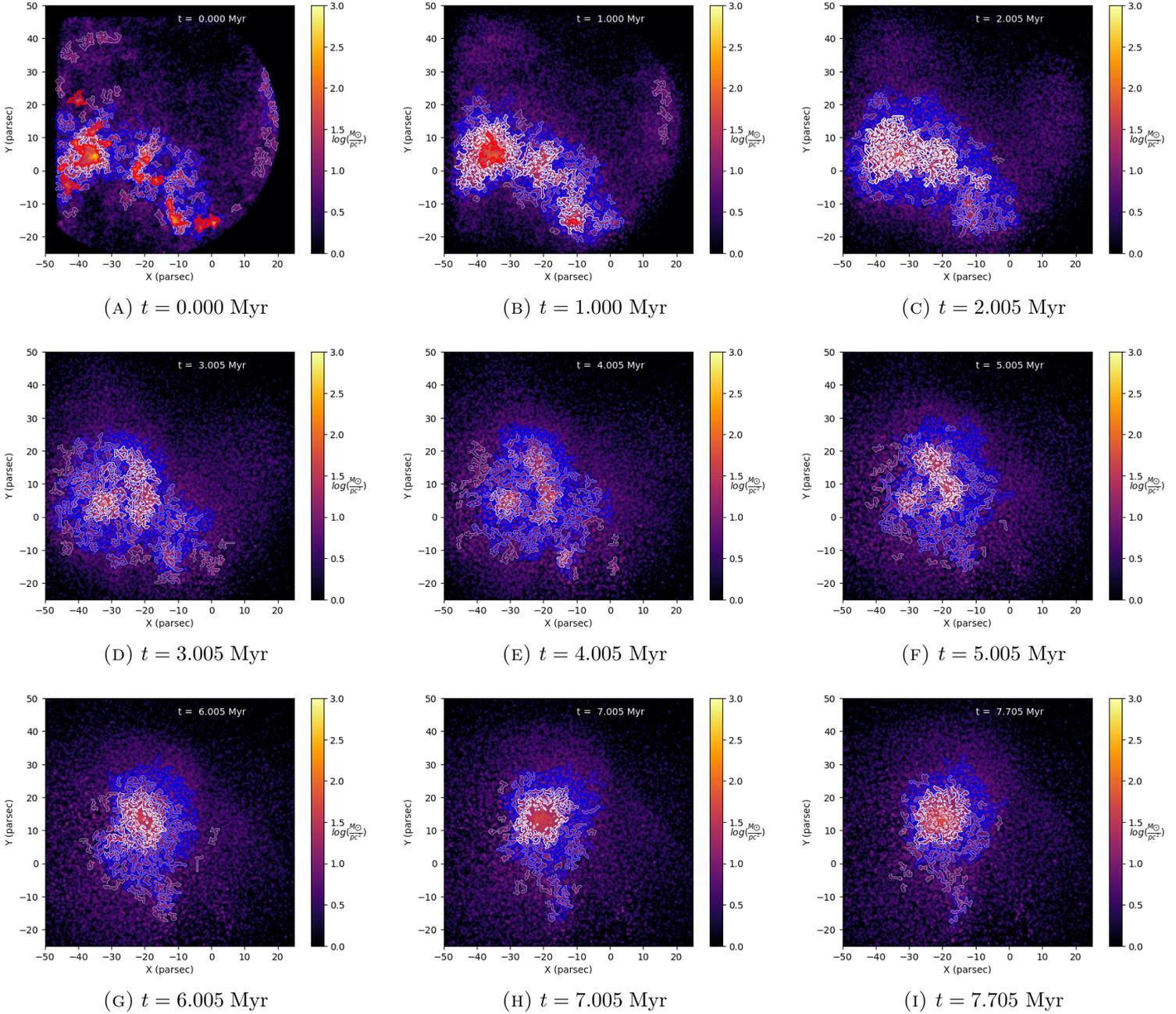
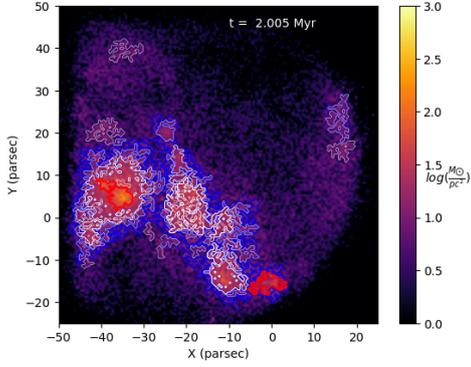
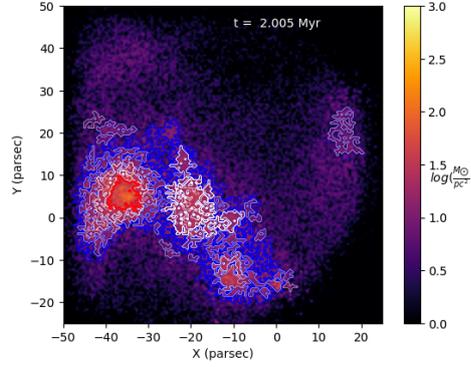


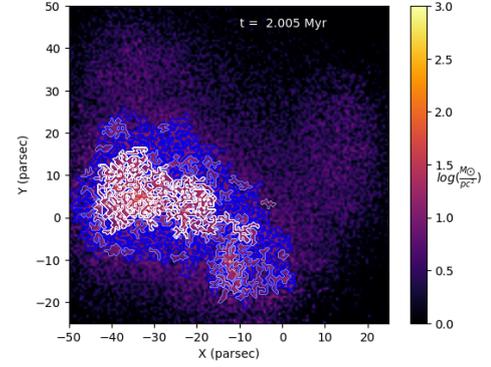
FIGURE 4.10: Gas distribution in Carina region over time with initial gas velocity dispersion of 5 km/s and new sampling. Over-plotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$.



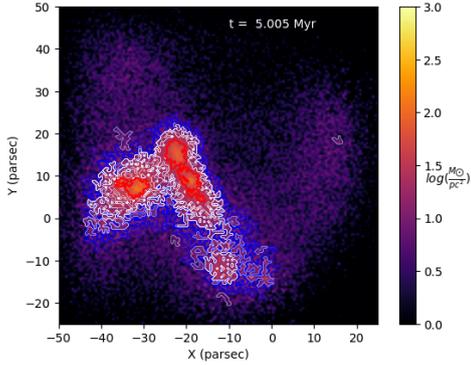
(A) Run I $t = 2.005$ Myr



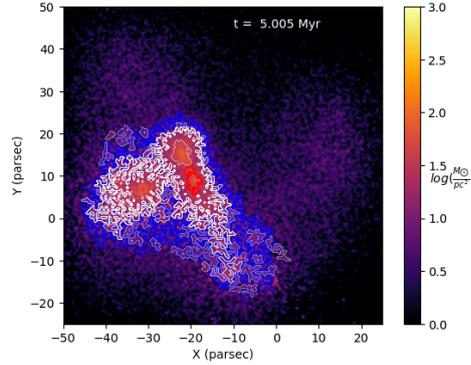
(B) Run V $t = 2.005$ Myr



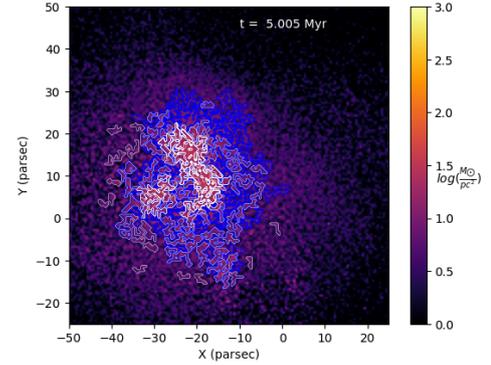
(C) Run VI $t = 2.005$ Myr



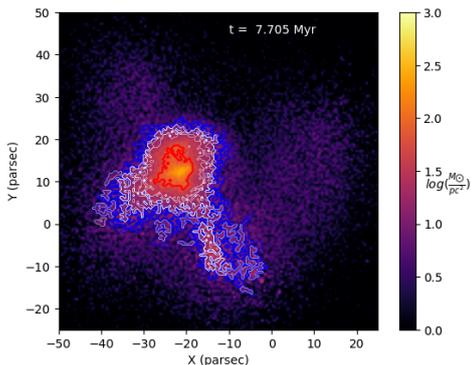
(D) Run I $t = 5.005$ Myr



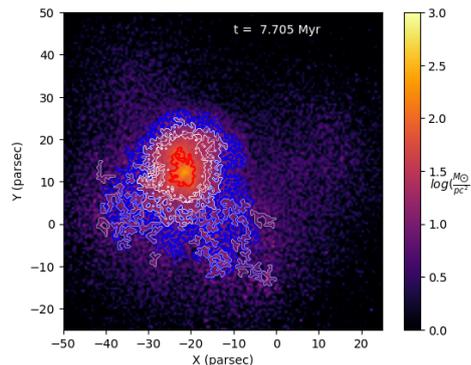
(E) Run V $t = 5.005$ Myr



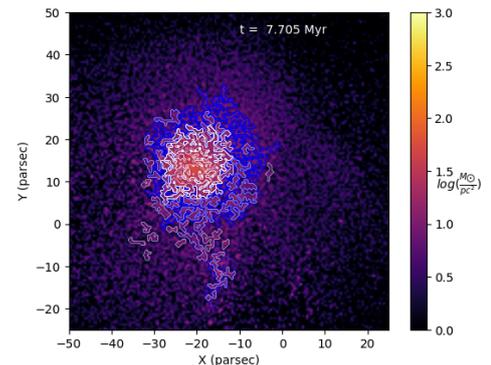
(F) Run VI $t = 5.005$ Myr



(G) Run I $t = 7.705$ Myr



(H) Run V $t = 7.705$ Myr



(I) Run VI $t = 7.705$ Myr

FIGURE 4.11: Comparison of gas distribution in Carina region over time between runs I, V and VI between $t = 2$ Myr to $t = 7.7$ Myr. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities between $100 - 1000 M_{\odot}/pc^2$

4.8 Method II Run I

This run (run VII in Table 4.1) initially assigned one particle to represent each pixel, used a varied Z depth obtained through the use of distance transforms and used an initial gas velocity dispersion of 1.4 km/s. The advantage of this gas particle creation technique is that all the observed data (within the region of interest) is used. As a result, the total gas mass is an order of magnitude higher than the runs performed with Method I and is also an order of magnitude higher than the total stellar mass. Therefore this run probes the regime where the gas mass is significantly greater than the stellar mass. The total gas mass in the Method II runs is a true representation of the total gas mass in the observed Carina region since the entirety of the image data was used (within the region of interest) and this is an improvement over the Method I runs. As well, the increase in gas particles as a result of the Method II technique slowed down our simulation by increasing the computation time. Therefore, this run was only evolved to about 2 Myr. The dynamical evolution of the stars and gas can be seen in Figures 4.12 and 4.13. Initially, we see that the overall density of the cloud is greater than the Method I runs which is expected as more mass is included due to the approach taken with Method II. In Figure 4.12, we see both the stellar and gas distributions contracting with notable contraction in the gas at $X = -40$ and $Y = 5$ parsecs which creates a region of higher density gas. In Figure 4.13, the stellar and gas distributions are also seen to contract over time. Initial straight filamentary structures in the gas become more curved and break up into clumps. It is known that the formation of stars follows patterns in the natal molecular cloud and it is interesting to see filaments in the gas break up in this way which could potentially

form stars.

If we consider the dendrogram analysis seen in Figure 4.14, initially there are regions of relatively lower density gas ($10 - 100 M_{\odot}/pc^2$) around regions of high density gas ($100 - 1000 M_{\odot}/pc^2$) which can be seen as white contours surrounding red contours ($\geq 100 M_{\odot}/pc^2$). Again, these figures show the entire dendrogram structure grouped into three density ranges. Since both branches and leaves are included, contours of the same color may be found within each other. We note that there are no regions initially with gas densities lower than $10 M_{\odot}/pc^2$ despite such regions being present in all the Method I runs. Therefore including all the observed pixel data significantly increases the initial density of the gas cloud compared to randomly sampling the data 100,000 times. The initial regions of high density gas, particularly in the main filamentary structure, are similar to those seen in the Method I runs. However, there are also high density regions around $X = -40$, $Y = 40$ parsecs and around $X = 15$, $Y = 10$ which were not as prominent in the Method I runs. Between $t = 0$ Myr and $t = 2$ Myr, we see an overall contraction in the gas with the initial dense regions merging slightly together. These regions persist which implies that star formation could be present according to the density threshold of roughly $129 M_{\odot}/pc^2$. Due to the higher overall gas densities in this run, we performed a second dendrogram analysis with higher density ranges which can be seen in Figure 4.15. Cyan corresponds to densities between $100 - 1000 M_{\odot}/pc^2$, yellow corresponds to densities between $1000 - 5000 M_{\odot}/pc^2$ and red corresponds to densities between $5000 - 10000 M_{\odot}/pc^2$. From this, we can more clearly see the contraction in the dense gas over time and that at $t = 2$ Myr, the region with the highest gas density is located at roughly $X = -35$, $Y = 7$ parsecs.

We also used DBSCAN to analyze the stars over time and found the results to be generally similar to those in the next run. Run II is where we sped up our simulation by changing the bridge timestep (from 0.005 Myr to 0.01 Myr) which dictates how often the interaction between the stars and gas are taken into account and these results can be seen in the following section.

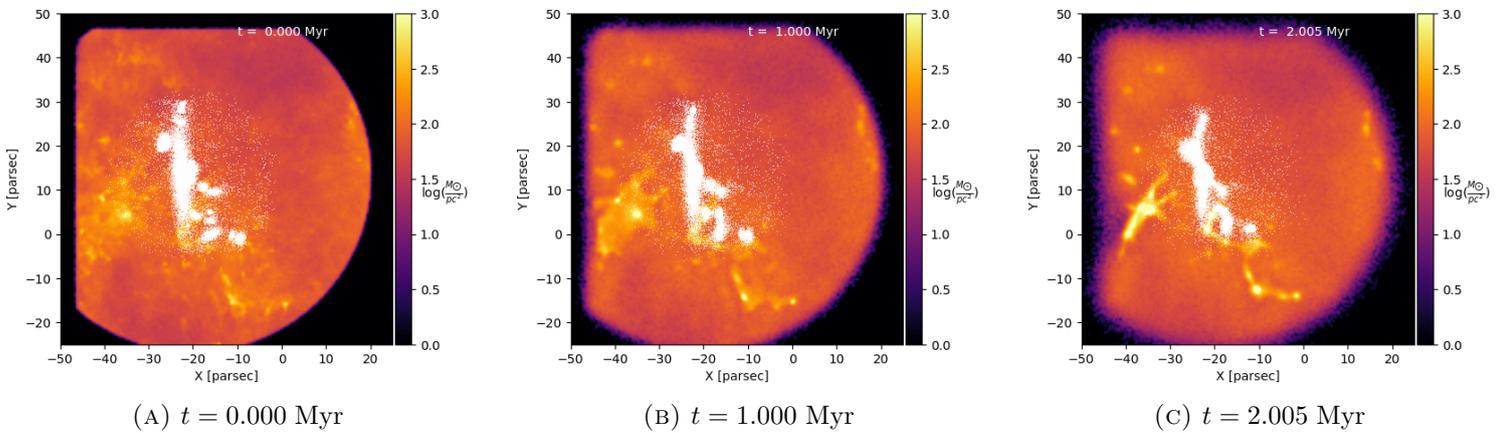


FIGURE 4.12: Evolution of Carina Region in XY plane using Method II with a varied Z depth and an initial gas velocity dispersion of 1.4 km/s.

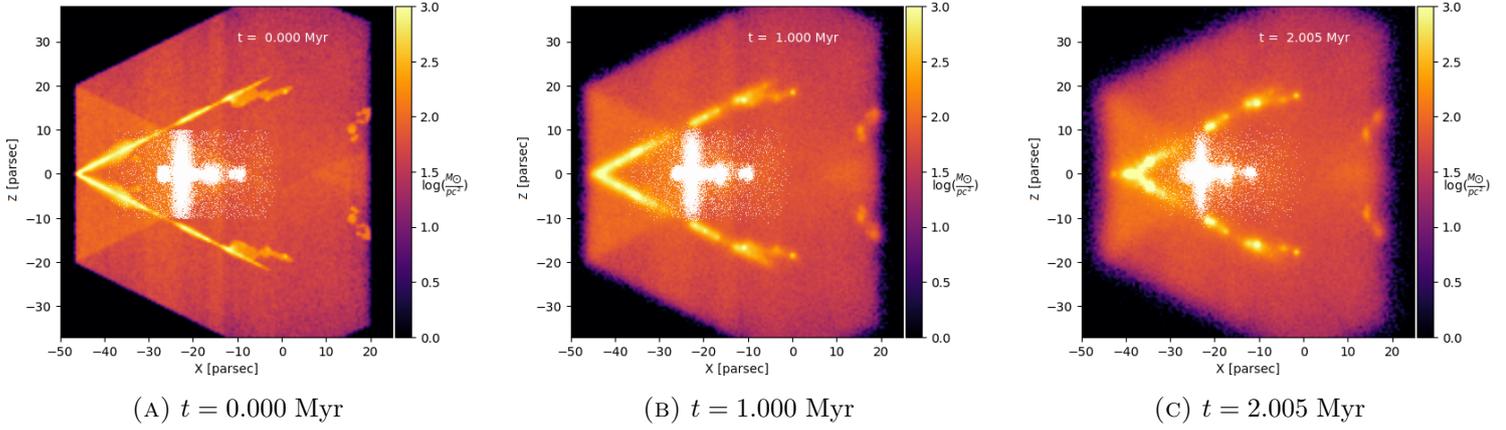


FIGURE 4.13: Evolution of Carina Region in XZ plane using Method II with a varied Z depth and an initial gas velocity dispersion of 1.4 km/s.

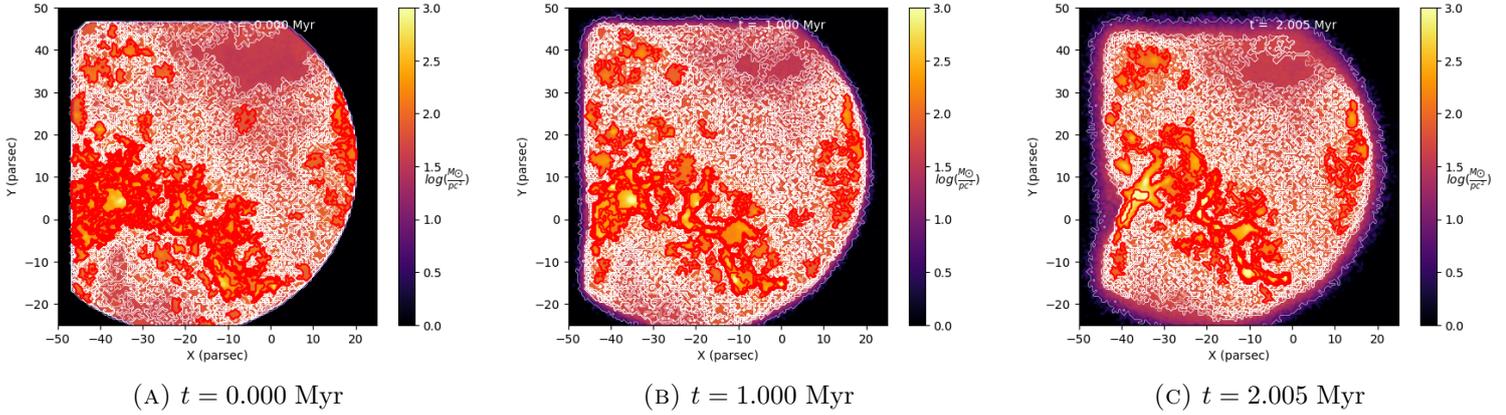


FIGURE 4.14: Gas distribution in Carina region over time using Method II with a varied Z depth, and an initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between 6 - 10 M_{\odot}/pc^2 , white corresponds to densities between 10 - 100 M_{\odot}/pc^2 and red corresponds to densities greater than 100 M_{\odot}/pc^2 .

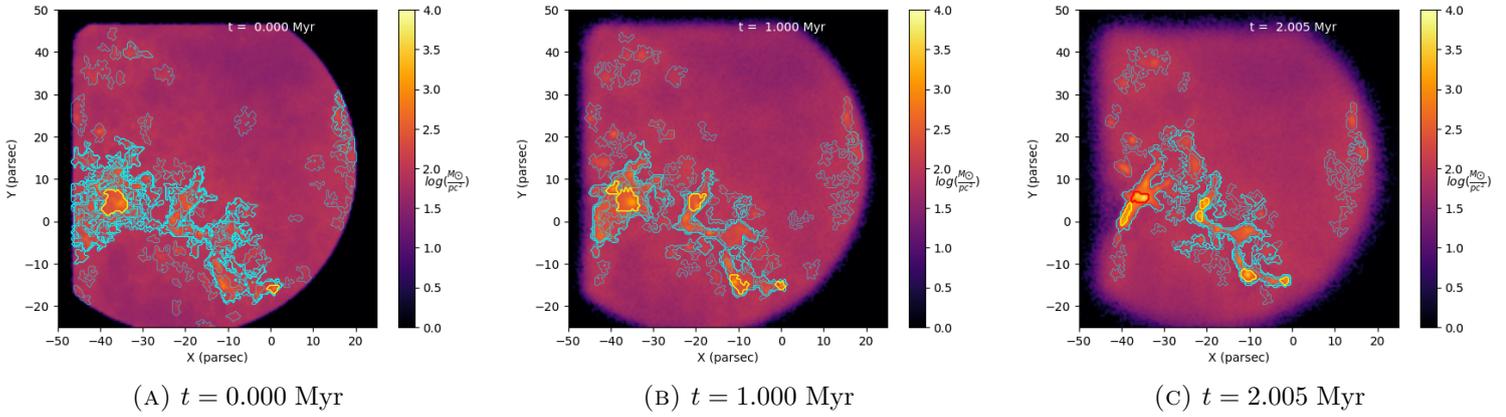


FIGURE 4.15: Gas distribution in Carina region over time using Method II with a varied Z depth, and an initial gas velocity dispersion of 1.4 km/s. Overplotted are contours from branches and leaves in the dendrogram. Cyan corresponds to densities between $100 - 1000 M_{\odot}/pc^2$, yellow corresponds to densities between $1000 - 5000 M_{\odot}/pc^2$ and red corresponds to densities between $5000 - 10000 M_{\odot}/pc^2$.

4.9 Method II Run II

This run (run 8 in Table 4.1) used the method where one particle was initially assigned to represent each pixel, a varied Z depth obtained through the use of distance transforms, an initial gas velocity dispersion of 1.4 km/s and an increased bridge timestep (0.01 Myr as opposed to 0.005 Myr) to allow the run to evolve to 4.6 Myr in a shorter amount of time. We see that the increased timestep results in fluctuations in the total energy between 1 - 4% not seen with the standard timestep and future work should be done to determine the best choice for the bridge timestep.

Although increasing this value decreases the accuracy of the simulation by reducing the interactions between the stars and gas, we see no significant difference between this and the previous run in terms of the overall cluster evolution. The dynamical evolution of the stars and gas can be seen in Figures 4.16 and 4.17. The evolution between $t = 0$ Myr and $t = 2$ Myr is similar to that in the previous run. Beyond that, the stars and gas continue to contract with the stars becoming concentrated at around $X = -30$, $Y = 10$ parsecs. In Figure 4.17, we see the initial filaments not only breaking apart into clumps but also feeding into the more concentrated star cluster at around $X = -30$, $Z = 0$ parsecs which increases the gas density in that region.

If we consider the dendrogram analysis seen in Figure 4.18, again initially there are regions of relatively lower density gas ($10 - 100 M_{\odot}/pc^2$) around regions of high density gas ($100 - 1000 M_{\odot}/pc^2$) which can be seen as white contours surrounding red contours. At $t = 0$ Myr, the dendrogram looks slightly different than that in

Run I (Method II), however these differences are attributed to the variations in the initial particle positions derived from drawing along the range of the X and Y dimensions of each pixel (see Chapter 3 for details). Overall, the differences are not significantly different between $t = 0$ Myr and $t = 2$ Myr. Beyond that, we see that the regions of high density gas merge together forming a structure that is elongated in the Y direction and that also curves upward from $X = -10$ parsecs to the right. Of the Method I runs that showed a similar filamentary structure, the curved part of it seen here was not observed in those runs. This highlights the necessity of including all the available information from the gas observations. We take regions marked in red to roughly contain high enough densities for star formation (density threshold of roughly $129 M_{\odot}/pc^2$) and if we only restrict ourselves to randomly sampling 100,000 times to obtain the initial conditions, we will miss potential sites of star-formation. A comparison between runs I and II from Method I and run II from Method II can be seen in Figure 4.19. Of these, the runs that used the observational gas data showed filamentary structures in the gas and regions of high density gas that could potentially lead to new star-formation. Of the runs that used observational gas, Method II run II is the most realistic as all the gas data in the region of interest were used and filamentary structures in the gas were incorporated.

Due to the higher overall gas densities in this run, we performed a second dendrogram analysis with higher density ranges which can be seen in Figure 4.20. Cyan corresponds to densities between $100 - 1000 M_{\odot}/pc^2$, yellow corresponds to densities between $1000 - 5000 M_{\odot}/pc^2$ and red corresponds to densities between $5000 - 10000 M_{\odot}/pc^2$. Again, we can more clearly see the contraction in the dense

gas over time and how the gas merges to form a curved structure by $t = 4.6$ Myr.

We analyzed the stars over time using DBSCAN and the results can be seen in Figure 4.21. Overall, the clusters seem to group together in a similar way as that seen in the Method I runs by merging together to create clusters elongated in Y before starting to form a somewhat spherical shape. However, in terms of the timescale on which this occurs, the formation of one main cluster occurs 1 Myr earlier (at roughly $t = 4$ Myr) than in the Method I runs. In this run at $t = 4.6$ Myr, one main cluster is present with its center at roughly $X = -26$, $Y = 5$ parsecs. In Method I run I, at $t = 5$ Myr, which is around the same time, one main cluster is also present. However this cluster is more elongated and its center is roughly at $X = -21$, $Y = 11$ parsecs. Therefore this shows that Methods I and II result in differences in the clustering of the stars over time. This difference is attributed to the greater amount of gas (and hence greater gas potential) present in the Method II runs compared to the Method I runs. A similar result was seen in the simulations described in Sills et al. 2018 which showed that a higher gas mass affects the stellar distribution.

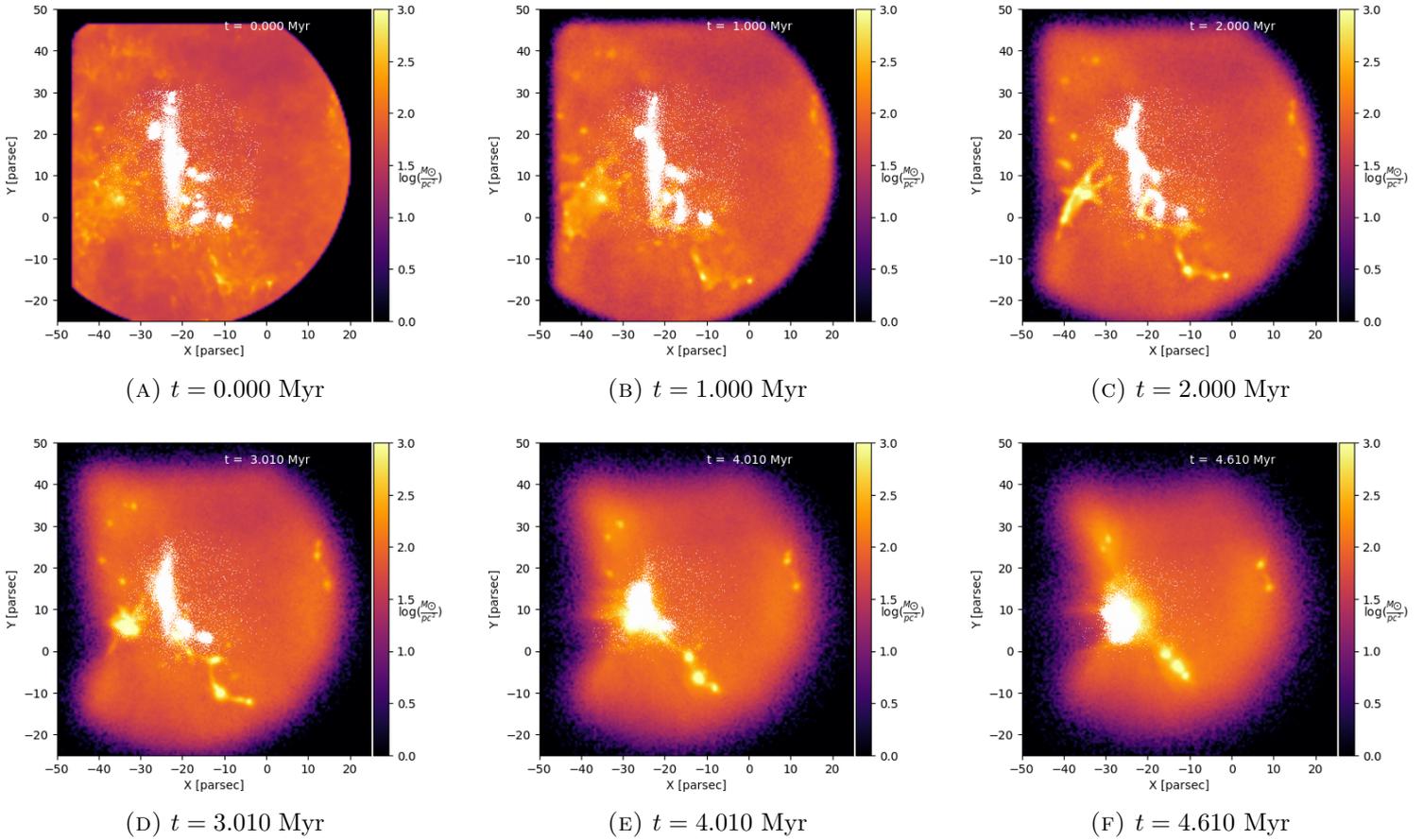


FIGURE 4.16: Evolution of Carina Region in XY plane using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep.

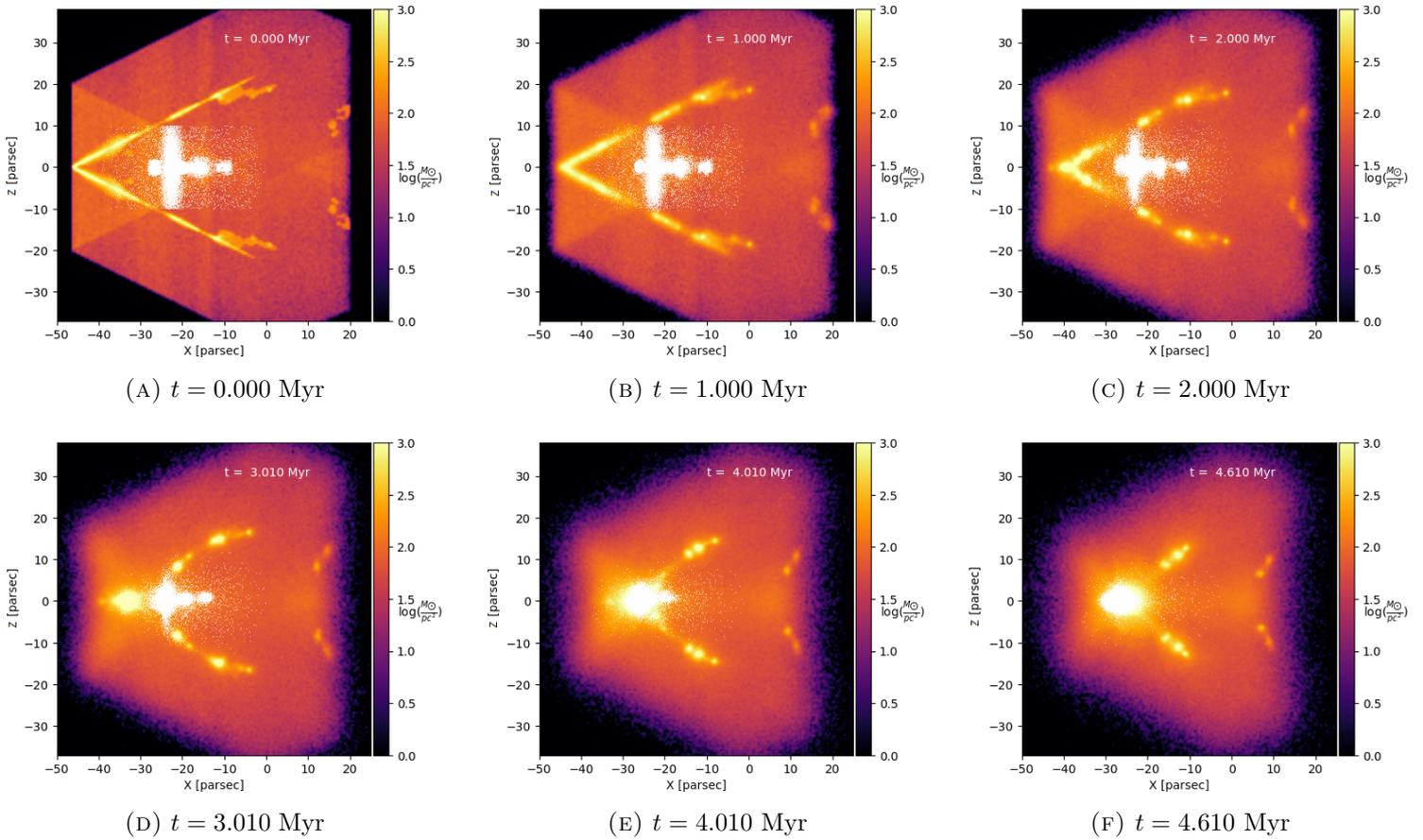


FIGURE 4.17: Evolution of Carina Region in XZ plane using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep.

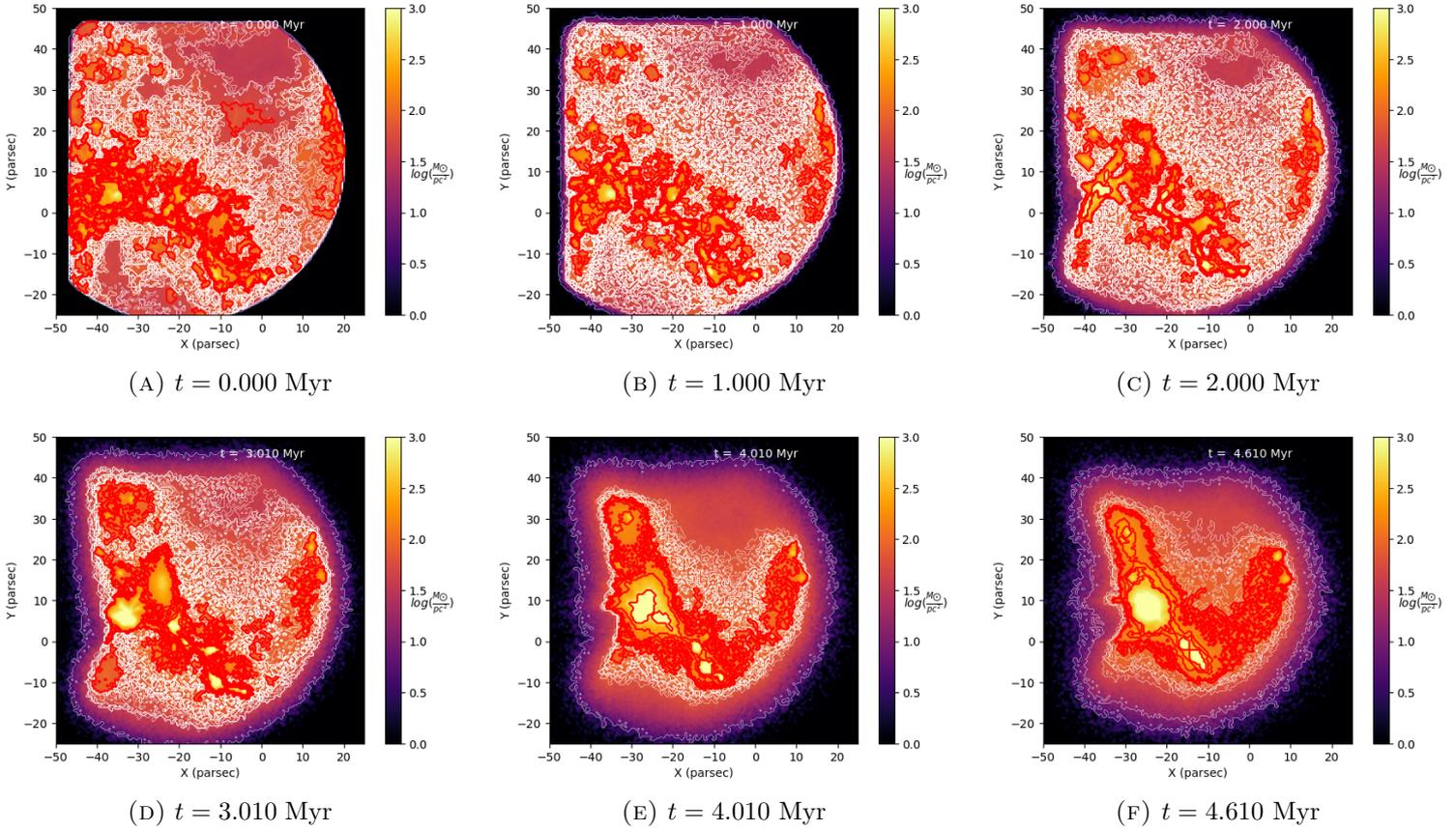


FIGURE 4.18: Gas distribution in Carina region over time using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep. Overplotted are contours from branches and leaves in the dendrogram. Blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities greater than $100 M_{\odot}/pc^2$.

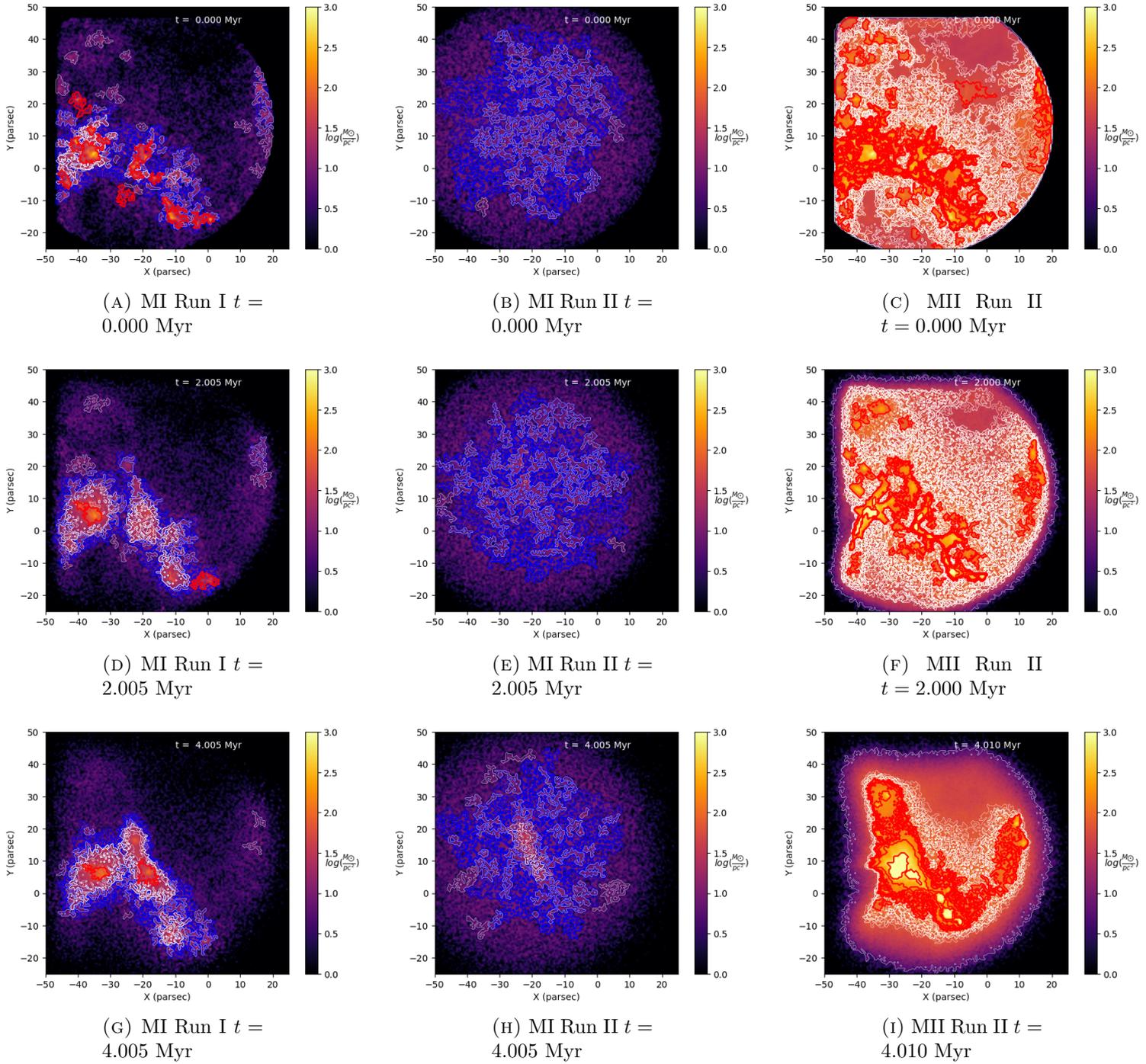


FIGURE 4.19: Comparison of gas distribution in Carina region over time between Method I (MI) runs I, II and Method II (MII) run II between $t = 0$ Myr to $t = 4$ Myr. Overplotted are contours from branches and leaves in the dendrogram where blue corresponds to densities between $6 - 10 M_{\odot}/pc^2$, white corresponds to densities between $10 - 100 M_{\odot}/pc^2$ and red corresponds to densities greater than $100 M_{\odot}/pc^2$

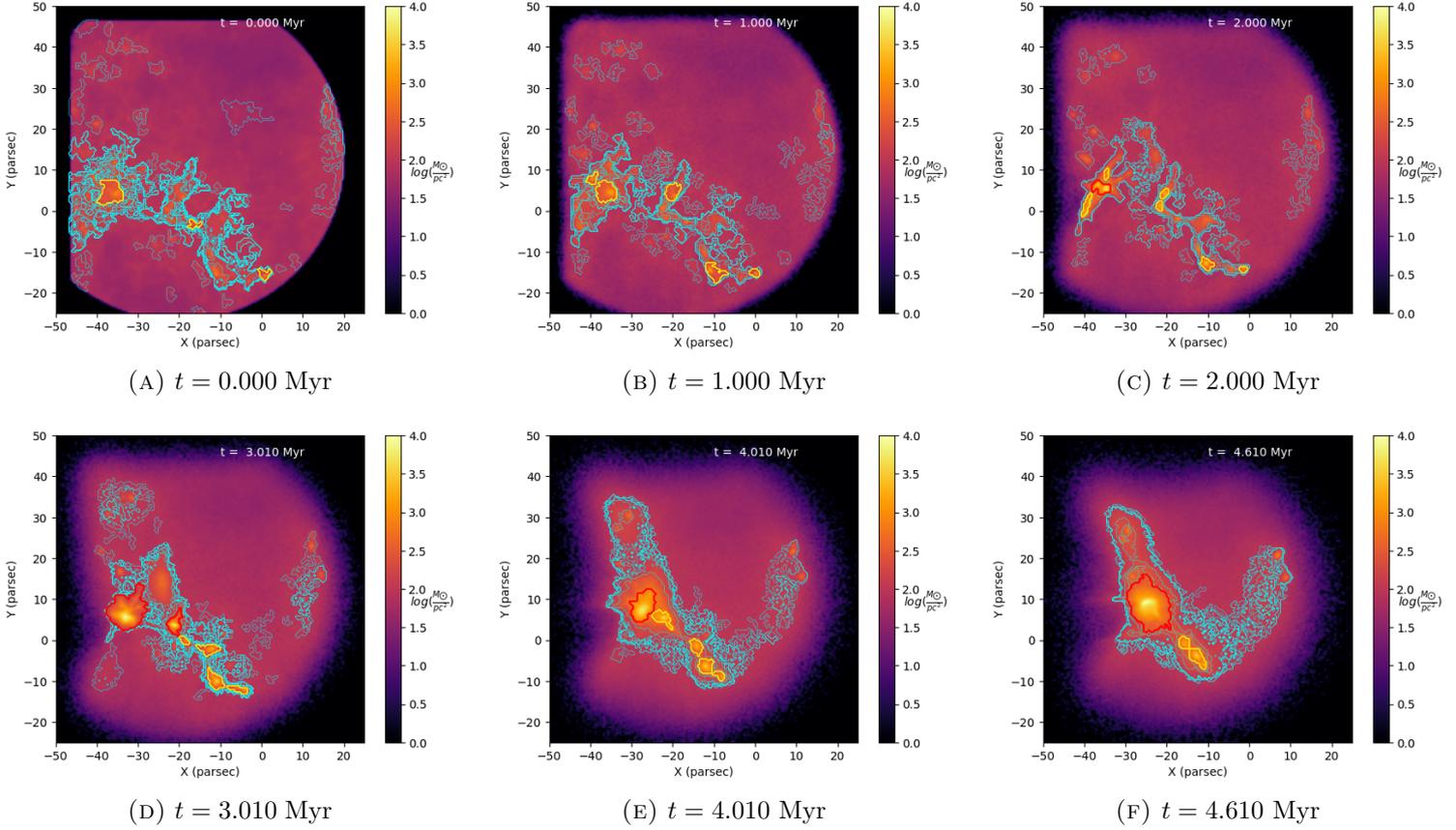


FIGURE 4.20: Gas distribution in Carina region over time using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep. Overplotted are contours from branches and leaves in the dendrogram. Cyan corresponds to densities between 100 - 1000 M_{\odot}/pc^2 , yellow corresponds to densities between 1000 - 5000 M_{\odot}/pc^2 and red corresponds to densities between 5000 - 10000 M_{\odot}/pc^2 .

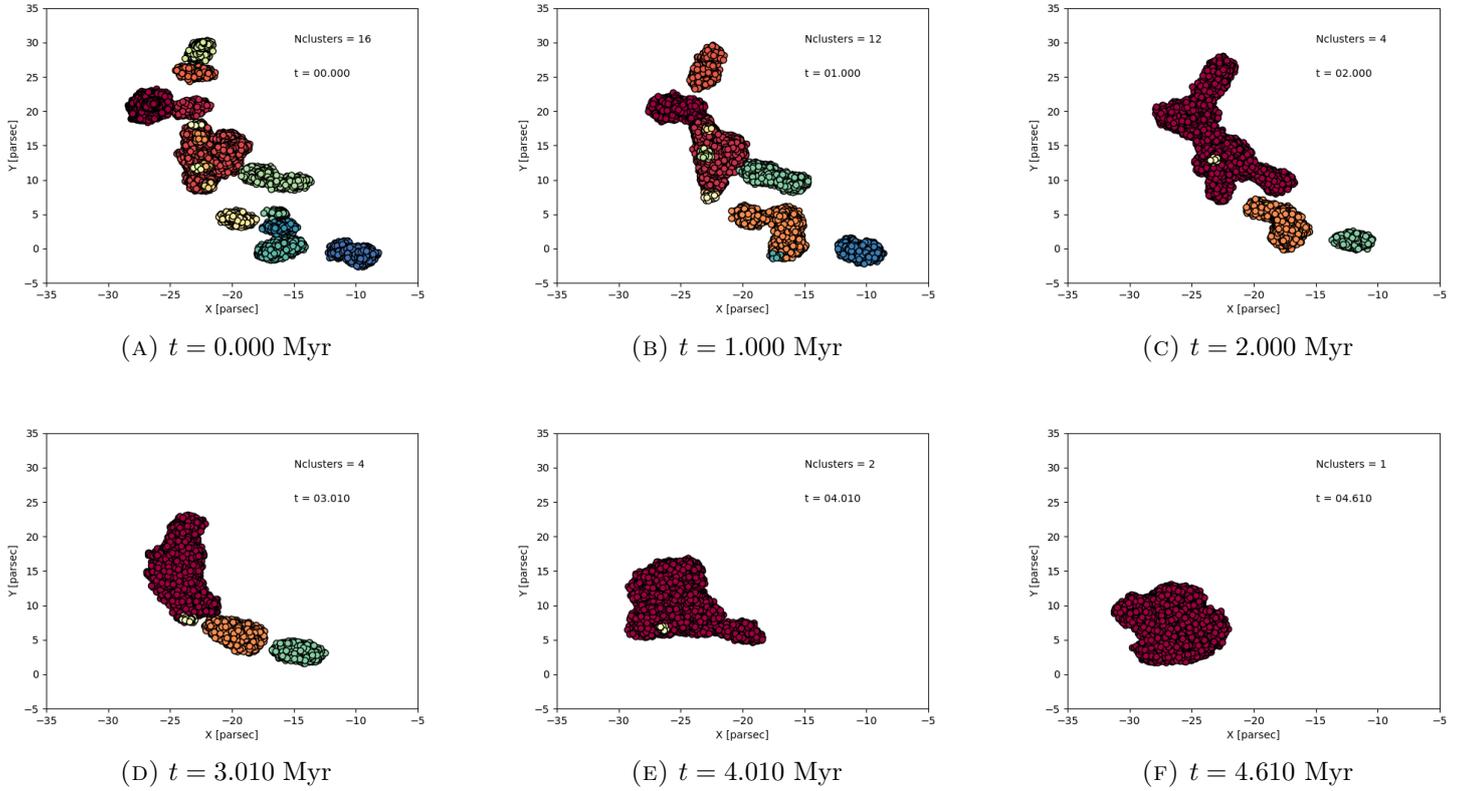


FIGURE 4.21: Clusters in Carina picked out by DBSCAN over time using Method II with a varied Z depth, an initial gas velocity dispersion of 1.4 km/s and an increased interaction timestep. Each cluster is denoted by a different colour. The colours do not necessarily correspond to the same stars across the images as they are only used to highlight the different clusters identified.

4.10 Stellar Distribution Comparisons

We analyzed the stellar distributions over time using DBSCAN and we also computed Λ_{msr} values (using AMUSE) to quantify any mass segregation that may be present in the clusters identified. As previously mentioned, we found no large differences between the Method I runs in terms of the overall evolution of the stellar distribution over time although there were slight variations in the number of clusters identified by DBSCAN. A plot of the number of clusters over time picked out using DBSCAN can be seen in Figure 4.22. Considering the Method I runs first, we see a general decrease in the number of clusters which reaches the lowest at $t = 5$ Myr. From there, the number of clusters increases with a sharp increase between $t = 5$ Myr and $t = 6$ Myr. Overall, the number of clusters is not drastically different between the Method I runs from $t = 0$ Myr to $t = 5$ Myr. Beyond $t = 5$ Myr, there is a greater spread in the number of clusters. The spread is due to the variation in the smaller clusters that are picked out although the main cluster appears similar throughout. These smaller clusters are due to noise and are transient. The results of DBSCAN on the Method I runs at $t = 6$ Myr, $t = 7$ Myr and $t = 7.7$ Myr can be seen in Figures 4.23, 4.24 and 4.25.

The Method II runs also show a general decrease in the number of clusters until $t = 4.6$ Myr when only one cluster is picked out. At $t = 4$ Myr, two clusters are picked out. However, one is a large main cluster while the other is significantly smaller which is again likely due to noise or a transient grouping of stars. Comparing Methods I and II, we find that a large main cluster is formed 1 Myr earlier with Method II (at $t = 4$ Myr) compared to Method I which demonstrates that the

gas conditions affects the stellar distribution. The slight differences between the two Method II runs are due to a combination of effects which includes variations in the original gas density arising from drawing from the X and Y dimensions of each pixel to determine particle positions and differences in the bridge timestep.

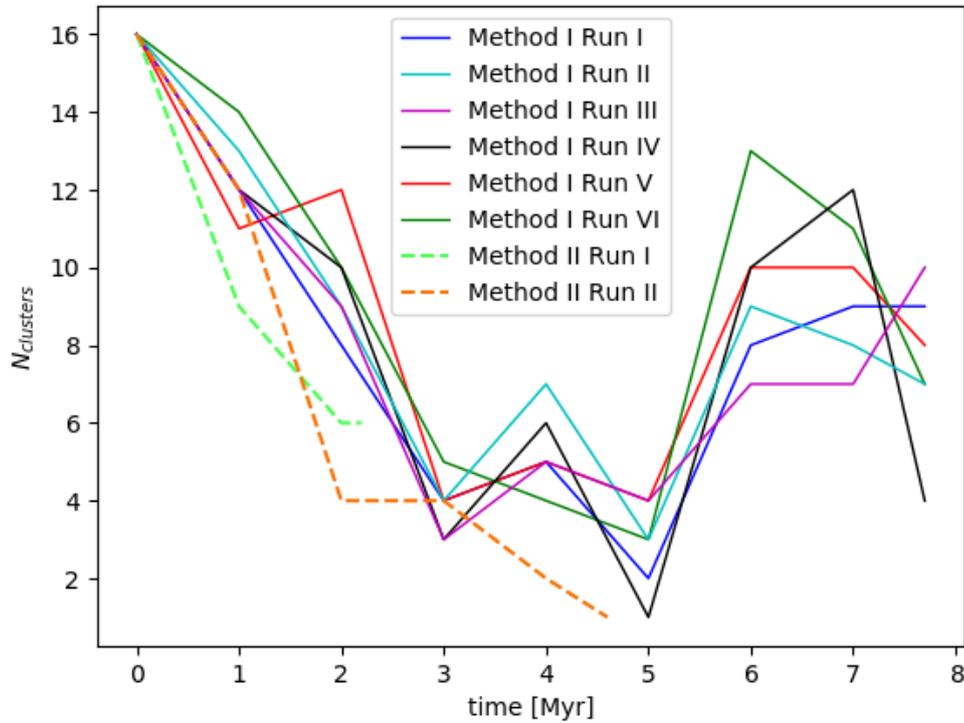


FIGURE 4.22: $N_{clusters}$ over time as identified by DBSCAN for all runs. Run details can be seen in Table 4.1.

Of the clusters identified by DBSCAN, the Λ_{msr} value was calculated for clusters that had at least 40 members. The maximum and minimum Λ_{msr} values (and their uncertainties) for the selected clusters can be seen in Table 4.2. Note that at $t = 0$ Myr for the three runs, the Λ_{msr} values are similar but not exactly the same. This is because 100 random sets are chosen to calculate the mass-segregation value which

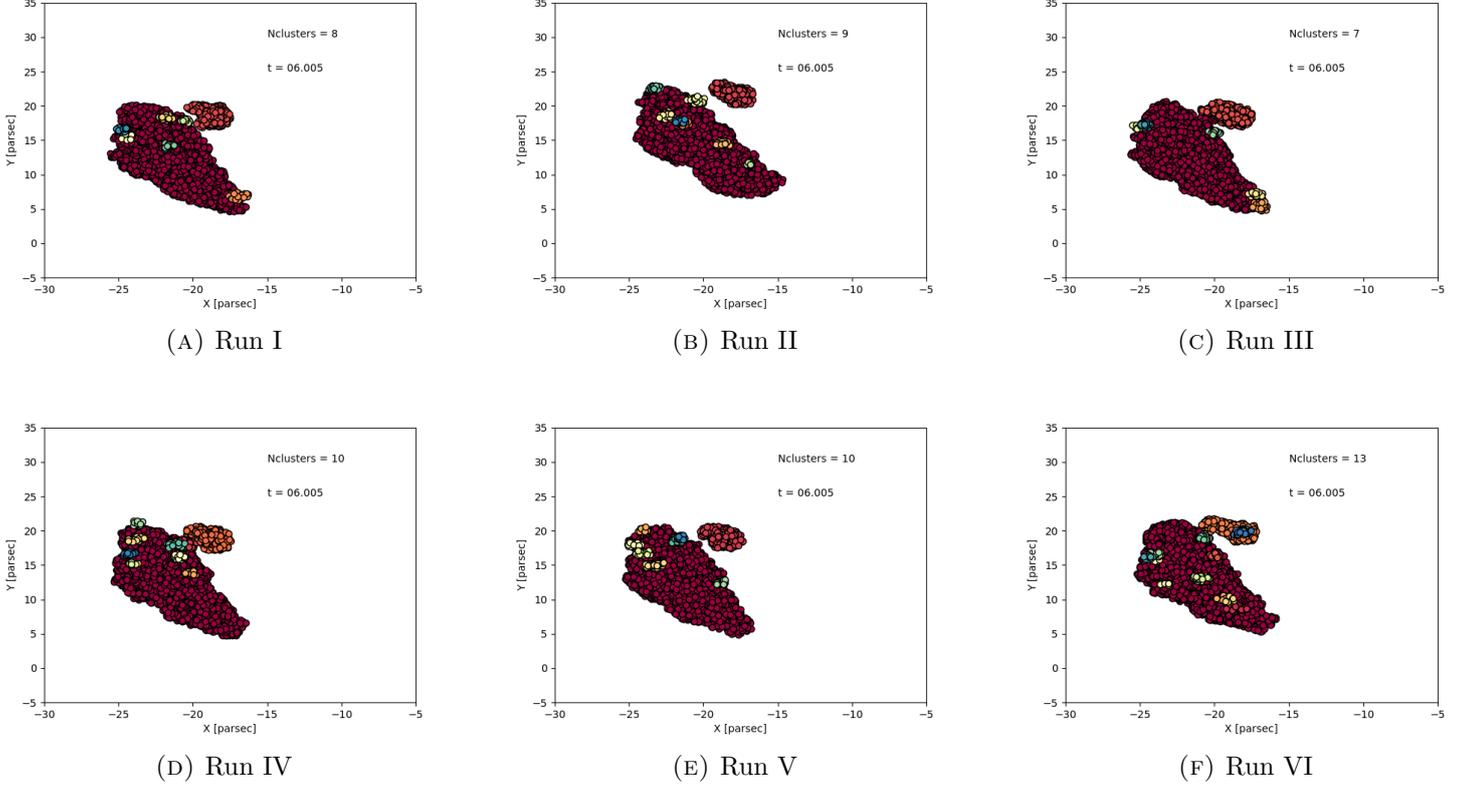


FIGURE 4.23: DBSCAN results for Method I runs I-VI at $t = 6$ Myr. Each colour corresponds to a different cluster. Colours do not necessarily correspond to the same stars between the images.

introduces a slight variation in the resulting value. However, the Λ_{msr} values at $t = 0$ are all still within the uncertainties of each other. Figures 4.26 and 4.27 show the results of DBSCAN on Method I run I and Method II run II, respectively, where the clusters with the greatest Λ_{msr} value (within uncertainties of each other) can be seen in black. The results from the Method I runs are all similar. As seen in Table 4.2, the Λ_{msr} values change over time and the maximum Λ_{msr} values are not always significantly greater than 1 (consider $\Lambda_{msr} \geq 2$ for mass segregation to be present). This means that even though the clusters in black have the greatest Λ_{msr} value, they may not be considered mass segregated.

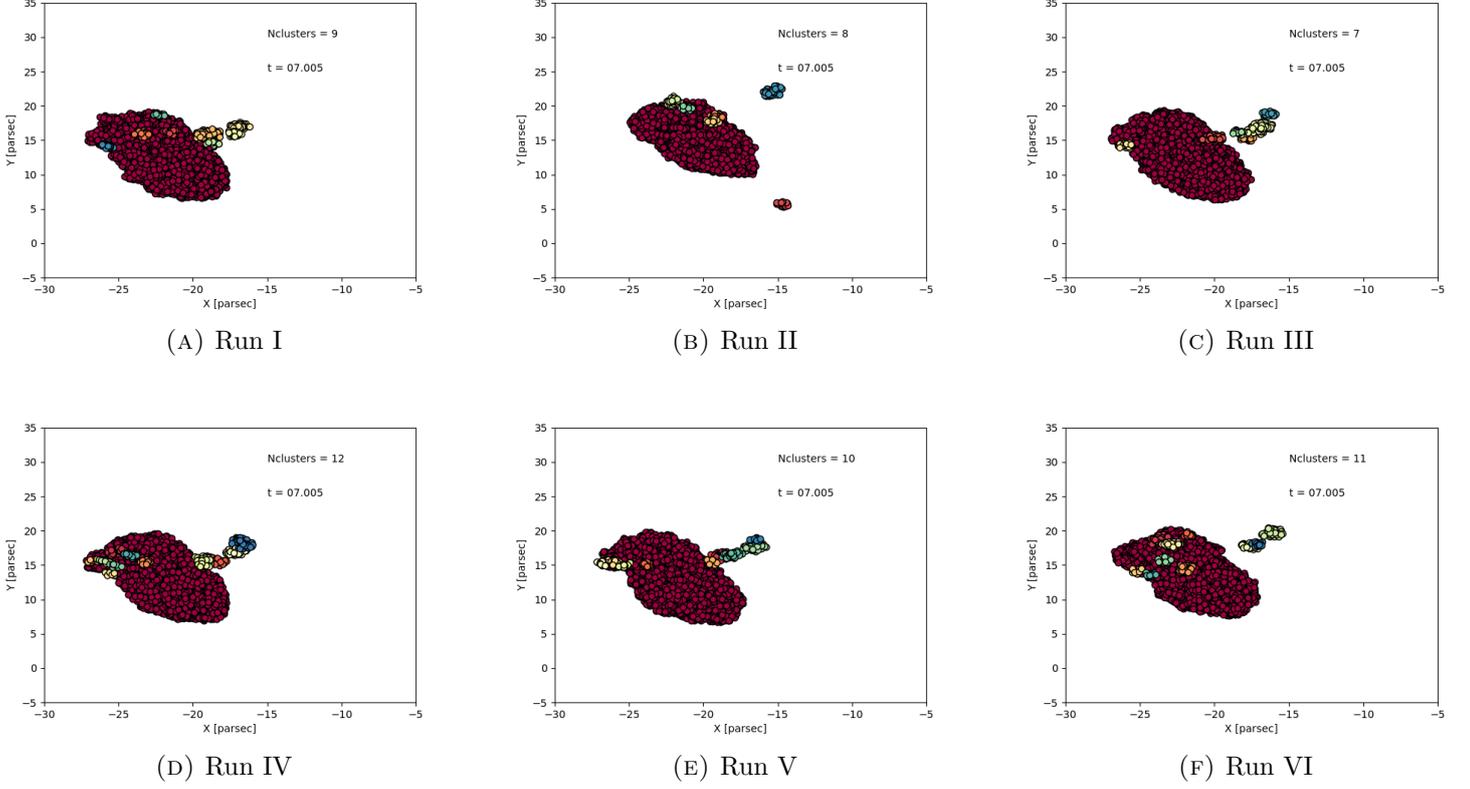


FIGURE 4.24: DBSCAN results for Method I runs I-VI at $t = 7$ Myr. Each colour corresponds to a different cluster. Colours do not necessarily correspond to the same stars between the images.

Initially at $t = 0$ Myr, the clusters are not mass segregated (see Sills et al. 2018). The maximum Λ_{msr} value is roughly 1.2 and the corresponding clusters can be seen in black which are in the middle of the stellar distribution. In the Method I runs, the clusters with the largest Λ_{msr} value fluctuate until $t = 5$ Myr when most of the clusters have merged together and the large main cluster then corresponds to the largest Λ_{msr} value. This large main cluster continues to have the largest Λ_{msr} value until $t = 7.7$ Myr. The same fluctuation is not seen in the Method II run, however the large main cluster that forms eventually does correspond to the largest Λ_{msr} value. The maximum Λ_{msr} values for each of the runs over time can be seen in

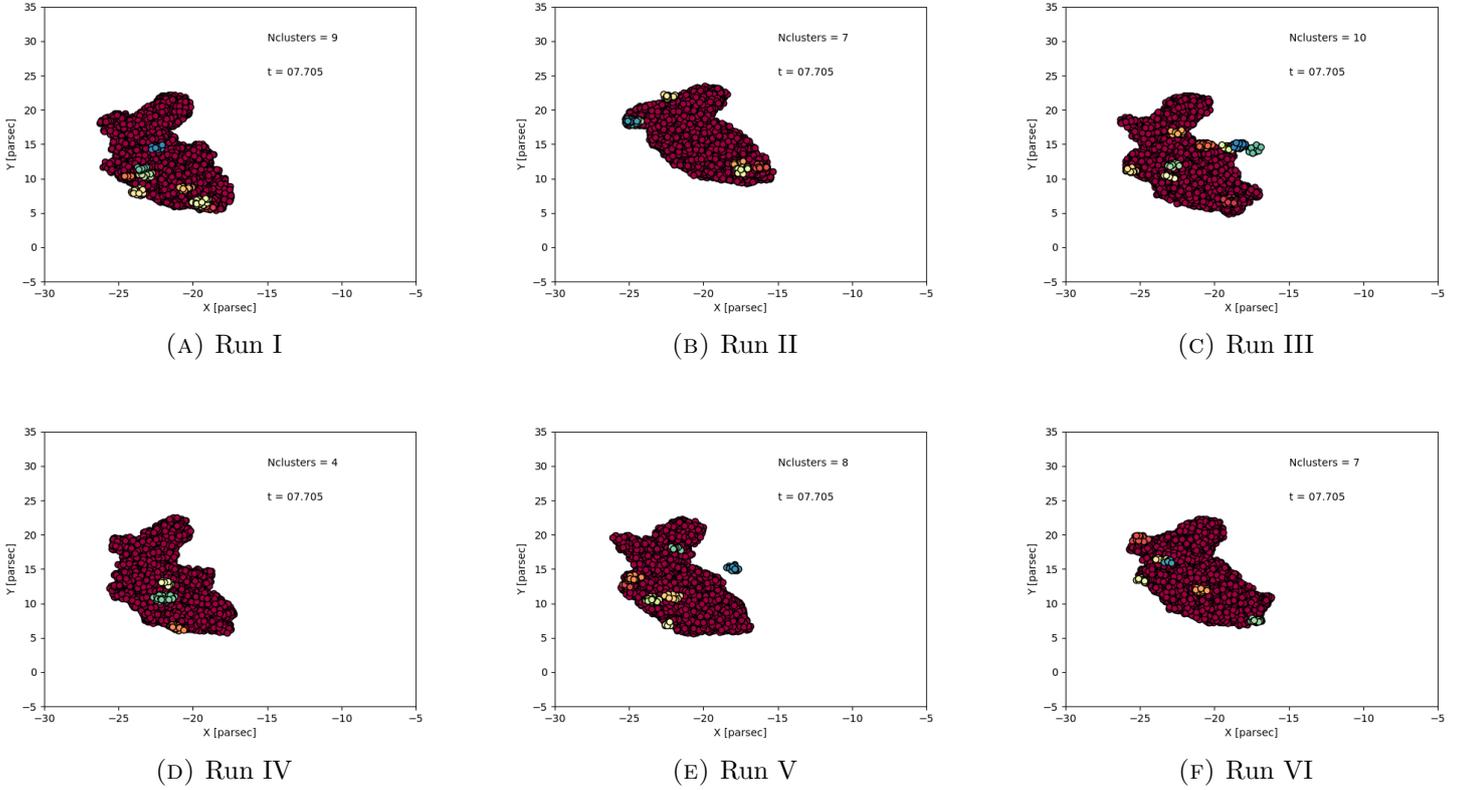


FIGURE 4.25: DBSCAN results for Method I runs I-VI at $t = 7.7$ Myr. Each colour corresponds to a different cluster. Colours do not necessarily correspond to the same stars between the images.

Figure 4.28. Overall we see that the maximum Λ_{msr} value ranges from roughly 1 to 5. We consider $\Lambda_{msr} \geq 2$ for mass segregation to be present which is true between $t = 1$ and $t = 4$ Myr and again between $t = 6$ and $t = 7$ Myr. We expect some fluctuation in the Λ_{msr} value between the mass segregated and non mass segregated case due to the merging of the subclusters. If a subcluster becomes mass segregated and then proceeds to merge with other subclusters, it will take time for the larger cluster to mass segregate due to dynamical interactions.

Over time, the runs are not significantly different except at $t = 3$ and at $t = 4$

Myr. At $t = 3$ Myr, run VI has a higher Λ_{msr} value than the other runs. It is unclear if it is due to the increase in the gas velocity dispersion or due to the new particle set generated from sampling. We see a spread in the maximum Λ_{msr} at $t = 4$ Myr before the clusters merge at $t = 5$ Myr. At $t = 4$ Myr, runs I and III are significantly different. These runs differ only in the initial temperature of the gas particles which shows that using a constant average gas temperature as opposed to varied temperatures results in a significantly lower Λ_{msr} value. Comparing Runs IV and V shows that differing particle sets from random sampling results in significantly different Λ_{msr} values. Runs I and IV have the highest Λ_{msr} values at this time with overlap in the errors. These runs differ only in the gas velocity dispersion which shows that increasing the initial gas velocity dispersion from 1.4 km/s to 2 km/s lowers the Λ_{msr} value but does not result in a significantly different value. Runs II and III only differ in the initial distribution of the gas and are not significantly different either. Run VI is significantly different from runs I and IV, however it is unclear if this is due to the increase in the gas velocity dispersion or the new particle set. Therefore significant differences between the Method I runs in the Λ_{msr} value arise due to different initial gas temperatures and different particle sets from the random sampling method. The Method II runs are not significantly different than the Method I runs between $t = 0$ Myr and $t = 3$ Myr although the Λ_{msr} are in general lower compared to the other runs. At $t = 4$ Myr, the Λ_{msr} decreases before rising slightly again at $t = 4.6$ Myr. This feature is similar to that seen in the Method I runs which dip at $t = 5$ Myr before slightly rising and can be attributed to the main cluster of stars forming 1 Myr earlier in the Method II run compared to the Method I runs. The Λ_{msr} values in the Method II run appear to be constrained between 1 and 2. There are not large fluctuations like those seen in

Method I. This suggests that environments with more gas prevents or limits mass segregation, however more work is needed in this area.

Run	Time (Myr)	$N_{clusters}$	$N_{clusters > 40stars}$	Min λ_{msr}	error	Max Λ_{msr}	Error
MI Run I	0.0	16	12	0.77	0.09	1.23	0.14
	1.0	12	6	1.19	0.10	1.91	0.21
	2.0	8	4	1.06	0.10	2.19	0.27
	3.0	4	3	1.17	0.16	2.33	0.23
	4.0	5	3	1.10	0.16	4.76	0.63
	5.0	2	1	N/A	N/A	1.33	0.16
	6.0	8	3	1.04	0.08	1.73	0.26
	7.0	9	3	1.16	0.09	1.61	0.20
7.7	9	1	N/A	N/A	1.76	0.27	
MI Run II	0.0	16	12	0.83	0.08	1.25	0.17
	1.0	12	6	1.18	0.11	1.89	0.24
	2.0	9	5	1.00	0.08	1.77	0.20
	3.0	4	4	1.10	0.10	1.77	0.19
	4.0	7	4	1.08	0.15	2.83	0.31
	5.0	3	2	1.07	0.12	1.35	0.13
	6.0	9	4	0.91	0.07	1.30	0.17
	7.0	8	2	1.27	0.16	1.33	0.14
7.7	7	1	N/A	N/A	1.22	0.17	
MI Run III	0.0	16	12	0.82	0.08	1.25	0.17
	1.0	12	6	1.19	0.10	1.90	0.21
	2.0	9	4	1.02	0.12	2.02	0.27
	3.0	3	3	1.12	0.13	1.90	0.20
	4.0	5	3	1.10	0.16	2.31	0.28
	5.0	4	2	1.01	0.08	1.12	0.13
	6.0	7	3	1.14	0.09	1.42	0.21
	7.0	7	3	0.92	0.09	1.44	0.21
7.7	10	2	1.05	0.10	1.32	0.18	
MI Run IV	0.0	16	12	0.80	0.08	1.23	0.15
	1.0	11	6	1.17	0.10	1.88	0.21
	2.0	12	4	1.01	0.11	1.96	0.28
	3.0	3	3	1.24	0.15	2.08	0.23
	4.0	6	4	1.07	0.10	3.93	0.48
	5.0	1	1	N/A	N/A	1.44	0.16
	6.0	10	3	1.04	0.12	1.69	0.23
	7.0	12	5	0.92	0.07	2.03	0.25
7.7	4	2	0.98	0.09	1.45	0.21	
MI Run V	0.0	16	12	0.88	0.10	1.21	0.15
	1.0	11	7	0.96	0.07	1.76	0.25
	2.0	12	5	1.11	0.12	2.06	0.25
	3.0	4	3	1.15	0.13	2.24	0.22
	4.0	5	3	1.07	0.16	1.95	0.26
	5.0	4	1	N/A	N/A	1.07	0.15
	6.0	10	2	1.34	0.12	1.88	0.30
	7.0	10	4	0.89	0.08	1.37	0.16
7.7	8	1	N/A	N/A	1.47	0.21	
MI Run VI	0.0	16	12	0.81	0.08	1.22	0.16
	1.0	14	6	1.19	0.10	1.90	0.18
	2.0	10	4	1.14	0.11	1.95	0.24
	3.0	5	3	1.03	0.12	3.06	0.30
	4.0	4	3	1.36	0.23	2.86	0.36
	5.0	3	1	N/A	N/A	1.48	0.19
	6.0	13	2	0.97	0.09	1.82	0.27
	7.0	11	5	0.86	0.07	1.57	0.19
7.7	7	2	1.19	0.08	1.22	0.18	
MII Run I	0.0	16	12	0.81	0.08	1.22	0.17
	1.0	9	5	1.09	0.10	1.58	0.20
	2.0	6	4	1.07	0.10	1.70	0.24
	2.2	6	3	1.22	0.12	1.91	0.25
MII Run II	0.0	16	12	0.81	0.07	1.24	0.15
	1.0	12	7	1.09	0.08	1.88	0.20
	2.0	4	3	1.34	0.17	1.81	0.24
	3.0	4	3	1.19	0.18	1.61	0.16
	4.0	2	1	N/A	N/A	1.18	0.18
4.6	1	1	N/A	N/A	1.56	0.16	

TABLE 4.2: Λ_{msr} values for clusters picked out by DBSCAN.

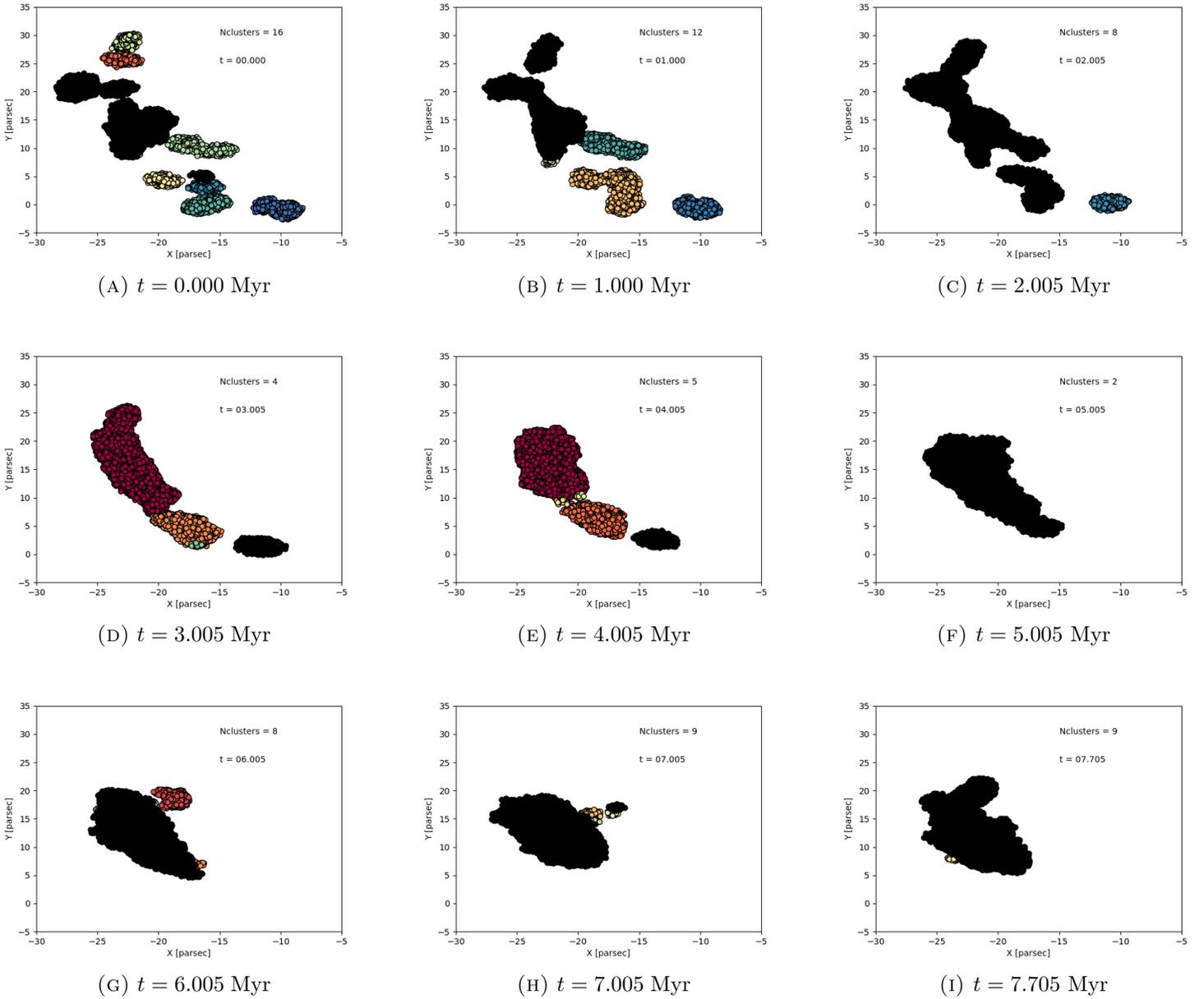


FIGURE 4.26: Clusters picked out by DBSCAN for Method I Run I with most mass-segregated clusters in black.

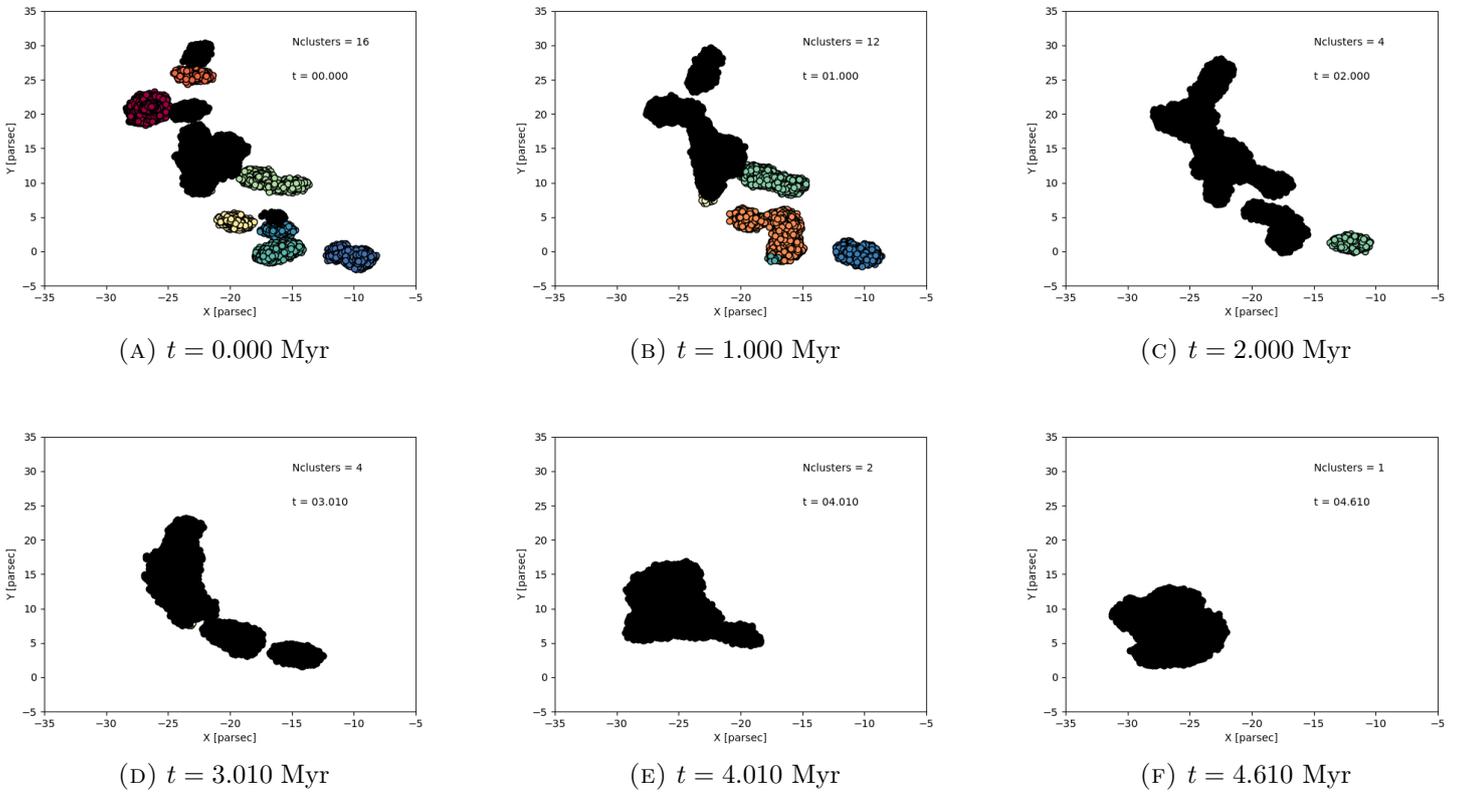


FIGURE 4.27: Clusters picked out by DBSCAN for Method II Run II with most mass-segregated clusters in black.

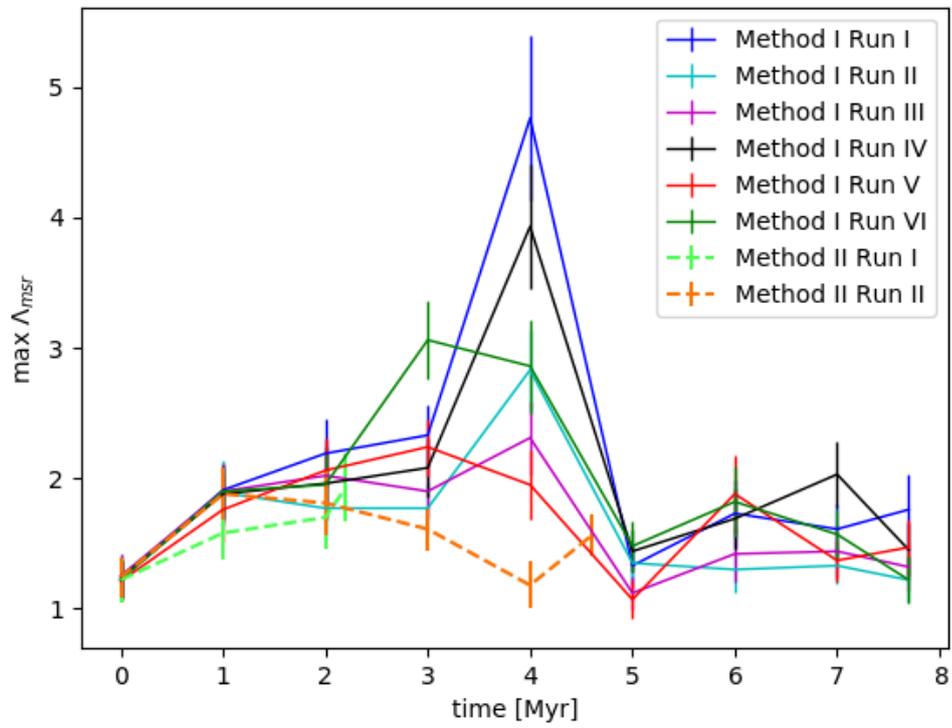


FIGURE 4.28: Max Λ_{msr} over time for clusters identified by DB-SCAN for all runs. Run details can be seen in Table 4.1.

4.11 Further Comparisons

In this section, we make further comparisons of different quantities over time between the runs. The purpose of comparing these quantities is to determine if there are any additional differences between the runs (and hence the initial conditions) that were not previously seen in the DBSCAN or dendrogram analysis. It also serves as another way to quantify how the initial gas properties affect the overall cluster evolution. A list of the specific quantities is as follows:

- Kinetic and potential energies of the gas
- Kinetic and potential energies of the stars
- Bound stellar and gas mass
- Lagrange radii of the stars and gas
- Average gas temperature over time

All of these quantities were tracked in the Method I runs. All of these quantities except the bound stellar and gas mass were tracked in the Method II runs due to the increased computational time required. Figure 4.29 shows the kinetic and potential energies of the stars and gas as a function of time for Method I runs I-VI. Runs I, II, and III all have similar gas kinetic energies at $t = 0$ Myr. This is expected as these runs have the same initial gas velocity dispersion of 1.4 km/s. Runs IV and V start with a higher gas kinetic energy and run VI has the greatest gas kinetic energy initially. This too is expected as the kinetic energy scales with the initial gas velocity dispersion (runs IV and V have initial gas velocity dispersions of 2

km/s and run VI has an initial gas velocity dispersion of 5 km/s). There is a drop in the gas kinetic energy across all runs between $t = 0$ Myr and $t = 1$ Myr with run VI experiencing the most significant drop. We did not see any large differences in the dendrograms between $t = 0$ Myr and $t = 1$ Myr which is consistent with the gas having slow movement during this time interval. We noticed significant changes in the gas structure becoming apparent at $t = 2$ Myr onwards which is consistent with the observed rise in the gas kinetic energy from this point. It is interesting to note that runs I and III begin with a lower gas kinetic energy but end up with higher values than runs IV and V. Despite having starting with the same initial gas velocity dispersion, run II deviates from runs I and III and only experiences a slight increase in the gas kinetic energy over time. Run II used an initial uniform sphere of gas while runs I and III used a gas distribution derived from the Carina observations. This shows that the initial gas distribution has an effect on the evolution of the gas kinetic energy over time. All the runs appear to have the gas kinetic energies level off beyond $t = 7$ Myr which is in line with the previously seen gas distributions becoming centrally concentrated and persisting. If we consider the gravitational potential energy of the gas, we see that all the runs begin with similar potentials except run II. All the runs except run II used a gas distribution derived from the Carina observations and so this result is expected. There is still some slight variation in the exact value at $t = 0$ Myr between these runs which is due to the differences in the particle sets generated with the sampling method. All the runs except run VI show a downward turn in the potential over time which implies that the gas becomes more bound. Runs I and III show the steepest decrease indicating that the gas is most bound in these runs. Compared to runs IV and V, we see that a difference in the initial gas velocity dispersion (with

the same initial gas distribution) affects the gas potential over time where the lower dispersion value leads to a more tightly bound gas distribution. Run II shows that gas initially distributed in a sphere with an initial velocity dispersion of 1.4 km/s also tends to become more bound over time. Run VI shows an upward trend in the potential indicating that the gas becomes less bound over time. This is in line with the results from the dendrogram analysis which showed that substructures in the gas did not persist as long as they did in runs I and III-V.

The kinetic energy of the stars over time appears similar between $t = 0$ Myr and $t = 4$ Myr. Beyond $t = 4$ Myr, we start to see a spread in the kinetic energies between the runs. This is interesting because we did not see any large significant differences between the runs in the DBSCAN analysis. However, here we see that the initial conditions of the gas do indeed affect the stellar properties over time. The two runs for which this is most noticeable are runs II and VI. Comparing run II to runs I and III shows that the initial uniform sphere of gas causes the kinetic energy of the stars to dip between $t = 4$ Myr and $t = 6$ Myr when the initial gas velocity dispersion is 1.4 km/s. Comparing run VI to runs I, III, IV and V shows that increasing the initial gas velocity dispersion to 5 km/s also causes a dip in the kinetic energy when the initial gas distribution is derived from the observations.

If we consider the gravitational potential energy of the stars over time, we again see that the runs appear similar between $t = 0$ Myr and $t = 4$ Myr. Beyond this, we again see a spread in the potential energies with runs II and VI being the most noticeable. We also see a spread in the stellar kinetic energies beyond $t = 4$ Myr with these same two runs being the most noticeable. The results from the DBSCAN analysis showed that the star clusters merged together until

$t = 4$ to $t = 5$ Myr when a predominant single cluster forms. We do not see large differences in the stellar kinetic or potential energies between the runs until $t = 4$ Myr which shows that the gas properties do not significantly affect the stars until they merge together. Once they merge, the surrounding gas has a significant effect. In runs II and VI, there were only regions of lower density gas distributed somewhat spherically from $t = 4$ Myr onwards. Our simulations take interactions between the stars and gas into account. Therefore the less dense gas results in less interactions taking place which could slow the movement and prevent the stars from reaching bound states similar to the other runs. Conversely, the other runs have higher density gas distributed in a filamentary structure from $t = 4$ onwards which results in more interactions with the stars causing the stellar kinetic energies to be higher and the stellar potential energies to be lower. We also see that as the initial gas velocity dispersion increases in the runs where the Carina observations were used (runs I, IV, V and VI), the less bound the stars are between $t = 4$ Myr and $t = 6$ Myr. The dendrogram analysis showed that increasing the initial velocity dispersion results in less dense regions of gas over time which ultimately contributes to less interactions taking place between the stars and the gas. As well, run II was the only run where the gas was initially distributed in a uniform sphere. This run differs from the other runs in that it reaches its most bound state (most negative potential energy and highest kinetic energy) at $t \approx 6.5$ Myr while the other runs reach this state at $t = 6$ Myr. Therefore, the spread seen in the stellar kinetic and potential energies of the different runs from $t = 4$ Myr onwards is attributed to the differing gas densities at this time which depends on the initial gas velocity dispersion and the initial spatial distribution of the gas.

Figure 4.30 shows the kinetic and potential energies of the stars and gas as a function of time for both Method I and Method II runs. Initially, the gas potential energy for the Method II runs is significantly lower (more bound) than the Method I runs. This is expected due to the increased gas mass used in the Method II runs. The gas potential energy decreases over time indicating significant contraction in the gas cloud from $t = 0$ to $t = 4.6$ Myr. Contraction was observed in Method I runs I-V although this occurred at a slower rate. Of all the runs, Method I run VI was the only one to show an increase in the gas potential energy over time (becoming less bound) which is attributed to the relatively high initial gas velocity dispersion (5 km/s). The gas kinetic energy of the Method II runs increases drastically over time which is consistent with the significant changes in the gas structure seen in the dendrogram analysis. In all the Method I runs, the gas kinetic energy levels off after the gas distribution becomes centrally concentrated. As previously mentioned, the Method II runs were evolved to a maximum of 4.6 Myr due to computational constraints (specifically the wall clock time to evolve the model and generate the output files, which contain the particle properties, at regular timesteps was increased to approximately 10 hours from 3-4 hours). However, it would be interesting to see if and when a centrally concentrated gas distribution forms in the Method II runs which would require further evolution of the system.

The initial gas conditions of the Method II runs results in significantly different trends in the potential and kinetic energies of the stellar distribution. The difference between the Method I and Method II runs starts to become apparent after $t = 1.5$ Myr. After this time, the stellar kinetic energy increases and the stellar

potential energy decreases compared to the Method I runs. This shows that the stellar distribution in the Method II runs becomes more bound over time compared to the Method I runs. Again, this shows that increased gas densities result in more interactions between the stars and gas resulting in higher stellar kinetic energies and lower stellar potential energies. To summarize, the gas distribution in the Method II run becomes more bound in a shorter period of time compared to the Method I runs as evidenced by the faster decrease in the value of the gas potential energy (more negative gas potential energy is more bound). This increased gas potential works to contract the stellar distribution in shorter amount of time as seen in the decrease in the value of the stellar potential energy.

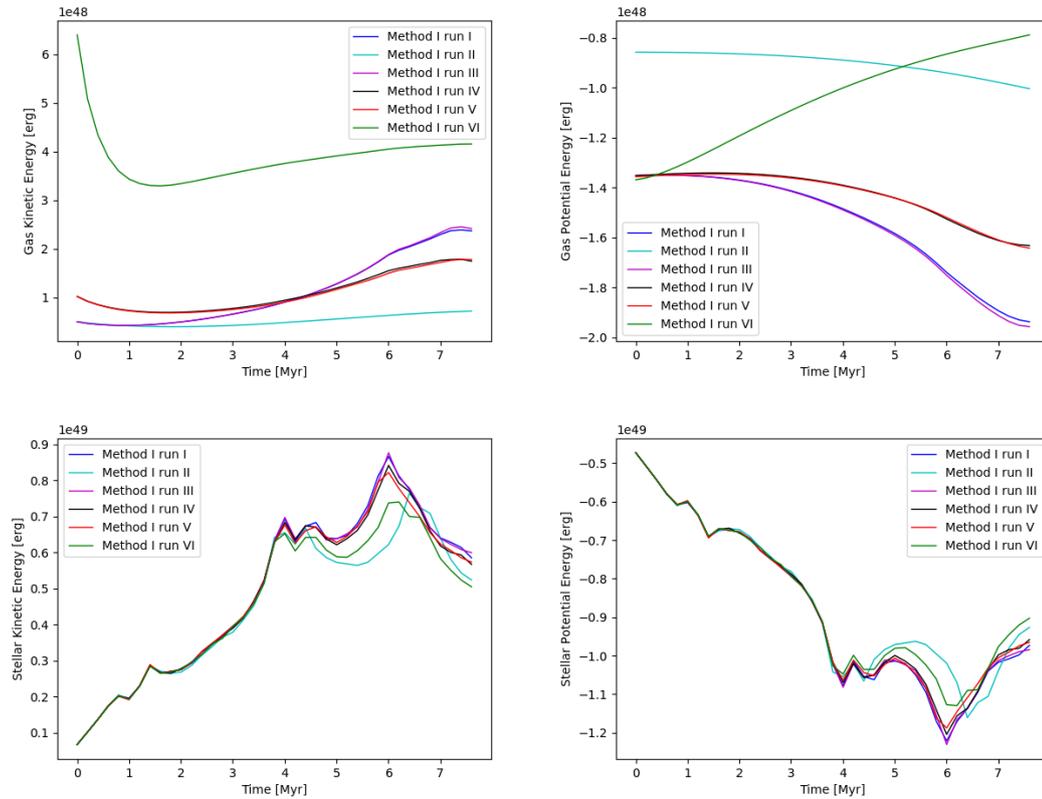


FIGURE 4.29: Kinetic and potential energies of the stars and gas over time for Method I runs I-VI.

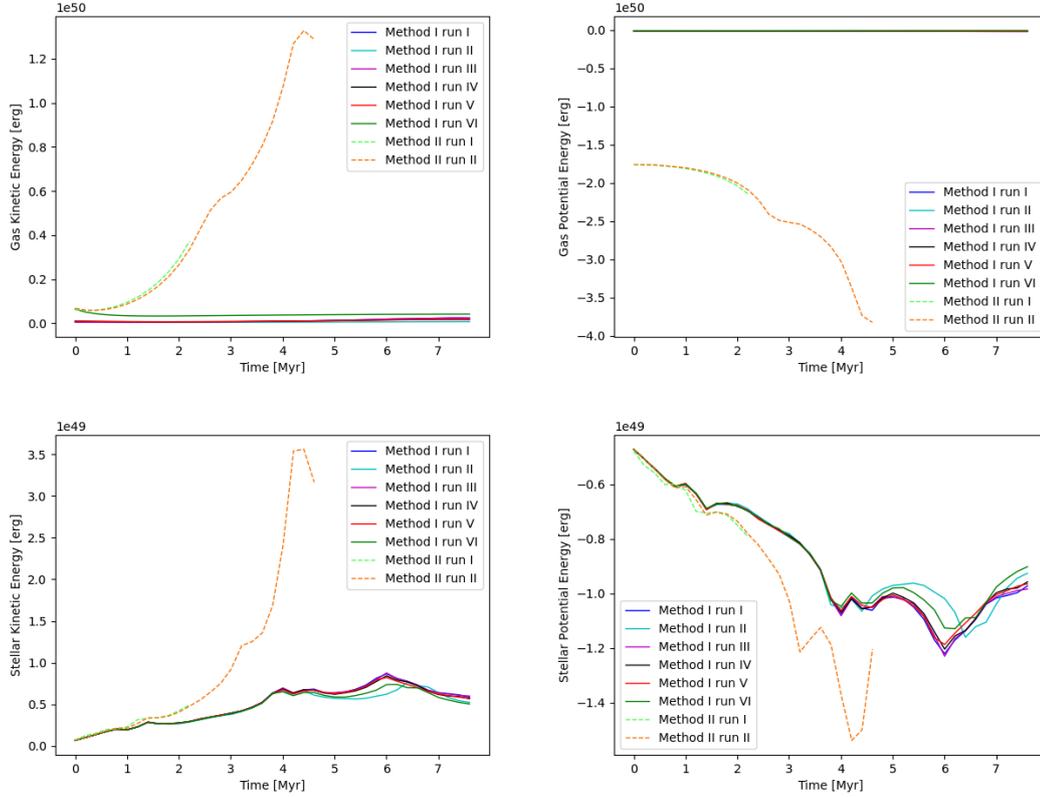


FIGURE 4.30: Kinetic and potential energies of the stars and gas over time for Method I and Method II runs.

The bound stellar and gas mass over time can be seen in Figure 4.31. The bound gas mass at $t = 0$ is similar for all runs except run VI. The slight differences in runs I-V are due to the different initial gas velocity dispersions and the initial gas distribution. Comparing runs I, III, IV and V, shows that an increase in the gas velocity dispersion (from 1.4 km/s to 2 km/s) results in a lower total bound gas mass. Increasing the gas velocity dispersion to 5 km/s (run VI) drastically decreases the total bound mass to about half. Between $t = 0$ Myr and $t = 1.5$ Myr, all runs show an increase in the bound gas mass which corresponds to the sub-cluster contraction seen in the XY and XZ images in the previous sections.

From $t = 2$ Myr onwards, the bound mass decreases with a prominent decrease occurring at around $t = 6$ Myr to $t = 7$ Myr for all the runs. It is also during the time interval that the stellar potential energy reaches its lowest (most negative) point indicating that the stellar distribution is the most bound. The bound stellar mass across all the runs shows a sharp decrease around $t = 3$ Myr to $t = 4$ Myr. The results from the DBSCAN analysis show that a few clusters are picked out that are in the process of merging at this time. There is also an increase in the stellar kinetic energy during this time interval as the merging occurs which leads to this decrease in the bound stellar mass. Beyond this, the bound stellar mass fluctuates without any distinct pattern between the runs.

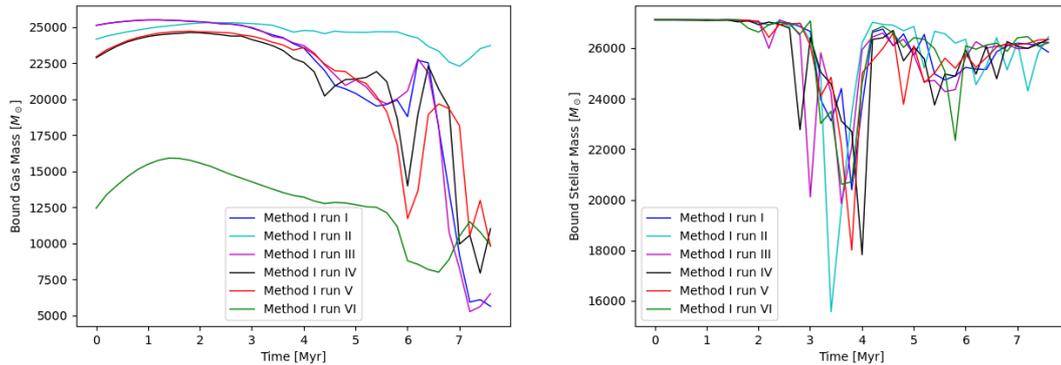


FIGURE 4.31: LEFT: Bound gas mass over time for Method I runs I-VI. RIGHT: Bound stellar mass over time for Method I runs I-VI.

Figure 4.32 shows the Lagrangian radii of the stars and gas over time. These radii describe a sphere centered on the stellar or gas distribution which contains a certain amount of mass. For example the 10% Lagrangian radius describes the radius of a sphere which contains 10% of the total mass in question. The 10% Lagrangian radius of the gas is initially the same for all Method I runs except for run II. This is expected as the Lagrangian radii depends on the spatial distribution

of the gas and run II is the only run where the gas is distributed in a significantly different manner (in a uniform sphere). Slight variations between the value can be seen in the other runs with Method I and this is attributed to the different particle sets created using the sampling method. Over time, all runs show a decrease in the 10% Lagrangian radius of the gas except for run VI. This is interesting as it shows that the gas distribution does not become more concentrated in the center than it was initially. By $t = 7.7$ Myr, run VI has the highest 10% Lagrangian radius of the gas. This shows that a greater initial gas velocity dispersion does not significantly affect the concentration of the gas near the center of the distribution. The 10% Lagrangian radius of the gas for the Method II runs is initially greater than most of the Method I runs except for Method I run II. This is expected as the gas distribution is more similar to Method I runs I, III-VI but with increased gas mass. Over time, the 10% Lagrangian radius of the gas for the Method II runs is seen to decrease significantly in a shorter period of time compared to the Method I runs with fluctuations occurring between $t = 2.5$ and $t = 4.6$ Myr. The significant decrease indicates a greater contraction of the gas which results in the higher densities observed in the dendrogram analysis. Considering the 50% Lagrangian radius, or the half-mass radius, Method I run VI shows a significant increase in this value with time and it tends to decrease with the other runs. Comparing Method I runs I and III to runs IV and V shows that an increase in the initial gas velocity dispersion from 1.4 km/s to 2 km/s slows the decrease of the half-mass radius. The significant increase in the gas half-mass radius value for Method I run VI physically corresponds to the dispersal or expansion of the gas distribution. This is interesting as this expansion is not able to be clearly seen in the dendrogram analysis. Method I run II used gas initially in a uniform sphere with a gas velocity

dispersion of 1.4 km/s. The gas half-mass radius for this run is different than the other runs at $t = 0$ but decreases over time similar to runs I, III, IV and V. The Method II runs begin with a 50% Lagrangian radius greater than most of the Method I runs except Method I run II and this value decreases significantly from $t = 2$ Myr onwards which again is indicative of increased contraction in the gas cloud over a short period of time compared to the Method I runs.

The stellar 10% Lagrangian radius appears to be similar across all the Method I runs. Initially, there is a significant decrease between $t = 0$ Myr and $t = 1$ Myr which corresponds to the initial sub-cluster contraction. After this point, the clusters expand slightly and then start to merge together which corresponds to the increase in the 10% Lagrangian radius value at around $t = 1.5$ Myr followed by the slow decrease over time. The Method II runs follow the same trend until $t = 2.5$ Myr when the stellar 10% Lagrangian radius begins to decrease more significantly than the other runs. Overall, the stellar half-mass radius (50% Lagrangian radius) value across all the Method I runs is similar over time with some separation at $t = 5$ Myr onwards seen in Method I runs II and VI. As previously mentioned, the gas distributions appear to have an effect on the stars after they have merged together and at $t = 5$ Myr, there is one predominant single cluster of stars. Method I Runs II and VI have regions of lower density gas compared to the other runs at this time. As stated before, this results in less significant interactions between the gas and the stars causing the stellar potential energies to be less bound (more positive) for these runs. This corresponds to an increase in the distances between the stars seen as an increase in the half-mass radius. The stellar half-mass radius for the Method II runs begins to significantly deviate from the other runs at $t = 1.5$

Myr. At this point the stellar half-mass radius decreases more rapidly compared to the Method I runs. At $t = 4.6$ Myr, there is a sharp increase which is due to some slight cluster expansion. The differences between the Method I and Method II runs demonstrate the effect of differing initial gas conditions as the increased gas mass causes earlier and increased contraction in both the stars and gas. Again, in our simulations we only focused on the region where the stars and gas overlap. If the other surrounding gas were to be included as well, this may have an effect on the collapse of the gas and stellar distributions. For example, the surrounding gas will likely result in pressure forces that keep the overall gas cloud from collapsing as quickly resulting in a slower decrease in the half-mass radius of the gas which would be more realistic.

The average gas temperatures for all the runs can be seen in Figure 4.33. Among the Method I runs, run VI is seen to have a drastic increase in the average gas temperature until about $t = 1.5$ Myr which then decreases beyond that while the other runs show a somewhat steady increase. The Method II runs initially show an increase similar to the Method I runs except run VI. After roughly $t = 2$ Myr, the temperature of the Method II run drastically increases compared to the other runs. The temperature values themselves seem quite high and there appears to be a dependence on the initial velocity dispersion and initial gas mass. This is certainly an area that can be further explored in the future.

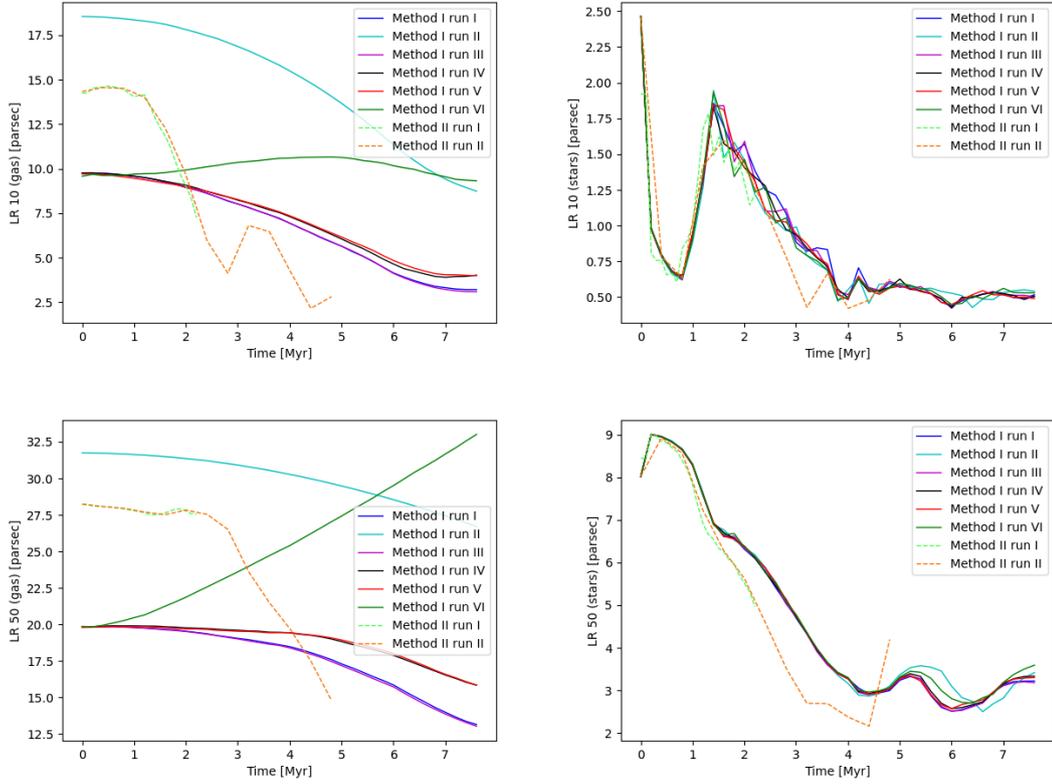


FIGURE 4.32: LEFT: Gas Lagrange radii for Method I and Method II runs. RIGHT: Stellar Lagrange radii for Method I and Method II runs.

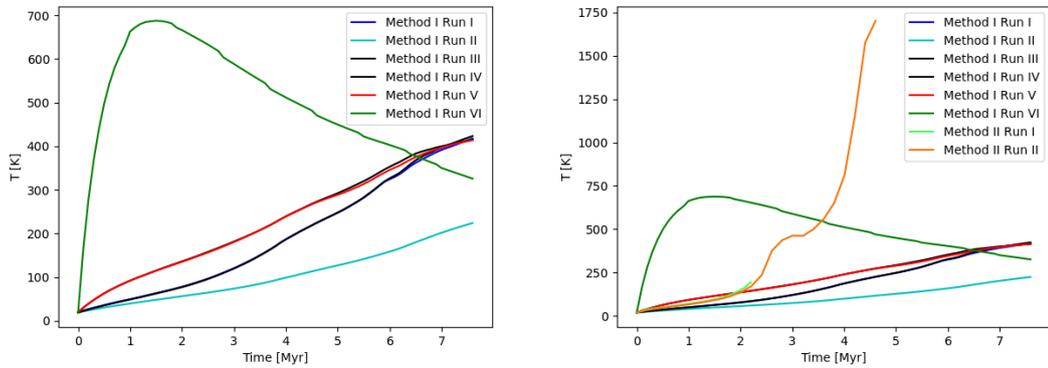


FIGURE 4.33: LEFT: Average gas temperature for Method I runs. RIGHT: Average gas temperature for Method I and Method II runs.

Chapter 5

Discussion and Conclusion

5.1 Incorporation of Observational Gas Data

We have successfully developed a technique to incorporate observational gas column density data into a star cluster simulation. We developed two versions where the first randomly samples the gas distribution and the second assigns at least one particle to represent each pixel. When we applied our techniques to a simulated data set as a test case, we found that Method II was better able to capture the spatial density distribution in 2D. We also explored different Z depth (cloud depth) estimates and found that using distance transforms of the high and low density regions gave rise to filamentary structures. This is more realistic than a constant cloud depth, however more work needs to be done in order to improve its semblance to a gas cloud. For example, the current technique forces the filaments to be symmetric which is not necessarily the case in real clouds. In our simulations, we paired Method I with a constant cloud depth of 20 parsecs and we paired Method II with a varied cloud depth determined using the distance transforms.

5.2 Effect of Initial Gas Properties on Star Cluster Evolution

We used stellar positions derived from the MYSTiX project and our gas particle creation techniques to model the evolution of the Carina region. We explored different initial gas velocity dispersions as well as gas initially distributed in a uniform sphere. We found that in all of our Method I runs, the stellar sub-clusters tended to merge together in a similar manner to form one large main cluster of stars. This merging process was not significantly affected by the initial properties of the gas. After the main cluster of stars was formed, the effects of the surrounding gas became noticeable. For example, in runs where the main cluster was surrounded by less dense gas, the kinetic energy of the stars was observed to be less compared to runs where the main cluster was surrounded by higher density gas. It is interesting that the differences between the initial gas properties only became apparent after the large main cluster was formed. As well, we explored three initial gas velocity dispersions of 1.4, 2 and 5 km/s. We found that the greater the initial gas velocity dispersion, the less likely it was for regions of dense gas to persist. This ultimately affects future star formation as the density threshold is not reached. Increasing the initial gas velocity dispersion also resulted in the gas distribution becoming more spherical in a shorter amount of time which shows that filamentary structures and clumps in the gas are more likely to remain with low gas velocity dispersions.

We performed two runs using Method II with an initial gas velocity dispersion of 1.4 km/s. However, due to the increased computation time from the significant

increase in gas particles, we were only able to evolve the system to 4.6 Myr. Compared to the Method I runs, we found differences in terms of the the filamentary structures in the gas. Regions of higher gas densities were also seen using Method II compared to Method I which has implications when considering future sites of star-formation.

We analyzed the stellar distribution over time using DBSCAN (a clustering algorithm) and tracked the mass segregation of the identified clusters over time by calculating Λ_{msr} values. We found that there in general there were no significant differences between the runs except at $t = 4$ Myr. This is right before the clusters merge together in the Method I runs and where the clusters merge in the Method II run. The Method II run showed that the clusters merge to form one main cluster approximately 1 Myr earlier than the Method I run. As well, the main cluster that forms with the Method II run is not at the same location as the main cluster that forms in the Method I runs. The Method II run began with a significantly greater initial gas mass which is more realistic. Further investigations of the Carina region could be done with Method II while varying other parameters (e.g initial gas velocity dispersion).

5.3 Future Work

In the future, our models can made increasingly realistic by incorporating stellar feedback processes. We would then be able to study their effects on the surrounding gas initially derived from observations. Our models can also be evolved for a longer duration of time with massive stars included to investigate the effects from

supernovae feedback. In addition, a treatment of binaries either those that are primordial or form dynamically would be interesting to consider as well. We can also incorporate more observational gas data as it becomes available. For example, we could include better measurements of the gas velocity dispersion and possibly extinction measurements as well to estimate the cloud depth. In terms of analyzing our simulations, instead of only using a density threshold to infer star formation, we could also determine if certain dense regions end up bound. We could also incorporate star formation in our simulations by including sink particles. As well, we could improve our analysis of the gas distributions by using filament finder codes in addition to constructing dendrograms. Filament finder codes typically find thinner dense regions which could then be used to determine localized sites of potential star formation.

Both of our gas particle creation techniques were based on randomly sampling to some extent. The first version relied on random sampling of the original gas density distribution to create particles. The second originally assigned one particle to represent each pixel and would introduce more particles if the mass of the particle was greater than $0.5 M_{\odot}$. The positions in X and Y of these particles were determined by drawing from the ranges of the pixels in the X and Y dimensions. We could modify the particle distribution to resemble a glass distribution to observe any effect that may have.

Going forward, the goal would be to simulate other star-forming regions in the MYSTiX catalogue. The gas particle creation techniques we have developed will allow us to accurately simulate such regions furthering our understanding of star-formation, the loss of stellar substructure and mass segregation over time.

As our models become increasingly complex with the inclusion of stellar feedback processes, we will also be able to investigate the effects of gas expulsion in realistic star cluster environments. This would allow us to probe the process by which young embedded clusters emerge from their natal molecular cloud and the extent to which they will be bound in the context of specific star-forming regions. Being able to incorporate observational gas distributions ensures that we consider the history of such regions and their complex gas morphologies as we evolve these systems forward in time.

Bibliography

- Allison, R. J., Goodwin, S. P., Parker, R. J., Portegies Zwart, S. F., and Grijs, R. de (2010a). The early dynamical evolution of cool, clumpy star clusters. *Monthly Notices of the Royal Astronomical Society* 407, 1098–1107.
- Allison, R. J., Goodwin, S. P., Parker, R. J., Portegies Zwart, S. F., and Grijs, R. de (2010b). The early dynamical evolution of cool, clumpy star clusters. *Monthly Notices of the Royal Astronomical Society* 407, 1098–1107.
- Allison, R. J., Goodwin, S. P., Parker, R. J., Portegies Zwart, S. F., Grijs, R. de, and Kouwenhoven, M. (2009). Using the minimum spanning tree to trace mass segregation. *Monthly Notices of the Royal Astronomical Society* 395, 1449–1454.
- André, P. et al. (2010). From filamentary clouds to prestellar cores to the stellar IMF: Initial highlights from the Herschel Gould Belt Survey. *Astronomy & Astrophysics* 518, L102.
- Arth, A., Donnert, J., Steinwandel, U., Böss, L., Halbesma, T., Pütz, M., Hubber, D., and Dolag, K. (2019). WVTICs – SPH initial conditions for everyone. *arXiv e-prints*, arXiv:1907.11250.
- Astropy-Collaboration, Price-Whelan, A., Contributors, A., et al. (Sept. 2018). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *The Astronomical Journal* 156, 123.

Bibliography

- Astropy-Collaboration, Robitaille, T. P., et al. (2013). Astropy: A community Python package for astronomy. *Astronomy & Astrophysics* 558, A33.
- Bally, J., Langer, W. D., Stark, A. A., and Wilson, R. W. (1987). Filamentary Structure in the Orion Molecular Cloud. *Astrophysical Journal Letters* 312, L45.
- Baumgardt, H. and Kroupa, P. (2007). A comprehensive set of simulations studying the influence of gas expulsion on star cluster evolution. *Monthly Notices of the Royal Astronomical Society* 380, 1589–1598.
- Beccari et al. (2020). Uncovering a 260pc wide, 35 Myr old filamentary relic of star formation. *Monthly Notices of the Royal Astronomical Society* 491, 2205–2216.
- Bertoldi, F. and McKee, C. F. (1992). Pressure-confined Clumps in Magnetized Molecular Clouds. *The Astrophysical Journal* 395, 140.
- Burton, M. G., Lazendic, J. S., Yusef-Zadeh, F., and Wardle, M. (2004). The Eye of the Tornado — an isolated, high-mass young stellar object near the Galactic Centre. *Monthly Notices of the Royal Astronomical Society* 348, 638–646.
- Cartwright, A. and Whitworth, A. P. (2004). The statistical analysis of star clusters. *Monthly Notices of the Royal Astronomical Society* 348, 589–598.
- Chen et al. (2019). Filamentary Accretion Flows in the Infrared Dark Cloud G14.225–0.506 Revealed by ALMA. *The Astrophysical Journal* 875, 24.
- Chira, R., Kainulainen, J., Ibañez-Mejía, J., Henning, T., and Mac Low, M. (2018). On the fragmentation of filaments in a molecular cloud simulation. *Astronomy & Astrophysics* 610, A62.
- Cortes-Rangel, G., Zapata, L. A., Toalá, J. A., Ho, P. T., Takahashi, S., Mesa-Delgado, A., and Masqué, J. M. (2020). ALMA Observations of the Extraordinary Carina Pillars: HH 901/902. *The Astrophysical Journal* 159, 62.

Bibliography

- De Grijs, R. and Goodwin, S. P. (2008). Star cluster ‘infant mortality’ in the Small Magellanic Cloud (Redivivus). *Monthly Notices of the Royal Astronomical Society* 383(3), 1000–1006.
- DeGioia-Eastwood, K., Throop, H., Walker, G., and Cudworth, K. (2001). The Star Formation History of Trumpler 14 and Trumpler 16. *The Astrophysical Journal* 549, 578–589.
- Dib, S., Schmeja, S., and Parker, R. J. (2018). Structure and mass segregation in Galactic stellar clusters. *Monthly Notices of the Royal Astronomical Society* 473, 849–859.
- Dinnbier, F. and Kroupa, P. (2020). Tidal tails of open star clusters as probes to early gas expulsion. II. Predictions for Gaia. *Astronomy & Astrophysics* 640, A85.
- Elmegreen, B. G. and Falgarone, E. (1996). A Fractal Origin for the Mass Spectrum of Interstellar Clouds. *The Astrophysical Journal* 471, 816–821.
- Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: *KDD*.
- Federrath, C., Banerjee, R., Clark, P. C., and Klessen, R. S. (2010). Modeling Collapse and Accretion in Turbulent Gas Clouds: Implementation and Comparison of Sink Particles in AMR and SPH. *The Astrophysical Journal* 713, 269–290.
- Feigelson, E. and Babu, G. (n.d.). *Astrostatistics and Astroinformatics Portal: Beware the Kolmogorov-Smirnov Test!* URL: <https://asaip.psu.edu/articles/beware-the-kolmogorov-smirnov-test/>.
- Feigelson, E. D. (2010). X-ray insights into star and planet formation. *Proceedings of the National Academy of Sciences* 107, 7153–7157.

Bibliography

- Feigelson, E. D. et al. (2013). Overview of the Massive Young Star-Forming Complex Study in Infrared and X-Ray (MYStIX) Project. *Astrophysical Journal Supplements* 209, 26.
- Frank, A. et al. (2014). Jets and Outflows from Star to Cloud: Observations Confront Theory. In: *Protostars and Planets VI*. Ed. by H. Beuther, R. S. Klessen, C. P. Dullemond, and T. Henning, 451.
- Fujii, M. and Portegies Zwart, S. (2016). The Formation and Dynamical Evolution of Young Star Clusters. *The Astrophysical Journal* 817, 4.
- Fujii, M., Iwasawa, M., Funato, Y., and Makino, J. (2007). BRIDGE: A Direct-Tree Hybrid N Body Algorithm for Fully Self-Consistent Simulations of Star Clusters and Their Parent Galaxies. *Publications of the Astronomical Society of Japan* 59(6), 1095–1106.
- Gaczkowski, B., Preibisch, T., Ratzka, T., Roccatagliata, V., Ohlendorf, H., and Zinnecker, H. (2013). Herschel far-infrared observations of the Carina Nebula complex. II. The embedded young stellar and protostellar population. *Astronomy & Astrophysics* 549, A67.
- Getman, K. V., Feigelson, E. D., Luhman, K. L., Sicilia-Aguilar, A., Wang, J., and Garmire, G. P. (2009). Protoplanetary disk evolution around the triggered star-forming region cepheus B. *The Astrophysical Journal* 699, 1454–1472.
- Gingold, R. A. and Monaghan, J. J. (1977). Smoothed particle hydrodynamics: theory and application to non-spherical stars. *Monthly Notices of the Royal Astronomical Society* 181, 375–389.
- Goldsmith, P. F., Heyer, M., Narayanan, G., Snell, R., Li, D., and Brunt, C. (2008). Large-Scale Structure of the Molecular Gas in Taurus Revealed by High Linear

Bibliography

- Dynamic Range Spectral Line Mapping. *The Astrophysical Journal* 680, 428–445.
- Goodwin, S. and Whitworth, A. (2003). The dynamical evolution of fractal star clusters: The survival of substructure. *Astronomy and Astrophysics* 413.
- Habibi, M., Stolte, A., Brandner, W., Hußmann, B., and Motohara, K. (2013). The Arches cluster out to its tidal radius: dynamical mass segregation and the effect of the extinction law on the stellar mass function. *Astronomy & Astrophysics* 556, A26.
- Heiderman, A., Evans Neal J., I., Allen, L. E., Huard, T., and Heyer, M. (2010). The Star Formation Rate and Gas Surface Density Relation in the Milky Way: Implications for Extragalactic Studies. *The Astrophysical Journal* 723(2), 1019–1037.
- Hernquist, L. and Katz, N. (1989). TREESPH: A Unification of SPH with the Hierarchical Tree Method. *Astrophysical Journal Supplement* 70, 419.
- Herschel John Frederick William, S. (1847). *Results of astronomical observations made during the years 1834, 5, 6, 7, 8, at the Cape of Good Hope; being the completion of a telescopic survey of the whole surface of the visible heavens, commenced in 1825.*
- Hillenbrand, L. A. and Hartmann, L. W. (1998). A Preliminary Study of the Orion Nebula Cluster Structure and Dynamics. *The Astrophysical Journal* 492, 540–553.
- Hills, J. (1980). The effect of mass loss on the dynamical evolution of a stellar system - Analytic approximations. *The Astrophysical Journal* 235, 986–991.

Bibliography

- Humphreys, R. M. and Martin, J. C. (2012). Eta Carinae: From 1600 to the Present. In: *Eta Carinae and the Supernova Impostors*. Vol. 384. Astrophysics and Space Science Library.
- Kashi, A. and Soker, N. (2010). Periastron Passage Triggering of the 19th Century Eruptions of Eta Carinae. *The Astrophysical Journal* 723, 602–611.
- Kauffmann, J., Pillai, T., and Goldsmith, P. F. (2013). Low Virial Parameters in Molecular Clouds: Implications for High-mass Star Formation and Magnetic Fields. *The Astrophysical Journal* 779(2), 185.
- Kim, B. et al. (2004). 13CO (J=1–0) Survey of Molecular Clouds toward the Monoceros and Canis Major Region. *Publications of the Astronomical Society of Japan* 56.
- Klaassen, P., Reiter, M., McLeod, A., Mottram, J., Dale, J., and Gritschneider, M. (2020). Carina’s pillars of destruction: the view from ALMA. *Monthly Notices of the Royal Astronomical Society* 491(1), 178–200.
- Kruijssen, J. D. (2012). On the fraction of star formation occurring in bound stellar clusters. *Monthly Notices of the Royal Astronomical Society* 426, 3008–3040.
- Krumholz, M. R. (2015). Notes on Star Formation. *arXiv e-prints*, arXiv:1511.03457.
- Krumholz, M. R., McKee, C. F., and Bland-Hawthorn, J. (2019a). Star Clusters Across Cosmic Time. *Annual Review of Astronomy & Astrophysics* 57, 227–303.
- Krumholz, M. R., McKee, C. F., and Bland-Hawthorn, J. (2019b). Star Clusters Across Cosmic Time. *Annual Review of Astronomy and Astrophysics* 57, 227–303.
- Kuhn et al. (2014). The spatial structure of young stellar clusters I. Subclusters. *The Astrophysical Journal* 787.

Bibliography

- Kuhn, M. A., Feigelson, E. D., Getman, K. V., Sills, A., Bate, M. R., and Borissova, J. (2015). The Spatial Structure of Young Stellar Clusters. III. Physical Properties and Evolutionary States. *The Astrophysical Journal* 812, 131.
- Lada, C. J. and Lada, E. A. (2003). Embedded Clusters in Molecular Clouds. *Annual Review of Astronomy and Astrophysics* 41, 57–115.
- Lallement et al. (2014). 3D maps of the local ISM from inversion of individual color excess measurements. *Astronomy & Astrophysics* 561.
- Larson, R. (1981). Turbulence and star formation in molecular clouds. *Monthly Notices of the Royal Astronomical Society* 194, 809–826.
- Larson, R. B. (2003). The physics of star formation. *Reports on Progress in Physics* 66, 1651–1697.
- Lin, C., Mestel, L., and Shu, F. (1965). The Gravitational Collapse of a Uniform Spheroid. *The Astrophysical Journal* 142, 1431.
- Littlefair, S. P., Naylor, T., Jeffries, R. D., Devey, C. R., and Vine, S. (2003). Mass segregation in the young open cluster NGC 2547. *Monthly Notices of the Royal Astronomical Society* 345, 1205–1211.
- Low, F. et al. (1984). Infrared cirrus: new components of the extended infrared emission. *The Astrophysical Journal* 278, L19–L22.
- Makino, J. and Aarseth, S. J. (1992). On a Hermite Integrator with Ahmad-Cohen Scheme for Gravitational Many-Body Problems. *Publications of the Astronomical Society of Japan* 44, 141–151.
- Marsh, K. et al. (2017). Multitemperature mapping of dust structures throughout the Galactic Plane using the PPMAP tool with Herschel Hi-GAL data. *Monthly Notices of the Royal Astronomical Society* 471, 2730–2742.

Bibliography

- Marsh, K., Whitworth, A., and Lomax, O. (2015). Temperature as a third dimension in column-density mapping of dusty astrophysical structures associated with star formation. *Monthly Notices of the Royal Astronomical Society* 454, 4282–4292.
- McClure-Griffiths, N., Dickey, J., Gaensler, B., Green, A., and Haverkorn, M. (2006). Magnetically Dominated Strands of Cold Hydrogen in the Riegel-Crutcher Cloud. *The Astrophysical Journal* 652, 1339–1347.
- McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley.
- McMillan, S. L., Vesperini, E., and Portegies Zwart, S. F. (2007). A Dynamical Origin for Early Mass Segregation in Young Star Clusters. *The Astrophysical Journal* 655, L45–L49.
- Mizuno, A., Onishi, T., Yonekura, Y., Nagahama, T., Ogawa, H., and Fukui, Y. (1995). Overall Distribution of Dense Molecular Gas and Star Formation in the Taurus Cloud Complex. *Astrophysical Journal Letters* 445, L161.
- Molinari et al. (2010). Hi-GAL: The Herschel Infrared Galactic Plane Survey. *The Astronomical Society of the Pacific* 122, 314–325.
- Monaghan (1992). Smoothed Particle Hydrodynamics. *Annual Review of Astronomy & Astrophysics* 30, 543–574.
- Monaghan, J. (1997). SPH and Riemann Solvers. *Journal of Computational Physics* 136, 298–307.
- Monahan, J. F. (2011). *Numerical Methods of Statistics*. 2nd. USA: Cambridge University Press.
- Nayak, O., Meixner, M., Indebetouw, R., De Marchi, G., Koekemoer, A., Panagia, N., and Sabbi, E. (2016). Studying Relation Between Star Formation

Bibliography

- and Molecular Clumps on Subparsec Scales in 30 Doradus. *arXiv e-prints*, arXiv:1608.05451.
- Pang, X., Grebel, E. K., Allison, R. J., Goodwin, S. P., Altmann, M., Harbeck, D., Moffat, A. F. J., and Drissen, L. (2013). On the Origin of Mass Segregation in NGC 3603. *The Astrophysical Journal* 764, 73.
- Parker, R. J., Dale, J. E., and Ercolano, B. (2015). Primordial mass segregation in simulations of star formation? *Monthly Notices of the Royal Astronomical Society* 446, 4278–4290.
- Pedregosa et al. (2011a). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011b). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Pelupessy et al. (2009). A multiphysics and multiscale software environment for modeling astrophysical systems. *New Astronomy* 14, 369–378.
- Pelupessy et al. (2013). The Astrophysical Multipurpose Software Environment. *Astronomy & Astrophysics* 557.
- Portegies-Zwart et al. (2013). Multi-physics Simulations Using a Hierarchical Interchangeable Software Interface. *Computer Physics Communications* 183, 456–468.
- Portegies-Zwart et al. (2018). Astrophysical Recipes: the art of AMUSE. *AAS IOP Astronomy publishing*.

Bibliography

- Povich, M. S. et al. (2011). Candidate X-ray emitting OB stars in the Carina nebula identified via infrared spectral energy distributions. *The Astrophysical Journal Supplement Series* 194, 6.
- Preibisch, T. et al. (2011). Near-infrared properties of the X-ray emitting young stellar objects in the Carina nebula. *The Astrophysical Journal Supplement Series* 194, 10.
- Preibisch, T., Zeidler, P., Ratzka, T., Roccatagliata, V., and Petr-Gotzens, M. (Dec. 2014). The Vista Carina Nebula Survey. I. Introduction and source catalog. *Astronomy Astrophysics* 572, A116.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Second. Cambridge, USA: Cambridge University Press.
- Price-Jones, N. and Bovy, J. (2019). Blind chemical tagging with DBSCAN: prospects for spectroscopic surveys. *Monthly Notices of the Royal Astronomical Society* 487, 871–886.
- Rebolledo, D. et al. (2016). The Carina Nebula and Gum 31 molecular complex - I. Molecular gas distribution, column densities, and dust temperatures. *Monthly Notices of the Royal Astronomical Society* 456, 2406–2424.
- Robitaille et al. (2013). URL: <https://dendrograms.readthedocs.io/en/stable/>.
- Roccatagliata, V., Preibisch, T., Ratzka, T., and Gaczkowski, B. (2013). Herschel far-infrared observations of the Carina Nebula complex. III. Detailed cloud structure and feedback effects. *Astronomy & Astrophysics* 554, A6.
- Sales, L. V., Marinacci, F., Springel, V., and Petkova, M. (2014). Stellar feedback by radiation pressure and photoionization. *Monthly Notices of the Royal Astronomical Society* 439(3), 2990–3006.

Bibliography

- Scally, A. and Clarke, C. (2002). Primordial substructure in the Orion Nebula Cluster. *Monthly Notices of the Royal Astronomical Society* 334, 156–166.
- Schneider, N., Csengeri, T., Bontemps, S., Motte, F., Simon, R., Hennebelle, P., Federrath, C., and Klessen, R. (2010). Dynamic star formation in the massive DR21 filament. *Astronomy & Astrophysics* 520, A49.
- Shu, F. (1992). *The physics of astrophysics. Volume II: Gas dynamics*.
- Sills et al. (2018). Dynamical evolution of stars and gas of young embedded stellar sub-clusters. *Monthly Notices of the Royal Astronomical Society* 477, 1903–1912.
- Smith et al. (2006). The Structure of the Homunculus. I Shape and Latitude Dependence from H₂ and (Fe II) Velocity Maps of eta Carinae. *The Astrophysical Journal* 644, 1151–1163.
- Smith, N., Egan, M. P., Carey, S., Price, S. D., Morse, J. A., and Price, P. A. (2000). Large-Scale Structure of the Carina Nebula. *The Astrophysical Journal* 532, L145–L148.
- Springel (2005). The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society* 364, 1105–1134.
- Springel, V. and Hernquist, L. (2002). Cosmological smoothed particle hydrodynamics simulations: the entropy equation. *Monthly Notices of the Royal Astronomical Society* 333, 649–664.
- Stelzer, B., Preibisch, T., Alexander, F., Mucciarelli, P., Flaccomio, E., Micela, G., and Sciortino, S. (2012). X-ray view of IC 348 in the light of an updated cluster census. *Astronomy & Astrophysics* 537, A135.
- Swain, M. and Ballard, D. (1991). Color Indexing. *International Journal of Computer Vision* 7, 11–32.

Bibliography

- Townsley, L. K. et al. (2011a). The CHANDRA Carina Complex Project: Deciphering the Enigma of Carina's Diffuse X-ray Emission. *The Astrophysical Journal Supplement Series* 194, 15.
- Townsley, L. K., Broos, P. S., Chu, Y.-H., Gruendl, R. A., Oey, M. S., and Pittard, J. M. (2011b). The integrated diffuse X-ray emission of the Carina nebula compared to other massive star-forming regions. *The Astrophysical Journal Supplement Series* 194, 16.
- ViaLactea (n.d.). URL: <http://www.astro.cardiff.ac.uk/research/ViaLactea/>.
- Wadsley et al. (2004). Gasoline: A flexible, parallel implementation of TreeSPH. *New Astronomy* 9, 137–158.
- Walt, S. van der et al. (2014). scikit-image: image processing in Python. *PeerJ* 2, e453.
- Wang, J. et al. (2011). A Chandra ACIS Study of the Young Star Cluster Trumpler 15 in Carina and Correlation with Near-infrared Sources. *The Astrophysical Journal Supplement* 194, 11.
- Ward, R. (2015). The Structure and Evolution of Unbound Star-forming Molecular Clouds. PhD thesis. McMaster University.
- Yonekura, Y., Asayama, S., Kimura, K., Ogawa, H., Kanai, Y., Yamaguchi, N., Barnes, P. J., and Fukui, Y. (2005). High-Mass Cloud Cores in the η Carinae Giant Molecular Cloud. *The Astrophysical Journal* 634, 476–494.