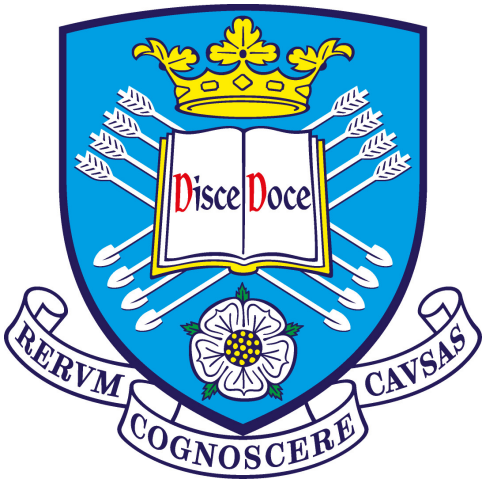
Quantifying Star Formation

George Arthur Blaylock-Squibbs

Department of Physics & Astronomy
The University of Sheffield



A dissertation submitted in candidature for the degree of Doctor of Philosophy at the University of Sheffield

September 2023

“

Alcohol is God’s apology for making us self-aware.

”

— James May

Contents

List of Figures

List of Tables

Declaration

I declare that, unless otherwise stated, the work presented in this thesis is my own. No part of this thesis has been accepted or is currently being submitted for any other qualification at the University of Sheffield or elsewhere.

The work presented in Chapter 3 can be found in Blaylock-Squibbs et al. (2022) and the work in Chapter 4 can be found in Blaylock-Squibbs & Parker (2023).

Acknowledgments

First off I would like to thank Richard for being an amazing supervisor whose guidance and generosity (both in time and pints) have made the last three years a truly memorable and immensely rewarding time. It is abundantly clear how much he cares, not only about the science, but also the well-being of those he works with. It's been a genuine pleasure working with him.

Simon, thanks for the stories (and even more pints!) and early feedback on some of my work. I'd like to thank the other students in E18, Summer, Luke, Adam, Dustin, Thaer and James. It was an *"experience"* sharing an office with you all, one that will be well and truly missed.

Thanks to past PhD students, Emma and Christina. Thanks to Rebecca and Alex for getting some of us into climbing early on, it was never something I thought I would enjoy as much as I do.

Special thanks to the postdocs, Mark, Martin and Joe. Martin your Git tutorial has saved me a lot of time and made writing this far less stressful than it would have been otherwise. Mark, for the brisket, and Joe, for the wonderful photos taken during Protostars and Planets VII.

Special thanks to Vik and Paul for the observing trip to La Palma. It's a highlight of my time at Sheffield and a memory I will cherish forever. Also, thanks to Paul for keeping Daedalus running and general computer wrangling/wizardry.

To the new PhD students starting this year, you've got a truly lovely bunch of people to help you if you need it. Best of luck and enjoy it while you can, it goes by far too quickly.

My first year was fully online due to Covid and I only got through it thanks to the support of my friends and family. Thanks to Mum, Dad, Granny and Abs for your love and support that I can never hope to repay. Thanks to Noona and Grandpa. They knew I started, and I hope that somehow they know I finished.

Summary

Observations of star-forming regions only allow us to view a single snapshot in time of their evolution. By combining observations of different star-formation regions we can build up a statistical picture of how stars form in groups, how these groups dynamically evolve and eventually disperse into the field. But to fully understand how these regions formed in the first place we need to make use of numerical simulations which allow us to follow the evolution of a single star-forming region. Then we can make quantitative comparisons using a suite of methods that can be used to characterise star-forming regions (i.e. the spatial distribution of stars, or if a region is mass segregated and to what degree). By using these methods we can make inferences on not only the initial conditions of the observed star-forming regions (degree of substructure, local surface density and virial state), but also the initial conditions of planet formation. As these methods are used to infer the physics of star and planet formation they must be extensively tested on synthetic data to ensure that the methods are robust. The work herein is my investigations into the robustness of these methods on simulated data, and if these new methods can be used to reliably make inferences on the properties and physics of star-forming regions.

I compare two new methods to more established and widely used methods to test their reliability. I show that INDICATE, a new clustering metric which has been used to investigate the star formation history of regions, can accurately identify areas of clustering in synthetic regions and gives results in agreement with other methods. However, INDICATE cannot be used to distinguish between star-forming regions with different morphologies, but it can be used to identify the presence of mass segregation.

I investigate the evolution of phase space densities (quantified using the Mahalanobis density) of simulated star-forming regions using N-body simulations. This work was performed to better understand how the 6D phase space density evolves with time, and if its evolution depends on the initial conditions of the simulations. The method has been used to infer the likely star-formation conditions of exoplanet host stars with hot Jupiters being very dense. I find that using the 6D (position-velocity) phase space density of star-forming regions does not allow their initial conditions to be reliably discerned.

I finish by investigating this possible link between the 6D phase space density of exoplanet host stars, quantified using the Mahalanobis density, and their initial formation conditions. I find that the phase space density of host stars and non-host stars does not evolve significantly differently for simulations with different initial conditions. I compare my results to previous works and find results in agreement with other works,

that the Mahalanobis density is not detecting traces of the initial conditions but is instead measuring the kinematics of the host stars.

1

Introduction

1.1 The Interstellar Medium

One of the earliest questions the first astronomers asked would likely have been "What are those points of light in the night sky?" We may never know what answers they came up with, but we do know the curiosity that undoubtedly led them to try in the first place. This same curiosity has sustained the field of astronomy over the millennia and has led to profound revelations of not only the stars themselves but of the origins of the elements that are vital to our very existence. Our investigations have given us answers to the questions of the very first astronomers, such as what stars are, but we are far from finished. The field of star formation still has many fundamental questions to answer such as determining if star formation is a truly universal process? There is also the question of how star formation begins in the first place; is it initiated via overdensities in the interstellar medium caused by turbulence or can supernovae trigger star formation?

The answers to these questions likely begin in the interstellar medium, which fills the space between stars in the Galaxy and primarily consists of gas. By mass the gas composition of the ISM is 70.4% hydrogen, 28.1% is helium and 1.5% is heavier elements (Ferrière, 2001).

The ISM has different phases where the density, temperature and different states of hydrogen are used to differentiate between them. The ISM has been split into six phases; molecular clouds, the cold neutral medium, regions of ionised hydrogen (HII), the warm neutral medium, the warm ionised medium and the very local interstellar medium within 100 pc of the Sun (Redfield & Falcon, 2008). Molecular clouds are the densest phase of the ISM with number densities between 200 cm^{-3} up to 10^5 cm^{-3}. The cold neutral medium is made up from neutral atomic hydrogen (HI), with temperatures of ~ 100 K and number densities of $\sim 10 - 100$ cm^{-3}. The warm neutral medium and warm ionised medium share similar temperatures (~ 8000 K) and densities ($\sim 0.1 - 0.5$ cm^{-3}). There is also the very local medium, with a temperature of ~ 6700 K and density of 0.11 cm^{-3}.

The hot interstellar medium has a temperature of 10^6 K and density of 5×10^{-3} cm^{-3} (Haverkorn & Spangler, 2013).

The different states of hydrogen in the ISM are important as it has implications for how the different ISM phases are observed. For example, H_2 can only exist in areas of sufficient density such that the H_2 is shielded from UV radiation that would otherwise break it back down into atomic hydrogen (HI). But as I will now discuss, observing H_2 is not practical. Instead, to understand the distribution of H_2, we have to use something we can observe, like the 21 cm emission line of HI (emission due to an electron's spin flipping, going from being in parallel with the proton's spin to antiparallel) (Ewen & Purcell, 1951).

We cannot directly observe molecular hydrogen (H_2) as it is relatively cold in the ISM (due to self-shielding) at $\sim$ 10s K; at this temperature H_2 cannot be observed using molecular emission lines in the optical or radio (the gas would need to be several 100s K for emission lines to be present).

The energy needed for a H_2 molecule to become excited is given by

$$E(J) = \frac{\hbar^2 J\,(J+1)}{2\mathcal{I}}, \tag{1.1}$$

where $\hbar$ is the reduced Planck's constant ($\hbar = h/2\pi$), J is the angular momentum quantum number, and $\mathcal{I}$ is the moment of inertia of the molecule. Because H_2 is a homonuclear diatomic molecule and therefore has no permanent electric dipole moment we instead have to use the quadrupole electric moment. The angular momentum transition rule for electric quadrupole moments is $\Delta J = 0, \pm 2$. The first excited rotational state of molecular hydrogen is calculated as follows,

$$\Delta E = E(J=2) - E(J=0) = \frac{\hbar^2}{2\mathcal{I}_{H_2}}, \tag{1.2}$$

the moment of inertia of H_2 is 4.46×10^{-48} kg m^2, giving a transition energy of 7.20×10^{-21} J. Using $T = E/k_b$, where k_b is the Boltzmann constant, a temperature of ~ 520 K is found. The typical temperature in the densest phase of the ISM (giant molecular clouds) is ~ 10 K, well below the calculated temperature required for emission from H_2 (Wilson et al., 1997).

Due to the lack of H_2 emission in the cold ISM we need to use other molecules, such as CO, as tracers. CO has an electric dipole moment with transition rule of $\Delta J = \pm 1$ and CO has moment of inertia, $\mathcal{I}_{CO} = 1.46 \times 10^{-46}$ kg m^2. Using Equation 1.1 but using the dipole transition rule ($\Delta E = E(J = 1) - E(J = 0)$) gives an energy of $E = 7.63 \times 10^{-23}$ J. Using $T = E/k_b$ the temperature required for this transition is ~ 5 K, below the typical temperatures within GMCs, meaning CO emissions are observable and can be used to trace the distribution of H_2. In addition, the continuum emission from dust particles are also used to map out the distribution of H_2 (Boulanger, 1999). This involves assuming that a certain amount of CO or dust corresponds to a certain amount of H_2 (Bohlin et al., 1978; Solomon et al., 1987). For example, we assume that the gas to dust ratio (gas:dust) is 100. The conversion from a measured CO intensity to N_{H2} is expressed as

$$N_{H_2} = X_{CO}\, I_{CO}, \tag{1.3}$$

where X_{CO} is the conversion factor and I_{CO} is the intensity of CO emissions (Hunt et al., 2023).

For stars to form, giant molecular clouds (GMCs) must exist for a suitable amount of time of around 10 Myr (Mouschovias et al., 2006). Estimating how long GMCs can exist for in the ISM can be determined using the following expression,

$$t = \frac{R_c}{\sigma_\nu}, \tag{1.4}$$

where R_c is the radius of the GMC and σ_ν is the total velocity dispersion. Using a radius of 100 pc and a velocity dispersion of 10 kms^{-1} we get a lifetime of $\sim$ 10 Myr, these values are taken from the high end of the size-velocity dispersion relation observed in Larson (1981). GMCs are of particular importance in our understanding of star formation as it is only within these regions that it can get cold enough for molecular hydrogen clouds to collapse, due to self-shielding.

Observations of molecular clouds in the ISM have also revealed them to be highly turbulent (Larson, 1981). The turbulence is detected using observed line broadening in emission spectra (Ho et al., 1977). The observed broadening of emission lines cannot be explained by thermal motions alone, meaning a turbulent bulk motion of the gas is needed to produce the observed broadening (Mac Low & Klessen, 2004). Figure 1.1 is taken from Caselli & Myers (1995) and shows the relationship between the non-thermal line width against the size of a molecular cloud. It shows that in low mass molecular clouds line broadening is dominated by thermal motions, whereas for the massive cores the line broadening is dominated by a non-thermal component. It is this turbulent bulk motion which gives rise to over and under densities in the ISM, causing the formation of structures.

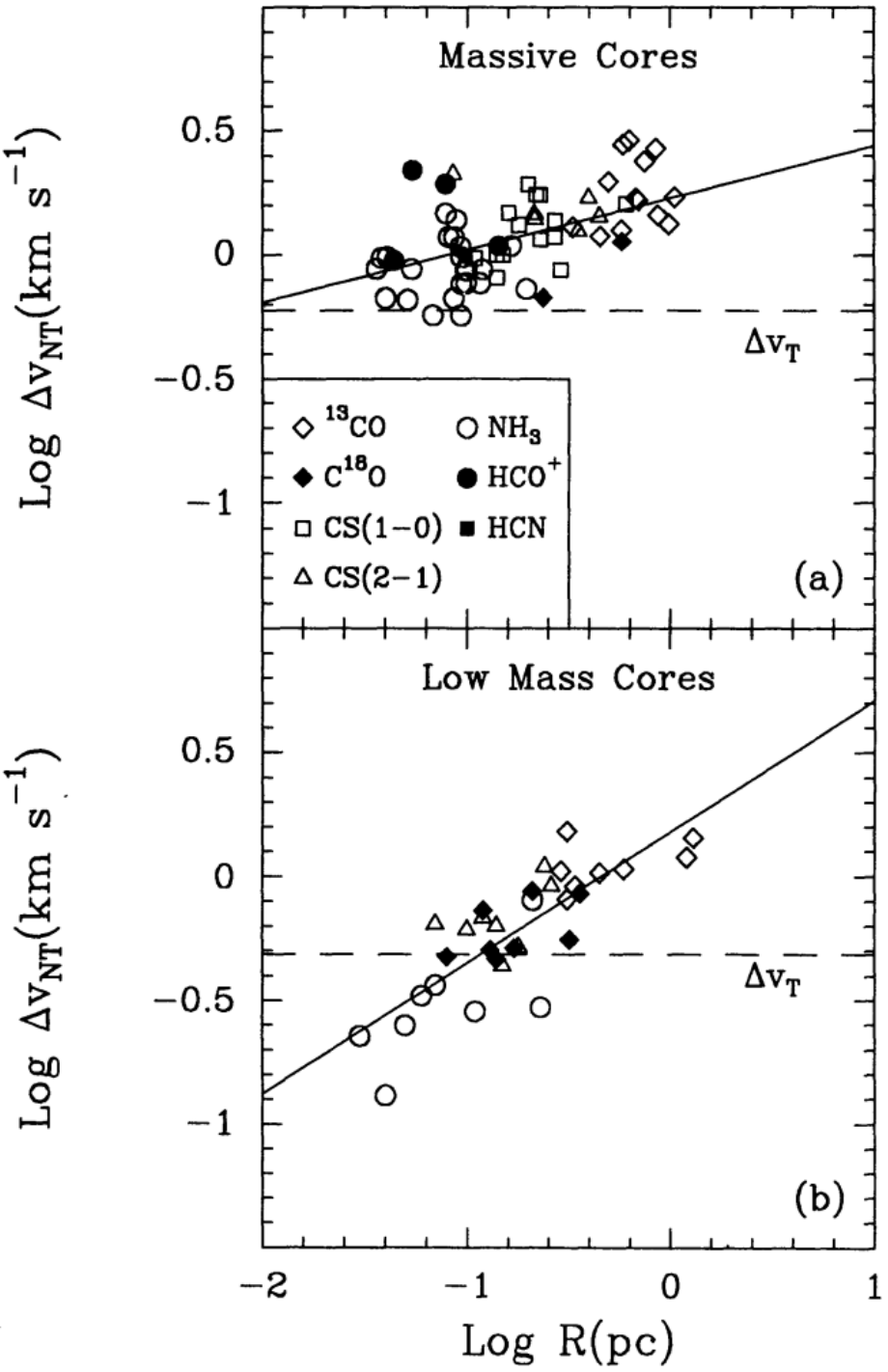


Figure 1.1: Figure from Caselli & Myers (1995). The plot shows the relation between the non-thermal line width and radii of high and low mass cores, shown in the top and bottom panels, respectively. The low mass cores line broadening has a significant contribution from thermal motion in the gas whereas in the massive cores the contribution from thermal motions is minimal compared to the non-thermal motions of the core.

1.2 Virial Theorem

To understand how the gas in the ISM can collapse and form structures we use the virial theorem. The virial theorem states that the time averaged kinetic energy is twice the time averaged negative potential energy of a bound system of particles (Clausius, 1870). What follows is a derivation of the virial theorem for a spherical gas cloud.

Let us assume our cloud of gas, which is isolated and in equilibrium, can be described by a spherical distribution of N point masses. Each of these points has a mass, m_i, and a position described by the vector, $\vec{r}_i$. The entire cloud's moment of inertia can be described as the sum of every individual point masse's moment of inertia, mathematically defined as

$$\mathcal{I} = \sum_{i}^{N} \left(m_i \ (\vec{r}_i \cdot \vec{r}_i) \right). \tag{1.5}$$

Since we are assuming the cloud is in equilibrium we know that the rate of change of the moment of inertia will be zero (i.e. $\dot{\mathcal{I}} = 0$). To calculate the time derivative of Equation 1.5 we make use of the product rule and obtain

$$\frac{d\mathcal{I}}{dt} = \sum_{i}^{N} \left(\cancelto{0}{\frac{d\, m_i}{dt} (\vec{r}_i \cdot \vec{r}_i)} + m_i \left(\frac{d\, \vec{r}_i}{dt} \cdot \vec{r}_i \right) + m_i \left(\vec{r}_i \cdot \frac{d\, \vec{r}_i}{dt} \right) \right), \tag{1.6}$$

and because $\dot{\vec{r}}_i = \vec{v}_i$ we get,

$$\dot{\mathcal{I}} = \sum_{i}^{N} \left(m_i \ (\vec{v}_i \cdot \vec{r}_i + \vec{r}_i \cdot \vec{v}_i) \right). \tag{1.7}$$

Taking a factor of two out of the summation we have

$$\dot{\mathcal{I}} = 2 \sum_{i}^{N} \left(m_i \ (\vec{v}_i \cdot \vec{r}_i) \right). \tag{1.8}$$

As the cloud is in equilibrium it follows that the second time derivative is also zero (i.e.

$\ddot{\mathcal{I}} = 0$). Using the product rule again we calculate the time derivative of Equation 1.8,

$$\frac{d\,\dot{\mathcal{I}}}{dt} = 2\sum_{i}^{N}\left(\cancelto{0}{\frac{d\,m_i}{dt}(\vec{v}_i\cdot\vec{r}_i)} + m_i\left(\frac{d\,\vec{v}_i}{dt}\cdot\vec{r}_i\right) + m_i\left(\vec{v}_i\cdot\frac{d\,\vec{r}_i}{dt}\right)\right). \tag{1.9}$$

Completing the calculation we get

$$\ddot{\mathcal{I}} = 2\sum_{i}^{N}\left(m_i\left(\dot{\vec{v}}_i\cdot\vec{r}_i\right)\right) + 2\sum_{i}^{N}\left(m_i\left(\vec{v}_i\cdot\vec{v}_i\right)\right). \tag{1.10}$$

We can simplify the above equation by letting

$$\ddot{\mathcal{I}}_1 = 2\sum_{i}^{N}\left(m_i\left(\dot{\vec{v}}_i\cdot\vec{r}_i\right)\right), \tag{1.11}$$

and since $\dot{\vec{v}}_i = \vec{a}_i$ we can use Newton's second law to rewrite Equation 1.11 as

$$\ddot{\mathcal{I}}_1 = 2\sum_{i}^{N}\left(\vec{F}_i\cdot\vec{r}_i\right). \tag{1.12}$$

Using Newton's third law we know that the force on particle i due to particle j is the same as the force on particle j due to i, but in the opposite direction. Therefore

$$\vec{F}_{ij} = -\vec{F}_{ji}. \tag{1.13}$$

We only care about the net force on i from all the other particles, so we can rewrite equation 1.12 as

$$\ddot{\mathcal{I}}_1 = 2\sum_{i}^{N}\sum_{j\neq i}^{N}\left(\vec{F}_{ij}\cdot\vec{r}_i\right). \tag{1.14}$$

Since the order in which we add forces does not matter we can express the above equation as,

$$\ddot{\mathcal{I}}_1 = \sum_{i}^{N}\sum_{j\neq i}^{N}\left(\vec{F}_{ij}\cdot\vec{r}_i + \vec{F}_{ji}\cdot\vec{r}_j\right). \tag{1.15}$$

We can write it this way if we remember Newton's third law, which we can also use again to simplify further,

$$\ddot{\mathcal{I}}_1 = \sum_i^N \sum_{j \neq i}^N \left(\vec{F}_{ij} \cdot (\vec{r}_i - \vec{r}_j) \right). \tag{1.16}$$

The gravitational force between two objects of m_i and m_j is

$$\vec{F}_{ij} = -\frac{G m_i m_j}{|\vec{r}_i - \vec{r}_j|^2} \frac{(\vec{r}_i - \vec{r}_j)}{|\vec{r}_i - \vec{r}_j|}, \tag{1.17}$$

where G is the gravitational constant. The fraction on the right-hand side is just the unit vector, $\hat{r}$. We substitute this back into Equation 1.16 and obtain

$$\ddot{\mathcal{I}}_1 = \sum_i^N \sum_{j \neq i}^N \left(-\frac{G\, m_i\, m_j}{|\vec{r}_i - \vec{r}_j|} \right), \tag{1.18}$$

which is just the gravitational potential energy of the cloud. Finally, we have the expression

$$\ddot{\mathcal{I}} = 2\, \Omega_{\mathrm{G}}. \tag{1.19}$$

The factor of two comes from the fact that our calculation double counts the force between pairs of particles (recall Newton's third law). We can define the second term of Equation 1.6 as the kinetic energy of all the particles. Firstly we let

$$\ddot{\mathcal{I}}_2 = 2 \sum_i^N \left(m_i\, \vec{v}_i^{\,2} \right), \tag{1.20}$$

which we then rewrite as

$$\ddot{\mathcal{I}}_2 = 4\mathcal{K}. \tag{1.21}$$

Substituting our expressions back into Equation 1.6 and dividing through by 2 we get,

$$\frac{\ddot{\mathcal{I}}}{2} = 2\mathcal{K} + \Omega_{\mathrm{G}}, \tag{1.22}$$

which is the virial theorem, with $\ddot{\mathcal{I}} = 0$ if the cloud is in equilibrium.

1.3 Jeans Criterion

Using the virial theorem we can calculate the mass and size a GMC, or dense core within a cloud, needs to begin collapsing under its own gravity. Using the virial theorem we know that if our molecular gas cloud is in equilibrium the following is true

$$2\mathcal{K} + \Omega_{\mathrm{G}} = 0, \tag{1.23}$$

and therefore, if the cloud is out of equilibrium, and collapsing, its gravitational potential energy is greater than its kinetic energy,

$$-\Omega_{\mathrm{G}} > 2\mathcal{K}, \tag{1.24}$$

which is the condition for collapse. Substituting in the equation of kinetic and potential energy for a gas into Equation 1.24 we obtain

$$\frac{GM^2}{5R} > Nk_{\mathrm{B}}T. \tag{1.25}$$

Here, R is the radius of the cloud, N is the number of molecules in the cloud, k_{B} is the Boltzmann constant and T is the temperature of the cloud. We can replace N with the mass of the cloud using

$$N = \frac{M}{m}, \tag{1.26}$$

where M is the cloud mass and m is the mean mass of the individual elements making it up (predominantly atomic or molecular hydrogen). Substituting for N we obtain

$$\frac{GM}{5R} > \frac{k_{\mathrm{B}}T}{m}, \tag{1.27}$$

assuming the cloud has constant density we can substitute for the remaining M,

$$\frac{4G\pi R^2\rho}{15} > \frac{k_{\mathrm{B}}T}{m}. \tag{1.28}$$

To find the Jeans length, R_{J}, we solve this inequality for R,

$$R_{\mathrm{J}}^2 = \frac{15k_{\mathrm{B}}T}{4\,\mu\,m_{\mathrm{H}}\,G\,\pi\,\rho}, \tag{1.29}$$

where k_{B} is the Boltzmann constant, T is the temperature of the gas, μ is the number of particles making up individual elements of the gas (i.e. for an atomic hydrogen cloud $\mu = 1$ and for molecular hydrogen clouds $\mu = 2$), m_{H} is the mass of a hydrogen atom, G is the gravitational constant and ρ is the mass volume density of the cloud.

To find the mass needed to overcome the thermal support we start again with Equation 1.27. However, this time we will substitute for R assuming a constant density of ρ. The density of a sphere is

$$\rho = \frac{3M}{4\pi R^3}. \tag{1.30}$$

We then solve for R:

$$R = \left(\frac{3M}{4\pi\rho}\right)^{\frac{1}{3}}. \tag{1.31}$$

We then substitute this into Equation 1.27:

$$\frac{GM^{\frac{2}{3}}}{5}\left(\frac{4\pi\rho}{3}\right)^{\frac{1}{3}} > \frac{k_{\mathrm{B}}T}{m}. \tag{1.32}$$

Solving for M gives us the Jeans mass,

$$M_{\mathrm{J}} = \left(\frac{5k_{\mathrm{B}}T}{G\mu m_{\mathrm{H}}}\right)^{\frac{3}{2}}\left(\frac{3}{4\pi\rho}\right)^{\frac{1}{2}}, \tag{1.33}$$

with all terms being the same as in Equation 1.29.

1.4 Structure in the turbulent ISM

In this section I will discuss a range of structures that have been observed in the ISM in descending order of their physical scales and describe the physical characteristics inferred from observations and discuss the theoretical work undertaken to determine their likely origins.

1.4.1 Bubbles

HI bubbles are observed in the cold ISM of our Galaxy using the 21 cm emission line of neutral atomic hydrogen. HI bubbles are caused by supernovae and initially the hydrogen is ionised (becoming HII) but as the bubble expands some HII recombines, cooling down the bubble (Tomisaka et al., 1981). Due to the driving force from supernovae, HI bubbles are observed to be expanding with velocities of between 5-20 $\mathrm{km\,s^{-1}}$ with typical radii of $\sim$ 100 pc (Tomisaka et al., 1981; Spitzer, 1998). They have been observed numerous times in the Galaxy (see Ehlerová & Palouš (2013)). They are also observed in other galaxies like NGC 628, shown in Figure 1.2, which highlights the prevalence of bubbles throughout it. The edges of these bubbles overlap and intersect with the edges of other bubbles and at these intersections Bracco et al. (2020) observed GMCs containing filamentary structures. The molecular gas appears at the edges and intersections of bubbles because as the bubbles expand they can sweep up material and compress it in a shock layer as shown in simulations from Inoue & Inutsuka (2016).

1.4.2 Giant Molecular Clouds

Giant molecular clouds (GMCs) are massive structures with masses $\sim 10^4\ \mathrm{M_\odot} < \mathrm{M_{GMC}} < \sim 10^6\ \mathrm{M_\odot}$ and can extend for $\sim$ 100s pc (Murray, 2011). Stars form within the densest regions of GMCs, sometimes called clumps or cores, located along the lengths of filaments.

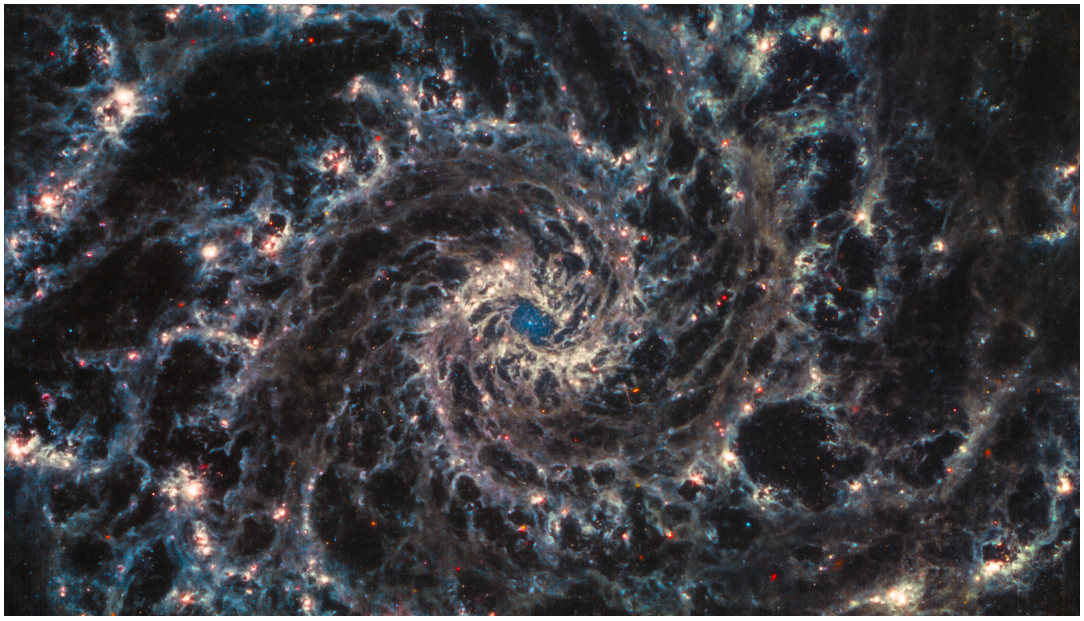


Figure 1.2: JWST MIRI observation of galaxy NGC628 showing large HI bubbles throughout the galaxy.

1.4.3 Filaments

Filaments are elongated strands of gas, usually with surface densities many times greater than the GMCs they are embedded in. Arzoumanian et al. (2019) find peak column densities of 9×10^{21} cm^{-2} with line masses of $\sim$ 11 M$_\odot$ pc^{-1} for filaments embedded in molecular clouds. Filaments in the Galaxy cover a wide range of physical scales, from $0.01 - 500$ pc, and are found in both atomic and molecular phases of the ISM. This scale range comes from the meta-analysis of Hacar et al. (2022) in which they looked at available filament survey data. The justification for this range is that at scales below 0.01 pc we would enter the regime of collapsing protostellar cores, and above 500 pc would involve extracting large scale filaments from the Galactic spiral arms. In this section I will primarily be focussing on the filaments inside giant molecular clouds as it is along these that prestellar cores are observed (Pilbratt et al., 2010; Könyves et al., 2015).

The origins of filaments may be due to the intersection of shock fronts, such as when

two expanding HI bubbles intersect (Tomisaka & Ikeuchi, 1983). It has also been shown in simulations that filaments can form from thermal instabilities in molecular clouds (MCs) with no intersecting sheets needed (Wareing et al., 2019).

Further complicating our view of filament formation is how the different environments affect the formation process. For example, the orientation of magnetic fields within a compressed layer and whether there is gas flowing onto the sheet during the intersection will affect the mode of formation (Hacar et al., 2022).

The challenge of linking observations and theory for filaments is further compounded as there is a lack of large scale simulations ($\sim$ 1000 pc) of cloud and filament formation that possess high enough spatial resolutions ($\lesssim$ 1 pc) (Hacar et al., 2022). Turbulent gas motions are observed within the filaments (like in the larger scale giant molecular clouds they are found in), with velocity gradients showing gas being accreted onto filaments (Clarke et al., 2018). This lends credence to the theory of hierarchical star cluster formation (see §1.7.3 of this chapter).

Determining if a filament is collapsing radially or not is important, as without this collapse prestellar cores will not be able to form. Prestellar cores are the immediate precursors to stars and are characterised observationally by looking for dense clumps of molecular gas that are collapsing with no IR sources within them. The virial state of a filament can be determined using the observed radial velocity dispersion. The virial line mass of a filament is

$$m_{\rm vir} = \frac{2\sigma_{\rm tot}^2}{G}, \tag{1.34}$$

where $\sigma_{\rm tot}$ is the total velocity dispersion (thermal and non-thermal components) and G is the gravitational constant (Hacar et al., 2022). The filament will collapse if its line mass exceeds the virial line mass, $m_{\rm vir}$. The line mass of a filament is defined as

$$m = \frac{M}{L}, \tag{1.35}$$

where M is the mass of the filament and L is the length of the filament, both of these parameters are determined observationally using column densities (Hacar et al., 2022).

Filaments undergo collapse and fragmentation when $m > m_{\rm vir}$, forming cores along their lengths, with the separations between the cores possibly signalling different fragmentation modes (Takahashi et al., 2013; Zhang et al., 2020).

There is some discussion about a possible fundamental core separation seen in the theoretical work of Inutsuka & Miyama (1992) where they show that an infinite isothermal uniform filament undergoing gravitational collapse will form cores with a separation given by

$$d_{\rm cs} \approx 4W_{\rm fil}, \tag{1.36}$$

where $W_{\rm fil}$ is the width of the filament in pc. Observations of regularly spaced cores in filaments has been observed in Henshaw et al. (2016). However, when comparing the cores to theoretical predictions they find a significant difference. This discrepancy is possibly resolvable with improvements in theoretical models, moving them away from idealised cylinders of gas (as formalised in Ostriker (1964)), to more physical models which include the influence of external pressure and magnetic support, which has been motivated by observations (Chapman et al., 2011).

1.4.4 Cores

Cores are overdensities of molecular gas observed along filaments within GMCs (Pilbratt et al., 2010). Cores are made out of the same material of the filaments they reside in, dense molecular hydrogen and dust, with typical radii of $\sim$ 0.1 pc and densities of $\sim 3 \times 10^{-17}$ kg m^{-3} (Onishi et al., 2002; Shirley, 2015; Pineda et al., 2022).

The cores are supported against gravitational collapse by their internal thermal pressure, which for stars to form needs to be overcome. To determine whether a cloud of gas or a core within it is collapsing we use the Jeans criterion, derived from the virial

theorem in §1.3 of this chapter.

The Jeans mass has some assumptions that are not representative of reality, such as the core being perfectly spherical, isothermal, of constant density and in isolation. Observations of molecular cores have shown this to be far from true, with cores having both prolate and oblate morphologies (Jones et al., 2001). Additionally, magnetic fields and turbulence within cores will act against a collapse, effectively raising the required mass needed (Pineda et al., 2022).

The Bonnor-Ebert mass is similar to the Jeans mass (same assumptions) but takes into account that a core will be embedded within another medium, which will exert a pressure on the core (Ebert, 1955; Bonnor, 1956). The Bonnor-Ebert mass for an isothermal sphere is

$$M_{\rm BE} = 1.182\,\frac{a^4}{\sqrt{G^3\,p_{\rm ext}}}, \tag{1.37}$$

where $p_{\rm ext}$ is the external pressure and a is

$$a = \sqrt{\frac{k_{\rm b}\,T_{\rm g}}{\mu\,m_{\rm H}}}, \tag{1.38}$$

which is the speed of sound in the gas, with μ being the mean molecular weight of the gas (2 in this case as the sphere is assumed to be made of molecular hydrogen) and $m_{\rm H}$ is the mass of atomic hydrogen (Galli et al., 2002).

The Bonnor-Ebert mass is lower than the Jeans mass due to the external pressure pushing material closer together increasing the density.

Determining the stability of a core observationally against gravitational collapse can be done by using virial analysis. We have already seen the virial theorem in §1.2 of this chapter for kinetic energy and gravitational energy. Below is a version showing the virial theorem but taking into account other sources of energy that will act against gravity,

$$\frac{\ddot{I}}{2} = 2\mathcal{K} + \Omega_{\rm G} + \Omega_{\rm M} + \Omega_{\rm P,}, \tag{1.39}$$

where $\mathcal{K}$ is the kinetic energy, $\Omega_{\rm G}$ is the gravitational potential energy, $\Omega_{\rm M}$ is the magnetic energy and $\Omega_{\rm P}$ is the external pressure (Pineda et al., 2022). In the literature it is common to perform a virial analysis using only the potential and kinetic energy, as calculating the magnetic field in regions observationally is challenging (Pattle et al., 2023).

Observationally, virial analysis is done using the one-dimensional velocity dispersion (i.e. radial velocity). We then compare the observed velocity dispersion to the predicted velocity dispersion for a cloud of mass M to be in virial equilibrium. The virial theorem states that the total energy of a system is the sum of twice the kinetic energy and the potential energy:

$$2\,\mathcal{K} + \Omega_{\rm G} = E_{\rm T}. \tag{1.40}$$

A system is said to be in virial equilibrium when $E_{\rm T} = 0$, therefore if $-\Omega_{\rm G} > \mathcal{K}$ the core is collapsing and conversely if $-\Omega_{\rm G} < \mathcal{K}$ the core is expanding (virial analysis is also used to determine if a group of stars is gravitationally bound or not). We can estimate the kinetic energy of a core using the following expression:

$$\mathcal{K} = \frac{1}{2} M \sigma_v^2, \tag{1.41}$$

where M is the mass of the core and σ_v is the one dimensional velocity dispersion needed for the core to be in virial equilibrium. The gravitational potential energy is defined as

$$\Omega_{\rm G} = -\frac{GM^2}{R}, \tag{1.42}$$

where G and M are the gravitational constant and mass of the core, respectively, and R is the radius of the core. Assuming the core is in virial equilibrium ($E_{\rm T} = 0$) and substituting Equations 1.41 and 1.42 into 1.40 gives,

$$M\sigma_v^2 - \frac{GM^2}{R} = 0. \tag{1.43}$$

Rearranging for the velocity dispersion of a core of mass, M, and radius, R, gives the following,

$$\sigma_v = \sqrt{\frac{GM}{R}}. \tag{1.44}$$

Therefore, if $\sigma_{\rm ob} < \sigma_v$ the core is not in virial equilibrium and is collapsing (the core is subvirial) and if $\sigma_{\rm ob} > \sigma_v$ the core is expanding (supervirial).

1.4.5 Thermodynamics of Core Collapse

A cloud of molecular gas can only collapse if it can cool down, reducing its thermal support. The collapse is only possible due to the interactions of H_2 and dust. If we assume that the gas and dust are in thermal equilibrium (same temperature) then the dust will preferentially gain kinetic energy from direct collisions with the gas, the dust will then re-emit the gained kinetic energy at longer wavelengths (infrared) which can pass through the dense core. This continuum emission removes energy from the core, resulting in a drop in the temperature which in turn reduces the thermal support causing the core to contract. However, this contraction causes the core to heat up again to around the same temperature it started with ($\sim$ 10 K). The core will continue to isothermally collapse until the core's density reaches a critical value of 10^{-13} g cm^{-3} at which point the core becomes opaque to its own radiation, stopping energy from being lost from the continuum emission of the dust and gas, halting any further fragmentation (Low & Lynden-Bell, 1976). Calculating the minimum Jeans mass for the critical density gives a mass of $\sim 0.004\,M_\odot$ ($\sim 4 M_{\rm Jupiter}$). A schematic of hierarchical and simultaneous fragmentation of a protostellar core is shown in Figure 1.3 from Guszejnov & Hopkins (2015), and shows how within protostellar cores there can be multiple over densities all collapsing simultaneously. In this particular case the protostellar core produces two protostars, represented by the red circles in the right most image of the figure.

If the radius of a cloud is larger than the Jeans length then gravity will overpower

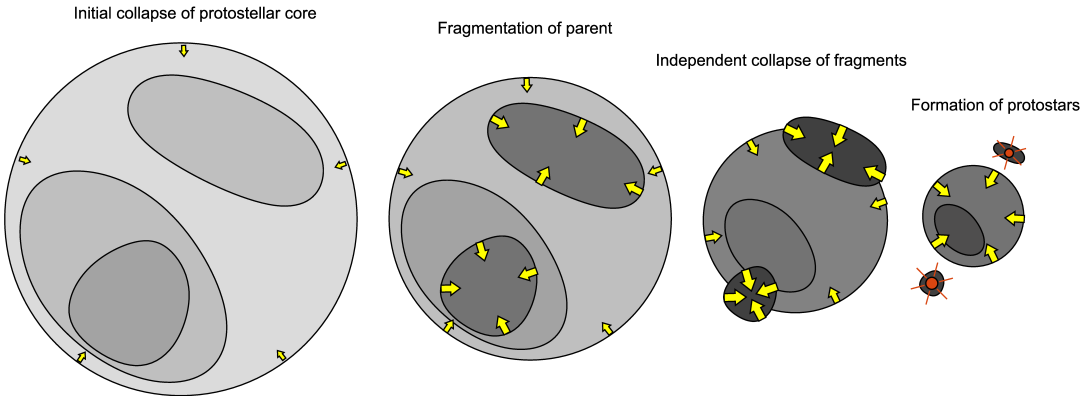


Figure 1.3: Figure from Guszejnov & Hopkins (2015) showing how a protostellar core collapses and fragments, producing multiple protostars. From left to right we see time advance and two main fragments forming. Yellow arrows represent that a fragment of the core is collapsing under its own self gravity. The darker the shade of the clouds the denser they are. The circles with lines radiating from them in the right most plot show the protostars themselves.

the cloud's thermal support and begins to collapse. Using a typical value for a dense core of $\rho = 3 \times 10^{-17}$ kg m^{-3}, temperature of 10 K with $\mu = 2$ as the core is made of H_2 we obtain a length of ~ 0.2 pc.

How the smallest cores fragment may have implications for the production of binary systems. For example in Figure 1.3 the densest regions that contain the protostars may undergo secondary fragmentation, producing very short period binaries with $a \leq 1$ AU (Bonnell & Bate, 1994). Working out the minimum fragmentation length using the Jeans length we find a length of ~ 5 AU, which if there is significant orbital decay the two protostars can move to within < 1 AU of each other (Korntreff et al., 2012).

1.5 Star Formation

In the previous sections I have described the structure present in the ISM and how larger structures feed mass down into increasingly smaller physical scales and how they influence these smaller scale structures. This section now looks at the final product at the lowest end of this physical scale, stars. Before the dense prestellar cores become

hydrogen burning stars they first must pass through the protostellar phase.

1.5.1 Protostars

Before a star becomes a hydrogen burning main sequence star it must first gain enough mass to achieve the required temperature and pressures within its core. A star's mass could be said to be its most important single parameter as it will determine how long it will live and even the manner of its death. It is during the protostellar phase that a star will attain its final mass.

The evolution from collapsing core to main sequence star is broken down into several stages. We empirically classify these young objects into Class 0, Class I, Class II and Class III objects. While all these objects are classified as protostars in the literature it is common to see that the embedded objects (Class 0 and Class I) are referred to as protostars and the non-embedded objects (Class II and Class III) are referred to as pre-main sequence (PMS) stars.

Throughout these phases the protostellar core is contracting, and if we were to place a new protostar on the HR diagram we would find that it is located in the top right. The protostar has a low temperature, but due to its large size on the order of $\sim 100\mathrm{s}\,\mathrm{AU}$ (due to the extended envelope) it will have a large luminosity. Because the collapse is isothermal it will have nearly the same temperature and its radius will decrease, lowering the surface area which in turn lowers its luminosity. This manifests as a near vertical downward movement on the HR diagram called the Hayashi track (Hayashi, 1961).

Different mass protostars will have different Hayashi tracks, with the lower mass protostars ($< 0.5\,\mathrm{M}_\odot$) immediately joining the main-sequence at the end of it. For more massive protostars $> 0.5\,\mathrm{M}_\odot$ they will follow a second track called the Henyey track (Henyey et al., 1955). This starts at the end of the Hayashi track and is due to the internal structure of the protostar becoming radiative, where a subsequent increase in

temperature evaporates any remaining dust within the protostellar core. The collapse is now adiabatic, and the protostar will then increase in temperature and luminosity all the way to the main sequence (manifesting as a diagonal movement heading towards to the top-left of the HR diagram).

Due to their nature the youngest protostars are incredibly difficult to observe directly as they are embedded within a cloud of dense gas and dust. This dust is the primary cause of extinction, either through scattering of short wavelength light or the absorption and re-emission at longer (redder) wavelengths. We must therefore observe them in the submillimetre and far-IR, specifically the thermal continuum emission of the dust. Figure 1.4 is taken from André (1994) and shows the spectral energy distributions (SEDs) for different classes of objects, and also a schematic view of the corresponding protostellar systems.

A Class 0 object has the beginnings of dense protostellar core at its centre. These objects are embedded deep within the molecular cores and are detected as embedded IR sources with peaks at $\sim$ 100 μm (Barsony, 1994). If we assume black body emission we find that the peak corresponds to a black body temperature of $\leq$ 30 K. The temperature corresponding to the peak emission is found using Wien's law,

$$T = \frac{b}{\lambda_{\mathrm{Max}}}, \tag{1.45}$$

where $b = 2.898 \times 10^{-3}$ m K and T is the temperature.

Most Class 0 objects contain multiple IR sources; the dense core may subfragment below the opacity limit of fragmentation, producing multiple protostellar systems (Low & Lynden-Bell, 1976). A schematic for a typical Class 0 protostellar system is shown in Figure 1.4 along with its corresponding SED. As the end of the Class 0 stage there is now a hydrostatic core; this is the seed that will gain its mass from accretion and eventually become a main sequence star.

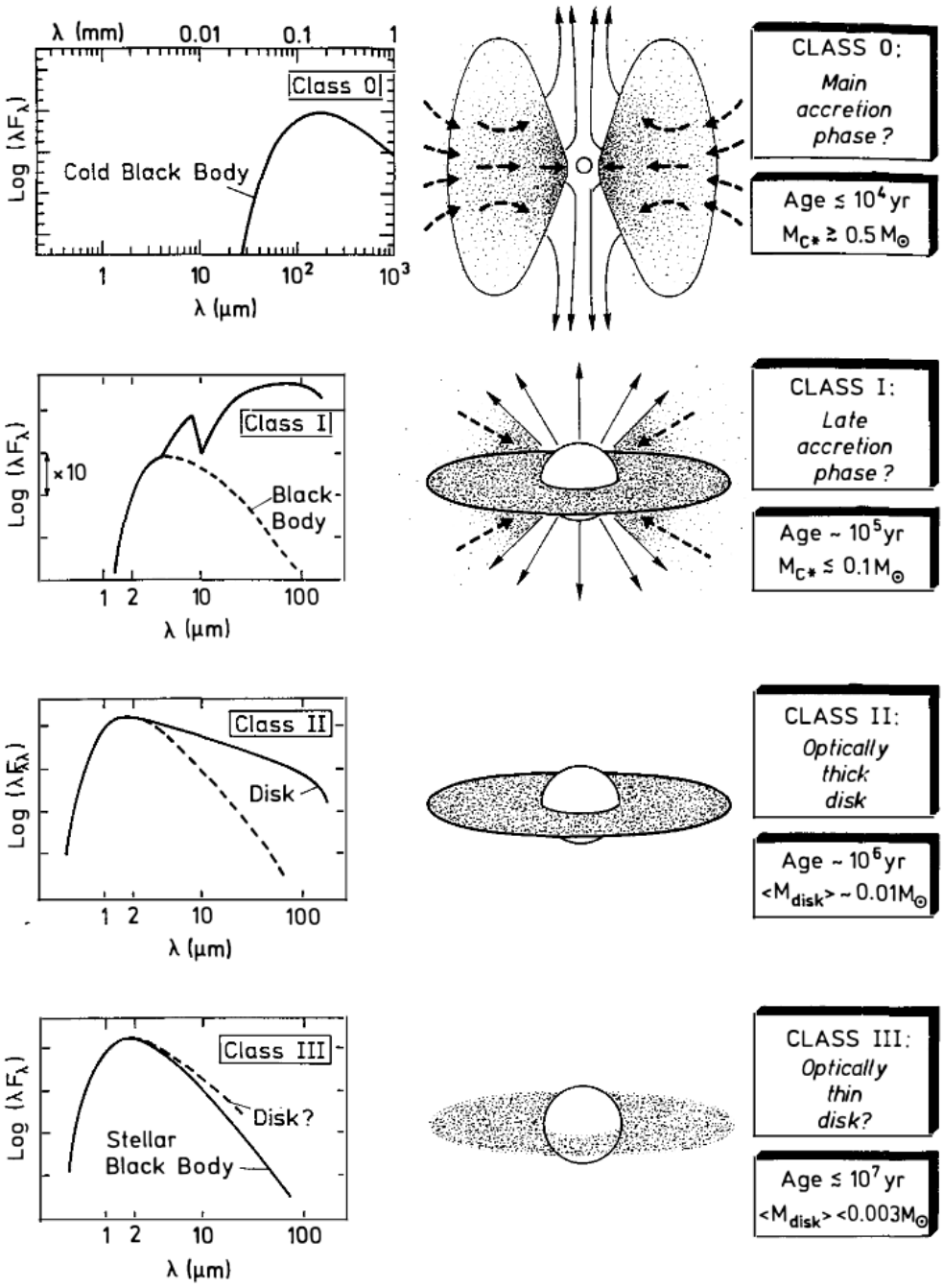


Figure 1.4: Observed SEDs for each of the protostellar classes. The time each protostar remains in each class increases from top to bottom. The left-hand column shows the SED expected for each class and the right-hand column shows a sketch of the corresponding protostellar system. Figure taken from André (1994).

A Class I object is still embedded within an envelope of gas (diameter of $\sim$ 1000 AU) and dust that it will accrete most of its final mass from. It is during the Class I phase that the protostar will accrete most of its mass. Compared Class 0 objects a Class I's SED has a strong IR excess due to the formation of a circumstellar disk. This disk likely forms when the object is still Class 0, and is made from material in the first hydrostatic core at the centre of the protostellar core (Machida et al., 2010). The circumstellar disk forms because of the conservation of angular momentum. The initial disk around the first hydrostatic core is thick and thermally supported and evolves into a thinner disk that becomes supported via the centrifugal force. The angular momentum is

$$\vec{\mathcal{L}} = m\,\vec{r} \times \vec{v} = \vec{\mathcal{I}}\omega, \tag{1.46}$$

where m is the mass of the protostar, $\vec{r}$ is the position and $\vec{v}$ is the angular velocity of the protostar. For completeness the other definition of the angular momentum is included in terms of the protostar's moment of inertia, $\vec{\mathcal{I}}$, and angular speed, ω.

As cores collapse they will begin to spin to conserve their angular momentum, and the protostar will inherit this spin. The surrounding material around the protostar will then become a circumstellar disk.

Material is accreted from the disk onto the protostellar core along magnetic field lines (see Figure 1.5 for a schematic of this accretion), depleting the surrounding envelope. Class I (and some Class 0) objects are observed to have bipolar molecular outflows, originating from opposite poles of the object (Eislöffel, 2000).

These outflows help carry away angular momentum from the infalling material. If this was not the case then as the material is added to the protostar it would have to spin faster and faster to conserve its angular momentum. This is one possible solution to the so-called angular momentum problem. The angular momentum problem is that the GMCs from which protostars form from have very high angular momentum (very high

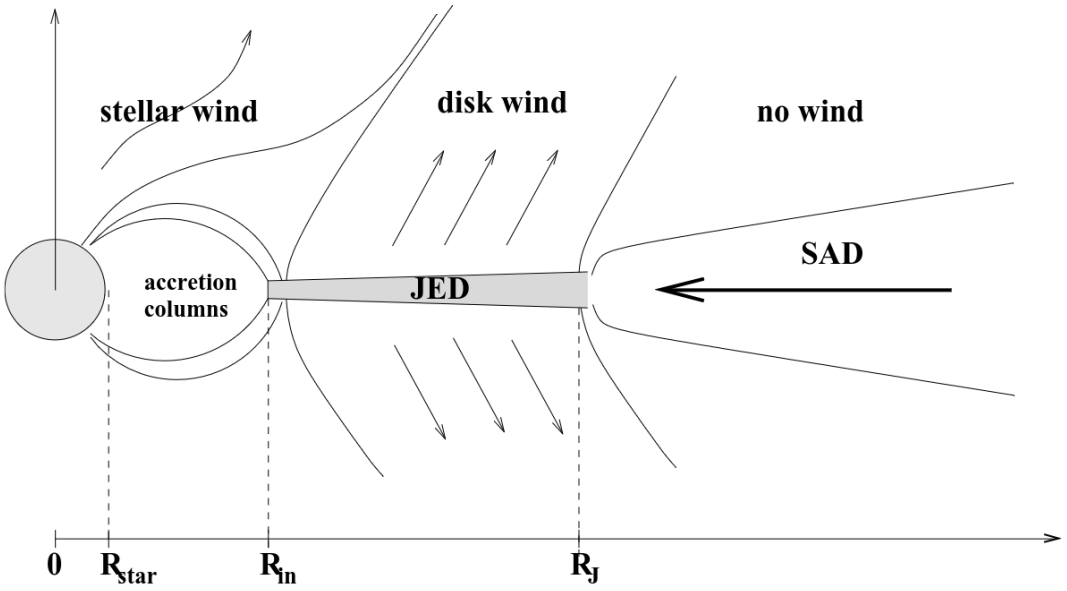


Figure 1.5: Schematic view of accretion onto a protostar from its circumstellar disk. From left to right the main features are the accretion columns, material flowing along magnetic field lines onto the star, the jet emitting disk (JED) and the standard accretion disk. Figure taken from Combet & Ferreira (2008).

mass, relatively low velocities and large size) and to ensure that the angular momentum is conserved as the GMC collapses the protostar would need to spin so fast that it would shred itself to pieces (Prentice & Ter Haar, 1971). Another way of removing the excess angular momentum may also be magnetic breaking within the protostellar cores, as the angular momentum can be transferred to the magnetic field via the twisting of the magnetic field (Mouschovias & Paleologou, 1979; Bodenheimer, 1995).

Observationally there is a strong link between outflows and presence of a circumstellar disk Ray & Ferreira (2021). Accretion is likely not constant and during episodes of greater accretion the outflows may increase in strength, giving rise to "gaps" in the molecular outflows and may allow the detection of protoplanets forming in the disk due to their interactions with the disk (Machida et al., 2011).

The sources of a protostar's luminosity are from accretion and internal luminosity from gravitational contraction. For both Class 0 and Class I objects the main source of

their luminosity is from accretion,

$$L_{Accretion} = \frac{GM_\star \dot{M}}{R_\star}, \tag{1.47}$$

where $M_\star$ is the mass of the protostar, $\dot{M}$ is the accretion rate and $R_\star$ is the radius of the protostar (Ray & Ferreira, 2021). From this relation we can calculate the accretion rate by assuming the values of $M_\star$ and $R_\star$, typically $\sim 1 - 5$ $M_\odot$ and $\sim$ 10s AU, respectively. The total bolometric luminosity is dominated by this accretion luminosity with $L_{\rm Accretion} > L_\star$.

Once the envelope has been accreted or expelled via outflows and stellar winds the protostar is no longer embedded. We now have a Class II protostar and is referred to as a pre-main sequence star (or a classical T-Tauri star). Now that the envelope is gone, and accretion has decreased the primary source of luminosity is the conversion of gravitational potential energy into luminosity via contraction, in other words $L_{\rm Accretion} < L_\star$. The timescale of this stage can be calculated by using the viral theorem ($E_{\rm T} = 2\mathcal{K} + \Omega_{\rm G}$) where we know that half of the gravitational potential energy will go into powering the luminosity. The luminosity liberated from the potential gravitational energy is

$$L_\star = \frac{1}{2}\left(-\frac{d\Omega_{\rm G}}{dt}\right). \tag{1.48}$$

The timescale for how long the luminosity is supplied via gravitational collapse is called the Kelvin-Helmholtz timescale and is

$$t_{\rm KH} = \frac{GM_\star^2}{R_\star L_\star}, \tag{1.49}$$

where $M_\star$ is the mass of the protostar, $R_\star$ is its radius and $L_\star$ its luminosity. From this equation we can see that the timescale will increase as the radius of a pre-main sequence star decreases during contraction. The SED of a Class II object peaks at a

shorter wavelength of around $\sim 1.5\,\mu$m corresponding to a temperature of $\sim$ 2000 K. There is still an IR excess that is likely from the presence of a protoplanetary disk. The H_2 will become dissociated at a temperature of $\sim$ 1800 K. Instead of the kinetic energy of gas heating the protostar up, a large portion of it will be used to break apart the H_2 and causes a second isothermal collapse (Cameron, 1962; Gould, 1964).

A Class III object has a similar SED to a Class II object with the main difference being a decrease in the IR excess. This is due to both material being lost from the disk via photoevaporation from the protostar its self or from nearby massive stars, and also material being accreted by planetesimals (ALMA Partnership et al., 2015). Because of the increase in $t_{\rm KH}$ when $R_\star$ decreases each subsequent stage of the protostars' evolution takes and order of magnitude longer.The lifetimes of Class 0, I, II and III objects are $\lesssim 10^5$ yr, $\sim 10^5$ yr, $\sim 10^6$ yr and $\sim 10^7$ yr, respectively. (Whitworth & Ward-Thompson, 2001)

In summary a pre-stellar core will become a Class 0 object when a hydrostatic object forms inside it. A Class 0 object is defined as a single (or multiple) IR sources peaking at $\sim 100\mu$ m. A Class I SED will peak at $\sim 4\mu$m, corresponding to a temperature of $\sim$ 700 K. There is now an IR excess with two components, the two components are the envelope that the Class I object is embedded within and the circumstellar disk. Once the envelope is accreted or blown away (or lost via outflows) we have a Class II object with a SED showing a shorter wavelength corresponding to a higher temperature, the IR excess has also decreased significantly due to the loss of the envelope. The remaining IR excess is likely the circumstellar or protoplanetary disk. A Class III object is the stage immediately before the main sequence and the stage a protostar will spend the longest amount of time in due to the Kelvin-Helmholtz contraction timescale increasing for objects with decreasing radii. The SED is nearly identical as a Class II object but with almost no IR excess due to material in the disk having been removed via stellar winds or being turned into protoplanets. A key revelation of star formation is that

planets can form around protostars long before they become main sequence stars (ALMA Partnership et al., 2015; Segura-Cox et al., 2020; Alves et al., 2020; Parker, 2020). This is discussed in more detail §1.12.

1.5.2 Main Sequence Stars

A class III object becomes a main sequence star once its core has started to fuse hydrogen at a temperature $\sim 10^7$ K and pressure $\sim 10^3$ g cm^{-3} (Chabrier & Baraffe, 1997). Hydrostatic equilibrium is now maintained via the energy released from nuclear fusion. The time spent on the main sequence is determined by how much mass the star has, with the lifetime of a star being strongly dependent on its mass. The mass luminosity relation is

$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^a, \tag{1.50}$$

where M is the stellar mass, L is the stellar luminosity and a depends on the mass range of the star (Kuiper, 1938). For a typical main sequence star $a = 4$ (Smith, 1983).

1.6 High-Mass Star Formation

Despite their short lifetimes ($\sim$ 10s Myr) and relative rarity there are few objects that can have such a profound impact on the ISM as high-mass stars do. They add energy to the ISM due to their strong stellar winds and produce heavier elements via nucleosynthesis when they die and explode (supernovae) (Burbidge et al., 1957; François et al., 2004). Understanding the conditions needed for high mass star formation can therefore shed light on how the ISM evolves with time, which in turn will affect any future star formation and galactic evolution (Zinnecker & Yorke, 2007).

From empirical evidence, like the initial mass function (IMF), high-mass stars are rare. In the literature a star is considered to be a high mass star if its mass is $\geq 8M_\odot$.

This is a physically motivated choice as any star with a mass greater than this will start fusing hydrogen while accretion is still occurring. This means that the star is already on the main sequence while material is still being accreted.

Observations of high mass stars show that they are often found near the centre of young star-forming regions, like the ONC (Allison et al., 2009). Loosely bound associations of massive stars are observed known as OB associations, named after their brightest members spectral types (Wright et al., 2023).

In their review Zinnecker & Yorke (2007) discuss three theories for massive star formation, competitive accretion, monolithic core collapse and stellar mergers. In the review they make it clear that high-mass star formation is unlikely to just be a scaled up version of low-mass star formation.

The timeline of high-mass star formation can be broken down into four stages (Zinnecker & Yorke, 2007). An initial compression of molecular gas, a collapse of gravitationally unstable overdensities to form protostars, accretion of material onto these protostars and lastly the disruption of the birth environment. Massive stars form rapidly, in $\sim 5 \times 10^5$ yr, and will have strong stellar winds and high UV fluxes.

Stars with masses $> 25\,M_\odot$ will not explode via supernovae when they die, instead they will collapse directly into a black hole. However, these massive stars will have a great effect on their surroundings via strong stellar winds before they collapse. Models have shown that stars between $9\,M_\odot - 25\,M_\odot$ will undergo a core-collapse supernovae, enriching the ISM via nucleosynthesis and clearing it away (Limongi & Chieffi, 2018). The current models show that the first supernovae occur after ~ 10 Myr in SFRs (Reiter & Parker, 2022). The models in Limongi & Chieffi (2018) attribute this to the rotational mixing within massive stars, which increases the duration of core hydrogen burning. This has implications as to how star clusters disperse into the Galactic field, as it is unlikely that supernovae can have a dominant role in removing gas from young clusters as most will have already dispersed in < 10 Myr, before the first supernovae have occurred.

1.6.1 Competitive Accretion

For a massive star forming via competitive accretion it (and its siblings) all start from cores of around a Jeans mass. In this model it is the cores located in the deepest gravitational potential wells that will produce the most massive stars, as they will have access to a greater reservoir of material to accrete compared to cores located in shallower potential wells (Bonnell, 2005).

The competitive accretion scenario implies that the most massive stars will be located at the centre of their respective clusters, surrounded by lower mass stars. This central spatial distribution of the most massive stars is called mass segregation and can either arise during the star formation process (primordial mass segregation) or later via dynamical interactions (dynamical mass segregation) (McMillan et al., 2007; Moeckel & Bonnell, 2009).

Determining if a region is mass segregated can be done quantitative using the mass segregation ratio, $\Lambda_{\rm MSR}$ (see Chapter 2 §2.3 (Allison et al., 2009)). One of the main questions in my research has been are newer methods robust at quantifying different parameters of SFRs? This is incredibly important to know as these methods are used to infer the likely formation mode of massive stars, as well as the overall state of SFRs (i.e. density, virial state and degree of substructure).

Mass segregation has been detected and quantified in the ONC using the mass segregation ratio (where $\Lambda_{\rm MSR} = 2.0 \pm 0.5$, see § 2.3.1 for details) (Allison et al., 2009). To use mass segregation as a marker for competitive accretion we need to determine if the mass segregation is dynamical or primordial. Determining which kind of mass segregation we are observing is done by comparing the age of a SFR with its crossing time, $t_{\rm c}$. The crossing time of a SFR is defined as,

$$t_{\rm c} = \frac{R}{\sigma_\nu}, \tag{1.51}$$

where R is the size of the region and σ_ν is the velocity dispersion of the region.

If the crossing time is less than the age of the region then dynamical mass segregation cannot be ruled out. However, if the crossing time is greater than the age of the SFR then it indicates that the most massive stars formed approximately where we observe them, meaning the observed mass segregation is primordial (McMillan et al., 2007). However, hydrodynamical simulations have shown that when feedback and stellar winds are taken into account that the distribution of the most massive stars is not significantly different compared to the lower mass stars in a region (Parker & Dale, 2017).

1.6.2 Monolithic Core Collapse

When star formation occurs within infrared-dark clouds (IRDCs) massive stars can form within "hot" cores. Hot cores consist of molecular hydrogen and have very high density ($10^5 - 10^8$ cm^{-3}), small size ($\leq$ 0.1 pc), relatively high temperature (50-250 K) and mass ($\sim 100 - 300$ M$_\odot$) (Rathborne et al., 2006). The moment protostars begin forming within dense dark cores inside the IRDCs the gas will begin to heat up. The greater the gas' temperature the greater its thermal support is against gravitational collapse, meaning more mass is needed for it to start collapsing. This manifests as an increase in the Jeans mass. If the core does not fragment and the star formation efficiency is high enough the core will produce a single high mass star (Rathborne et al., 2006).

Observing massive star formation is difficult due to the high extinction caused by the dust and is further compounded by their rarity as they are more likely to be at greater distances from us (mean distance of several kpc) (Zinchenko, 2022). Despite these challenges observations have found massive protostars within the dark cores of IRDCs (Rathborne et al., 2005; Henshaw et al., 2016; Barnes et al., 2023). Additional challenges arise due to the rapid formation rate and evolution of massive stars. It makes it incredibly difficult to find large samples of massive stars not already on the main

sequence due to their rapid formation timescales.

Evidence for monolithic collapse would be the observation of massive stars in isolation, with no lower mass stars nearby. The formation of massive stars happens very quickly with their luminosity from contraction dominating the luminosity from accretion. The young massive star will have strong stellar winds which will clear its surroundings of gas; halting further star formation in the vicinity. There are some observations that support the monolithic collapse scenario. Both Lamb et al. (2010) and Bressert et al. (2012) observe high mass stars in relative isolation. In Lamb et al. (2010) they find both high mass stars in isolation, some of which are massive runaway stars. They also find some high mass stars surrounded by small groups of lower mass stars. Similarly, Bressert et al. (2012) find isolated high mass stars (defined as $M \gtrsim 30$ $M_\odot$ in their work) which are unlikely to be runaway stars.

In summary, there is evidence for both monolithic and competitive high mass star formation. This leads to the likely scenario that both of these formation modes occur. This could mean high mass star formation is dependent on the initial conditions of giant molecular clouds and cores, and is not just a scaled up version of low mass star formation.

1.7 Star-Forming Regions

So far I have focussed on the formation of a single star within a prestellar core and briefly discussed high-mass star formation. Observations have led us to the discovery that stars form in groups of tens to thousands of members (Lada & Lada, 2003; Bressert et al., 2010).

SFRs undergo rapid dynamical evolution, erasing any initial spatial information that may shed light on their initial conditions. Most SFRs will expand and due to the increasing influence of the Galactic tidal field (as the gravitational potential of the group decreases due to ejected stars and the loss of gas), stars will become unbound from their

natal group, and disperse into the Galaxy. There has been work that uses the spectra of stars to chemically tag them, meaning that stars with the same chemical tags likely formed from the same cloud at approximately the same time (Kos et al., 2017). The low-age spreads observed in SFRs, supports the idea that groups of stars form around the same time (Jeffries et al., 2011).

In this section I will discuss the various definitions of a cluster in the literature and some key observations of SFRs.

1.7.1 What is a cluster?

Due to projection effects and contamination of the foreground stars determining which stars are a part of a cluster is difficult. The fact that clusters are not single objects, but ensembles of 10s to 1000s of stars make defining boundaries challenging (i.e. where does the cluster end and the field begin?). As a consequence of this we have many definitions of a what a cluster is. This is important to highlight as if clusters are defined differently then we cannot compare them to one another as the physics we infer about clusters may depend on how we have defined a cluster in the first place. Another compounding factor is that clusters of stars can look very different from one another, see Figure 1.7 and Figure 1.8 showing the ONC and Taurus, respectively.

Regardless of the preferred definitions or naming conventions (clusters and SFRs) there is a clear distinction that can be made, that there can be gravitationally bound "clusters" and gravitationally unbound "associations" of stars (Gieles & Portegies Zwart, 2011).

Definitions that use a physical parameter such as the surface density of a region makes sense to use where there is an obvious distinction between the clustered stars and the background field. But the issue with defining a cluster based on surface density is that there are multiple different density thresholds used to define a cluster. In Bressert

et al. (2010) they discuss the different surface densities (in terms of young stellar objects (YSOs)) used in the literature to define stellar clusters which is summarised in Figure 1.6. The range of surface densities spans from 3 to 60 YSOs pc^{-2} (Carpenter, 2000; Lada & Lada, 2003; Jørgensen et al., 2008; Gutermuth et al., 2009). In contrast Kruijssen (2012) finds no critical density of YSOs that would imply a fundamental star-formation surface density.

There is an entire suite of cluster finding algorithms, again with slightly different definitions of what a cluster is. For example the widely used clustering algorithm `DBSCAN` defines clusters as areas of high density surrounded by a lower density population (Ester et al., 1996).

To better understand clusters (or SFRs) a range of methods have been developed to quantify the spatial distributions of stars within them. Since stars form in substructured environments they will inherit their initial spatial distribution from the collapsing natal cloud, which in turn may affect the amount of material protostars can accrete. Constraining the initial conditions could therefore help us understand not only star formation but also the effects on the IMF (if any), the observed number of binary or higher order systems and even the effect dense SFRs will have on planet formation occurring around stars within them.

SFRs are highly dynamic; both observations and simulations have shown that stars can be ejected from their natal group (Blaauw, 1961; Oh & Kroupa, 2016; Schoettler et al., 2020). These runaway stars can help constrain the initial stellar density of SFRs (Schoettler et al., 2020). The dynamics of SFRs are crucial to understanding the observed distribution of multiple systems in the Galactic field; most stars are observed to be in binary and higher order systems (Eggleton & Tokovinin, 2008; Raghavan et al., 2010). Simulations have shown that the field population of multiple systems in the field cannot be reproduced via dynamical capture; instead the field population we observe must have mostly formed in binary or higher order systems. Then in the dense star-forming regions,

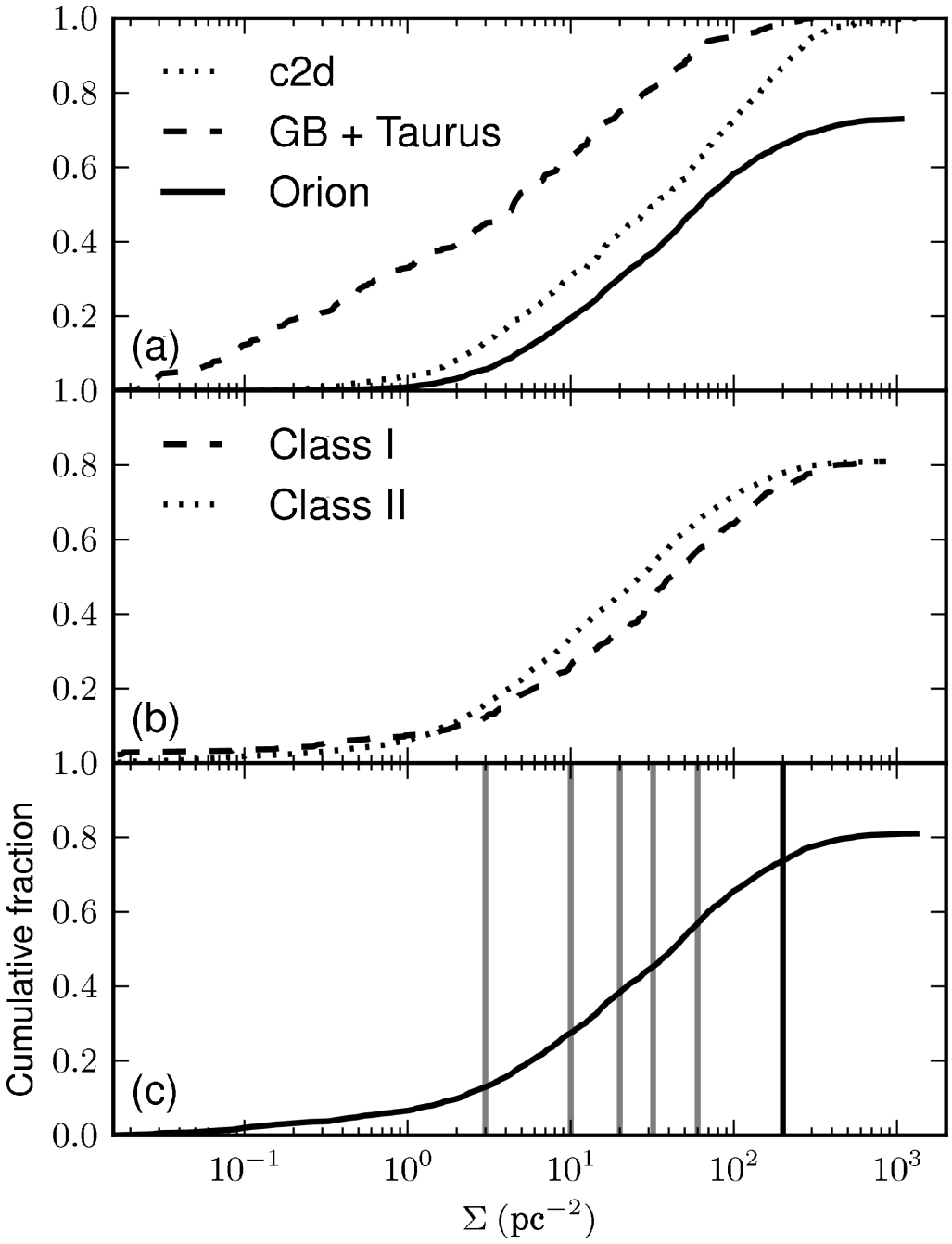


Figure 1.6: Cumulative fraction of surface densities for surveys of GB+Taurus, c2d and Orion, taken from Bressert et al. (2010). Panel (c) is especially important as it shows (via the vertical lines) the different surface density thresholds used in the literature to define a stellar cluster. The observations of YSO surface densities show a smooth continuous distribution, such that the chosen thresholds in the literature do not constrain "monolithic" or "hierarchical" star cluster formation.

via direct interactions, the systems will lose members (Parker et al., 2009; Goodwin, 2010; Marks & Kroupa, 2011).

1.7.2 Monolithic Cluster Formation

In monolithic cluster formation all the gas is within the final volume of the cluster before stars begin to form (Longmore et al., 2014; Williams et al., 2022). Regions of the gas cloud will become over dense, exceeding the Jeans mass and will collapse, becoming prestellar cores. Once the protostars have formed at the centres of the cores their stellar winds will expel any remaining gas in the region, removing the gravitational potential which keeps the stars bound. This removal of gas decreased the gravitational potential of the stellar cluster and is thought to be why stellar clusters are dispersed into the galactic field (Fellhauer & Kroupa, 2005; Moeckel et al., 2012).

Observationally we see what looks like the product of monolithic cluster formation in the ONC shown in Figure 1.7. The ONC shows a smooth, radial distribution of stars, as expected from monolithic formation.

1.7.3 Hierarchical Cluster Formation

In the hierarchical cluster formation model there is a lot happening at the same time. Filaments within the GMCs are collapsing and fragmenting along their lengths forming prestellar cores, while at the same time the GMC is also collapsing (Williams et al., 2022). Filaments can overlap forming hub-and-spoke systems, with the filaments funnelling material into the hubs Anderson et al. (2021). Massive star formation has been observed within these hubs (Xu et al., 2023). Star formation is occurring throughout the GMC and eventually, through mutual gravitational attraction, all the subgroups of stars can merge into one cluster (Bate, 2012, 2014; Grudić et al., 2018).

Observationally we expect the stars within regions that form via a hierarchical pro-

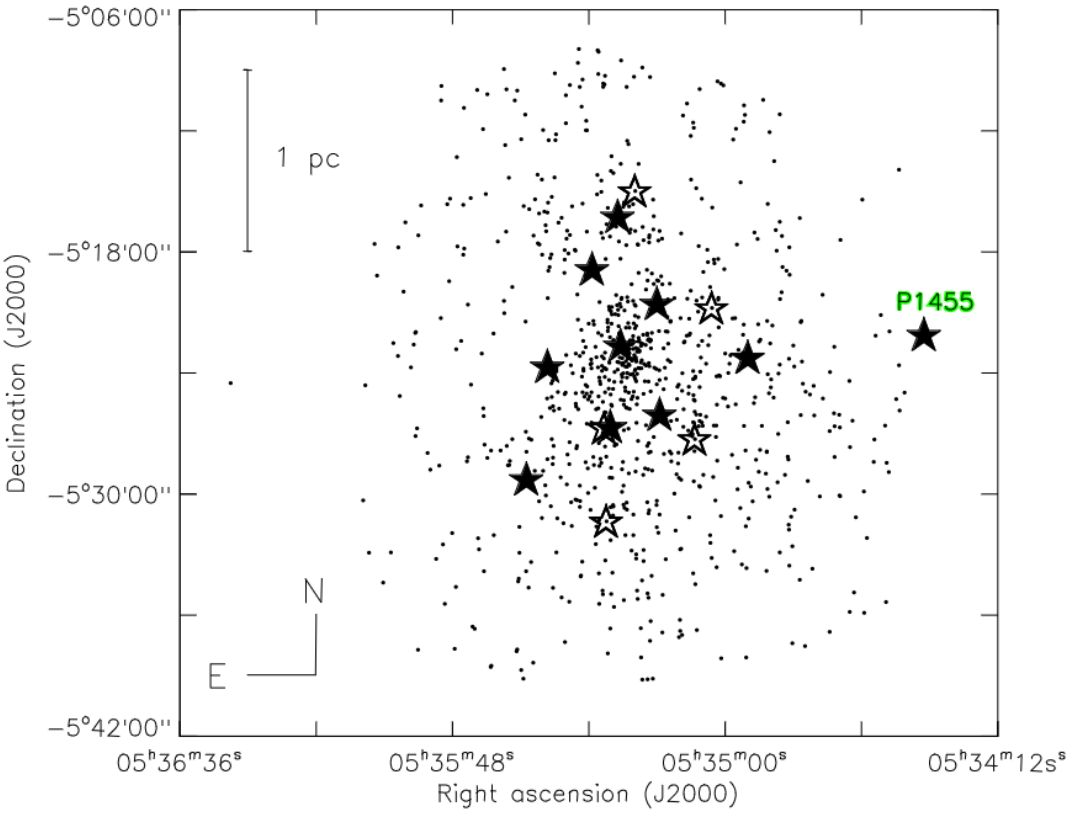


Figure 1.7: Locations of stars (black points) in the ONC (Hillenbrand & Hartmann, 1998). The ONC is an example of a region with little to no substructure. Parker (2014) finds a Q-value of ~ 0.9, which while marginal, is consistent with a lack of substructure in the ONC. Figure from Biazzo et al. (2011) where they look at the chemical properties of the filled stars. The empty stars are targets originally analysed in D'Orazi et al. (2009).

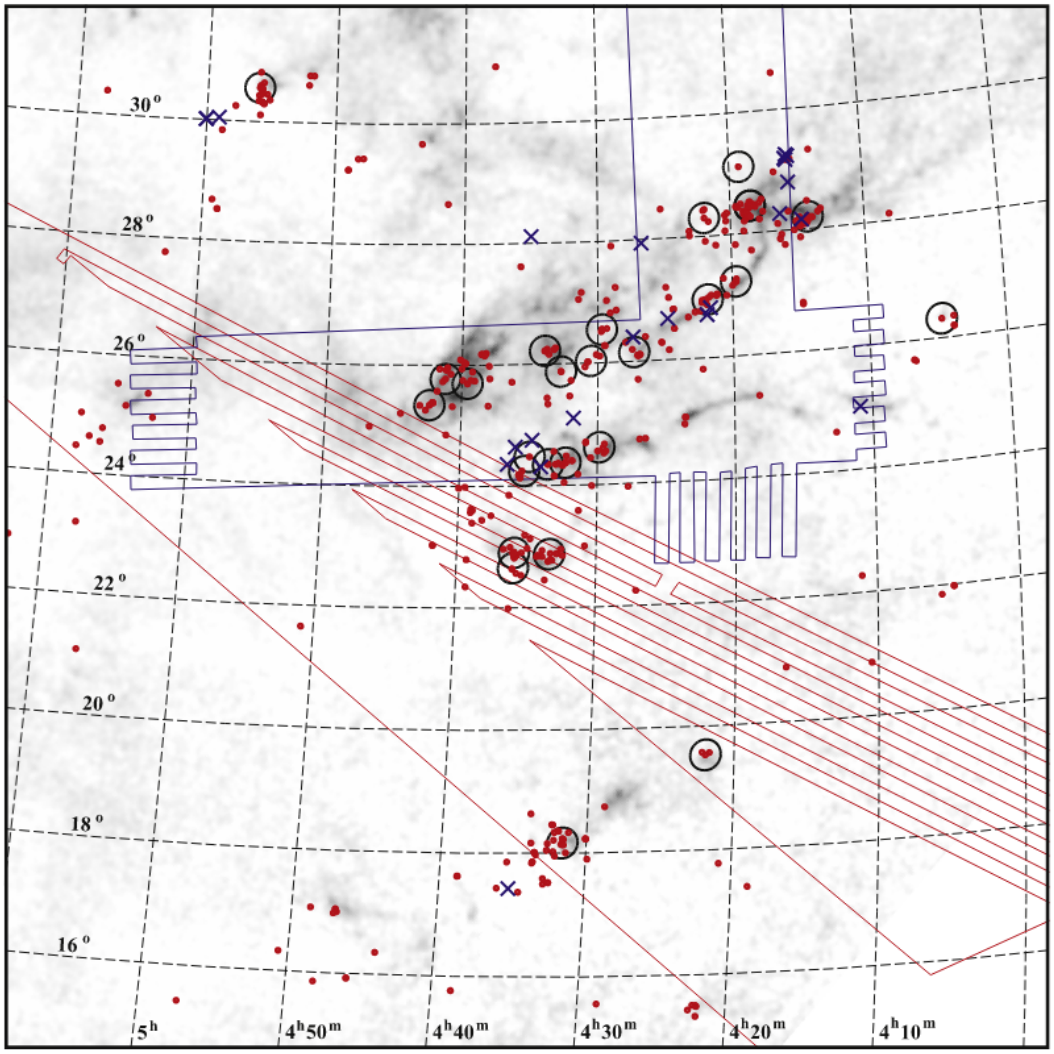


Figure 1.8: Plot of previously known members of Taurus (filled red circles) and new members found in the survey (blue crosses) by Luhman et al. (2017). The regions outlined in blue are red are the areas surveyed using SDSS (Finkbeiner et al., 2004). The grey shading represents the dark clouds in Taurus, with darker shades corresponding to greater extinction (Dobashi et al., 2005). Taurus is an example of a substructured region with Cartwright & Whitworth (2004) finding it to have a Q-value of 0.45, corresponding to a fractal dimension of 1.6 (see § 2.3.2).

cess to inherit substructure from the collapsing gas cloud. This substructure is frequently quantified using the Q-parameter (see Chapter 2 §2.3.2) (Cartwright & Whitworth, 2004). This is evident in young regions, or in low density regions that are expanding (supervirial). The expansion and low initial stellar density acts to reduce the number of dynamical interactions a star can have. This decrease in dynamical processing preserves the initial spatial distributions stars inherit from their formation. If a region is collapsing (subvirial) the stars will undergo far more dynamical processing, rapidly erasing traces of the initial substructure, as measured using the Q-parameter, in $< 5\,\mathrm{Myr}$. (Scally & Clarke, 2002; Goodwin & Whitworth, 2004; Allison et al., 2010). Methods need to be developed and tested to see if substructure, either spatial or kinematic can be measured for longer (see Arnold et al. (2022) in which kinematic substructure is quantified). Any such method that quantifies substructure will need extensive testing as they can be used to infer the likely formation mode of star-forming regions.

The rate of substructure erasure depends on the initial density, with high density regions undergoing a faster rate of substructure erasure than low density regions (Parker, 2014). This is unfortunate as it means that SFRs can share similar morphologies but different initial conditions.

This can be ameliorated by combining two methods that quantify the spatial distribution of stars. Two such methods, the Q-parameter and the local surface density ratio (see § 2.3) were used together in Parker (2014) to determine the initial densities of SFRs (Cartwright & Whitworth, 2004; Maschberger & Clarke, 2011). The desire to quantify the substructure of SFRs comes from the idea that the dynamical evolution of a region can be used as a proxy for its age (i.e. a dynamical age) helping further constrain the initial conditions of SFRs.

1.8 The Initial Mass Function

The initial mass function (IMF) is an empirical relation between the number of stars within certain mass ranges and was first expressed a power law in Salpeter (1955) and is

$$\Phi(\log(m)) = \frac{d\,N}{d\,\log(m)} \propto m^{-\Gamma} \tag{1.52}$$

where N is the number of objects in the mass range $\log(m) + d\log(m)$. The power law exponent $\Gamma \approx 1.35$.

Subsequent observations have revealed a turnover in the IMF at $M \lessapprox 0.1 M_{\odot}$. This turnover prevents the entire mass distribution from being described using a single power-law. As such a multi-power law IMF has been proposed by Kroupa (2001) in the form of the piece wise function

$$p_{\mathrm{Kroupa}}(m) = \begin{cases} Ak_0 m^{-0.3}, & 0.01\,\mathrm{M}_{\odot} \leq m < 0.08\,\mathrm{M}_{\odot} \\ Ak_1 m^{-1.3}, & 0.08\,\mathrm{M}_{\odot} \leq m < 0.5\,\mathrm{M}_{\odot} \\ Ak_2 m^{-2.3}, & 0.5\,\mathrm{M}_{\odot} \leq m < 1\,\mathrm{M}_{\odot} \\ Ak_3 m^{-2.3}, & 1\,\mathrm{M}_{\odot} \leq m < 150\,\mathrm{M}_{\odot} \end{cases}, \tag{1.53}$$

where A is a normalisation constant is and k_i is a coefficient (see Pflamm-Altenburg & Kroupa (2006) for details on calculating k_i) (Kroupa, 2001). The IMF is described using these four different power laws across different mass intervals. A functional form of the IMF was developed in Maschberger (2013) and is

$$p_{\mathrm{Maschberger}}(m) = \left(\frac{m}{\mu}\right)^{-\alpha} \left(1 + \left(\frac{m}{\mu}\right)^{(1-\alpha)}\right)^{-\beta}, \tag{1.54}$$

where α is the low mass index, β is the high mass index and μ is the mean stellar mass. In § 2.1.4 I go into detail on how I use the Maschberger IMF to pick masses for the work

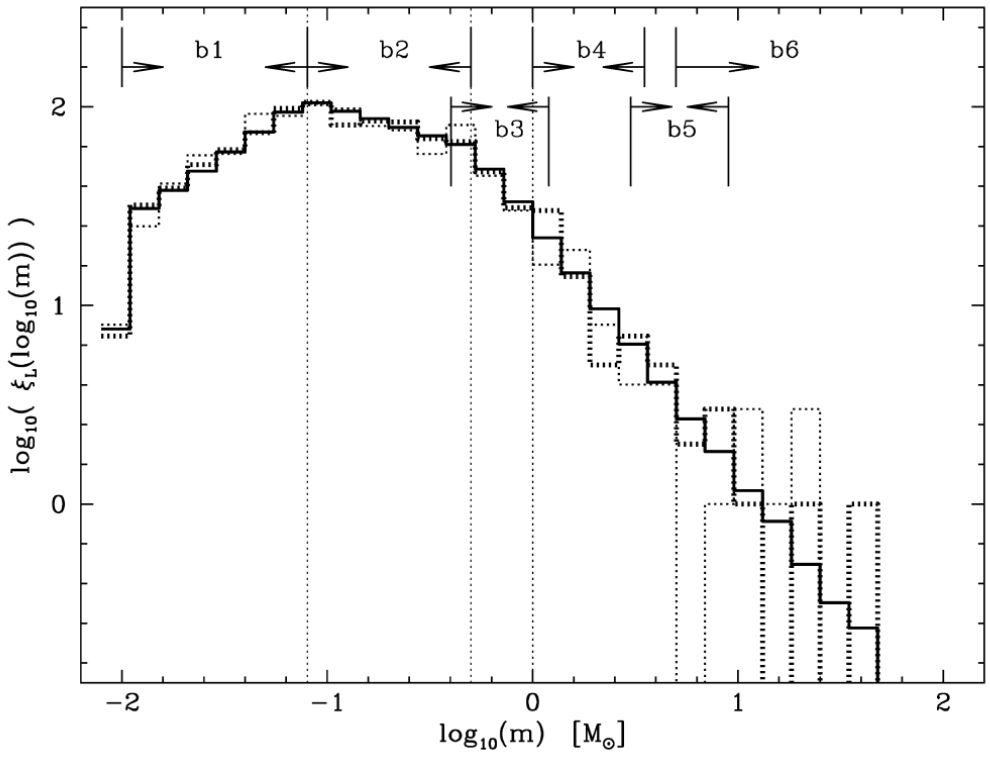


Figure 1.9: Figure taken from Kroupa (2001) showing the IMF for 10^6 stars. The thin and thick dotted lines show two random distributions drawn from the IMF. The intervals labelled along the top show what regions are described using different power laws.

performed in the science chapters. The probability density function from Maschberger (2013) is shown in Figure 1.10.

The initial mass functions with various power laws overlaid is shown in Figure 1.9. An apparent link between the IMF and the birth environment can be seen when looking at the core mass function (CMF) (shown in Figure 1.11), this is defined the same as the IMF but instead of star masses we are instead looking at the masses of clumps of dense molecular gas. This has a similar shape to the IMF of stars but slightly offset to higher masses. This difference implies a star formation efficiency of $\sim 35\%$. However, the SFE in GMCs can be as low as $\sim 2\%$, with some simulations finding a time averaged SFE of $\sim 4\%$ (up to $\sim 40\%$ in the densest simulated clusters) (Myers et al., 1986; Bonnell et al., 2011).

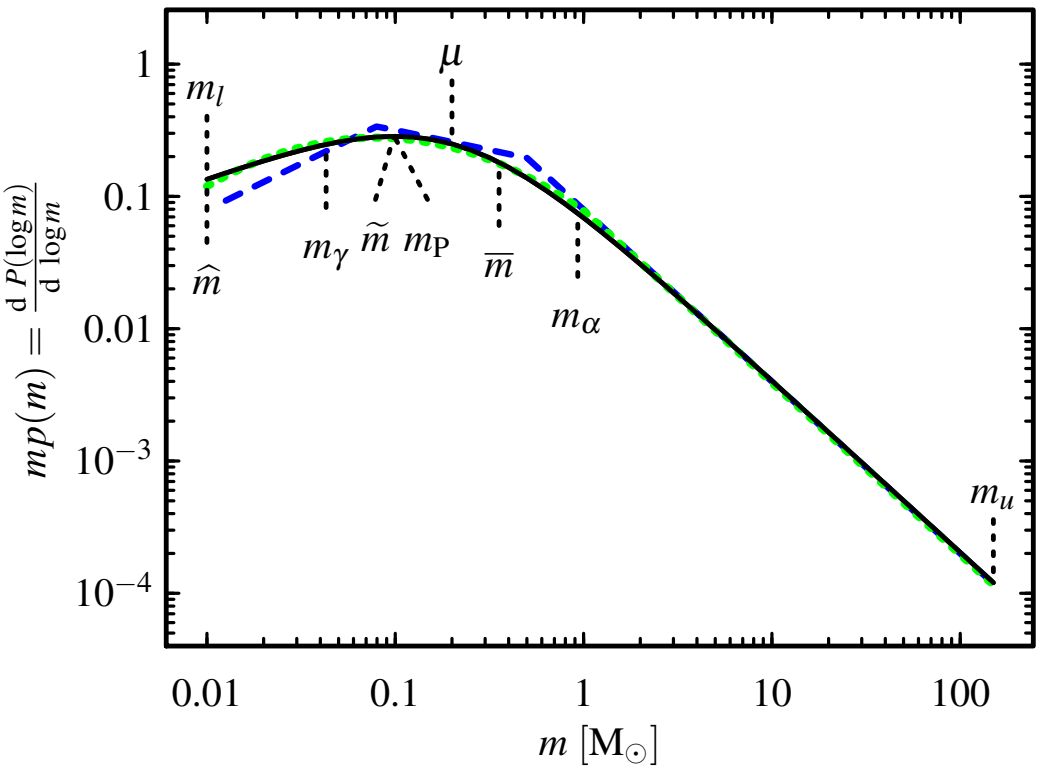


Figure 1.10: The probability density function (Equation 2.4) from Maschberger (2013) is shown by the solid black line. The blue dashed line shows a comparison to the Kroupa IMF and the green dashed line shows the Chabrier IMF (Chabrier & Baraffe, 1997). Figure taken from Maschberger (2013).

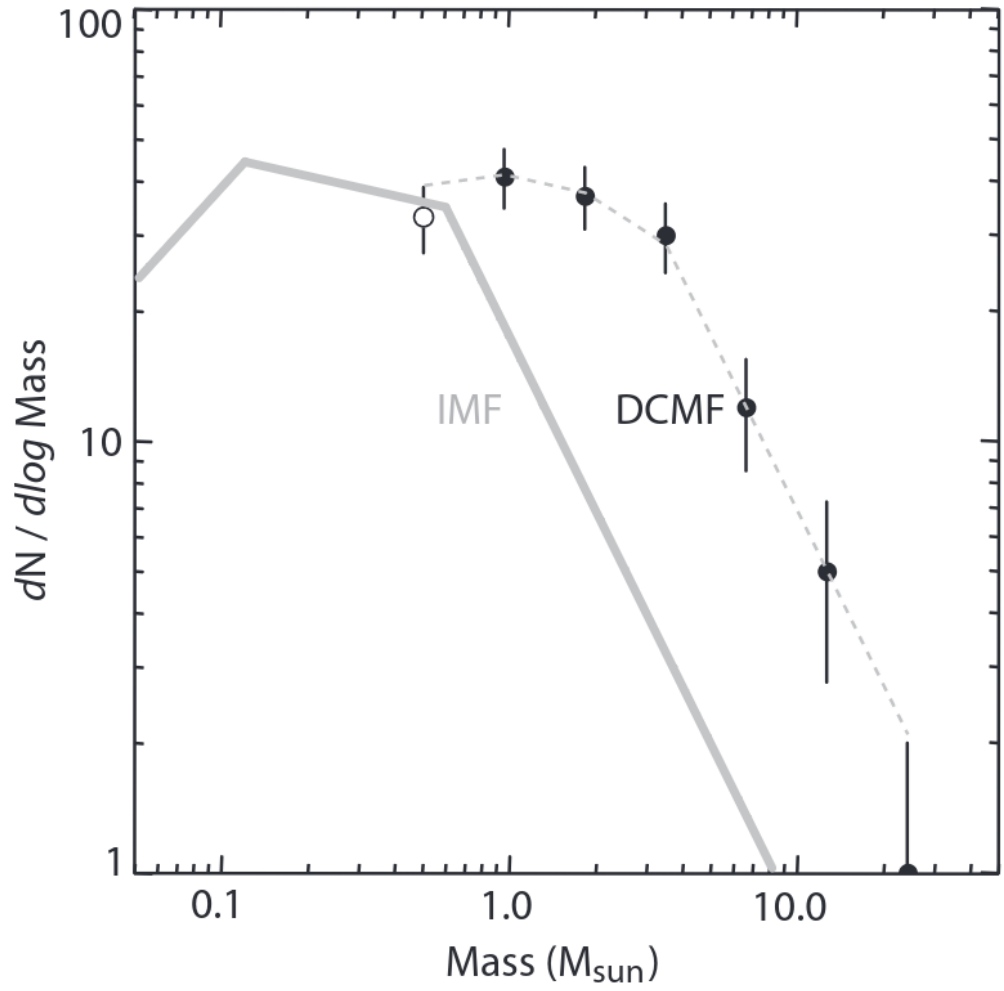


Figure 1.11: Plot showing the IMF of stars and the dense core mass function (DCMF). The difference between the DCMF and the IMF can be explained as star formation is not efficient, so the peak of the IMF will correspond to a lower mass compared to the DCMF. Figure taken from Alves et al. (2007).

From the IMF we have learnt that the Sun has a relatively high mass compared to most stars in the Galaxy. We have also observed that massive stars are relatively rare, with the average stellar mass being ~ 0.5 $M_{\odot}$. Many objects formed from collapse of GMCs will be substellar objects called brown dwarfs. They are colloquially referred to as "failed stars" due to their mass being below the limit for hydrogen fusion, which is ~ 80 M_{Jup}.

Understanding the origin of the IMF is of great importance in determining if star formation is a universal process (Padoan et al., 1997). The primary theories on the origin of the IMF focus on turbulence or competitive accretion (Klein et al., 2007).

If the mass distribution of stars is independent of the star formation environment then star formation can be said to be universal. Theory suggests that the IMF may be affected by the metallicity of SFRs, with lower metallicities producing a top-heavy IMF (Larson, 1998; Dopcke et al., 2011). This dependence on the IMF on the metallicity comes from the greater cooling that can occur in high metallicity gas, which decreases the thermal support a molecular core can have, which in turn reduces the Jeans mass (less mass needed to overcome the thermal support).

The binary fraction of regions can also be used to test the universality of star formation. In King et al. (2012) they find that if star formation is universal then all SFRs must be initially cold and clumpy (subvirial and highly substructured).

1.8.1 Where are all the old bound SFRs?

Very few star forming regions survive for longer than 10 Myr (with most stars forming in unbound regions), there are several theories as to why (Lada & Lada, 2003). First, is the idea that star formation happens over a range of densities and as bound clusters must have had a very high star formation efficiency the region must have been very dense initially (Bressert et al., 2010; Kruijssen, 2012). Very dense regions are thought to be

rare, with Kruijssen (2012) estimating only $\sim 35\%$ of stars formed in dense regions with other estimates putting it much lower, with only $\sim 10\%$ of stars being in bound clusters today (Krumholz et al., 2019).

For stars to remain bound after the gas potential is removed they need to have a sufficiently high enough SFE (ϵ). The models of Shukirgaliyev et al. (2017) finds that $\epsilon \geq 15\%$ allows a cluster of stars to remain bound, if we assume that both the gas and stars have the same spatial distribution and that the stars are in virial equilibrium with the gas.

In Kim et al. (2021) they calculate the effective star formation efficiency (eSFE) using the virial parameter of simulated regions and find it to be low, $\sim 2\%$, which agrees within a factor of two of observations (Utomo et al., 2018).

Investigations by Goodwin & Bastian (2006) defined the eSFE using the virial ratio and is

$$eSFR = \frac{1}{2Q_\star}, \tag{1.55}$$

where

$$Q_\star = \frac{\mathcal{K}_\star}{\Omega_\star}, \tag{1.56}$$

where $\mathcal{K}_\star$ is the kinetic energy of the stars and $\Omega_\star$.

The probability of a SFR surviving gas expulsion is strongly dependent on the initial virial state at the instant of gas expulsion (Goodwin, 2009). Using the equation from Goodwin (2009) we can see that very small changes in the SFE dictate the dynamical behaviour of star clusters. For example, taking the eSFE to be 1% corresponds to a virial ratio of 0.5 (virial equilibrium) but at a eSFE of 5% the stars have a virial ratio of 0.1 (subvirial).

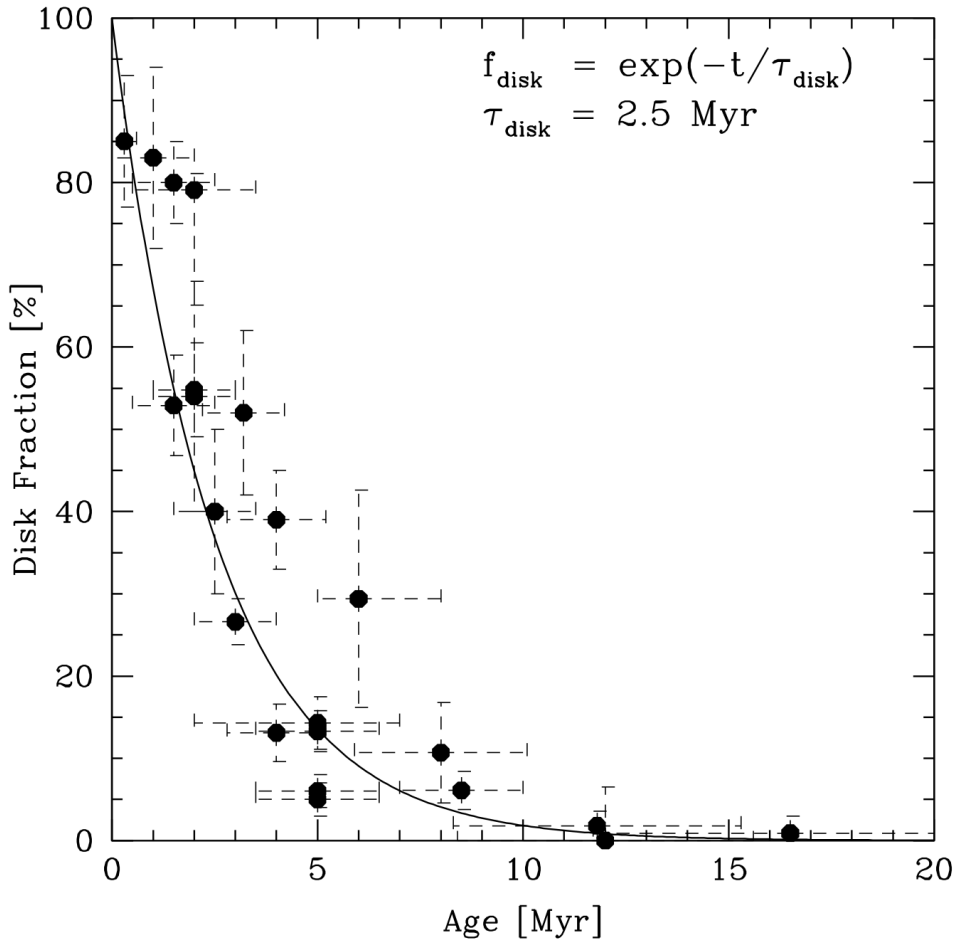


Figure 1.12: Plot showing the percentage of stars in a group that have protoplanetary disks around them plotted against their ages. The solid black line shows the fitted exponential model. A disk half-life of 2.5 Myr is found. Figure taken from Mamajek (2009).

1.9 Planet Formation in Star-Forming Regions

To date we have found 5483 exoplanet candidates, of which 3699 are confirmed [1]. Their ubiquity in the Milky Way is evident, meaning that each star is likely a host to multiple planets.

Planets can form around protostars while the star is still accreting material from its natal envelope (Greaves & Rice, 2010; Vorobyov, 2011; Richert et al., 2018). Radiation

[1] https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html, accessed 2nd August 2023.

from nearby massive stars can remove material from its protoplanetary disk via photoevaporation (Scally & Clarke, 2001; Nicholson et al., 2019; Concha-Ramírez et al., 2019; Parker et al., 2021). The density of the regions will dictate which physical process will dominate disk disruptions.

Figure 1.13 shows the wide range of morphologies that protoplanetary disks can have (Andrews et al., 2018). This has implications for how we model both star and planet formation. The role of direct interactions via fly bys in dense star-forming regions may happen for a large portion of the cluster members, leaving distinct perturbations in the circumstellar disk (Cuello et al., 2023).

1.9.1 Dynamical Effects on Exoplanetary Systems

In dense ($> 10^4$ $M_\odot\,pc^{-3}$) SFRs it has been shown in simulations that direct interactions between stars during fly bys can truncate the circumstellar disk (Vincke & Pfalzner, 2016).

Planets can be affected during their formation in their host protostar's protoplanetary disk Vincke & Pfalzner (2016). Observations of protoplanetary disks using ALMA have shown that specific structures seen in them, with spirals being caused by flybys in simulations (Bae et al., 2023).

In lower density simulations ($100 - 1000$ $M_\odot\,pc^{-3}$) planets can become perturbed to varying degrees, with some becoming free-floating "rogue" planets or stolen outright from their host star (Daffern-Powell et al., 2022).

1.9.2 Photoevaporation

The photoevaporation of circumstellar disks can happen in very low density SFRs (~ 10 $M_\odot\,pc^{-3}$) and can have a significant effect on not only planet formation but also planet migration (Clarke et al., 2001; Johnstone et al., 2004; Kley & Crida, 2008; Nelson, 2018).

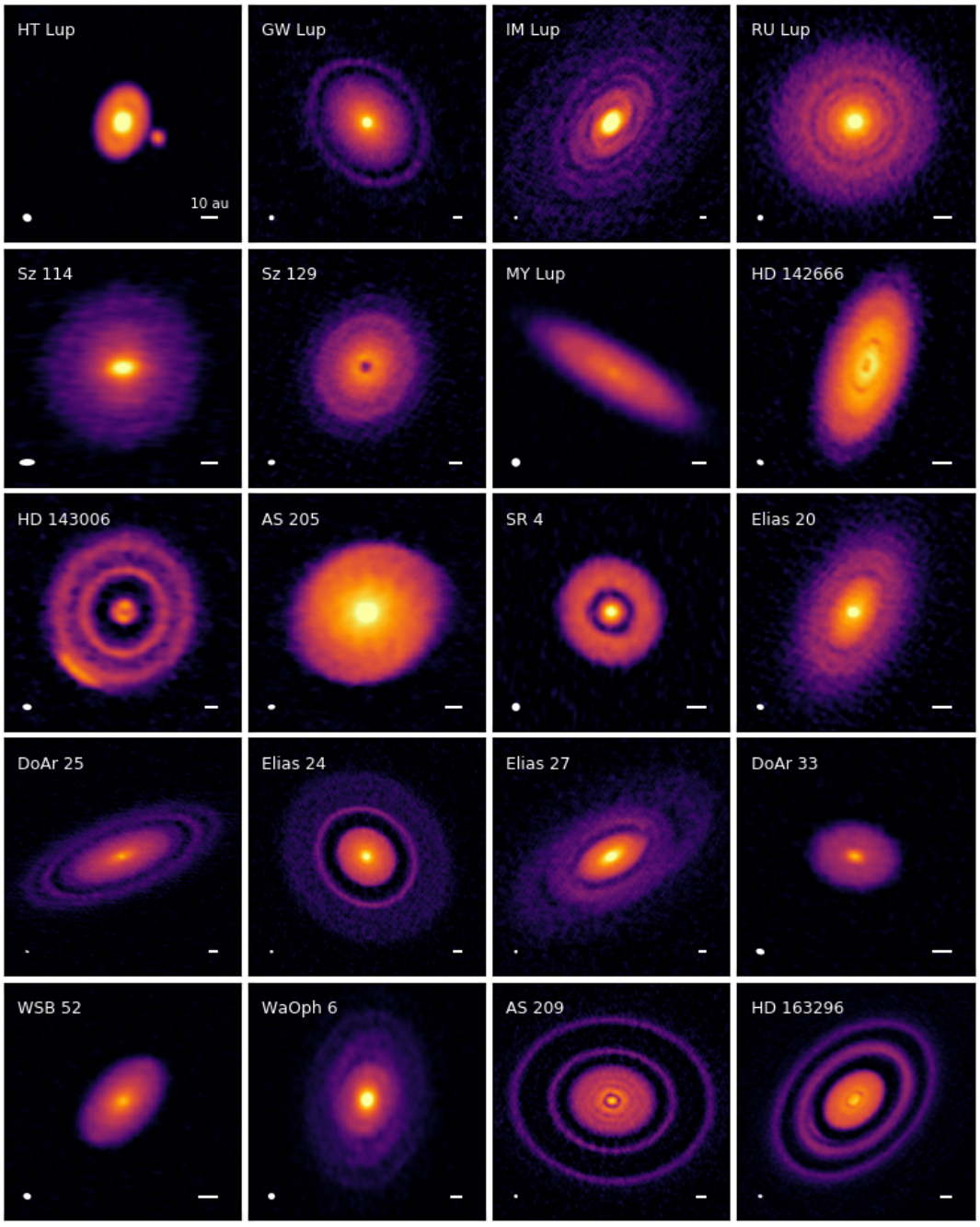


Figure 1.13: Figure taken from Andrews et al. (2018) demonstrating the wide range of different morphologies that protoplanetary disks can have. If certain morphologies are linked to specific dynamic events is under active investigation and could help further constrain the initial conditions of star formation.

Massive stars (> 5 $M_\odot$) are sources of far and extreme ultraviolet radiation, with $\nu_{\mathrm{FUV}} < 200$ nm and $\nu_{\mathrm{EUV}} < 100$ nm, respectively. It is worth stating explicitly that planet formation is occurring at the same time as star-formation with SFRs and so this radiation can have an effect on the amount of material available for planets to form from, allowing time constraints on planet formation to be determined (Concha-Ramírez et al., 2019; Daffern-Powell & Parker, 2022). This radiation, particularly the FUV, can cause protoplanetary disks to lose gaseous material, thus restricting gas giant formation. The radiation is absorbed by the dust in the disk and is remitted at longer wavelengths which heats up the dust and gas of the disk. The gas has a low escape velocity and so it is easily lost from the disk. The findings of Nicholson et al. (2019) show a correlation between the rate of circumstellar disc destruction and the initial conditions of the star-forming regions. This further highlights the importance of understanding the initial conditions of SFRs and how their evolution affects planet formation.

1.10 Future Observations

The James Webb Space Telescope (JWST) is already revealing new information about SFRs and the planets within them (see Lustig-Yaeger et al. (2023)).

JWST is characterising the atmospheres of exoplanets (Lustig-Yaeger et al., 2023). JWST will also allow the atmospheres of hot super-Earths (massive terrestrial planets with radii ≤ 2 $R_\oplus$) to be measured (Zilinskas et al., 2020).

Figure 1.14 shows an image of ρ Ophiuchi taken by JWST and shows the star formation process at different stages, included embedded Class I and II objects. Their outflows clearly visible in the right half of the image. The bright new star in the lower middle of the image has cleared a cavity in its natal cloud.

Figure 1.14: JWST image of ρ Ophiuchi. The image shows objects at different stages in the star formation process, with several embedded sources with clearly visible bipolar outflows emanating from them. The main outflow originates from an embedded source in the upper right quarter of the image. In the middle bottom is a new main sequence star that has cleared away its natal cloud, creating a cavity. Credit: NASA, ESA, CSA, STScI, Klaus Pontoppidan (STScI), Image Processing: Alyssa Pagan (STScI).

2

Methods

2.1 Constructing Synthetic Star-Forming Regions

Due to our incomplete knowledge of the star formation process and the long timescales involved we need to make use of simulations to help us constrain the initial conditions of SFRs. These simulations allow us to compare the spatial, kinematic and mass distributions to the observed distributions of stellar properties in SFRs and the field (e.g. the IMF and multiplicity fraction). The initial spatial distribution of stars in the synthetic regions are chosen to match the observed parameters of real star forming regions as closely as possible, such as density, number of members, degree of substructure and virial state. Due to the stochastic nature of star forming regions (regions with the same statistical initial conditions can evolve differently) we must make use of many simulations to allow for statistical analysis to be performed.

In this section I will discuss the spatial and kinematic construction of radially concentrated regions, Plummer spheres, substructured regions and uniform control regions. I will conclude the section with a discussion on how the masses and initial velocities of stars are chosen in the synthetic regions.

2.1.1 Smooth and Centrally Concentrated Regions

Some SFRs are observed to have smooth, centrally concentrated distributions that can be modelled using a radial density profile. The density profile I use is

$$n \propto r^{-\alpha}, \tag{2.1}$$

where n is the number density, r is the radial distance from the origin of the region and α is the radial density exponent, where higher values of α will produce more centrally concentrated regions (Cartwright & Whitworth, 2004).

Plummer Spheres

Plummer spheres have been used extensively in the literature. In fact, they became the default spatial distribution used in N-body simulations (Aarseth et al., 1974; Kroupa, 2008). A Plummer sphere of stars has a three-dimensional density distribution of

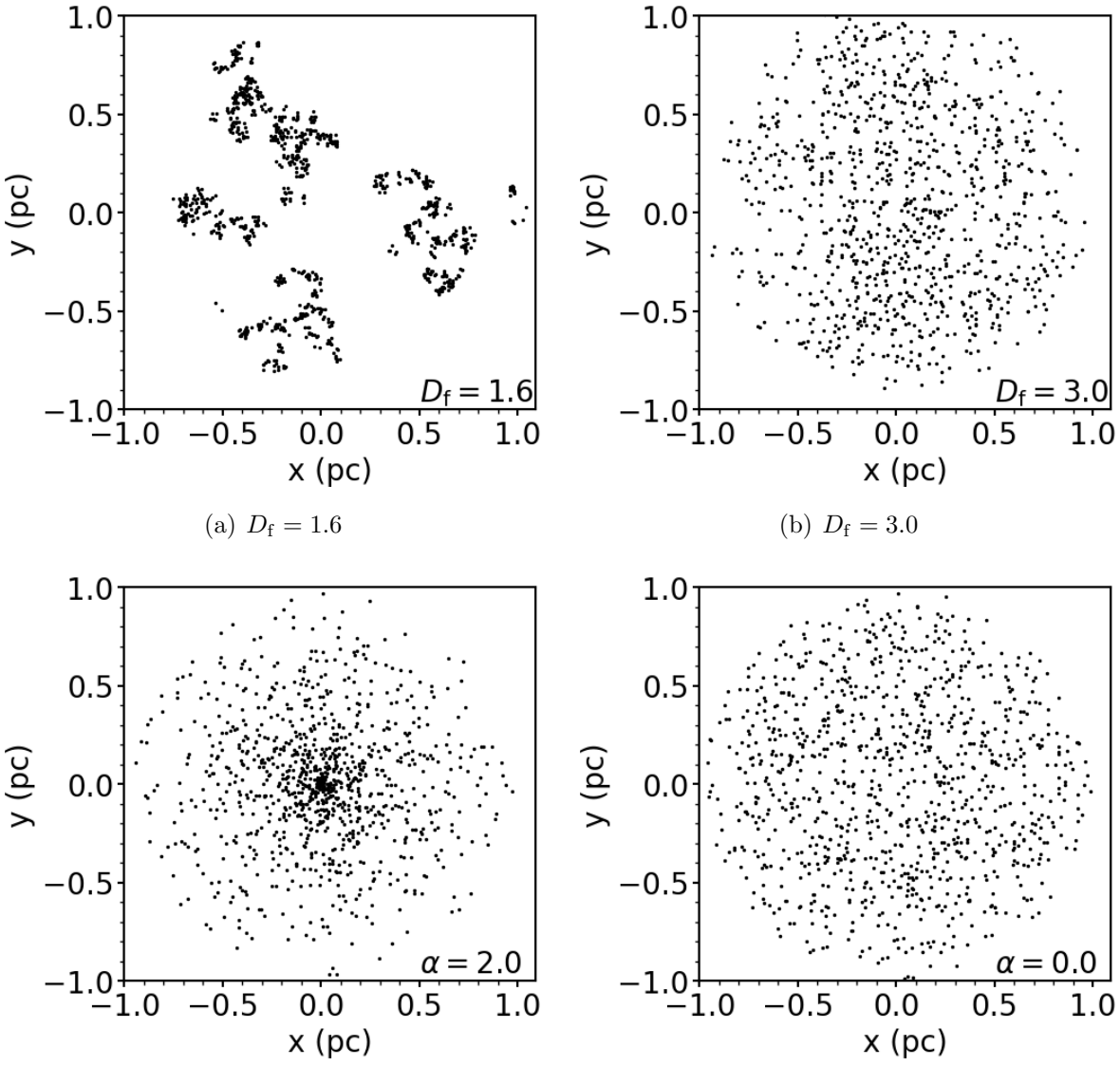
$$\rho_{\rm p}(r) = \frac{3M_0}{4\pi a^3}\left(1+\frac{r^2}{a^2}\right)^{-\frac{5}{2}}, \tag{2.2}$$

where M_0 is the total mass of all stars in the sphere, r is the distance from the centre of the region and a is the Plummer radius which sets the physical scale of the sphere (Plummer, 1911; Kroupa, 2008).

2.1.2 Substructured Regions: Box Fractal Method

Some SFRs are observed to be highly substructured. To recreate this substructure in my simulations I have used a fractal generator developed in Goodwin & Whitworth (2004) and Cartwright & Whitworth (2004), which has been used extensively in the literature (Allison et al., 2009; Parker & Goodwin, 2015; Portegies Zwart, 2016; Lomax et al., 2018; Daffern-Powell & Parker, 2020).

The box fractal method can generate many regions with different degrees of substructure by changing a single parameter, the fractal dimension $D_{\rm f}$. The lower this fractal dimension, the more substructure a region will have. Panels (a) and (b) of Figure 2.1 show two example regions, one with a high degree of substructure and one with no substructure, respectively. The method works as follows. A single star is placed at the centre of a cube of side length $N_{\rm Div} = 2$. This cube is then subdivided down into $N_{\rm Div}^3$ (in this case it is 8) sub-cubes. A star is placed at the centre of each sub-cube and each of the star's corresponding cubes has a probability of being subdivided again into 8 more sub-cubes. The probability of this happening is given by $N_{\rm Div}^{(3-D_{\rm f})}$ where $D_{\rm f}$ is the fractal



(a) $D_f = 1.6$

(b) $D_f = 3.0$

(c) $\alpha = 2.0$

(d) $\alpha = 0.0$

Figure 2.1: Examples of fractal and radially concentrated regions with 1000 stars with 1 pc radii. The top left panel shows a highly substructured region with a fractal dimension of $D_f = 1.6$, the top right panel shows a region with no substructure with fractal dimension $D_f = 3.0$. The bottom left panel shows a region that is centrally concentrated with a density index of $\alpha = 2.0$ and the bottom right panel shows a region with no central concentration with $\alpha = 0.0$. Panels (a) and (b) taken from Blaylock-Squibbs & Parker (2023).

dimension of the region. Stars that do not have their cubes subdivided are removed from the region along with any previous generations of stars that preceded them. A small amount of noise is added to each of the stars to stop them having a regular grid-like structure. The process of subdivision and adding new stars is done until the target number of stars is reached or exceeded in the most recent iteration of the method. It is this final generation of stars which become the $N_\star$ of the region. The method will often produce more stars than desired. Excess stars are removed at random until the target number of stars is reached (Daffern-Powell & Parker, 2020).

2.1.3 Uniformly Distributed Star-Forming Regions

In addition to the fractal and centrally concentrated synthetic regions I also generate uniform distributions. Whilst not physical these distributions are useful as a control distribution. Since they have no substructure, by implication any method that detects structure in them is unreliable. The uniform distribution is

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}, \tag{2.3}$$

where a and b are the lower and upper bounds, respectively. I pick x and y coordinates from a uniform distribution with lower and upper limits of -1 pc and 1 pc, respectively, to generate regions with an area of 4 pc^2.

2.1.4 Selecting Masses for Synthetic SFRs

I use the Maschberger IMF to assign masses to synthetic stars in this work (Maschberger, 2013). The Maschberger IMF's key advantage is its ease of use, in that it describes the IMF with a single function.

The probability distribution is as follows,

$$p(m) \propto \left(\frac{m}{\mu}\right)^{-\alpha} \left(1 + \left(\frac{m}{\mu}\right)^{1-\alpha}\right)^{-\beta}, \tag{2.4}$$

where μ is the mean stellar mass, m is the stellar mass, $\alpha = 2.3$ is the high mass index and $\beta = 1.4$ is the low mass index for the power law. These indices where chosen to match the canonical IMF from Kroupa (2008).

The function to generate stellar masses is called the quantile function and is

$$m(u) = \mu \left(\left[u\Big(G(m_u) - G(m_l)\Big) + G(m_l) \right]^{\frac{1}{1-\beta}} - 1 \right)^{\frac{1}{1-\alpha}}, \tag{2.5}$$

where m_l is the lower mass and m_u is the upper mass. For N-body simulations I use a lower mass of 0.1 $M_\odot$, an upper mass of 50 $M_\odot$ and a mean mass of 0.2 $M_\odot$. In Chapter 3 I use static regions and use a lower mass of $0.01 M_\odot$, upper mass of $150 M_\odot$ and the same mean mass of $0.2 M_\odot$. $G(m)$ is called the auxiliary function and has the form

$$G(m) = \left(1 + \left(\frac{m}{\mu}\right)^{1-\alpha}\right)^{1-\beta}. \tag{2.6}$$

The quantile function takes a value u as the input which is randomly picked from a uniform probability distribution where $0 < u < 1$. This will output a mass that has a probability, u, of being picked.

2.2 Simulating Star-Forming Regions

The selection of initial conditions in simulations needs to be considered carefully to allow for meaningful comparison between the physics we think is occurring in our simulations and real physics we observe in SFRs. To this end I generate subvirial simulations to match observations that indicate prestellar cores may be subvirial with respect to one

another, and may then virialise to form bound smooth, centrally concentrated star clusters (Foster et al., 2015; Kuznetsova et al., 2015; Parker & Wright, 2018). I also run sets of simulations that are initially supervirial. This is done to mimic the observations that some young SFRs ($\sim 1-5$ Myr) are supervirial and therefore appear to be expanding (Bravi et al., 2018; Kuhn et al., 2019; Kounkel et al., 2022).

2.2.1 Picking Velocities for Different Regions

For synthetic regions to be used in N-body simulations each star must also have an initial velocity. The initial velocities of stars will have an effect on the evolution of the region as a whole (i.e. sub- or supervirial). In this section I discuss how velocities are chosen for each of the different synthetic regions I have used.

Initial Velocities for Smooth, Centrally Concentrated Regions

The velocities of stars in smooth, centrally concentrated regions are picked from a Gaussian distribution with mean 0 $\mathrm{km\,s^{-1}}$ and variance 1 $\mathrm{km\,s^{-1}}$. From observations, we know that some star-forming regions can have quite large velocity dispersions of $\sim 3-4\,\mathrm{km\,s^{-1}}$ like in the ONC, whereas in some regions they can be as low as $\sim 0.6\,\mathrm{km\,s^{-1}}$ like in IC 348 (Tobin et al., 2009; Cottaar et al., 2015). Regardless these initially picked velocities are scaled to ensure that the regions have the chosen virial ratio (see §2.2.1).

Initial Velocities for Substructured Regions

For substructured regions the initial velocities of the parent stars are picked from a Gaussian distribution with a mean of 0 $\mathrm{km\,s^{-1}}$ and variance 1 $\mathrm{km\,s^{-1}}$. Subsequent child stars in the box-fractal method will inherit their parents' velocity along with a random velocity component drawn from the same Gaussian. The random velocity component is then multiplied by a factor of $\left(\frac{1}{N_{\mathrm{Div}}}\right)^{g}$,where g is the current generation of stars. Because

of this term the noise component added to subsequent generations of stars decreases with increasing g (Daffern-Powell & Parker, 2020).

The noise term scales with $N_{\rm Div}^{D_{\rm f}-3}$ which results in stars in proximity to each other being able to have similar velocities, and stars far apart from one another having different velocities. This is done to better match the observed velocity-scale relationship, $v(L) \propto L^{0.38}$ (Larson, 1981). The box-fractal velocity-scale relation is described as $v(L) \propto L^{3-D_{\rm f}}$ where $D_{\rm f}$ is the desired fractal dimension. From this expression we can see that for highly substructured regions with $D_{\rm f} = 1.6$ the velocity scaling relation will depart from the observed one from Larson (1981), with $v(L) \propto L^{1.4}$. For regions with less substructure (i.e. $D_{\rm f} = 2.6$) the scaling relation is $v(L) \propto L^{0.4}$ which is much closer to the observed relation (Parker & Wright, 2018).

Scaling Velocities to Get the Desired Virial Ratio

To get SFRs with different initial virial states the velocities of all stars are scaled using the virial ratio defined as

$$\alpha_{\rm vir} = \frac{\mathcal{K}}{|\Omega_{\rm G}|}, \tag{2.7}$$

where $\mathcal{K}$ is the total kinetic energy of the region and $\Omega_{\rm G}$ is the total potential energy of the region.

In Chapter 4 for subvirial simulations I use a virial ratio of $\alpha_{\rm vir} = 0.1$ and for supervirial simulations I use a virial ratio of $\alpha_{\rm vir} = 0.9$. A virial ratio of $\alpha_{\rm vir} = 0.5$ corresponds to the simulation being in virial equilibrium.

2.2.2 Two-Body Problem

Before discussing the integration scheme used in this work I will discuss using numerical methods to solve the two-body problem. The two-body problem can be solved analytically (see Foong (2008)), but for problems with $N \geq 3$ the use of numerical methods

is required. The methods used here can be expanded to solve N-body simulations, but they will be very time inefficient. To calculate the gravitational interactions between two bodies (with masses m_1 and m_2, respectively) and their resulting motions involves using Newton's second and third laws. The force the two bodies exert on each other is

$$\vec{F}_{1,2} = -G\frac{m_1 m_2}{|\vec{R}|^2}\,\hat{R}, \tag{2.8}$$

where $\vec{F}_{1,2}$ is the force acting on 1 due to 2, $|\vec{R}|$ is the magnitude of the vector pointing from 1 to 2 and $\hat{R}$ is the unit vector showing the direction the force is acting. Using Newton's third law we can state

$$\vec{F}_{1,2} = -\vec{F}_{2,1}. \tag{2.9}$$

A time step is defined which is dt. The time of the next time step is therefore

$$t_{\mathrm{n+1}} = t_{\mathrm{n}} + dt, \tag{2.10}$$

where t_{n} is the time of the current snapshot.

To find the positions of the two bodies at time $t_{\mathrm{n+1}}$ firstly the momentum of the bodies in the next snapshot needs to be calculated using

$$\vec{p}_{i,\mathrm{t_{n+1}}} = \vec{p}_{i,\mathrm{t_n}} + \vec{F}_{i,t_{\mathrm{n}}}\,dt, \tag{2.11}$$

where $\vec{p}_i = m_i\vec{v}_i$ is the momentum of body i. The position of the two bodies is then calculated using

$$\vec{r}_{i,t_{\mathrm{n+1}}} = \vec{r}_{i,t_{\mathrm{n}}} + \frac{\vec{p}_{i,t_{\mathrm{n+1}}}}{m_i}\,dt. \tag{2.12}$$

The choice of dt will affect the accuracy of the simulation and the time it will take to complete a run of a simulation. These steps are repeated until the system has evolved to the desired time.

2.2.3 N-Body Simulations: `Starlab`

N-body simulations allow us witness the dynamics of SFRs, which take place over millions of years, over the course of a few tens of seconds. The N-body simulations used in this work are run using the `Kira` integrator, one part of the `Starlab` [1] package (Zwart et al., 1999; Portegies Zwart et al., 2001). The other component of `Starlab` is `SeBa`, the stellar and binary evolution package (Portegies Zwart et al., 2001).

Each simulation has a population of 1000 stars and is run for a simulated time of 10 Myr as bound SFRs older than this are rare. My choice of 1000 stars comes from Lada & Lada (2003) in which they find the following relation between the number of clusters and the mass interval of those clusters,

$$N_{\rm cl} \propto M_{\rm cl}^{-2}, \tag{2.13}$$

where $N_{\rm cl}$ is the number of clusters and $M_{\rm cl}$ is the mass of the cluster. This power-law is obeyed for star clusters between masses of $10 < M_{\rm cl}/M_{\odot} < 10^5$ and so my choice of 1000 stars puts the simulations close to the middle of this distribution. To generate the initial snapshot that `Kira` can integrate I make use of `initials`, a program that takes user input about the desired cluster, such as the number of stars, degree of substructure, IMF and binary population. The initial positions and velocities of stars in the generated cluster are in units of pc and km s^{-1}, respectively. `initials` will convert these to N-body units (see §2.2.3) to improve efficiency of the simulation and allow for greater accuracy. `Kira` will simulate this initial distribution of stars for the desired amount of time of 10 Myr.

An N-body simulation that directly calculates the forces between all stars will have quadratic time complexity O(N^2), meaning the time taken to complete each simulation will scale with the square of the number of stars in the simulation. This will make

[1] https://www.sns.ias.edu/~starlab/index.html

simulations of massive clusters with large numbers of stars impractical without specialist hardware (i.e. graphical processing units) to reduce the time taken. However, there are methods to increase efficiency and decrease the amount of time it takes to run simulations, allowing them to be run on moderately powerful workstations. The `Kira` integrator uses a predictor-evaluator-corrector (PEC) method with different time steps for each particle, chosen such that the time intervals, Δt_n, are all powers of two. This allows for blocks of many stars to be advanced simultaneously. The time step for a block is defined as

$$\Delta t_n = \frac{\Delta t_0}{2^n}, \tag{2.14}$$

where Δt_0 is the largest time step in the simulation at the current time, n is the time step "rung", with lower rungs having smaller time steps. Each star's n is determined using the predicted future positions of stars (see Equation 2.17 below). The block time step scheme schematic is shown in Figure 2.2. The initial value of Δt_0 is calculated using

$$\Delta t_0 = \sqrt{\eta \frac{|\vec{a}_{1,i}|\,|\ddot{\vec{a}}_{1,i}| + |\dot{\vec{a}}_{1,i}|^2}{|\dot{\vec{a}}_{1,i}||\dddot{\vec{a}}_{1,i}| + |\ddot{\vec{a}}_{1,i}|^2}} \tag{2.15}$$

where $\vec{a}, \dot{\vec{a}}, \ddot{\vec{a}}$ and $\dddot{\vec{a}}$ are the acceleration, the first time derivation of the acceleration (the jerk) and the second and third time derivatives. η is a parameter to control the accuracy of the simulations, nominally $\eta \approx 0.02$, giving a good balance between computation time and accuracy (Aarseth, 1985; Dehnen & Read, 2011).

To further improve time efficiency `Kira` makes use of a tree structure, where nodes in the tree represent either stars or the centres of mass of multiple systems. If stars move within a certain distance of each other they are treated as a single node and if they move apart again they are treated as two separate nodes. This distance is the close encounter distance and is

$$R_{\rm close} = \frac{r_{\rm vir}\,(m_1 + m_2)}{2M_{\rm tot}}, \tag{2.16}$$

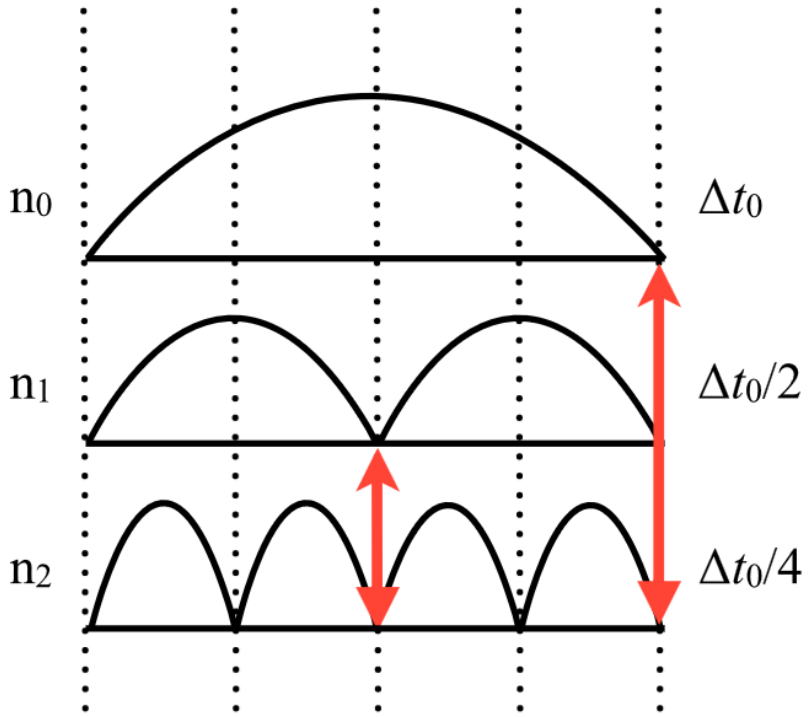


Figure 2.2: Schematic of a block time step scheme. Red arrows show the times when different rungs have the same time and are in sync, it is at these points that stars can move up and down rungs depending on the accuracy needed in their calculations. The further down the hierarchy the smaller the time step is, which results in more accurate calculations. Figure taken from Dehnen & Read (2011).

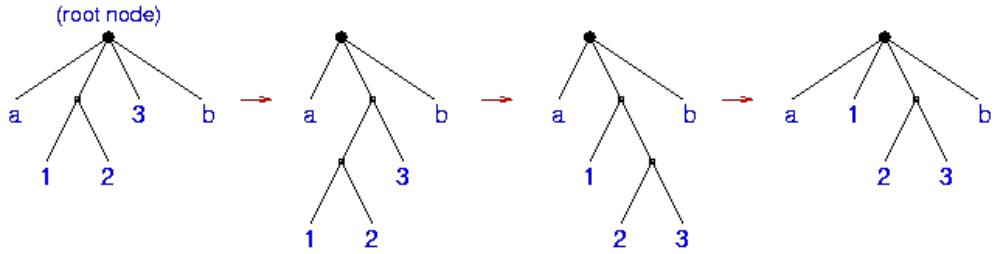


Figure 2.3: Diagram showing the tree structure that `Kira` utilises. The very top node (the root node) represents the entire simulation, the next layer down and below is where the direct force calculations are performed. Top nodes are either individual stars or higher order systems. This diagram shows a three body exchange interaction with time increasing from left to right. Bodies 1 and 2 are in a binary pair, approaching body 3. Body 3 becomes bound to 2 and as a result body 1 is ejected and is represented as a single node in the tree. Credit: `Starlab`, https://www.sns.ias.edu/~starlab/index.html

where m_1 and m_2 are the masses of the two stars, $r_{\rm vir}$ is the virial radius of the cluster and $M_{\rm tot}$ is the total mass of the cluster.

The motion of the nodes is calculated directly with respect to its parent node. The tree structure of a simulation is shown in Figure 2.3.

The integration is performed using a Hermite individual time step scheme (Makino, 1991). Each star in a simulation has the following parameters: its own time and its own time step, its position, present velocity and acceleration, and the time derivative of acceleration called the "jerk". The use of interpolation to predict the locations of stars in the next time step means fewer direct force calculations need to be done. Comparing to previous integration schemes (see Aarseth (1985)) the Hermite individual time step (HITS) schemes allow for longer time steps while maintaining the same accuracy (Makino & Aarseth, 1992).

The predicted position and velocities of the stars are calculated using

$$\vec{x}_{\rm p} = \vec{x}_0 + \dot{\vec{x}}_0 \Delta t + \frac{1}{2}\vec{a}_0 \Delta t^2 + \frac{1}{6}\dot{\vec{a}}_0 \Delta t^3, \tag{2.17}$$

and

$$\dot{\vec{x}}_{\rm p} = \dot{\vec{x}}_0 + \vec{a}_0 \Delta t + \frac{1}{2}\dot{\vec{a}}_0 \Delta t^2. \tag{2.18}$$

The subscript 0 in these equations represents that those quantities are in the current time step. Using the predicted positions and velocities we can then predict the associated acceleration, $\vec{a}$, of each star and its time derivative, $\dot{\vec{a}}$, using

$$\vec{a}_i = -G \sum_{j \neq i}^{N} m_j \frac{\vec{x}_i - \vec{x}_j}{|\vec{x}_i - \vec{x}_j|^3}, \tag{2.19}$$

and

$$\dot{\vec{a}}_i = -G \sum_{j \neq i}^{N} m_j \frac{\vec{x}_{ij}^2 \, \dot{\vec{x}}_{ij} - 3\vec{x}_{ij} \left(\vec{x}_{ij} \cdot \dot{\vec{x}}_{ij} \right)}{|\vec{x}_{ij}|^5}, \tag{2.20}$$

where $\vec{x}_{ij} = \vec{x}_i - \vec{x}_j$. The corrected position and velocity of the stars are then calculated in the next time step (represented using a subscript 1) using

$$\vec{x}_1 = \vec{x}_0 + \frac{1}{2}\left(\dot{\vec{x}}_1 + \dot{\vec{x}}_0 \right) \Delta t + \frac{1}{12} \left(\vec{a}_0 - \vec{a}_1 \right) \Delta t^2 + O\left(\Delta t^5 \right) \tag{2.21}$$

and

$$\dot{\vec{x}}_1 = \dot{\vec{x}}_0 + \frac{1}{2}\left(\vec{a}_1 + \vec{a}_0 \right) \Delta t + \frac{1}{12} \left(\dot{\vec{a}}_0 - \dot{\vec{a}}_1 \right) \Delta t^2 + O\left(\Delta t^5 \right), \tag{2.22}$$

where $O\left(\Delta t^5\right)$ is the error (Makino & Aarseth, 1992; Dehnen & Read, 2011).

In summary the algorithm proceeds as follows:

a) Determine which star has the smallest $t_i + \Delta t_i$ and set the global simulation time to this value.

b) Using Equations 2.17 and 2.18 predict the positions and velocities of all stars at the new global time.

c) Compute the new acceleration ($\vec{a}_i$) and its jerk ($\dot{\vec{a}}_i$) using the predicted positions and velocities.

d) Using the predicted positions, velocities, acceleration and jerk, find the corrected positions and velocities using Equations 2.21 and 2.22.

e) Go back to step a).

To get usable data from `Kira` the file it outputs is run through an extraction program. The extraction program allows the user to pick the number of desired snapshots from the simulation. For example, in my simulations I extract 100 snapshots over the 10 Myr simulated time resulting in the time between each snapshot being 0.1 Myr.

Calculations for Binary Systems

Binary systems (i.e. star-star or star-planet) are treated differently depending on the degree to which they are perturbed by external nearby stars for a particular time step. It is useful to define a hard- and soft-binary regime, where the binding energy of a binary (or higher order system) is compared to the typical energy of a star in the cluster. The gravitational binding energy of a binary system is

$$E_{\text{Binary}} = m_1\, m_2 \left(\frac{|\vec{v}|^2}{2\,(m_1 + m_2)} - \frac{G}{|\vec{R}|} \right), \tag{2.23}$$

where m_1 and m_2 are the primary and secondary stellar masses, respectively. $|\vec{v}|$ is the absolute magnitude difference between the two stars' velocities and $|\vec{R}|$ is the absolute magnitude distance between them.

The typical energy of a star in a cluster is described using a Maxwellian distribution

$$E_{\text{Maxwellian}} = \bar{m}\, \sigma_v^2, \tag{2.24}$$

where $\bar{m}$ is the mean stellar mass in the cluster and σ_v is the velocity dispersion. A

binary system is "soft" if the following is true

$$|E_{\text{Binary}}| \ll \bar{m}\,\sigma_v^2, \tag{2.25}$$

and as "hard" when

$$|E_{\text{Binary}}| \gg \bar{m}\,\sigma_v^2. \tag{2.26}$$

The behaviour of hard and soft binary systems can be summarised with the "Heggie-Hills" law, which states that hard binaries get harder and soft binaries get softer (Heggie, 1975; Hills, 1975). The degree of perturbation is quantified by comparing the ratio of the binding energy of the binary system with the typical kinetic energy of nearby stars that will perturb the system.

The `Kira` integrator solves unperturbed binary systems as a two-body problem, for moderately perturbed binary systems it will artificially slow the velocity components down and increase the perturbation (see Mikkola & Aarseth (1998)), the positions of the stars are then extrapolated forward in time to determine future positions of the two stars. Highly perturbed systems are decomposed into their constituent parts and solved directly (Portegies Zwart et al., 2001).

N-Body Units

N-body units are a system of rescaled physical units that allow for more accurate calculations without using more memory. `Kira` uses the *N*-body units scheme defined in Heggie & Mathieu (1986). A simulation has its units scaled so that the following is true

$$G = 1 \tag{2.27}$$

$$M = 1 \tag{2.28}$$

$$E = -\frac{1}{4} \tag{2.29}$$

where G is the gravitational constant, M is the total mass of the stars in the simulation and E is the total energy of the simulation. The value of E comes from the radius of the cluster in virial equilibrium being set to 1. Computers cannot represent decimal numbers exactly using floating point representation. For simple calculations this can be accounted for by simply rounding the value. However, when these calculations are done multiple times (like numerically integrating in an N-body simulation) with values that require large exponents to accurately describe them, the error can become significant. By using N-body units all the parameters of the stars are now all of order unity, which frees up memory previously used to store the exponent. This extra space can now be used to represent stellar parameters more accurately, resulting in a reduction of truncation and round-off errors. At the end of each calculation the N-body units are converted back into their original physical units (pc and $\mathrm{km\,s^{-1}}$).

2.3 Spatial and Kinematic Metrics

To better understand the star formation process we need to be able to make quantitive comparisons between observations and models. The focus of my research has been testing new methods and assessing their ability at determining the initial conditions and morphologies of simulated SFRs. Before detailing the methods I have used in my work I will provide some background on earlier methods to help contextualise the work I will present in the following sections.

Star formation is a rapid process, occurring within a few crossing times (Elmegreen, 2000), which is often less than 1 Myr. During this process, stars are forming and moving (Alcock & Parker, 2019), further muddying the formation picture. And whilst observations of the earliest stages have improved greatly with e.g. ALMA, observations of

star-forming regions are often at older ages, where significant dynamical evolution may have taken place (Klessen & Kroupa, 2001; Allison et al., 2010; Parker et al., 2014; Schoettler et al., 2019; Daffern-Powell & Parker, 2020). Dynamical evolution alters the spatial and kinematic distributions of young stars, erasing the signature of the initial conditions, but can be used as a proxy for age and used to converge on a likely set of initial conditions for a given SFR (Parker, 2014). To enable comparisons of observations and simulations, methods are needed that allow different parameters of SFRs to be quantified, such as the degree of substructure and mass segregation (Cartwright & Whitworth, 2004; Allison et al., 2009; Sánchez & Alfaro, 2009; Alfaro & González, 2016; González & Alfaro, 2017; Jaffa et al., 2017; Buckner et al., 2019; Arnold et al., 2022; Joncour et al., 2018; Kuhn et al., 2014; Gouliermis et al., 2014).

Early methods such as the auto-correlation function and two-point correlation function compared the number of excess star pairings to a random distribution of stars as a function of scale (Gomez et al., 1993; Larson, 1995). These methods where used extensively to determine the degree of substructure, with early work suggesting breaks in the two-point correlation function corresponded to the Jeans length (Simon, 1997) (though see Bate et al. (1998)) and the size of the widest stellar binaries in the regions in question (Kraus & Hillenbrand, 2008; Joncour et al., 2017).

Subsequent work made extensive use of minimum spanning trees (MSTs) to quantify structures in star-forming regions. An MST is a graph of points connected to each other in such a way that the total edge length of the graph is minimised and that all points are connected to at least one other point with no closed loops. Cartwright & Whitworth (2004) introduced the Q-parameter to quantify spatial substructure, and Allison et al. (2009) introduced the $\Lambda_{\rm MSR}$ (mass segregation ratio) method to quantify mass segregation.

Parker (2014) showed that the initial conditions of a region can be inferred from the spatial information, if a suitable number of metrics are combined, including the

relative stellar surface density around the most massive stars (Maschberger & Clarke, 2011; Küpper et al., 2011).

The majority of the following methods are designed to operate on two or three-dimensional spatial data, whereas recent observational data (e.g. from Gaia and associated ground-based surveys) has provided high resolution spatial and kinematic data, which will be six-dimensional.

2.3.1 The Mass Segregation Ratio - $\Lambda_{\rm MSR}$

The mass segregation ratio $\Lambda_{\rm MSR}$ was first introduced in Allison et al. (2009) to quantify the degree of mass segregation in a star-forming region. The definition of mass segregation in this case is that the most massive stars are closer to each other than expected from the average separation of all of the stars in the region.

This method generates an MST for the chosen subset of stars; frequently this subset consists of the 10 most massive stars. It will then pick 10 random stars from the region and make an MST for these random stars. This is done 200 times to calculate the mean edge length of the randomly chosen trees. The ratio is calculated using the following equation,

$$\Lambda_{\rm MSR} = \frac{\langle l_{\rm average} \rangle}{l_{10}}^{+\sigma_{5/6}/l_{10}}_{-\sigma_{1/6}/l_{10}}, \tag{2.30}$$

where $\langle l_{\rm average} \rangle$ is the average edge length found for all the randomly constructed MSTs and l_{10} is the edge length of the subset's MST. It is important to note that the random MSTs can also contain members of the chosen subset.

If the ratio is > 1 then the region's 10 most massive stars are mass segregated, if the ratio is ~ 1 then the most massive stars are not mass segregated and if the ratio is less than 1 they are inversely mass segregated (the most massive stars are further apart than the average stars in the region). In this work I mark the value 1 to show the boundary between mass segregation and inverse mass segregation. Following Parker & Goodwin

(2015) I only take ratios above 2 to signify mass segregation, to avoid false positives.

Following Parker (2018) I calculate the uncertainty using the randomly constructed MSTs. Firstly, I order the lengths of the random MSTs and find the values that lie 1/6 and 5/6 of the way through this list. This gives the values which correspond to a 66 per cent deviation from the median MST length.

2.3.2 Q-Parameter

The Q-Parameter was introduced in Cartwright & Whitworth (2004) to quantify and distinguish between different cluster morphologies. The Q-parameter also makes use of MSTs and proceeds as follows. Firstly, the normalised correlation length is found. This is the mean separation between all stars in a region which is then divided by the region's radius to normalise it.

The mean edge length of the region is found by constructing an MST for the region and then finding the mean edge length. The mean edge length is normalised by dividing it by $\frac{\sqrt{N_{\rm total}A}}{N_{\rm total}-1}$, where $N_{\rm total}$ is the number of stars in the region and A is the area of the region. I use the circular area (see Schmeja & Klessen (2006) and Parker (2018) for a discussions on normalisation techniques), with the radius defined as the distance from the centre of mass to the most distant star. The Q-parameter is then defined as

$$Q = \frac{\bar{m}}{\bar{s}}, \tag{2.31}$$

where $\bar{m}$ is the normalised mean edge length of the MST and $\bar{s}$ is the normalised correlation length between stars. Regions with substructured morphologies have $Q < 0.8$ whereas regions with smooth, centrally concentrated morphologies have $Q > 0.8$.

2.3.3 INDICATE

INDICATE is a tool introduced by Buckner et al. (2019) to quantify the clustering tendencies of individual stars, where each star is assigned its own index from which a star's spatial distribution is determined to be random or spatially clustered. The higher this index, the greater the affiliation, or clustering of a star to others. Previous methods of quantifying structure, such as the Q-parameter (e.g. Cartwright & Whitworth, 2004) involve calculating a value for the entire region which quantifies the amount of substructure present.

The INDICATE algorithm proceeds as follows. Firstly, an evenly spaced control field is generated with the same number density as the dataset. For each point, j, in the dataset, the Euclidean distance to the $N^{\rm th}$ nearest neighbour in the control field is measured, and then the mean of those distances, $\bar{r}$, calculated. Then for j, the algorithm counts how many other points from the dataset are within a radius of $\bar{r}$. The index for point j is the defined as

$$I_j = \frac{N_{\bar{\rm r}}}{N}, \tag{2.32}$$

where I_j is the (unit-less) index, $N_{\bar{\rm r}}$ is the number of points within $\bar{r}$ of point j and N is the nearest neighbour number (i.e. if $N = 5$ I measure the distance from data point j to its $5^{\rm th}$ closest neighbour in the control field). The index is independent of the shape, size and density of the dataset (Buckner et al., 2019).

To determine the index above which a star's spatial distribution is not random and therefore clustered INDICATE is applied to a uniformly distributed set of points (see Appendix 3.6), with the same number density as the dataset, and using the same control field.

In Buckner et al. (2020) this is repeated 100 times to remove any statistical fluctuations. In this work I present the results for running this once but I also show results for 100 repeats in Appendix 3.7. For small sample sizes of 50 stars I run 100 repeats as

these small samples are subject to fluctuations that can make a significant difference to the number of stars with indexes above the significant index.

The significant index is defined as $I_{\rm sig} = \bar{I} + 3\sigma$ where $\bar{I}$ is the mean index of the uniform distributions and σ is the standard deviation of mean indexes. Any star in the dataset with an index greater than $I_{\rm sig}$ is clustered above random.

To ensure that the correct distance to the nearest neighbour is found for points on the outskirts of the regions the control grid is extended beyond the dataset. If the control grid is not extended then edge effects can make a very small change to the index of those stars (see Appendix B in Buckner et al. (2019)).

2.3.4 The Local Surface Density Ratio - $\Sigma_{\rm LDR}$

The local surface density ratio $\Sigma_{\rm LDR}$ was introduced in Maschberger & Clarke (2011) to quantify the differences between the surface densities of subsets of stars within groups of stars and for this work I choose the 10 most massive stars as the subset of interest. The algorithm proceeds as follows. Firstly, for each star the distance to its $N^{\rm th}$ nearest neighbour is calculated, then calculate the area of a circle with a radius equal to the $N^{\rm th}$ nearest neighbour distance. To find the surface density of the stars the nearest neighbour number is divided by the area of the circle. In this work a nearest neighbour number of 5 is used. The ratio is defined as

$$\Sigma_{\rm LDR} = \frac{\tilde{\Sigma}_{\rm subset}}{\tilde{\Sigma}_{\rm all}}, \tag{2.33}$$

where $\tilde{\Sigma}_{\rm subset}$ is the median surface density found for the 10 most massive stars and $\tilde{\Sigma}_{\rm all}$ is the median surface density found for the entire region. Therefore, if $\Sigma_{\rm LDR} > 1$ the 10 most massive stars are found in areas of higher than average stellar surface density, and conversely, if $\Sigma_{\rm LDR} < 1$ then they are located in areas of lower than average surface density. The significance of any difference is quantified using a two-sample Kolmogorov-

Smirnov test. Where if $p << 0.01$ I reject the null hypothesis that the 10 most massive stars share the same underlying distribution of surface densities compared to the entire region.

2.3.5 The Mahalanobis Distance

The Mahalanobis distance is a metric that measures the distances of points in a distribution to the mean of that distribution where the distribution exists in an arbitrary number of phase space dimensions (Mahalanobis, 1936).

The Mahalanobis distance does this by removing any correlations in the data by multiplying the distances between points and the average of the region by the inverse of the covariance matrix; this also has the effect of re-scaling the data. Once this re-scaling has been done the Euclidean distances are found in this rescaled phase space; this is the Mahalanobis distance, $M_{\rm d}$.

Each point in a dataset is described using a vector where each element is a measured parameter of that point,

$$\vec{x} = (x_1, x_2, x_3, ..., x_{\rm N})^{\rm T}, \tag{2.34}$$

where $x_1, x_2, x_3, ..., x_n$ are the parameters. For example, if each point has the three parameters, (x, y, z) then this is simply its physical position in three-dimensional (3D) space.

The Mahalanobis distance between a point in a distribution and the mean of that distribution in an N dimensional phase space is defined as

$$M_{\rm d}(\vec{x}, \vec{\mu}) = \sqrt{(\vec{x} - \vec{\mu})^{\rm T}\, \mathbf{S}^{-1}\, (\vec{x} - \vec{\mu})}, \tag{2.35}$$

where the $\vec{x}$ is the vector of a point describing all its parameters, $\vec{\mu}$ is a vector of the averages of the parameters of interest and $\mathbf{S}^{-1}$ is the inverse of the covariance matrix for

all the parameters in the region.

Mahalanobis Density

The Mahalanobis distance, $M_{\rm d}$ has been also used to define a phase space density called the Mahalanobis density, $\rho_{\rm m,N}$ (Winter et al., 2020)[2].

To calculate the Mahalanobis density, firstly, calculate the Mahalanobis distance between points in the phase space (i.e. distance between $\vec{x}$ and $\vec{y}$). Following Winter et al. (2020) I calculate the Mahalanobis distance between two points using

$$m_{\rm d}(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^{\rm T}\,\mathbf{S}^{-1}\,(\vec{x}-\vec{y})}, \tag{2.36}$$

where $\vec{x}$ is the vector describing the measurements of one point, $\vec{y}$ is the vector describing the measurements of another and $\mathbf{S}^{-1}$ is the inverse of the covariance matrix of all the parameters.

The calculation of the Mahalanobis density proceeds as follows. Firstly, find the Mahalanobis distance to the N^{th} nearest neighbour, then divide the nearest neighbour number by the volume whose side length is defined as the Mahalanobis distance to the N^{th} nearest neighbour. The Mahalanobis density is then defined as

$$\rho_{\rm m,N} = N m_{\rm d,N}^{-D_{\rm p}}, \tag{2.37}$$

where $\rho_{\rm m,N}$ is the Mahalanobis density, N is the nearest neighbour number, $m_{\rm d,N}$ is the Mahalanobis distance to the N^{th} nearest neighbour and $D_{\rm p}$ is the number of dimensions in the phase space (Winter et al., 2020). Once the Mahalanobis densities have been calculated they are then normalised so that the median Mahalanobis density is unity.

[2]I have used different notation for the Mahalanobis distance to avoid confusion with the fractal dimension and also the number of dimensions. I instead use $M_{\rm d}$ instead of D as used in Winter et al. (2020) to avoid confusion with the fractal dimension. The number of dimensions in the phase space has been changed from D to $D_{\rm p}$ to avoid confusion with the fractal dimension, which is now represented as $D_{\rm f}$.

In this work I apply the Mahalanobis density to two different phase spaces, the positional phase space (3D) and the position-velocity phase space (6D). For this work I find the Mahalanobis distance to the 20^{th} nearest neighbour in the phase space (i.e. $N = 20$), the same as in Winter et al. (2020).

3

Investigating the Structure of Star-Forming Regions Using INDICATE

Summary

The ability to make meaningful comparisons between theoretical and observational data of star-forming regions (SFRs) is key to understanding the star formation process. In this chapter I test the performance of INDICATE, a method developed by Buckner et al. (2019) to quantify the clustering tendencies of individual stars in a region, on synthetic SFRs with sub-structured, and smooth, centrally concentrated distributions. INDICATE works by quantifying the degree of stellar affiliation of each individual star, and also determines whether this is above random expectation for the SFR in question.

INDICATE has been used to infer the star formation history of NGC 3372 and to investigate the spatial distribution of the most massive stars (Buckner et al., 2019). In this chapter I compare INDICATE to current methods in the literature that quantify both the clustering and mass segregation of stars. In this chapter I perform and show the results of these tests, and conclude that INDICATE is a robust clustering metric that can be reliably used to quantify the clustering of SFRs.

I show that INDICATE cannot be used to quantify the overall structure of a region due to a degeneracy when applied to regions with different morphologies. I also test the ability of INDICATE to detect differences in the local stellar surface density and its ability to detect and quantify mass segregation. I then compare it to other methods such as the mass segregation ratio $\Lambda_{\rm MSR}$, the local stellar surface density ratio $\Sigma_{\rm LDR}$ and the cumulative distribution of stellar positions. INDICATE detects significant differences in the clustering tendencies of the most massive stars when they are at the centre of a smooth, centrally concentrated distribution, corresponding to areas of greater stellar surface density. When applied to substructured regions INDICATE finds significant differences between the clustering of the most massive stars when they are located in areas of greatest clustering, and when the centre of the regions corresponds to an area of higher than average clustering. When applied to a subset of the 50 most massive stars

I show INDICATE can detect signals of mass segregation. I also apply INDICATE to the following nearby SFRs: Taurus, ONC, NGC 1333, IC 348 and ρ Ophiuchi and find a diverse range of clustering tendencies in these regions.

The work in this chapter shows that INDICATE is a robust clustering metric that gives results generally in agreement with established methods, such as the $\Lambda_{\rm MSR}$ for mass segregation and $\Sigma_{\rm LDR}$ for local surface densities of the most massive stars. The consistency of INDICATE with the established methods lends credence to the physics of SFRs inferred by using it.

3.1 Introduction

Star formation is observed to occur in giant molecular clouds where the stars form in embedded groups (Lada & Lada, 2003). These embedded groups are often part of SFRs that have a range of different morphologies (i.e. smooth centrally concentrated spherical distributions or more complex substructured distributions) and densities (Bressert et al., 2010; Kruijssen, 2012). Quantifying the amount of spatial (and kinematic) substructure is key to determining whether star formation is universal (i.e. the same everywhere) or whether it is dependent on local environmental factors.

Young SFRs are often observed to be substructured and subvirial, but this substructure can be erased over a very short time period (the order of a few crossing times within substructured regions) due to dynamical interactions. These interactions lead to dynamical mass segregation (e.g. McMillan et al., 2007; Allison et al., 2009, 2010; Moeckel & Bonnell, 2009; Parker et al., 2013; Domínguez et al., 2017) where the most massive stars migrate to the centre of the region on the order of a crossing timescale (Bonnell & Davies, 1998). This has an important implication as the observed locations of the most massive stars are not necessarily where they formed.

INDICATE is a new method proposed in Buckner et al. (2019) to quantify the clus-

tering tendencies of points in a distribution (e.g. stars in SFRs), which they used to characterise the clustering behaviour of stars in the Carina Nebula (NGC 3372). The method was further employed in Buckner et al. (2020) to investigate the clustering tendencies of YSOs, and thus star formation history of NGC 2264 (see also Nony et al. (2021)).

3.2 Applying INDICATE to Synthetic Star-Forming Regions

Here I present the results of testing INDICATE on sets of synthetic SFRs of 1000 stars. I show three example regions to highlight the clustering differences between the morphologies tested. I show the results for a single fractal distribution of $D_{\rm f} = 1.6$, a single radial distribution with $\alpha = 2.0$ (see Figure 2.1) and a single uniform distribution, which acts as a control.

The creation of synthetic regions (how positions and masses are picked) is described in Chapter 2 §2.1. In this chapter the lower mass limit used in the synthetic clusters is $0.01\,M_{\odot}$, the upper mass is $150\,M_{\odot}$ and a mean mass of $0.2\,M_{\odot}$.

3.2.1 Can INDICATE Determine the Structure of Star-Forming Regions?

To test INDICATE's ability to differentiate between SFRs with different initial morphologies I run it over 100 different synthetic SFRs of 1000 stars with fractal dimensions of $D_{\rm f} = 1.6, 2.0, 2.6$ and 3.0 and smooth, centrally concentrated distributions with radial density exponents $\alpha = 0.0$, 0.5, 1.0, 1.5, 2.0, 2.5 and 2.9.

INDICATE quantifies the association (clustering) of stars relative to other stars by making use of a control grid of evenly spaced points that has the same number density

as the data being investigated. For each star the distance to its 5^{th} nearest neighbour in the control grid is found. The mean 5^{th} nearest neighbour distance, $\bar{\text{r}}$, is found. Then, for each star count how many other stars are within this distance, $N_{\bar{\text{r}}}$ (see also Chapter 2 §2.3.3). The INDICATE index for each star is found by dividing the number of stars within $N_{\bar{\text{r}}}$ by the nearest neighbour number N.

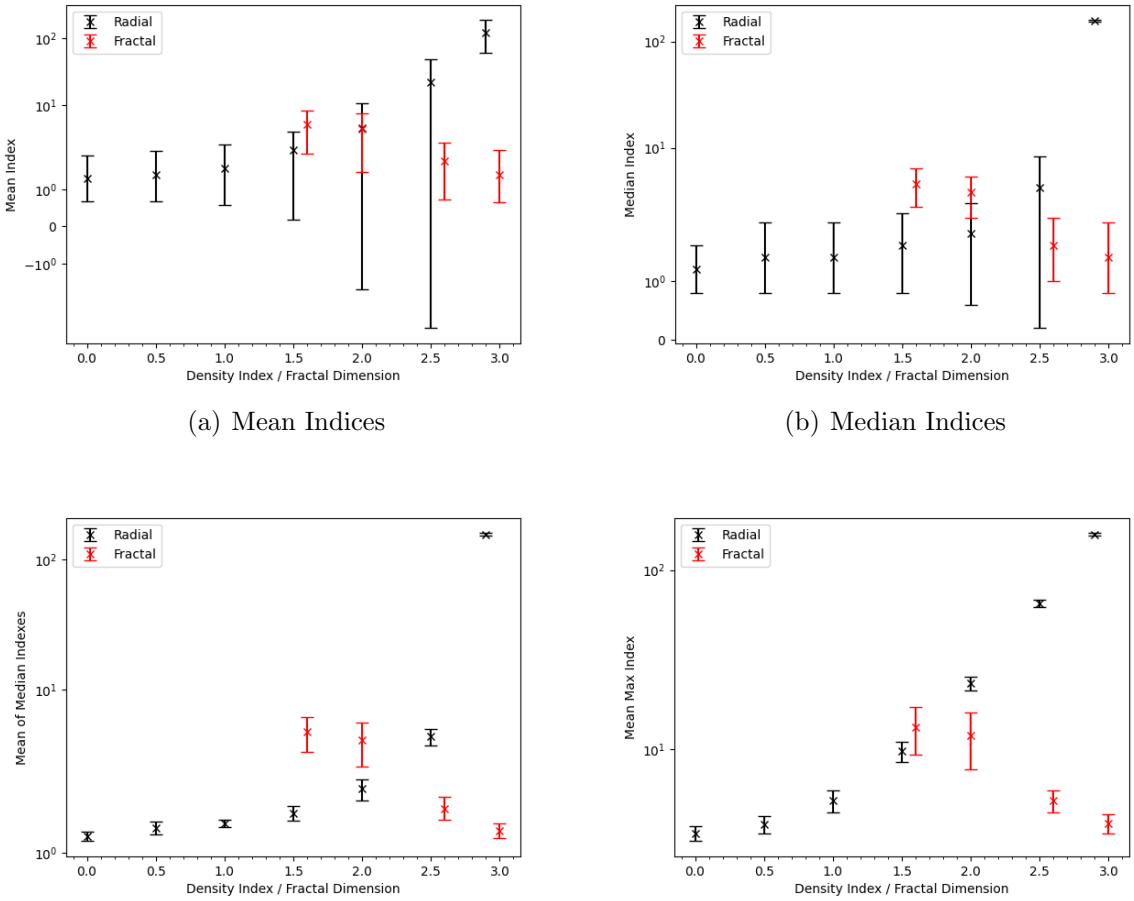
These results are shown in Figure 3.1, which shows the mean, median, mean median and mean maximum index for each set of these distributions. The mean and median INDICATE index is calculated by taking all the stars' indexes in the 100 different regions and finding the mean and median of them. The mean median is calculated by finding the median INDICATE index for each individual region then finding the mean of these 100 median values. The mean maximum INDICATE index is calculated by finding the maximum index is each individual region then finding the mean of these 100 values.

Figure 3.1 clearly shows that the index is degenerate because different smooth and fractal distributions can have the same index. For fractal distributions this degeneracy is present across all fractal dimensions and the index value overlaps with the centrally concentrated distributions with $\alpha < 2.0$. Because the index is similar across SFRs with different geometries and levels of substructure, INDICATE cannot be used to quantify the type of geometry in the same way that the Q-parameter can.

I present the index distributions for three realisations of synthetic regions with different degrees of substructure in § 3.8 and show that clustering tends to be higher for regions with more substructure.

3.2.2 Can INDICATE be used to quantify mass segregation?

To test the ability of INDICATE to detect and quantify if the most massive stars are in regions of localised above-average stellar surface density I apply it to all 1000 stars in the fractal, smooth centrally concentrated and uniform synthetic SFRs where the



(a) Mean Indices

(b) Median Indices

(c) Mean of Median Indices

(d) Mean Maximum Indices

Figure 3.1: INDICATE for regions with different spatial distributions. Panel (a) shows the mean indices found for 100 different realisations of ideal SFRs of differing fractal dimension and radial densities with the error bars representing the standard deviation of the mean indices. Panel (b) shows the median indices for the same regions and the error bars here represent the median absolute deviation. Panel (c) shows the mean of the median indices found for each of the 100 regions with the error bars representing the standard deviation of the mean. In panel (d) the mean maximum indices are shown with the error bars being the standard deviation of this value.

masses are randomly assigned to stars using the IMF from Maschberger (2013). The mass configurations of these regions are then changed by swapping the 10 most massive stars with the 10 stars of highest INDICATE index and by swapping the 10 most massive stars with the 10 most central stars (i.e. 10 stars closest to the origin (0,0)). Comparing the median INDICATE index of all the stars in the region with the median index of the 10 most massive stars allows the relative clustering tendencies of the most massive stars to be determined. If the median index of the most massive stars is greater than the median index for the entire region then the most massive stars are more clustered than the typical star in a region, and as a consequence of this are found in locations of higher than average local stellar surface density. If the opposite is true then the most massive stars are less clustered than the typical star and in areas of lower than average local stellar surface density. To determine if any detected difference in the spatial distribution, according to INDICATE, of the most massive stars in these tests is significant I employ a two sample KS test with a significance threshold of 0.01, below which I reject the null hypothesis that the 10 most massive stars and the entire population of stars are spatially distributed the same way. If the p-value $\ll 0.01$ then there is a significant difference in the index distributions (and therefore clustering tendencies) of the 10 most massive stars compared to the entire region.

The criteria for INDICATE to detect mass segregation is from Buckner et al. (2019) and requires that the 10 most massive stars are non-randomly clustered with respect to a subset of the 50 most massive stars (i.e. $\tilde{I}_{10} > I_{\rm sig}$). To see if the 10 most massive star are spatially distributed differently than the entire subset a two sample KS test with significance threshold of 0.01 is used. The INDICATE results for the 50 most massive stars in each of example synthetic regions is shown in Table 3.1. I also show the results of measuring the mass segregation in 100 realisations of each cluster and mass configuration in Table 3.2 which shows that for most synthetic regions that INDICATE detects mass segregation in, so does $\Lambda_{\rm MSR}$.

Table 3.1: Results of applying INDICATE to just the 50 most massive stars in the synthetic SFRs. From left to right the columns are: the median index for all 50 stars, the median index for the 10 most massive stars, the significant index, the percentage of stars with indexes greater than the significant index and the p-value from a KS test between all 50 stars and the 10 most massive stars.

Region	$\tilde{I}_{50}$	$\tilde{I}_{10}$	$I_{\rm sig}$	$\% > I_{\rm sig}$	p
$D = 1.6$, m	$1.4^{+0.6}_{-0.6}$	$1.6^{+0.4}_{-0.8}$	2.1	22	1.00
$D = 1.6$, hmhi	$1.3^{+2.3}_{-0.5}$	$3.6^{+0.0}_{-0.0}$	2.1	38	$\ll 0.01$
$D = 1.6$, hmc	$1.4^{+0.8}_{-0.4}$	$2.6^{+0.0}_{-0.2}$	2.1	28	$\ll 0.01$
$\alpha = 2.0$, m	$2.6^{+1.8}_{-2.0}$	$1.1^{+2.0}_{-0.7}$	2.1	52	0.48
$\alpha = 2.0$, hmhi	$5.7^{+0.3}_{-5.0}$	$6.0^{+0.2}_{-0.0}$	2.1	64	$\ll 0.01$
$\alpha = 2.0$, hmc	$5.6^{+0.4}_{-4.9}$	$6.0^{+0.0}_{-0.0}$	2.1	62	$\ll 0.01$
Uniform, m	$0.8^{+0.2}_{-0.4}$	$0.8^{+0.3}_{-0.2}$	2.1	0	1.00
Uniform, hmhi	$0.7^{+0.5}_{-0.3}$	$1.1^{+0.1}_{-0.4}$	2.1	0	0.67
Uniform, hmc	$0.8^{+0.4}_{-0.4}$	$2.0^{+0.4}_{-0.4}$	2.1	10	0.01

Table 3.2: INDICATE results when applied to only the 50 most massive stars across 100 realisations of each morphology. From left to right the columns are: the median of the median indexes found for all 50 stars across all 100 realisations, the number of times the median index for the 50 most massive stars is above the significant index in a realisation, the median of the median index of the 10 most massive stars found across all regions, the number of times the median index for the 10 most massive stars is greater than the significant index for a realisation, the number of times that $\Lambda_{\rm MSR}$ detects mass segregation in the realisations that INDICATE has detected mass segregation, the median of the median indexes found for 10 randomly chosen stars across all regions, the number of times the median index of a realisation is greater than the significant index.

Region	$\tilde{I}_{50}$	$\# > I_{\rm sig}$	$\tilde{I}_{10,\rm mm}$	$\# > I_{\rm sig}$	# MS	$\tilde{I}_{10,\rm ran}$	$\# > I_{\rm sig}$
$D = 1.6$, m	1.5	12	1.5	14	0	1.5	16
$D = 1.6$, hmhi	1.7	28	3.6	94	90	1.7	35
$D = 1.6$, hmc	1.6	23	3.0	95	95	1.8	29
$\alpha = 2.0$, m	1.4	14	1.5	29	1	1.6	30
$\alpha = 2.0$, hmhi	2.6	54	4.8	100	100	2.7	65
$\alpha = 2.0$, hmc	2.6	57	4.8	100	100	2.8	62
Uniform, m	0.8	0	0.7	0	0	0.7	0
Uniform, hmhi	0.8	0	1.2	15	11	0.8	0
Uniform, hmc	0.8	0	2.4	87	87	0.8	3

Table 3.3: INDICATE results of 100 different realisations for each of the presented morphologies. From left to right the columns are: the median of the median indexes found across all 100 realisations of clusters with different morphologies and mass configurations, the number of times a realisation's median index for all stars is above its significant index, the median of the median indexes of the 10 most massive stars, the number of times a realisation's median index for the 10 most massive stars is above its significant index, the median of the median index for 10 randomly chosen stars and the number of times a realisation's median index for 10 random stars is above its significant index.

Region	$\tilde{I}_{\rm all}$	$\# > I_{\rm sig}$	$\tilde{I}_{\rm 10,mm}$	$\# > I_{\rm sig}$	$\tilde{I}_{\rm 10,ran}$	$\# > I_{\rm sig}$
$D = 1.6$, m	4.4	100	4.4	97	4.5	98
$D = 1.6$, hmhi	4.4	100	11.8	100	4.6	98
$D = 1.6$, hmc	4.4	100	3.5	84	4.5	97
$\alpha = 2.0$, m	1.8	2	2.0	38	2.0	41
$\alpha = 2.0$, hmhi	1.8	2	22.7	100	2.0	32
$\alpha = 2.0$, hmc	1.8	2	22.2	100	2.1	32
Uniform, m	1.0	0	0.9	0	0.9	0
Uniform, hmhi	1.0	0	2.2	34	0.9	0
Uniform, hmc	1.0	0	0.9	0	0.9	0

Table 3.3 shows how many times across 100 realisations for a substructured region ($D = 1.6$), a smooth, centrally concentrated region ($\alpha = 2.0$) and a uniform distribution the median index of the entire region of 1000 stars, the 10 most massive stars and 10 random stars are above the significant index. Table 3.4 shows the results of applying $\Lambda_{\rm MSR}$ and $\Sigma_{\rm LDR}$ to the different sets. It clearly shows that when the most massive stars are swapped with the most clustered stars according to INDICATE both $\Lambda_{\rm MSR}$ and $\Sigma_{\rm LDR}$ in the majority of cases detect mass segregation or that the 10 most massive stars are in areas of higher than average surface density, respectively.

For all of these tests the $\Sigma_{\rm LDR}$, $\Lambda_{\rm MSR}$ and cumulative distribution of stellar position (CDF) methods are also applied for comparison. I also employ the two sample KS test to determine if any $\Sigma_{\rm LDR}$ or CDF results are significantly different between the 10 most massive stars and the entire population in a region. I use the same threshold value of 0.01, below which the null hypothesis is rejected that the 10 most massive stars and all stars in a region are distributed the same way. I show INDICATE plots for the 50 most

Table 3.4: Results of applying $\Lambda_{\rm MSR}$ and $\Sigma_{\rm LDR}$ to all 1000 stars in 100 different realisations of each morphology and mass configuration. From left to the right the columns are the number of times $\Lambda_{\rm MSR} < 0.5$ (which would indicate significant inverse mass segregation, as is observed in the Taurus star-forming region), the number of times $\Lambda_{\rm MSR} > 2$ which counts how many times $\Lambda_{\rm MSR}$ detects strong signals of mass segregation and the number of times the ratio $\Sigma_{\rm LDR}$ is found to be > 1 and significant according to a KS test with a threshold p-value $\ll 0.01$.

Region	$\#\Lambda_{\rm MSR} < 0.5$	$\#\Lambda_{\rm MSR} > 2$	$\#\Sigma_{\rm LDR,Sig} > 1$
D=1.6, m	0	0	1
D=1.6, hmhi	0	94	71
D=1.6, hmc	0	100	6
$\alpha = 2.0$, m	0	1	2
$\alpha = 2.0$, hmhi	0	100	100
$\alpha = 2.0$, hmc	0	100	100
Uniform, m	0	0	0
Uniform, hmhi	0	34	100
Uniform, hmc	0	100	9

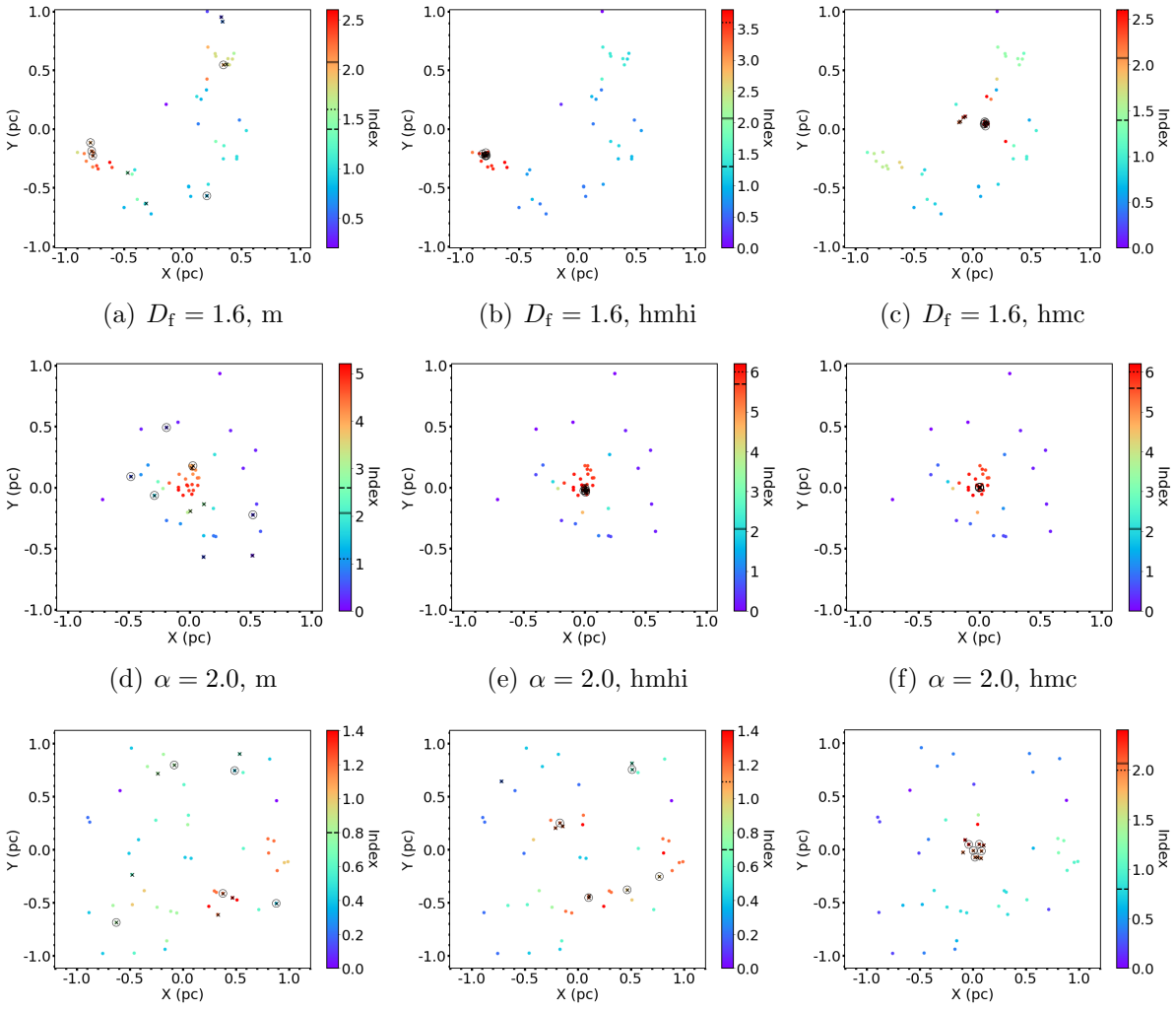
massive plots in Figure 3.2, where the significant index is calculated 100 times.

3.2.3 Fractal Star-Forming Regions

INDICATE is first applied to all 1000 stars in a substructured SFR with a fractal dimension of $D_{\rm f} = 1.6$ and the results are shown in Figure 3.3 and Figure 3.4.

Random masses

Firstly, INDICATE is applied to all 1000 stars in a substructured SFR with a fractal dimension of $D_{\rm f} = 1.6$ with randomly assigned masses. Figure 3.3(a) shows that INDICATE clearly identifies areas of high spatial clustering, and finds that 82.2% of stars have an index greater than the significant index of 2.3. The median index of the entire region is 4.4 and for the 10 most massive stars it is 4.5; a KS test returns a p-value of 0.9, suggesting that this difference is not significant and that both high and low mass stars share the same spatial distribution. Two of the most massive stars are located in the region with the highest INDICATE indexes, with a third massive star just outside this

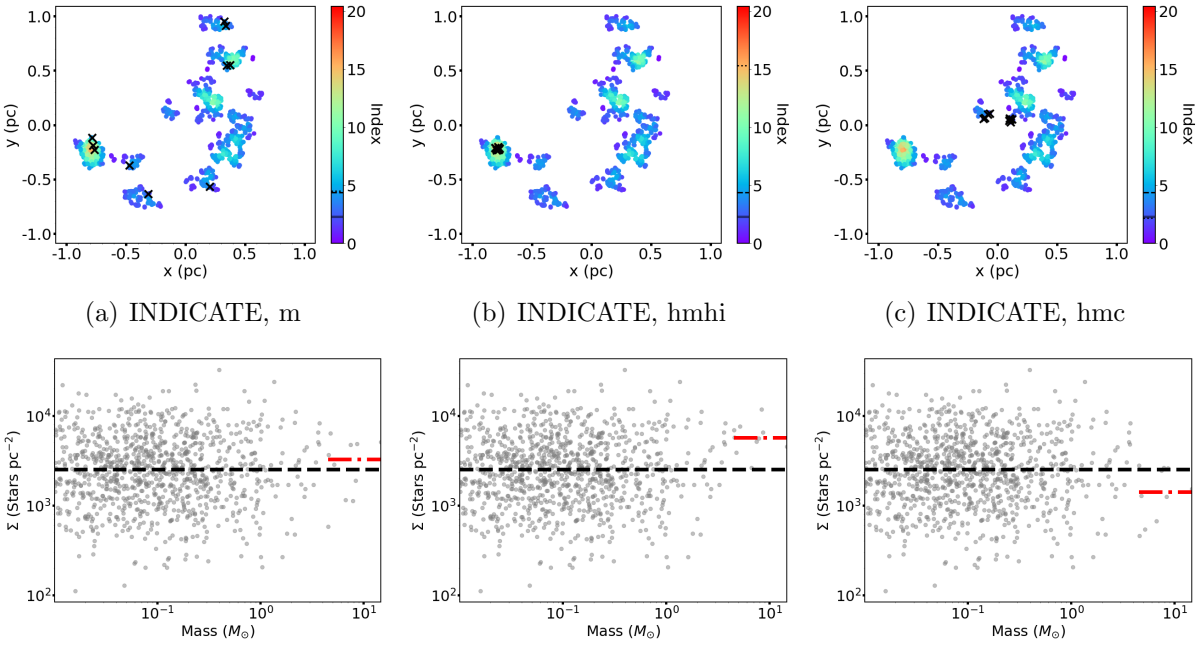


(a) $D_{\mathrm{f}} = 1.6$, m (b) $D_{\mathrm{f}} = 1.6$, hmhi (c) $D_{\mathrm{f}} = 1.6$, hmc

(d) $\alpha = 2.0$, m (e) $\alpha = 2.0$, hmhi (f) $\alpha = 2.0$, hmc

(g) Uniform, m (h) Uniform, hmhi (i) Uniform, hmc

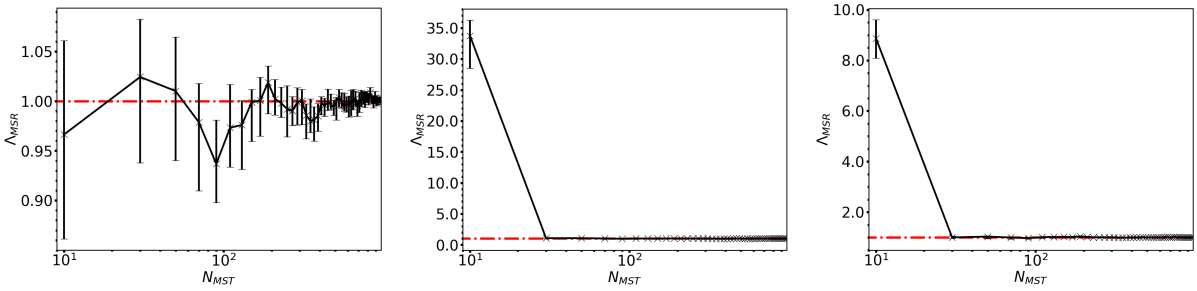
Figure 3.2: From left to right the columns show the different mass configurations for each of the synthetic regions' 50 most massive stars and their respective INDICATE results. From top to bottom shows the 50 most massive stars from the substructured region with fractal dimension of $D_{\mathrm{f}} = 1.6$, smooth centrally concentrated with density exponent $\alpha = 2.0$ and the uniformly distributed region. The solid black line in the colour bar shows the significant index (calculated using 100 repeats), the dashed black line is the median INDICATE index for the entire subset and the dotted black line is the median INDICATE index for the 10 most massive stars.



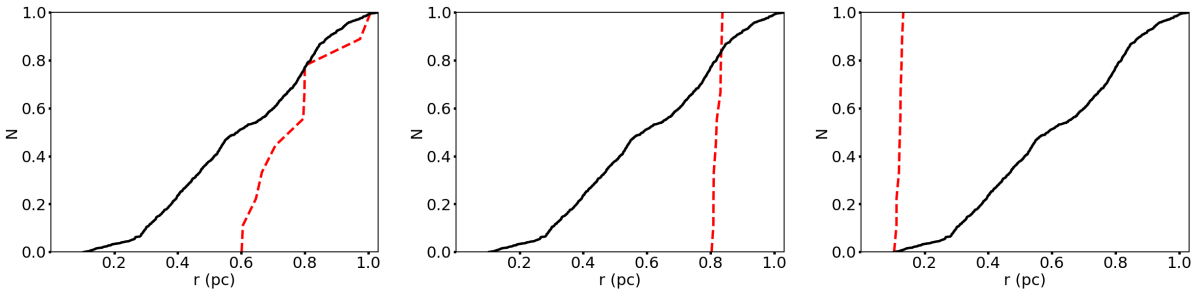
(a) INDICATE, m (b) INDICATE, hmhi (c) INDICATE, hmc

(d) Σ_{LDR}, m (e) Σ_{LDR}, hmhi (f) Σ_{LDR}, hmc

Figure 3.3: Substructured synthetic SFR of 1000 stars with $D_{\mathrm{f}} = 1.6$. The top row shows the INDICATE values and the bottom row shows the Σ-m plots. From left to right the columns show the different mass configurations, random masses (m), high mass high index (hmhi) and high mass centre (hmc). The colours in the top row show the INDICATE indexes for each star with the scale being the same as the radial region to allow comparisons using the colour alone. The solid black line, dashed black line and the dotted black line in the colour bar are the significant index, median index of all stars and the median index of the 10 most massive stars, respectively. In the second row, the median surface density of the stars is shown by the black dashed line, the median surface density of the 10 most massive stars is shown by the red dash-dotted line.



(a) Λ_{MSR}, m (b) Λ_{MSR}, hmhi (c) Λ_{MSR}, hmc



(d) Radial distribution, m (e) Radial distribution, hmhi (f) Radial distribution, hmc

Figure 3.4: Substructured synthetic SFR of 1000 stars with $D_{\mathrm{f}} = 1.6$. The top row shows Λ_{MSR} results and the bottom row shows the CDF of all the stars radial distance from the origin. From left to right the columns show the different mass configurations, random masses (m), high mass high index (hmhi) and high mass centre (hmc). The top row shows the Λ_{MSR} values and their associated errors with the black lines, the red dash dotted line is the canonical value used to signify mass segregation. In the bottom row the black line represents the CDF of radial distance from the centre for all the stars, the red dashed line is the CDF for the 10 most massive stars.

region. The remaining stars are located in regions of lower spatial clustering according to INDICATE.

Applying INDICATE to the 50 most massive stars I find that 22% of stars in the subset have indexes above the significant index of 2.1. The median index for the subset is 1.4 and the median index for the 10 most massive stars is 1.6, which is below the significant index meaning that INDICATE is not detecting mass segregation. The significant index was calculated using 100 repeats to eliminate statistical fluctuations which are present in a sample of this size. A KS test returns a p-value = 1.00, correctly identifying that there is no difference in clustering tendencies of the 10 most massive stars and the rest of the subset.

Figure 3.3(d) shows local stellar surface density against mass, with $\Sigma_{\rm LDR} = 1.3$, and a p-value of 0.7, meaning no significant difference between the local stellar surface density of high mass stars and the entire region. This is in agreement with the INDICATE result that the most massive stars are distributed similarly to the other stars in the region.

Figure 3.4(a) shows the $\Lambda_{\rm MSR}$ result, with $\Lambda_{\rm MSR} = 0.97^{+0.09}_{-0.11}$ for the 10 most massive stars which is consistent with no significant mass segregation and is in agreement with INDICATE.

Figure 3.4(d) shows the cumulative radial distribution of the 10 most massive stars and all the stars in the region. The radial distribution starts at 0.6 pc for the 10 most massive stars; this is due to the randomly assigned stellar masses, which happen to mainly be in clumps a relatively large distance away from the centre. When comparing the two distributions using a KS test a p-value $\ll 0.01$ is returned, meaning that the most massive stars could be mistakenly inferred to have been drawn from a different underlying radial distribution than all the stars.

High Mass High Index

The 10 most massive stars are now swapped with the 10 stars that have the highest INDICATE index and these results are shown in the middle column of Figure 3.3 and Figure 3.4. Figure 3.3(b) shows the positions of the 10 most massive stars (the black crosses), which are now in positions of the highest INDICATE index.

The median index of the 10 most massive stars has increased from 4.5 to 15.3, with the median index for the entire region staying the same at 4.4, as does the significant index of 2.3 and the percentage of stars with indexes greater than it. These parameters stay the same as the overall geometry of the region has not changed, just the masses assigned to 20 of the stars that have been swapped. Comparing the 10 most massive stars to the rest of the population using a KS test gives a p-value $\ll 0.01$, implying a significant difference in the clustering tendencies of the most massive stars. Figure 3.3(b) shows this difference clearly, as the 10 most massive stars are now located in the most clustered locations according to INDICATE, as a consequence of this they are also now visibly more spatially concentrated.

Applying INDICATE to just the 50 most massive stars I find now that the percentage of stars with indexes above the significant index of 2.1 is 38%, with a median index for all 50 stars of 1.3 and a median index of 3.6 for the 10 most massive stars. As the median index for the 10 most massive stars is above the significant index mass segregation has been detected in the region. The reason the amount of stars with significant indexes has changed is due to the fact that the swap is made in the full region of 1000 stars of which I then pick the 50 most massive stars. A KS test returns a p-value $\ll 0.01$ confirming that the tendency of the 10 most massive stars to cluster with high mass stars is significantly different to that of the entire subset of the 50 most massive stars.

Σ_{LDR} has increased from 1.3 to 2.3, with a p-value $\ll 0.01$. Figure 3.3(e) shows the 10 most massive stars are now above the median surface density of all of the stars (shown

by the horizontal dashed black line). In this case the reason for this is that swapping stars to the most clustered areas as measured using INDICATE results in them also being swapped into areas with higher than average local stellar surface density.

I now measure mass segregation according to $\Lambda_{\rm MSR}$ of the region, and find the 10 most massive stars have a mass segregation ratio of $\Lambda_{\rm MSR} = 33.74^{+2.54}_{-5.27}$. This implies significant mass segregation. In this particular region this is because all the most massive stars are located in a single clump with a high INDICATE index.

Because all the most massive stars have been moved to the same region the average distance between them is shorter than when looking at the average distances between random stars in the region. Figure 3.4(b) shows the peak value of $\Lambda_{\rm MSR}$ for the 10 most massive stars is ~ 34 which then rapidly decreases to ~ 1, meaning no mass segregation for lower mass stars, which is to be expected because these stars have not been swapped.

When comparing the cumulative distributions of the positions of the 10 most massive stars and all the stars (see Figure 3.4(e)), a clear difference can be seen. The distribution of positions for the 10 most massive stars is very narrow because they are in a very concentrated location therefore they are all a similar distance away from the origin. A KS test between the cumulative distribution of positions for all the stars and the 10 most massive stars returns a p-value $\ll 0.01$.

High Mass Centre

The 10 most massive stars are now swapped with the 10 most central stars (the results for this are shown in the right-hand column of Figure 3.3) and Figure 3.4. Because of the box-fractal construction of the region the origin (at (0,0)) is in empty space, so the most massive stars are split into two groups around the origin.

The median INDICATE index for the 10 most massive stars is 2.2 (decreasing from 4.5 for the 10 most massive stars in the original region with randomly assigned masses), below the median index for the region which is 4.4. A KS test returns a p-value $\ll 0.01$

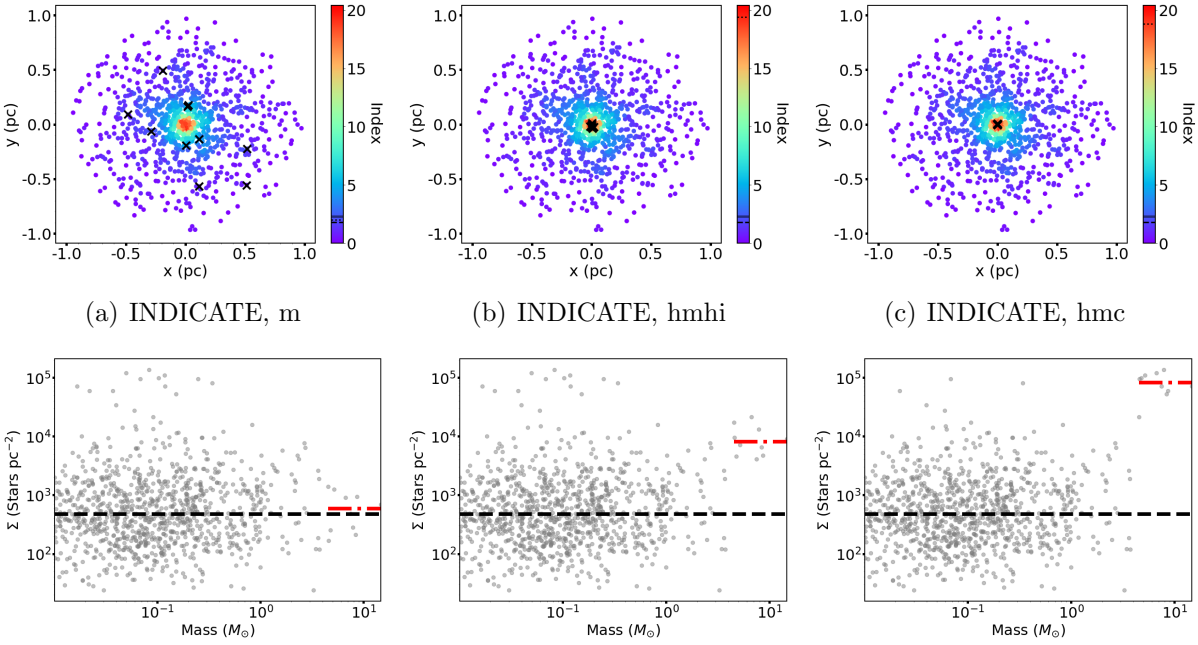
implying a significant difference in the distributions, in this case the most massive stars are in locations of lower clustering than the rest of the stars in the region. Figure 3.3(c) shows the two main groups of the most massive stars either side of the centre of the SFR and the stars here have a relatively low INDICATE index. The central locations of this region happen to be of relatively low index compared to the rest of the SFR.

Applying INDICATE to just the 50 most massive stars I find that 28% of stars have indexes above the significant index of 2.1. With a median index for the entire subset of 1.4 and 2.6 for the 10 most massive stars, which is above the significant index of 2.1 meaning the region is mass segregated. A KS test returns p-value $\ll 0.01$ confirming that the tendency of the 10 most massive stars to cluster with high mass stars is significantly different to the entire subset.

Figure 3.3(f) shows the local surface density against mass plot. The 10 most massive stars are no longer above the median local stellar surface density (black dashed line) because the points nearest to the centre of this region have a lower local surface density. $\Sigma_{\rm LDR} = 0.56$ with a p-value $= 0.03$, meaning no significant difference in the local surface density of the 10 most massive stars compared to all the stars. The surface densities therefore display similar behaviour to INDICATE, which also shows a decrease in the measured index.

When $\Lambda_{\rm MSR}$ is applied to this region a significant amount of mass segregation is detected with $\Lambda_{\rm MSR} = 8.87^{+0.74}_{-0.78}$, (Figure 3.4(c)). This is much lower than when swapping the most massive stars with the most clustered as measured with INDICATE, decreasing from over 30 to 8.87 in Figure 3.4(b) and Figure 3.4(c), respectively. This is due to the areas of highest INDICATE index being highly concentrated in one area, whereas in this case the most central region is in empty space, so the most massive stars are spread out around this point.

The cumulative distribution of the positions of stars is shown in Figure 3.4(f), which also shows the massive stars to be mass segregated and much closer to the centre than



(a) INDICATE, m (b) INDICATE, hmhi (c) INDICATE, hmc

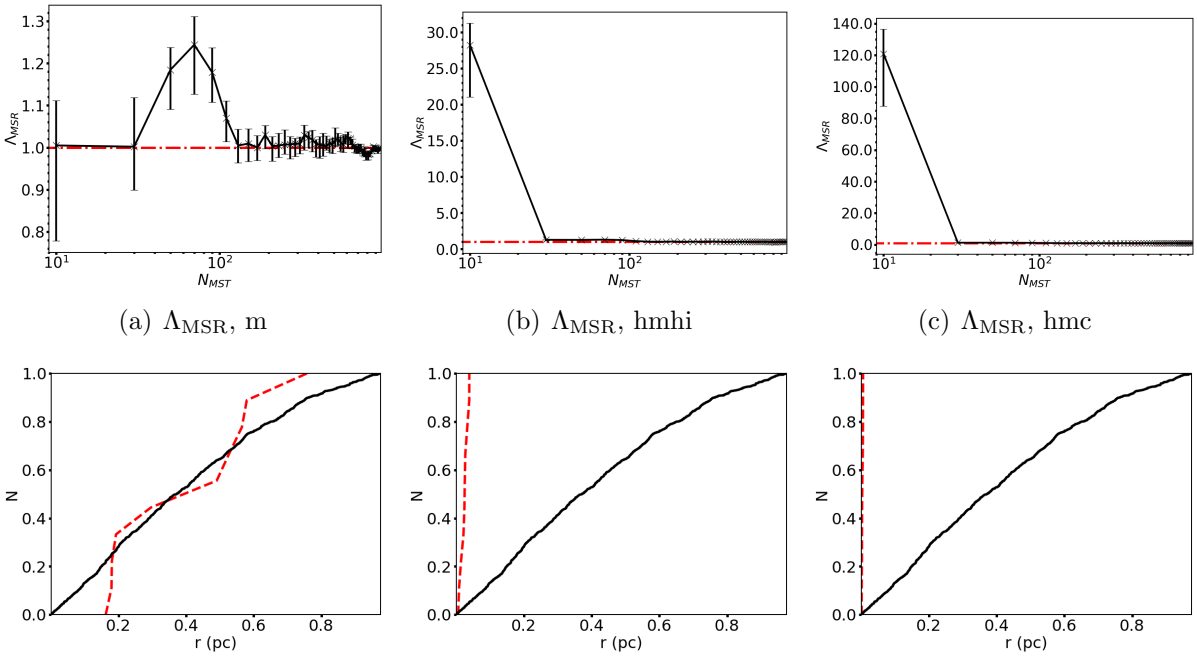
(d) Σ_{LDR}, m (e) Σ_{LDR}, hmhi (f) Σ_{LDR}, hmc

Figure 3.5: A synthetic, centrally concentrated star-forming region of 1000 stars with radial density exponent $\alpha = 2.0$. The top row shows the INDICATE values and the bottom row shows the Σ-m plots. From left to right the columns show the different mass configurations, random masses (m), high mass high index (hmhi) and high mass centre (hmc). The colours in the top row show the INDICATE indexes for each star with the scale being the same as the radial region to allow comparisons using the colour alone. The solid black line, dashed black line and the dotted black line in the colour bar are the significant index, median index of all stars and the median index of the 10 most massive stars, respectively. In the second row, the median surface density of the stars is shown by the black dashed line, the median surface density of the 10 most massive stars is shown by the red dash-dotted line.

the average star. A KS test returns a p-value $\ll 0.01$.

3.2.4 Smooth, Centrally Concentrated Star-Forming Regions

INDICATE is now applied to a geometrically smooth, centrally concentrated distribution consisting of 1000 stars with a radial density profile exponent $\alpha = 2.0$. The results are presented in Figure 3.5 and Figure 3.6.



(a) Λ_{MSR}, m (b) Λ_{MSR}, hmhi (c) Λ_{MSR}, hmc

(d) Radial distribution, m (e) Radial distribtuion, hmhi (f) Radial distribution, hmc

Figure 3.6: A synthetic, centrally concentrated star-forming region of 1000 stars with radial density exponent $\alpha = 2.0$. The top row shows Λ_{MSR} results and the bottom row shows the CDF of all the stars radial distance from the origin. From left to right the columns show the different mass configurations, random masses (m), high mass high index (hmhi) and high mass centre (hmc). The top row shows the Λ_{MSR} values and their associated errors with the black lines, the red dash dotted line is the canonical value used to signify mass segregation. In the bottom row the black line represents the CDF of radial distance from the centre for all the stars, the red dashed line is the CDF for the 10 most massive stars.

Random Masses

The smooth, centrally concentrated region shown in Figure 3.5(a) has a clear central region where INDICATE has detected high levels of spatial clustering, finding that 44.1% of stars are spatially clustered above random, with indexes greater than the significant index of 2.3. In this case the most massive stars are spread out across the region with none of the 10 most massive stars located in the central area. A median index of 1.8 is found for all stars and a median index of 2.0 is found for the 10 most massive stars: neither the massive stars nor rest of the population is typically in areas of non-random stellar affiliation as both have median indexes below the significant index. A KS test confirms that there is no significant difference in their spatial clustering with a p-value = 0.55.

Applying INDICATE to just the 50 most massive stars in the region I find that 52% of stars have indexes above the significant index of 2.1, with a median index for the entire subset of 2.6 and 1.1 for the 10 most massive stars, which is below the significant index of 2.1, meaning INDICATE detects no mass segregation in the region. A KS test returns a p-value = 0.48 implying no difference in the clustering tendencies of the 10 most massive stars compared to the rest in the subset.

Figure 3.5(d) shows the local stellar surface density against mass plot for this region with the red dashed-dotted line showing the median surface density for the 10 most massive stars and the black dashed line showing the median surface density of all the stars. A similar result is seen with $\Sigma_{\rm LDR} = 1.24$ with a p-value = 0.63 from the KS test, indicating the difference is not significant.

Figure 3.6(a) shows the mass segregation ratio for the 10 most massive stars is $\Lambda_{\rm MSR} = 1.00^{+0.11}_{-0.23}$ meaning that $\Lambda_{\rm MSR}$ finds no significant mass segregation for the 10 most massive stars in this region.

The cumulative distribution of positions are very similar between the most massive

stars and the rest, with p-value = 0.67. Figure 3.6(d) shows the radial distribution of the 10 most massive stars in red; it closely matches the radial distribution of the entire region.

High Mass High Index

As before, the 10 most massive stars are swapped with the 10 stars with the greatest clustering as measured by INDICATE. Figure 3.5(b) shows the 10 most massive stars are now located in the centre of the region. The median index for the 10 most massive stars has increased from 2.0 to 19.4, with a p-value $\ll 0.01$ indicating a significant difference of the clustering tendencies between all the stars (which have a median index of 1.8) and the 10 most massive stars. Therefore, the 10 most massive stars are found in areas of above average stellar affiliation.

INDICATE is applied to just the 50 most massive stars in the region and finds that 64% of stars have indexes above the significant index of 2.1, with the subset having a median index of 5.7 and a median index of 6.0 for the 10 most massive stars. As the median index for the 10 most massive stars is larger than significant index INDICATE has detected mass segregation. A KS test returns a p-value $\ll 0.01$ implying a significant difference in the clustering tendencies of the 10 most massive stars despite the median indexes being similar.

Figure 3.5(e) shows the clear difference in the median surface density of the entire region (black dashed line) compared to the median surface density of the 10 most massive stars (red dashed-dotted line). $\Sigma_{\rm LDR} = 17$ (increasing from 1.24) with $p \ll 0.01$, in agreement with INDICATE that the 10 most massive stars are found in areas of greater stellar clustering compared to the rest of the population.

$\Lambda_{\rm MSR}$ detects significant mass segregation with $\Lambda_{\rm MSR} = 28.23^{+3.05}_{-7.20}$, increasing from 1.00 when compared to this region with randomly assigned masses and is in agreement with INDICATE when applied to the 50 most massive stars. Like in the fractal region

in Figure 3.3, there is one area of high clustering, so the most massive stars are moved closer together as a result. This is also reflected in the cumulative distribution of positions (Figure 3.6(e)) with a much steeper function for the 10 most massive stars with p-value $\ll 0.01$, this is very similar to the results in Figure 3.4(e) for a fractal distribution.

High Mass Centre

Figure 3.5(c) highlights that the most massive stars are now closer to each other after being swapped with the stars closest to the origin. The median index for the 10 most massive stars is 18.8, larger than the median index for the region of 1.8, meaning the 10 most massive stars find themselves in locations of greater than average stellar affiliation. A KS test returns a p-value $\ll 0.01$ finding a significant difference in the clustering tendencies of the 10 most massive stars compared to the rest.

INDICATE is applied to the 50 most massive stars and finds that 62% of them have indexes above the significant index of 2.1, with a median index of 5.9 for the entire region and 6.0 for the 10 most massive stars. The median index for the 10 most massive stars is above the significant index and so the region is found to be mass segregated by INDICATE. A KS test returns a p-value $\ll 0.01$ implying a significant difference between the 10 most massive stars' spatial clustering and the rest.

In Figure 3.5(f) the most massive stars find themselves in areas of much higher surface density than when they were moved to areas of greatest INDICATE index. $\Sigma_{\rm LDR}$ increases from 17 to $\Sigma_{\rm LDR} = 174.72$ with a p-value $\ll 0.01$. As the median INDICATE index for the 10 most massive stars has decreased to 18.8 from 19.4 it demonstrates that INDICATE is not measuring the exact same quantity as the local stellar surface density, as if it was one would expect the median index of the 10 most massive stars to be larger than 19.4.

$\Lambda_{\rm MSR} = 120.90^{+15.71}_{-33.05}$, signifying significant mass segregation for the 10 most massive stars (see Figure 3.6(c)). This is the strongest signal that $\Lambda_{\rm MSR}$ returns for all the

synthetic SFRs shown in this chapter.

Figure 3.6(f) shows that the cumulative distribution of the positions of the 10 most massive stars are now much closer to the centre of the region than before. A KS test returns a p-value $\ll 0.01$, implying that there is a significant difference in the spatial distribution of the 10 most massive stars compared to the rest.

3.2.5 Uniform Star-Forming Regions

Figure 3.7 shows the results of INDICATE being applied to a uniform distribution of 1000 stars located between $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$.

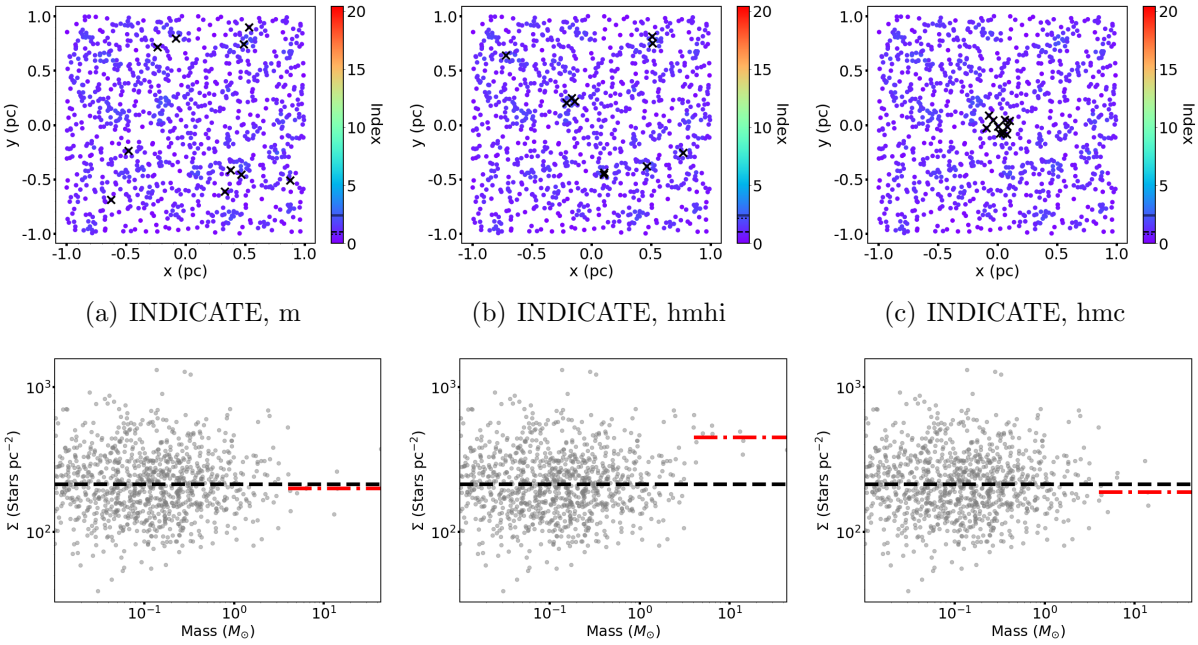
Random Masses

Figure 3.7(a) shows a uniform distribution with randomly assigned masses. The median INDICATE index for the entire region is 1.0 and for the 10 most massive stars it is 0.8 and there is no significant difference between the 10 most massive stars and the rest of the region with a p-value = 0.68. In this region no star has an index above the significant index of 2.4 meaning the clustering in this region is random, as is expected.

INDICATE is applied to just the 50 most massive stars and finds that no stars have an index above the significant index of 2.1. The median index for the region is 0.8, as is the median index of the 10 most massive stars. As the median for the 10 most massive stars is below the significant index INDICATE detects no mass segregation in the region. A KS test returns a p-value = 1.00, confirming that the 10 most massive stars and the entire subset have similar clustering tendencies.

In Figure 3.7(d) the most massive stars find themselves in similar surface density areas as the rest of the stars in the region with $\Sigma_{\mathrm{LDR}} = 0.94$. No significant difference is detected with a KS test returning a p-value = 0.59.

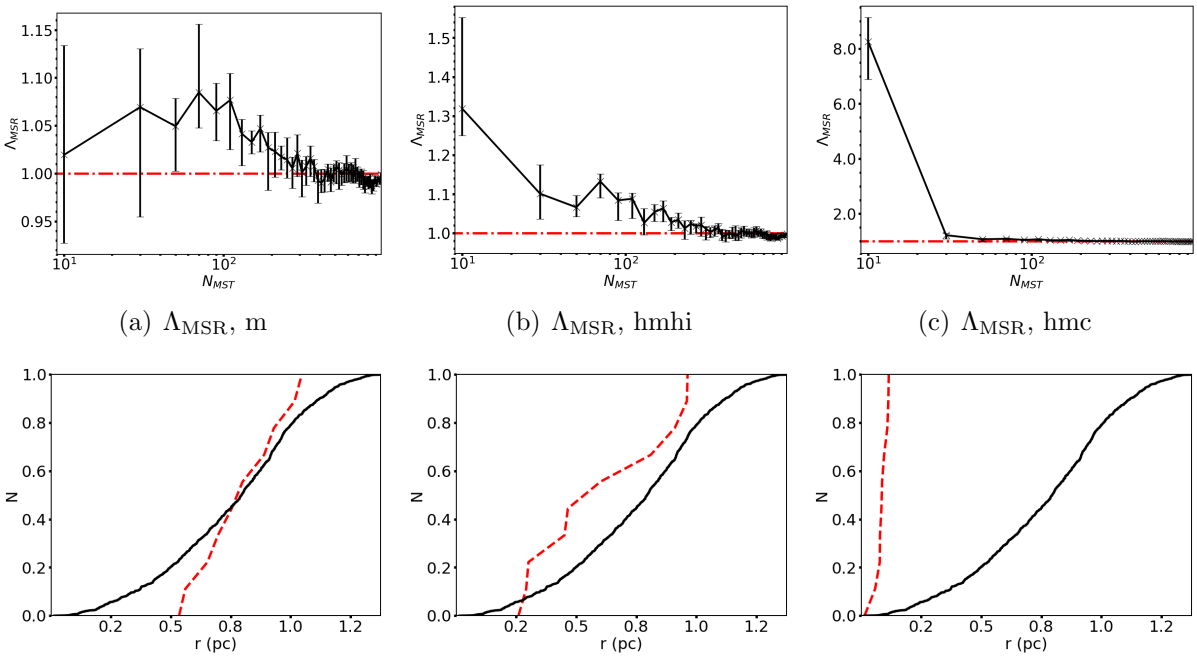
Figure 3.8(a) shows a very weak signal as expected of $\Lambda_{\mathrm{MSR}} = 1.02^{+0.11}_{-0.09}$ meaning there is no mass segregation.



(a) INDICATE, m (b) INDICATE, hmhi (c) INDICATE, hmc

(d) Σ_{LDR}, m (e) Σ_{LDR}, hmhi (f) Σ_{LDR}, hmc

Figure 3.7: A synthetic uniform distribution of 1000 stars. The top row shows the INDICATE values and the bottom row shows the Σ-m plots. From left to right the columns show the different mass configurations, random masses (m), high mass high index (hmhi) and high mass centre (hmc). The colours in the top row show the INDICATE indexes for each star with the scale being the same as the radial region to allow comparisons using the colour alone. The solid black line, dashed black line and the dotted black line in the colour bar are the significant index, median index of all stars and the median index of the 10 most massive stars, respectively. In the second row, the median surface density of the stars is shown by the black dashed line, the median surface density of the 10 most massive stars is shown by the red dash-dotted line.



(a) Λ_{MSR}, m (b) Λ_{MSR}, hmhi (c) Λ_{MSR}, hmc

(d) Radial distributions, m (e) Radial distributions, hmhi (f) Radial distributions, hmc

Figure 3.8: A synthetic uniform distribution of 1000 stars. The top row shows Λ_{MSR} results and the bottom row shows the CDF of all the stars radial distance from the origin. From left to right the columns show the different mass configurations, random masses (m), high mass high index (hmhi) and high mass centre (hmc). The top row shows the Λ_{MSR} values and their associated errors with the black lines, the red dash dotted line is the canonical value used to signify mass segregation. In the bottom row the black line represents the CDF of radial distance from the centre for all the stars, the red dashed line is the CDF for the 10 most massive stars.

Figure 3.8(d) shows the cumulative distribution of positions starting further out for the 10 most massive stars than for the entire region but quickly matches the overall distribution. No significant difference is detected with a p-value = 0.60.

High Mass High Index

Now the 10 most massive stars are swapped with the stars with the 10 highest INDICATE indexes (Figure 3.7(b)). Unlike in the centrally concentrated and fractal SFRs there is more than one region with relatively high indexes. The median index for the 10 most massive stars has increased from 0.8 to 2.2, with the entire region having a median index of 1.0, with a p-value $\ll 0.01$ when comparing the 10 most massive stars to the entire region, suggesting a significant difference. As both the median index for the region and the median index for the 10 most massive stars are below the significant index of 2.4 all the stars still have a random spatial distribution according to INDICATE.

INDICATE is applied to just the 50 most massive stars and finds that no stars have an index above the significant index of 2.1. The median index of the entire subset is 0.7 and for the 10 most massive it is 1.1. As the median index for the 10 most massive stars is below the significant index INDICATE detects no mass segregation. A KS test returns a p-value = 0.67 implying no difference in the distribution of the 10 most massive stars compared to the entire subset.

Figure 3.7(e) shows the median local stellar surface density of the 10 most massive stars is greater than the median local stellar surface density of the region, with $\Sigma_{\mathrm{LDR}} = 2.11$ (increasing from 0.94) and a KS test returns a p-value $\ll 0.01$, meaning a significant difference in the local stellar surface density of the 10 most massive stars and all the stars.

Figure 3.8(b) shows that $\Lambda_{\mathrm{MSR}} = 1.32^{+0.23}_{-0.07}$ for the 10 most massive stars suggesting that weak mass segregation is being detected, before decreasing as more stars are added to the subset. This is due to the random picking of stars for the subset MST. In Parker

& Goodwin (2015) they suggest ignoring results of $\Lambda_{\rm MSR} < 2$ to avoid false positives such as this.

In Figure 3.8(e) the cumulative distributions of positions are shown for the 10 most massive stars in red and all stars in black. A KS test returns a p-value = 0.19 showing no significant difference.

High Mass Centre

Now the 10 most massive stars are swapped with the 10 most central stars; shown in Figure 3.7(c). The median index for 10 most massive stars is 0.8 (the same as when masses are randomly assigned) and similarly there is no significant difference in the spatial clustering of the most massive stars and the rest of the stars with a p-value = 0.64.

INDICATE is applied to just the 50 most massive stars and finds that 10% of stars have an index above the significant index of 2.1. The median index for the region is found to be 0.8 and 2.0 for the 10 most massive. As the median index for the 10 most massive stars is below the significant index according to INDICATE they are randomly distributed and so no mass segregation has been detected. A KS test returns a p-value $\ll 0.01$ meaning the 10 most massive stars and the rest are distributed differently as the 10 most massive stars are closer together than would be expected from a random uniform distribution.

Figure 3.7(f) shows the local surface density against mass plot. A value of $\Sigma_{\rm LDR} = 0.88$ is found (lower than when masses are swapped with stars of greatest INDICATE index) with a KS test giving a p-value = 0.49 implying no significant differences in the surface density of the most massive stars compared to all the stars.

Figure 3.8(c) shows the $\Lambda_{\rm MSR}$ results and a lower value is found than for other examples with the highest masses moved to the centre. $\Lambda_{\rm MSR} = 8.25^{+0.88}_{-1.36}$ meaning mass segregation is detected for the 10 most massive stars and this quickly drops off as the rest of the stars are uniformly distributed. This is opposite to the INDICATE result which

finds no mass segregation using the given criteria but does find a significant difference in the clustering of the 10 most massive stars compared to the entire subset.

Figure 3.8(f) shows the radial cumulative distributions of positions. The 10 most massive stars show a similar trend as the fractal and smooth star-forming regions, with a much steeper function when the most massive stars are swapped with the most central stars. A KS test returns a p-value $\ll 0.01$.

3.2.6 Summary

The INDICATE method has clearly identified regions of clustering in the synthetic datasets. INDICATE gives results that are in agreement with $\Sigma_{\rm LDR}$ when applied to the entire region and results that are generally in agreement with $\Lambda_{\rm MSR}$ when applied to only the 50 most massive stars in the synthetic star forming regions. The INDICATE results when applied to all 1000 stars in the example regions are summarised in Table 3.5 the results of applying INDICATE to only the 50 most massive stars are shown in Table 3.1. The results of applying $\Sigma_{\rm LDR}$, $\Lambda_{\rm MSR}$ and CDF methods to the example synthetic regions are summarised in Table 3.6.

Table 3.5: Results of INDICATE being applied to all stars in the synthetic SFRs. From left to right the columns are: the median index for all stars in the region, the median index for the 10 most massive stars, the significant index and the p-value returned from a KS test comparing the indexes between the 10 most massive stars and all stars in the region. The null hypothesis is rejected when p-value $\ll 0.01$.

Region	$\tilde{I}_{\rm all}$	$\tilde{I}_{10}$	$I_{\rm sig}$	$\% > I_{\rm sig}$	p
$D = 1.6$, m	$4.4^{+2.6}_{-1.4}$	$4.5^{+3.3}_{-0.6}$	2.3	82.2	0.90
$D = 1.6$, hmhi	$4.4^{+2.6}_{-1.4}$	$15.3^{+0.2}_{-0.4}$	2.3	82.2	$\ll 0.01$
$D = 1.6$, hmc	$4.4^{+2.6}_{-1.4}$	$2.2^{+0.0}_{-0.0}$	2.3	82.2	$\ll 0.01$
$\alpha = 2.0$, m	$1.8^{+3.0}_{-1.0}$	$2.0^{+2.4}_{-0.8}$	2.3	44.1	0.55
$\alpha = 2.0$, hmhi	$1.8^{+3.0}_{-1.0}$	$19.4^{+0.8}_{-0.0}$	2.3	44.1	$\ll 0.01$
$\alpha = 2.0$, hmc	$1.8^{+3.0}_{-1.0}$	$18.8^{+0.2}_{-0.0}$	2.3	44.1	$\ll 0.01$
Uniform, m	$1.0^{+0.2}_{-0.4}$	$0.8^{+0.4}_{-0.0}$	2.4	0.0	1.00
Uniform, hmhi	$1.0^{+0.2}_{-0.4}$	$2.2^{+0.0}_{-0.2}$	2.4	0.0	$\ll 0.01$
Uniform, hmc	$1.0^{+0.2}_{-0.4}$	$0.8^{+0.1}_{-0.0}$	2.4	0.0	0.64

Table 3.6: Results of the other methods being applied to all stars in the synthetic SFRs. From left to right the columns are: the local stellar surface density ratio, the p-value from a KS test comparing the median local stellar surface density of the 10 most massive stars to the median local stellar surface density of the entire region, the mass segregation ratio and the p-value of a KS test comparing the CDF of positions of the 10 most massive stars and all the stars in each region.

Region	$\Sigma_{\rm LDR}$	$\Sigma_{\rm LDR}$ (p)	$\Lambda_{\rm MSR}$	CDF (p)
$D = 1.6$, m	1.30	0.69	$0.97^{+0.09}_{-0.11}$	$\ll 0.01$
$D = 1.6$, hmhi	2.30	$\ll 0.01$	$33.74^{+2.54}_{-5.27}$	$\ll 0.01$
$D = 1.6$, hmc	0.56	0.03	$8.87^{+0.74}_{-0.78}$	$\ll 0.01$
$\alpha = 2.0$, m	1.24	0.63	$1.00^{+0.11}_{-0.23}$	0.67
$\alpha = 2.0$, hmhi	17.00	$\ll 0.01$	$28.23^{+3.05}_{-7.20}$	$\ll 0.01$
$\alpha = 2.0$, hmc	174.72	$\ll 0.01$	$120.90^{+15.71}_{-33.05}$	$\ll 0.01$
Uniform, m	0.94	0.59	$1.02^{+0.11}_{-0.09}$	0.60
Uniform, hmhi	2.11	$\ll 0.01$	$1.32^{+0.23}_{-0.07}$	0.19
Uniform, hmc	0.88	0.49	$8.25^{+0.88}_{-1.36}$	$\ll 0.01$

3.3 Testing INDICATE on 100 realisations

This section shows additional tests showing that INDICATE is robust at quantifying stellar clustering. This is done by applying INDICATE to 100 different realisations of the substructured, smooth centrally concentrated and uniform distributions presented in this work. Table 3.7 shows ranges and interquartile ranges (IQR) of INDICATE indexes of all 1000 stars over 100 different realisations of just the different cluster morphologies and shows that the spread of clustering is greater in substructured regions with an IQR of 1.6 compared to the centrally concentrated and uniform regions which have the same IQR of 0.2. Table 3.8 shows the INDICATE index ranges for the 10 most massive stars, and that when the most massive stars are moved to the centre of centrally concentrated regions the indexes increase. Table 3.9 shows the INDICATE index ranges for 10 randomly chosen stars, and shows that when randomly picking stars across the different sets of clusters the range of indexes is different, but that within a set with a certain morphology, the results are similar regardless of the mass configuration of the region.

Table 3.7: INDICATE was applied to all 1000 stars in 100 different realisations of the SFRs presented in this chapter. The distribution of INDICATE indexes is summarised here for all stars. From left to right the columns are: the 25^{th} quantile, 75^{th} quantile, the interquartile range (IQR), minimum index, maximum index, the range between the minimum and maximum index and the median significant index found across all realisations.

Region	25^{th} Quantile	75^{th} Quantile	IQR	Min I	Max I	I Range	$\tilde{I}_{\text{sig}}$
$D = 1.6$	3.8	5.4	1.6	2.4	10.0	7.6	2.3
$\alpha = 2.0$	1.8	2.0	0.2	1.6	2.4	0.8	2.3
Uniform	0.8	1.0	0.2	0.8	1.0	0.2	2.3

Table 3.8: INDICATE was applied to all 1000 stars in 100 different realisations of the SFRs presented in this chapter. The distribution of INDICATE indexes is summarised here for the 10 most massive stars. From left to right the columns are: the 25$^{\text{th}}$ quantile, 75$^{\text{th}}$ quantile, the IQR, minimum index, maximum index and the range between the minimum and maximum index.

Region	25$^{\text{th}}$ Quantile	75$^{\text{th}}$ Quantile	IQR	Min I	Max I	I Range
$D = 1.6$, m	3.4	5.7	2.3	2.2	8.3	6.1
$D = 1.6$, hmhi	9.8	14.5	4.8	6.0	27.0	21.0
$D = 1.6$, hmc	2.6	5.3	2.7	0.9	11.5	10.6
$\alpha = 2.0$, m	1.4	2.6	1.2	0.7	7.7	7.0
$\alpha = 2.0$, hmhi	21.4	24.4	3.0	18.2	27.7	9.5
$\alpha = 2.0$, hmc	20.7	23.6	2.9	17.8	27.4	9.6
Uniform, m	0.8	1.0	0.2	0.4	1.5	1.1
Uniform, hmhi	2.1	2.4	0.3	1.8	2.8	1.0
Uniform, hmc	0.7	1.2	0.5	0.4	1.6	1.2

Table 3.9: INDICATE was applied to all 1000 stars in 100 different realisations of the SFRs presented in this chapter. The distribution of INDICATE indexes is summarised here for 10 randomly chosen stars in each realisation. From left to right the columns are: the 25$^{\text{th}}$ quantile, 75$^{\text{th}}$ quantile, the IQR, minimum index, maximum index and the range between the minimum and maximum index.

Region	25$^{\text{th}}$ Quantile	75$^{\text{th}}$ Quantile	IQR	Min I	Max I	I Range
$D = 1.6$, m	3.7	5.7	2.0	1.9	13.6	11.7
$D = 1.6$, hmhi	3.6	5.6	2.0	1.9	13.6	11.7
$D = 1.6$, hmc	3.6	5.7	2.1	1.9	13.4	11.5
$\alpha = 2.0$, m	1.5	2.5	1.0	0.4	6.6	3.2
$\alpha = 2.0$, hmhi	1.5	2.5	1.0	0.4	6.6	6.2
$\alpha = 2.0$, hmc	1.5	2.6	1.1	0.4	8.0	7.6
Uniform, m	0.8	1.0	0.2	0.5	1.3	0.8
Uniform, hmhi	0.8	1.0	0.2	0.5	1.4	0.9
Uniform, hmc	0.8	1.0	0.2	0.5	1.3	0.8

3.4 Applying INDICATE to Observational Data

INDICATE is applied to the following real SFRs: Taurus, ONC, NGC 1333, IC 348 and ρ Ophiuchi. For the ONC, NGC1333 and IC348 stars without known masses are ignored when performing KS tests between the 10 most massive stars and the rest of the stars in the region. The results of applying INDICATE to all points in the observational data sets are presented in table 3.10. In §3.5 I show the results of using INDICATE to detect mass segregation in the real star-forming regions.

3.4.1 Taurus

The Taurus SFR is located 140 pc away with an estimated age of around 1 Myr (Bell et al., 2013).

I use the dataset from Parker et al. (2011), which has 361 objects and appears to be substructured, with a central region surrounded by small groups of stars (Gomez et al., 1993). Taurus has been investigated previously and has a corresponding fractal dimension, $D_{\rm f}$, inferred from the Q-parameter value 0.45 of 1.55 ± 0.25 (Cartwright & Whitworth, 2004). Figure 3.9 shows Taurus after INDICATE is applied, and it finds that 86% of stars are spatially clustered above random with indexes greater than the significant index of 2.1. Taurus has a median index of 6.6 for the entire region and 3.6 for the 10 most massive stars (masses are calculated in § 2 of Parker et al. (2011)). The most massive stars are highlighted with crosses in Figure 3.9, they are spread out across the SFR (see also Parker et al. (2011)), with most lying in less clustered regions. INDICATE detects no significant difference in distribution of indexes between the most massive stars and all stars with a p-value = 0.07.

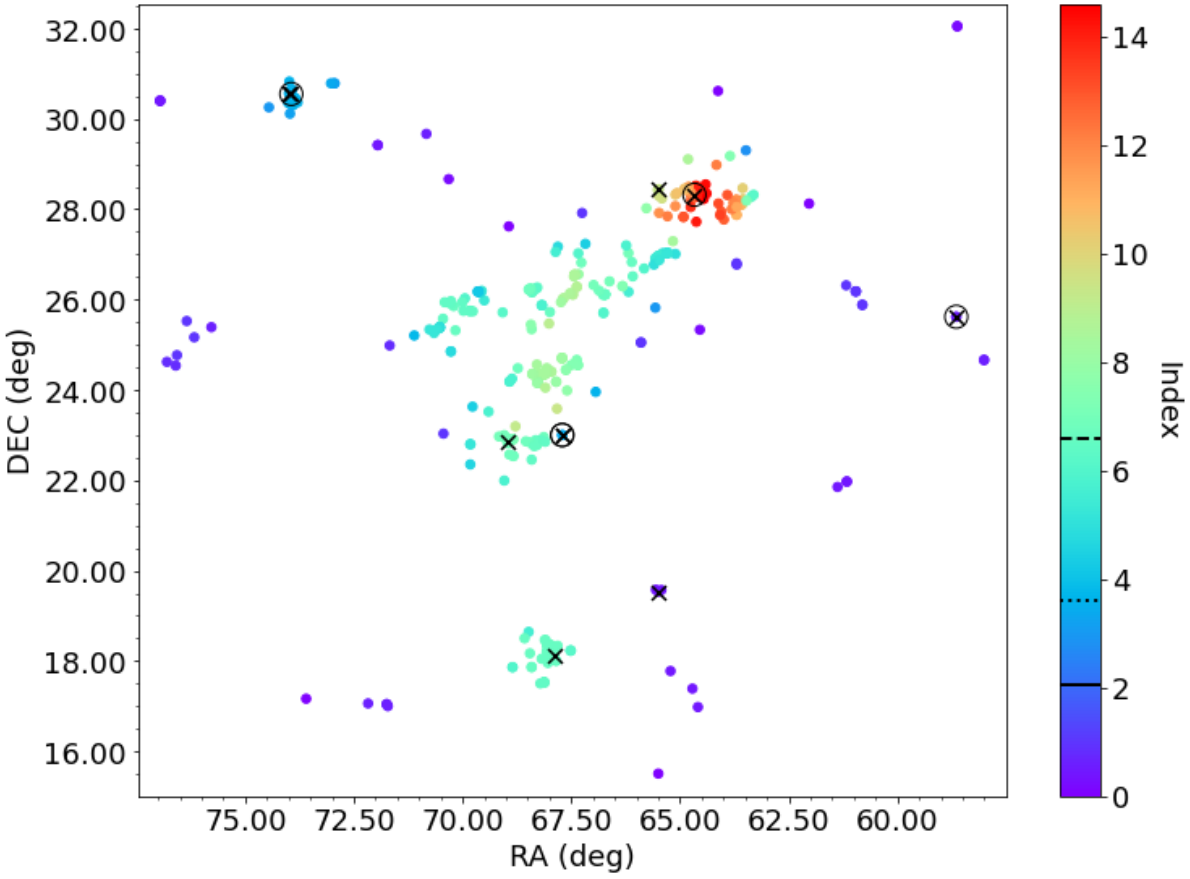


Figure 3.9: The Taurus SFR. The ten most massive stars are highlighted with crosses, the five most massive are circled. The significant INDICATE index is shown by the solid black line, the median index for all the stars is shown by the dashed black line and the median index for the 10 most massive stars is shown using the dotted black line in the colour bar.

Table 3.10: Table showing the percentage of stars in each observed region above the significant index and the results of two sample KS tests for each region. The 10 most massive stars indices are compared to all indices. A significance threshold of 0.01 is chosen to disprove the null hypothesis: that all stars in the region are distributed the same way. The ONC is the only region in which significant difference in the clustering tendencies of the 10 most massive stars is found compared to all the stars. From left to right the columns are: the name of the region, the median INDICATE index for the 10 most massive stars in the region ($\tilde{I}(10)$), the median INDICATE index for the entire region ($\tilde{I}(all)$), the percentage of stars in the region above the significant INDICATE index and the p-value result from comparing the spatial clustering of the 10 most massive stars to all stars in the region.

Name	$\tilde{I}(10)$	$\tilde{I}(all)$	% above I_{sig}	p-value
Taurus	3.6	6.6	85.9	0.07
ONC	10.3	1.4	46.1	0.003
NGC 1333	5.9	5.1	73.9	0.57
IC 348	2.6	3.2	59.2	0.63
ρ Ophiuchi	1.3	1.8	39.6	0.82

3.4.2 ONC

The Orion Nebula Cluster (ONC) is a very dense centrally concentrated region shown in Figure 3.10. I use the dataset from Hillenbrand & Hartmann (1998) which contains 1576 objects. The line of empty space to the south of the area of highest index is due to a band of extinction. 641 objects do not have an assigned mass in the dataset and so are removed when comparing the indexes between the 10 most massive stars and the entire region. The distance to the ONC is around 400 pc away with an estimated age of around 1 Myr (Jeffries et al., 2011; Reggiani et al., 2011). I apply the Q-parameter to the stars with masses, stars without masses and all stars and find that there is no significant difference in the Q-parameter, suggesting that the three subsets follow the same spatial distribution. The median index for the ONC is 1.4, the median index for the 10 most massive stars is 10.3 with 46.1% of stars spatially clustered above random (the ONC has a significant index of 2.4). These results are similar to the synthetic region shown in Figure 3.5. A KS test gives a p-value of 0.003, below the chosen threshold of 0.01 meaning that the 10 most massive stars have different clustering tendencies when

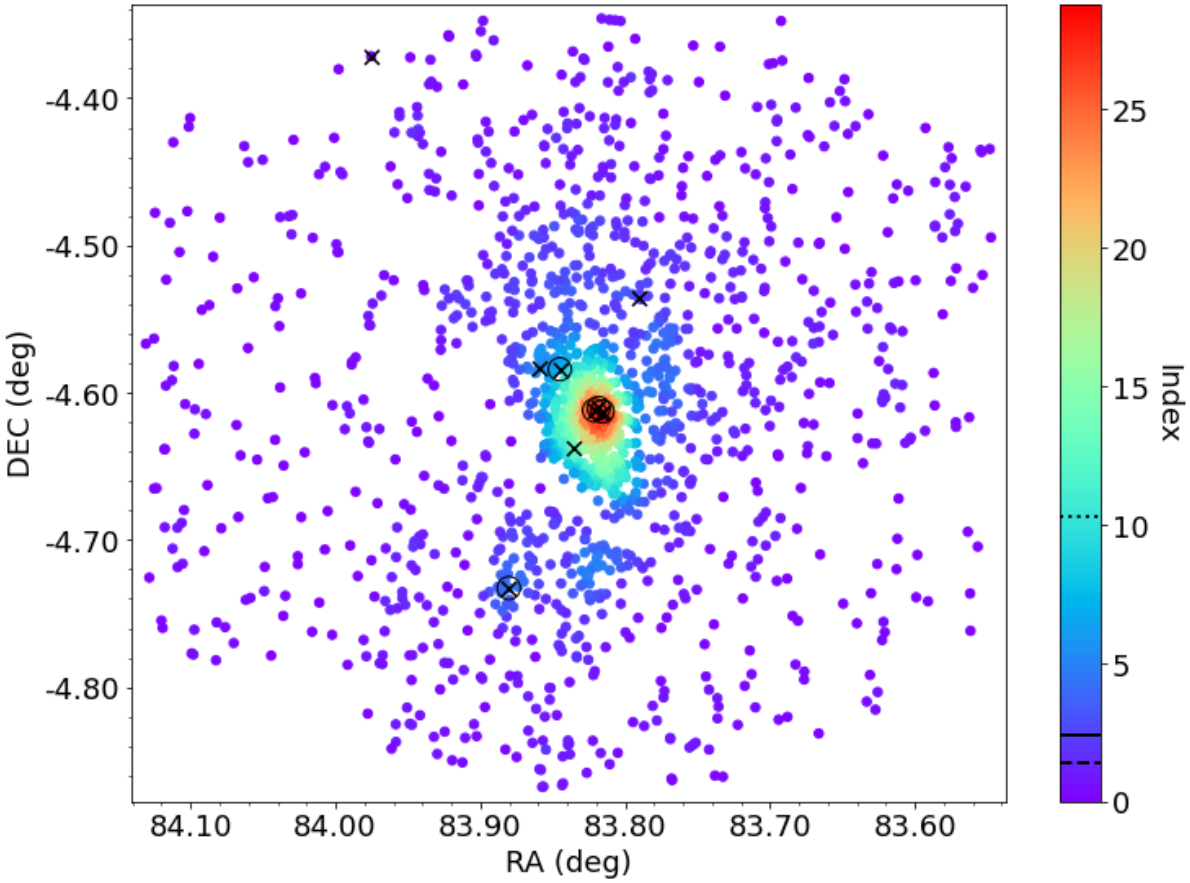


Figure 3.10: INDICATE plot of the ONC. The ten most massive stars are highlighted with crosses, the five most massive are circled. The significant index from INDICATE is shown with the solid black line in the colour bar, the median index of all the stars is shown by the dashed black line and the median index for the 10 most massive stars is shown by the dotted black line.

compared to the entire region.

3.4.3 NGC 1333

The NGC 1333 SFR (shown in Figure 3.11) contains 203 objects, 162 of which have an assigned mass in the dataset used by Parker & Alves de Oliveira (2017) that allows us to determine any relation between the mass and clustering. The distance to the region is 235 pc with an age of around 1 Myr (Parker & Alves de Oliveira, 2017; Pavlidou et al., 2021). INDICATE finds that 74% of stars are spatially clustered above random (the

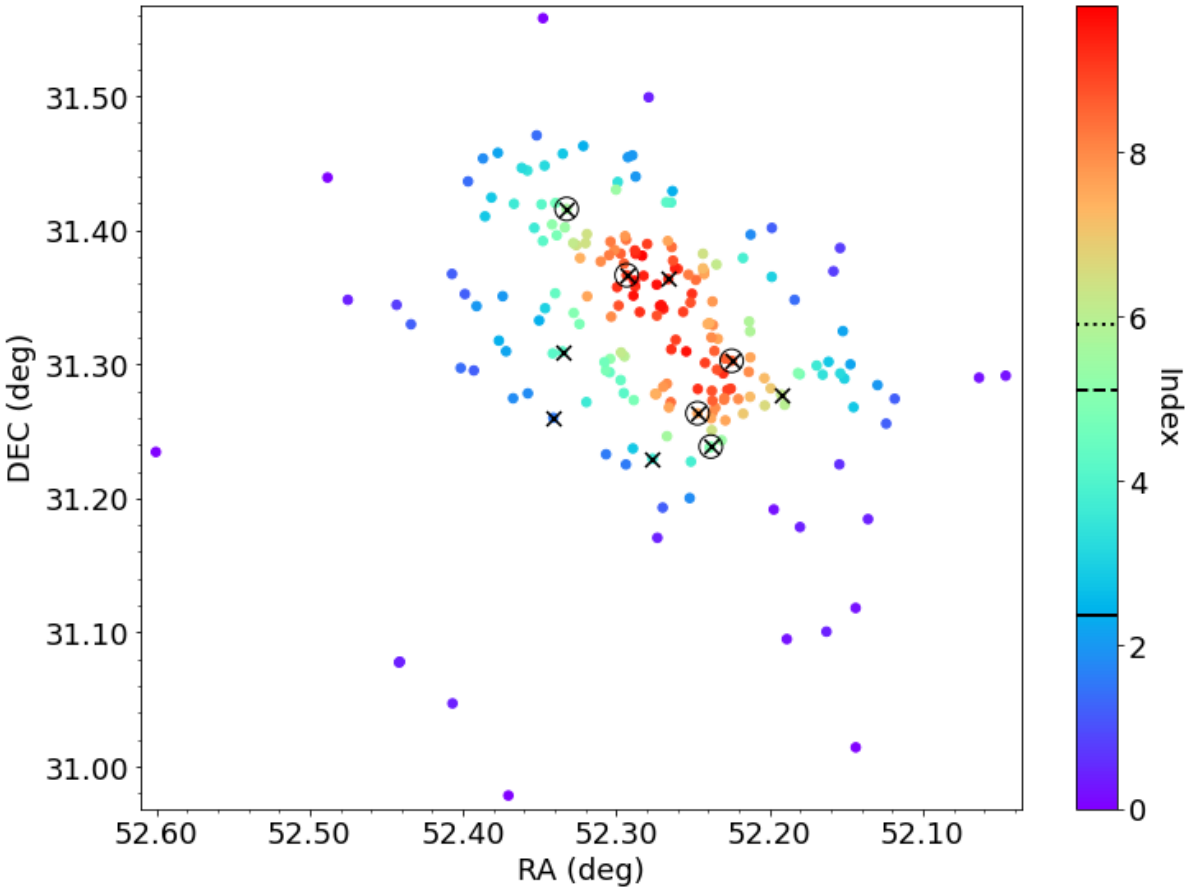


Figure 3.11: Plot of NGC 1333. The ten most massive stars are highlighted with crosses, the five most massive are circled. The significant index from INDICATE is shown in the colour bar by the solid black line, the median index for all the stars is shown by the dashed black line and the median index for the 10 most massive stars is shown by the dotted black line.

region has a significant index of 2.4). INDICATE has highlighted an extended central region of relatively high spatial clustering, with the most massive stars spread out around this region. A median index of 5.1 is found for all stars and for the 10 most massive stars a median index of 5.9 is found with a p-value = 0.57. This implies no significant difference in the spatial clustering of the most massive stars compared to all the stars.

3.4.4 IC 348

The data from Parker & Alves de Oliveira (2017) contains 478 objects for IC 348, 19 of which do not have an assigned mass in the dataset and are ignored when comparing the clustering tendencies of the most massive stars and all stars.

The results of running INDICATE on this region are shown in Figure 3.12, clearly showing a central region of relatively higher spatial clustering. The distance to IC 348 is around 300 pc (Parker & Alves de Oliveira, 2017) with an age between $2 - 6$ Myr (Cartwright & Whitworth, 2004; Bell et al., 2013). IC 348 has been previously investigated in Parker & Alves de Oliveira (2017) using the Q-parameter to determine its overall structure. It was found to have a Q-value of 0.85, corresponding to a smooth and centrally concentrated distribution with a radial density exponent of $\alpha = 2.5$. INDICATE is applied to IC 348 and it finds 59.2% of stars are spatially clustered above random and the region has a significant index of 2.3. A median index of 3.2 is found for the region and a median index of 2.6 is found for the 10 most massive stars. A KS test between the 10 most massive stars and the rest of the stars gives a p-value = 0.63, meaning no significant difference in the distribution of the most massive stars and the rest. This is because the most massive stars are spread out over the entire SFR, 5 of them are located within the central region, 4 are found between the edge of this region and the outskirts of the region, with one of the most massive stars found right at the edge of the plot.

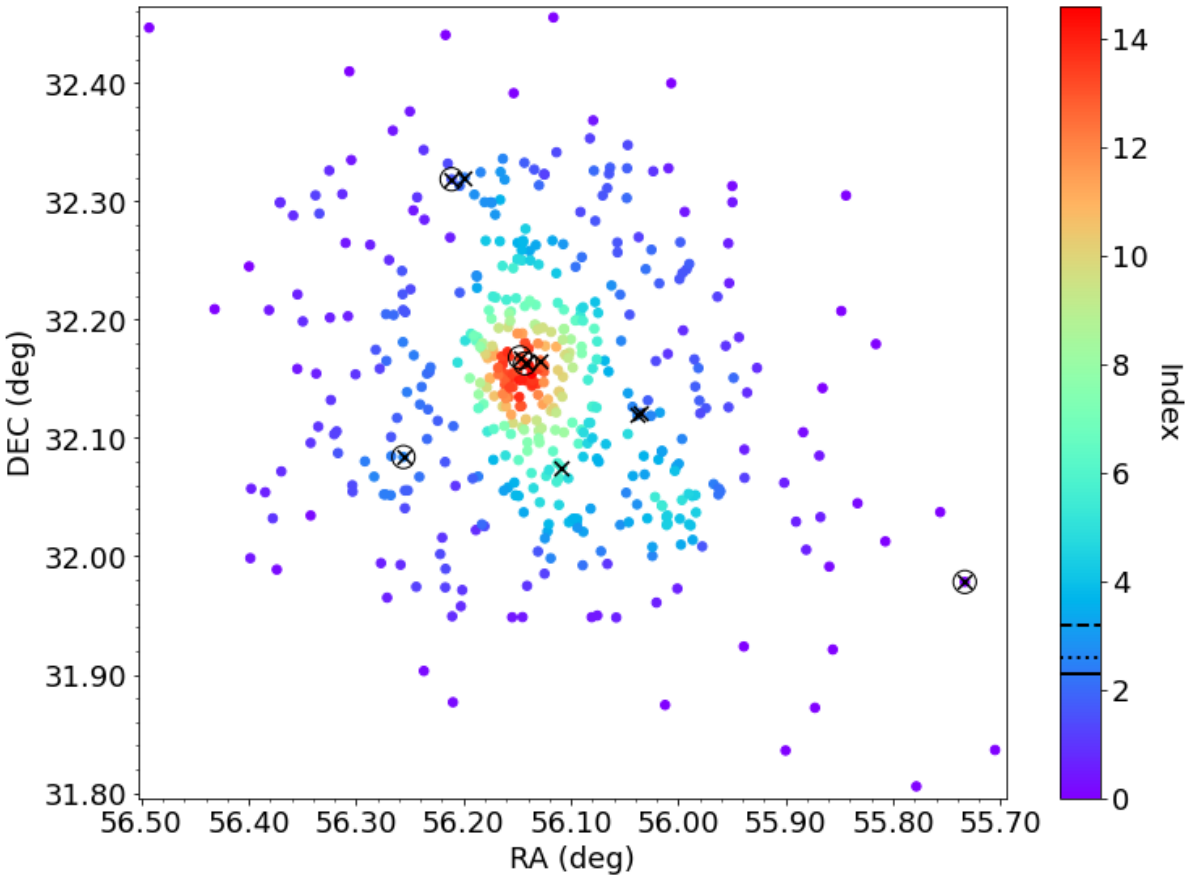


Figure 3.12: Plot of IC 348. The ten most massive stars are highlighted with crosses, the five most massive are circled. The significant index from INDICATE is shown by the solid black line in the colour bar, the median index for all the stars is shown with the dashed black line and the median index for the 10 most massive stars is shown with the dotted black line.

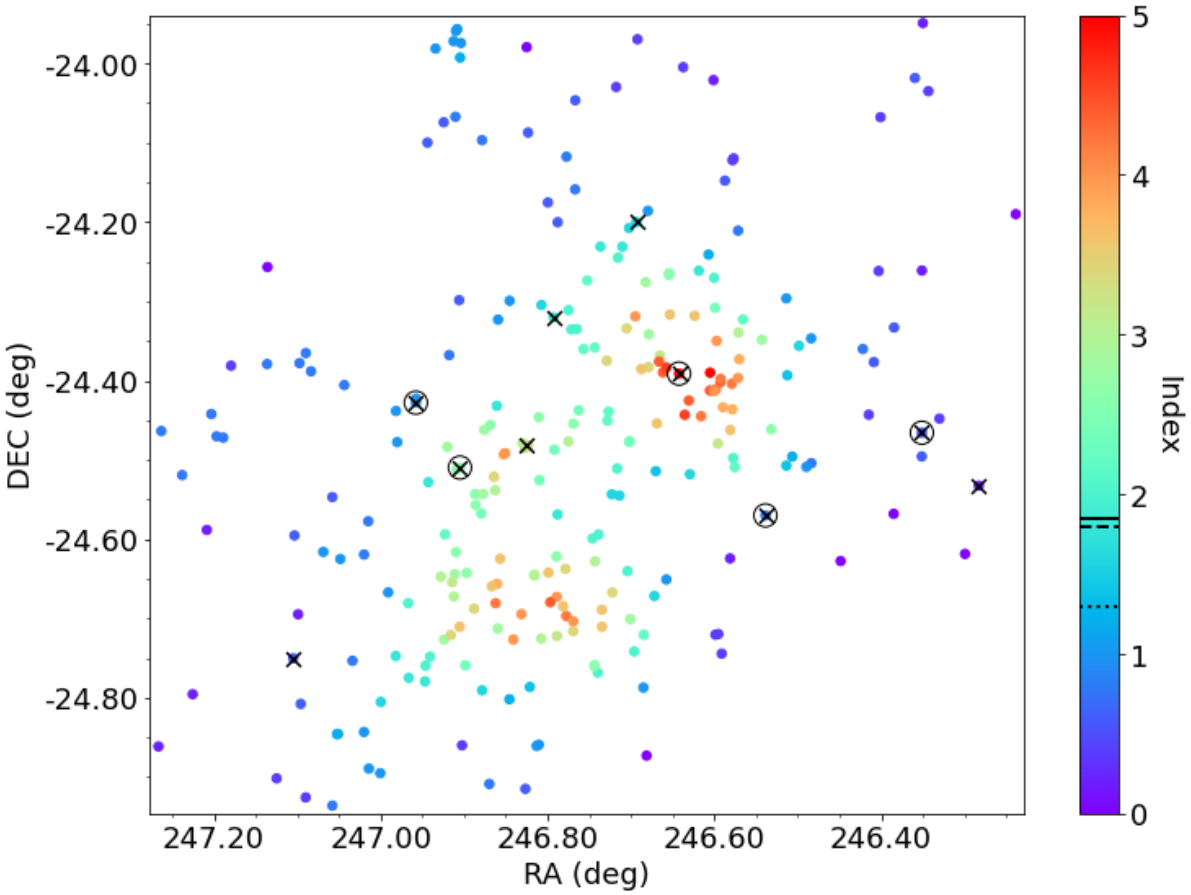


Figure 3.13: Plot of ρ Ophiuchi. The ten most massive stars are highlighted with crosses, the five most massive are circled. The significant index from INDICATE is shown by solid black line in the colour bar, the median index for all the stars is shown by the dashed black line and the median index for the 10 most massive stars is shown using the dotted black line.

3.4.5 ρ Ophiuchi

I use the dataset from Parker et al. (2012) which contains 255 objects. ρ Ophiuchi is located around 130 pc away with an age of around $0.3 - 2.0$ Myr (Parker et al., 2012; Bontemps et al., 2001).

The region is shown in Figure 3.13. INDICATE finds that 39.6% of stars are spatially clustered above random, with a significant index of 2.2. A median index of 1.8 is found for the entire region and 1.3 is found for the 10 most massive stars, with a p-value = 0.82 meaning no significant difference between clustering tendencies of the most massive stars

and the rest of the stars. These results are similar to IC 348 which also has its most massive stars spread out over the SFR. One of the most massive stars is located in an area of high clustering, but the rest have been spread out over relatively lower clustered locations in ρ Ophiuchi.

3.5 Detecting Mass Segregation In Observational Data

INDICATE is applied to the 50 most massive stars in the observational data, repeating the significance calculation 100 times for each region.

I use the same criteria to determine if INDICATE has detected mass segregation as for the synthetic star-forming regions in §3.2.2. A summary of these results is shown in table 3.11.

3.5.1 Taurus

The 50 most massive stars in Taurus have a median index of 1.3 and the 10 most massive a median index of 1.2.

INDICATE finds that 12% of stars have a index above the significant index of 2.5. As the median index for the 10 most massive stars is below the significant index, so no mass segregation is detected.

A KS test returns a p-value = 0.86 meaning no significant difference between the 10 most massive and the entire subset of the 50 most massive stars.

3.5.2 ONC

The median index for just the 50 most massive stars is 1.4 and the median index for the 10 most massive stars is 4.3.

INDICATE finds that 46% of stars have a index above the significant index of 2.1. As the median index for the 10 most massive stars is larger than the significant index

Table 3.11: Results of applying INDICATE to only the 50 most massive stars in the observed star-forming regions. From left to right the columns are: the median index for the entire subset of the 50 most massive stars, the median index for the 10 most massive stars in the subset, the significant index, the percentage of stars with indexes above the significant index and the p-value from a KS test between all 50 stars and the 10 most massive stars INDICATE indexes.

Name	$\tilde{I}(50)$	$\tilde{I}(10)$	$I_{\rm sig}$	$\% > I_{\rm sig}$	p
Taurus	$1.3^{+0.7}_{-0.5}$	$1.2^{+0.2}_{-0.2}$	2.5	12	0.86
ONC	$1.4^{+2.8}_{-1.0}$	$4.3^{+0.1}_{-0.8}$	2.1	46	0.15
NGC1333	$2.9^{+0.9}_{-1.1}$	$2.5^{+0.9}_{-0.6}$	2.1	66	1.00
IC348	$4.2^{+1.0}_{-3.4}$	$2.0^{+2.6}_{-1.6}$	2.0	62	0.67
ρ Ophiuchi	$1.0^{+0.6}_{-0.4}$	$0.8^{+0.3}_{-0.2}$	2.1	0	0.39

INDICATE is detecting signals of mass segregation, i.e. the 10 most massive stars are more affiliated with other high mass stars than the typical high mass star.

A KS test returns a p-value = 0.15 implying no significant difference between the 10 most massive stars and the 50 most massive stars' clustering tendencies.

3.5.3 NGC 1333

A median index of 2.9 is found for all the stars in the subset and a median of 2.5 is found for the 10 most massive stars.

INDICATE finds that 66% of stars have indexes above the significant index of 2.1. The median index for the 10 most massive stars is above the significant index but below the median index for the entire subset meaning that the 10 most massive stars are less clustered with respect to other massive stars than is typical for the region.

A KS test returns a p-value = 1.00 implying no significant difference in the clustering tendencies of the 10 most massive stars and the 50 most massive stars.

3.5.4 IC 348

A median index of 4.2 is found for all stars in the subset and a median index of 2.0 is found for the 10 most massive stars.

INDICATE finds that 62% of stars have an index above the significant index of 2.0. The median index for the 10 most massive stars is the same as the significant index but below the median index for the subset; this may be INDICATE detecting a borderline signal of mass segregation.

A KS test returns a p-value = 0.67 implying no significant difference in the clustering tendencies of the 10 most massive stars and the entire subset.

3.5.5 ρ Ophiuchi

A median index of 1.0 is found for all the stars in the subset and a median index of 0.8 is found for the 10 most massive stars.

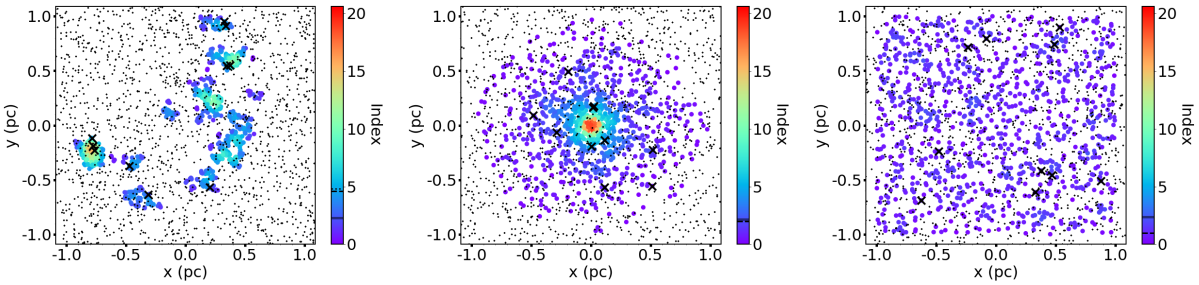
INDICATE finds that no stars have an index above the significant index of 2.1. As the median index of the 10 most massive stars is below the significant index no mass segregation is detected.

A KS test returns a p-value = 0.39 implying no significant difference in the clustering tendencies of the 10 most massive stars compared to the overall region.

3.6 Poisson Control Field

The effect of changing the control field is shown in Figure 3.14. Instead of using an evenly spaced control grid with a uniform field to find the significant index I instead used a Poisson control field, then found the significant index using a Poisson distribution.

The INDICATE results using a Poisson control field as shown in Figure 3.14 are similar to when an evenly spaced control field is used with a uniform distribution to



(a) INDICATE, $D_{\rm f} = 1.6$ (b) INDICATE, $\alpha = 2.0$ (c) INDICATE, uniform

Figure 3.14: INDICATE using a Poisson distribution as the control field for the synthetic regions used in the main work. From left to right (a) is the substructured region with fractal dimension $D_{\rm f} = 1.6$, (b) is the centrally concentrated, smooth distribution with density exponent $\alpha = 2.0$ and (c) shows the uniform distribution. The control field is extended beyond the data to remove edge effects when measuring to the $5^{\rm th}$ nearest neighbour. The most massive stars are shown by the black crosses in all the panels and the small black points are the Poisson control field. The colour map has been scaled based on the index results of the smooth, centrally concentrated distribution.

determine the significant index. I calculated the significant index for each region using 20 different Poisson distributions of the same number density as the data.

The number of stars clustered above random in the fractal distribution increased to 86.8% from 82.2%.

The number of stars clustered above random for the radial and uniform distributions is now 46.4% and 0.1%, increasing from 44.4% and 0.0%, respectively.

3.7 Significant Index Calculations

I calculated each of the synthetic regions' significant index using 100 repeats and found that the difference between single run calculations or using repeats is negligible, unless the overall sample size is small such as restricting the sample size to the 50 most massive stars when searching for classical mass segregation using INDICATE. In these cases it is strongly encouraged to use repeats to reduce statistical fluctuations.

For the substructured synthetic region (of fractal dimension $D_{\rm f} = 1.6$) with 100

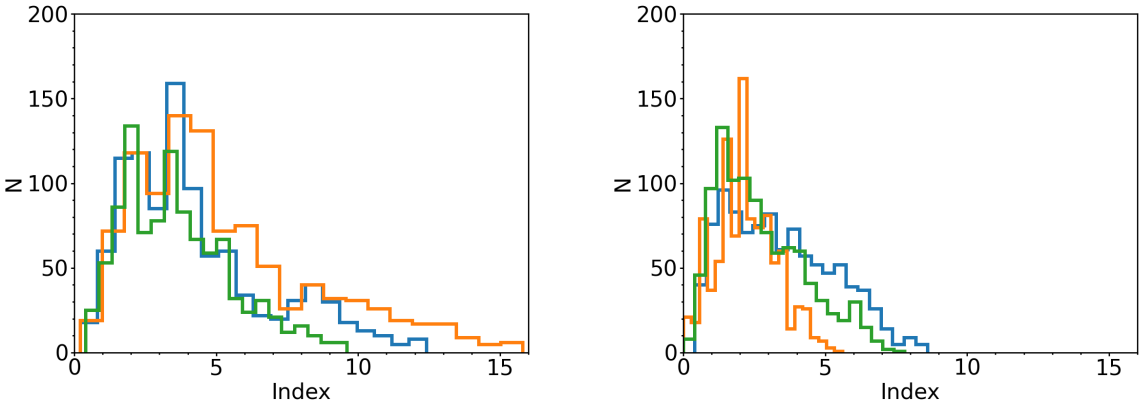
repeats a significant index of 2.3 is found, which is the same as for one run. The percentage of stars clustered above random also stays the same at 82.2%. The median INDICATE index for all the stars in the region is also the same at 4.4. For the median index of the 10 most massive stars it has increased from 4.5 to 4.6 with a p-value = 0.86. This is basically the same as for one iteration. When swapping the most massive stars with the stars with the highest index I find the same results as for one iteration and the same results are also found when the most massive stars are swapped with the most central stars.

The smooth, centrally concentrated synthetic region (with density exponent $\alpha = 2.0$) has a significant index of 2.3 when calculated using 100 repeats, which is the same as when using one iteration. The percentage of stars with indexes above the significant index has also stayed the same at 44.1%, as have the median indexes for the entire region and the 10 most massive stars, with respective values of 1.8 and 2.0. The results of the KS test are also the same, returning a p-value = 0.55. I find the same results as a single significance calculation when running repeats after swapping the 10 most massive stars with stars that have the greatest INDICATE index and also when the 10 most massive stars are swapped with the 10 most central stars.

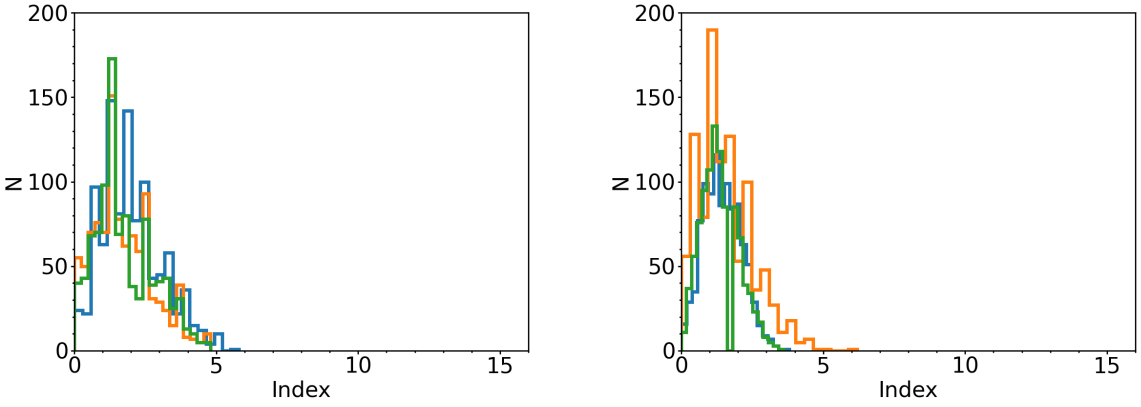
The uniform synthetic region has a significant index of 2.3 with 100 repeats which is different for the significant index for one iteration which was 2.4. The percentage of stars above the significant index has gone up from 0% to 0.1%. The median index for the region is 1.0 and for the 10 most massive stars is 0.8, the same as for 1 iteration. A KS test also returns the same results as for one iteration with a p-value = 0.68. Running 100 repeats to calculate the significant index after the 10 most massive stars are swapped with the stars which have the greatest INDICATE indexes gives the same result as for one iteration (just with a different significant index of 2.4), INDICATE also finds the same result when swapping the 10 most massive stars with the 10 most central stars.

3.8 Index Distributions

The distribution of the INDICATE index was also investigated. This is shown in Figure 3.15 and shows the index distributions for different fractal dimensions. The more substructured a region the larger their range of index values. Figure 3.16 shows the distribution of INDICATE indexes for the synthetic SFRs in this chapter. The substructured and radial distributions show a wider range of clustering tendencies when compared to the uniform region.

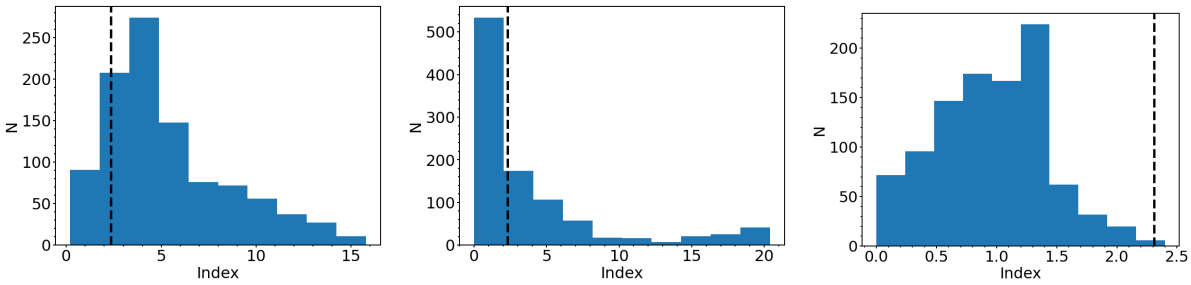


(a) $D_f = 1.6$ (b) $D_f = 2.0$



(c) $D_f = 2.6$ (d) $D_f = 3.0$

Figure 3.15: Histograms showing the INDICATE index distribution for 3 different realisations of different fractal distributions. (a) shows 3 different $D_f = 1.6$ distributions, (b) 3 different $D_f = 2.0$ distributions, (c) 3 different $D_f = 2.6$ distributions and (d) shows 3 different $D_f = 3.0$ distributions. The range of index values increases with more substructure as stars are more likely to have a higher INDICATE index in highly substructured regions.



(a) Substructured distribution with fractal dimension $D_{\rm f} = 1.6$

(b) Smooth, centrally concentrated distribution with density exponent $\alpha = 2.0$

(c) Uniform Distribution

Figure 3.16: Histograms showing the distribution of the INDICATE indexes for each of the synthetic SFRs shown in this chapter. The vertical dashed line represents the significant index for each region.

3.9 Conclusions

I have investigated the performance of the INDICATE method to detect the spatial clustering tendencies in young SFRs. I have also assessed its ability to quantify mass segregation, and have applied it to pre-main sequence stars in nearby SFRs.

I have shown in Figure 3.1 that whilst INDICATE can be used to quantify the clustering tendencies for individual stars in a region it cannot be used to provide any further information on the overall structure of a SFR due to the degeneracy of the INDICATE index across different morphologies. I confirm that when INDICATE is applied to an entire region it can detect significant differences in the local stellar surface density between the 10 most massive stars and the entire population and will find results that are in agreement with $\Sigma_{\rm LDR}$.

When INDICATE is applied to the subset of the 50 most massive stars only it will detect when the 10 most massive stars are more clustered with respect to other massive stars and in most cases will agree with the $\Lambda_{\rm MSR}$ method. However, INDICATE did not detect any significant clustering for the 10 most massive stars in a uniform field when the 10 most massive stars are swapped with the 10 most central stars. Using the chosen criteria for mass segregation INDICATE finds that the 10 most massive stars

are spatially distributed differently than the entire subset of the 50 most massive stars, which is confirmed using a KS test. For this region mass segregation was detected by $\Lambda_{\rm MSR}$.

When applied to nearby SFRs, INDICATE shows that there is a wide range of indexes within individual SFRs (e.g. Taurus in Figure 3.9) and between SFRs (e.g. ρ Oph and the ONC, which have small ranges of indexes and larger ranges of indexes respectively). INDICATE finds the areas of clustering reliably in these regions.

I also quantify the clustering tendencies of the most massive stars in these regions compared to all the stars. In the ONC, I find significant differences in the clustering tendencies of the 10 most massive stars when compared to all the stars' finding that the 10 most massive stars are in areas of greater local stellar surface density than the average star in the region.

The other observed regions show no significant differences in the clustering tendencies of the 10 most massive stars compared to all stars in the region. This is due to the 10 most massive stars in these regions being spread out (see figures 3.9, 3.11, 3.12 and 3.13), resulting in a wider range of INDICATE indexes.

In summary INDICATE is a robust clustering metric that allows the significance of any clustering to be determined on a star-by-star basis, giving results that are consistent with other methods. When applied to just a subset of the 50 most massive stars INDICATE is able to detect mass segregation, with results in general agreement with $\Lambda_{\rm MSR}$. Due to the robustness of INDICATE at quantifying the clustering of stars in synthetic sets, and its agreement with other methods, it can be reliably used to infer the physics of SFRs.

4

The Evolution of Phase Space Densities in Star-Forming Regions

Summary

The multi-dimensional phase space density (both position and velocity) of star-forming regions (SFRs) may encode information on the initial conditions of star and planet formation. Winter et al. (2020) developed a new metric based on the Mahalanobis distance has been used to claim that hot Jupiters are more likely to be found around exoplanet host-stars in high 6D phase space density, suggesting a more dynamic formation environment for these planets.

However, later work showed that this initial result may be due to a bias in the age of hot Jupiters and the kinematics of their host stars (discussed in Chapter 5) (Adibekyan et al., 2021; Mustill et al., 2022).

The origin of hot Jupiter host stars' having a high Mahalanobis density in 6D needs to be determined. If the origin of the overdensity can be linked to the initial conditions of star formation then it could offer an exciting new way of investigating the star formation process. Firstly, the physical interpretation of the Mahalanobis density needs investigating as due to how it renormalises and rescales the data (which removes units) any physical interpretation of the Mahalanobis metrics is inherently difficult (De Maesschalck et al., 2000; Mustill et al., 2022).

In this chapter I test the ability of the Mahalanobis distance and density to differentiate more generally between SFRs with different morphologies by applying it to static regions that are either substructured or smooth, and centrally concentrated. I find that the Mahalanobis distance is unable to differentiate between different morphologies.

I analyse the Mahalanobis density evolution of N-body simulations and show that the initial conditions of the N-body simulations cannot be constrained using only the Mahalanobis distance or density. Furthermore, I find that the more dimensions in the phase space the less effective the Mahalanobis density is at distinguishing between different initial conditions. I show that a combination of the mean three-dimensional (x, y,

z) Mahalanobis density and the Q-parameter for a region can constrain its initial virial state. However, this is due to the discriminatory power of the Q-parameter and not from any extra information imprinted in the Mahalanobis density.

4.1 Introduction

Recently, in an attempt to quantify the phase space densities of exoplanet host stars Winter et al. (2020) developed the Mahalanobis density, a new application of the Mahalanobis distance (Mahalanobis, 1936).

The Mahalanobis distance has been used in astronomy for classifying objects, for example in Siegal & Griffiths (1974) it is used to analyse and classify different types of asteroid impact craters and in Jakimiec et al. (1991) it was used to classify sunspots into groups.

Due to the differing dimensions, and units of very different scale (i.e. length in pc and velocity in $\mathrm{km\,s^{-1}}$) making multivariate comparisons can be difficult. However, the Mahalanobis distance makes multivariate comparisons possible over wide ranges of dynamical scales by rescaling the axes and removing the units. Winter et al. (2020) used this method to develop the Mahalanobis density and use it to propose the hypothesis that host stars in high phase space densities are more likely to have hot Jupiter planets (planet with mass $> 50\,M_{\oplus}$ and $a < 0.2$ AU) around them compared to the lower phase space densities. However, Mustill et al. (2022) shows that this result may be due to a bias from the peculiar velocities of the stars. When the peculiar velocities of the stars are accounted for, there is no longer an excess of hot Jupiters in high 6D (x, y, z, Vx, Vy, Vz) phase space densities.

Irrespective of the ongoing debate surrounding the application of the Mahalanobis distance to exoplanet host stars, in this chapter I aim to test this metric more generally by applying it to synthetic static regions and assess its performance in quantifying phase

Table 4.1: This table shows the different initial conditions of the simulations. For each of these initial conditions 10 simulations are run for 10 Myr. From left to right the columns are the initial fractal dimension of the region, the number of stars, the initial virial ratio and the initial radius of the simulations in pc.

Fractal Dimension	$N_\star$	Virial Ratio	Radius (pc)
$D_\mathrm{f} = 1.6$	1000	0.1, 0.9	1, 5
$D_\mathrm{f} = 3.0$	1000	0.1, 0.9	1, 5

space structures of N-body simulations of SFRs.

4.1.1 Simulation Setup

The simulations are setup as described in §2.1. For the subvirial simulations I scale the velocities of stars, so the simulation has a virial ratio of $\alpha_\mathrm{vir} = 0.1$. Supervirial simulations have stellar velocities scaled such that the virial ratio is $\alpha_\mathrm{vir} = 0.9$, with a virial ratio of $\alpha_\mathrm{vir} = 0.5$ corresponds to virial equilibrium. The initial conditions for the simulations are summarised in Table 4.1.

Masses are assigned using the Maschberger IMF (see §2.1.4) with a lower mass limit of 0.1 $M_\odot$, upper mass limit of 50.0 $M_\odot$ and a mean mass of 0.2 $M_\odot$ (Maschberger, 2013).

4.2 Results

In this section I show the results of the Mahalanobis distance applied to both static and N-body simulations of SFRs with various initial conditions. I present the 3D and 6D Mahalanobis densities calculated in the N-body simulations and compare the evolution of the Mahalanobis density over time to the other methods for quantifying spatial and kinematic distributions in SFRs.

4.2.1 Static Regions

First the Mahalanobis distances between stars and the average point in a region is calculated, $\vec{\mu}$, I then calculate the average Mahalanobis distance for each respective region ($\bar{M}_{\rm d}$) in the sets of synthetic and static star clusters, with each set having a different structural parameter. Each region in the set consists of 1000 stars. I calculate $\bar{M}_{\rm d}$ for substructured regions with fractal dimensions $D_{\rm f} = 1.6, 2.0, 2.6, 3.0$ and clusters with radial density profile indexes, $\alpha = 0.0, 1.0, 2.0, 2.5$. I also show the results of a set of 100 Plummer spheres which have a radial density profile described by Equation 2.2 which have an equivalent radial density index of 2.5.

Figure 4.1 shows the mean of the means for the $M_{\rm d}$ (the Mahalanobis distance of each star to its region's averages in the 3D phase space) for the 100 clusters in each set of initial conditions. I first calculate the mean Mahalanobis distance in each of the 100 regions in the set and then calculate the mean of those means. The error bars show the standard deviation of the mean of the mean Mahalanobis distances found in each of the regions in a particular set. Figure 4.1 clearly shows that $\bar{M}_{\rm d}$ is degenerate across a wide range of morphologies, and therefore, $\bar{M}_{\rm d}$ calculated in the 3D phase space is unable to differentiate between the different morphologies. There is much more spread in the values for the Plummer sphere models compared to the radial and fractal models, with the fractals having the smallest spread of $\bar{M}_{\rm d}$ and the radial regions sitting between the two. This is due to Plummer spheres being formally infinite in extent, so the calculation occasionally has to normalise over very distant stars.

4.2.2 *N*-body Results

Figure 4.2 shows the mean of the mean Mahalanobis distances (calculated in both 3D and 6D) found for 10 different N-body simulations with initial fractal dimension $D_{\rm f} = 1.6$ and 1 pc radii. The left-hand panel shows the results for subvirial (collapsing) regions,

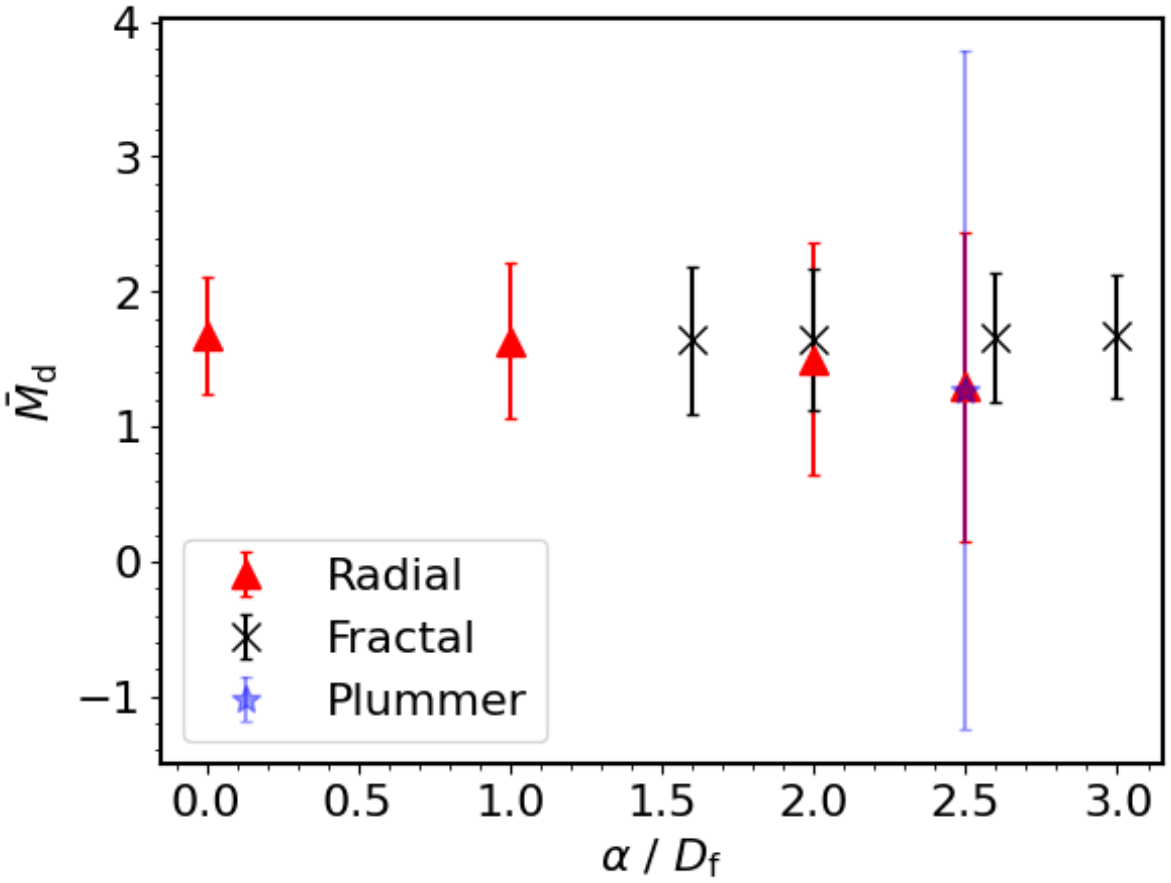
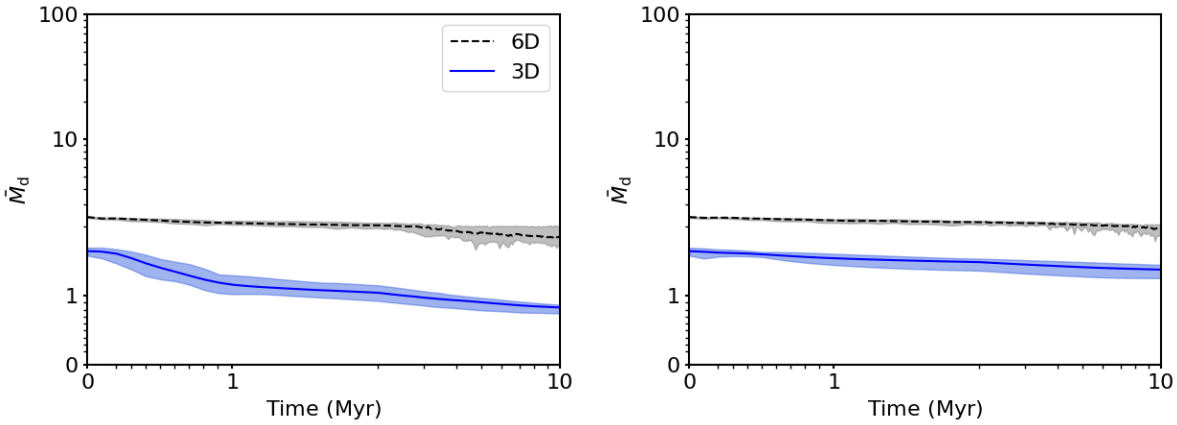


Figure 4.1: Mean of the mean Mahalanobis distances calculated in the 3D phase space for sets of 100 different SFRs plotted against the structural parameter used to make the sets. The red triangles are the smooth, centrally concentrated radial regions, the purple star (on top of the red triangle with the structural parameter equal to 2.5) is the Plummer sphere and the black crosses are the fractal regions. The error bars show a single standard deviation.

and the right-hand panel shows the results for the supervirial (expanding) regions. In both the sub- and supervirial cases there is a decrease in the Mahalanobis distance over time when calculated in both the 3D and 6D phase spaces. For the initially subvirial simulations the 3D Mahalanobis distance swiftly decreases at the start and then continues to decrease for the rest of the simulation but at a slower rate. For the supervirial regions there is a less pronounced decrease in $\bar{M}_{\rm d}$ compared to the initially subvirial regions. The $\bar{M}_{\rm d}$ calculated in the 6D phase space shows more modest decrease over time for both initially sub- and supervirial simulations.

Figure 4.3 shows the mean Mahalanobis density ($\bar{\rho}_{\rm m,20}$) calculated in both the 3D and 6D phase spaces for two sets of 10 (one subvirial and the other supervirial) simulations with an initial fractal dimension of $D_{\rm f} = 1.6$ and initial radii of 1 pc. The highest Mahalanobis densities are found in the 3D phase space (x, y, z) for the supervirial simulations; however these large final values are only present for a few of the simulations. In the 3D phase space the Mahalanobis density increases in the first 2-4 Myr, after which the density stays the same for the rest of the run time. This is most likely due to the early dynamical interactions of stars; as they move closer to each other the stars' Mahalanobis densities will increase. In §4.2.4 I show the relationship between the Mahalanobis distance and density for the high density simulations with and without substructure. I show this for both the 3D and 6D phase spaces.

Initially, in the subvirial simulations, the 6D Mahalanobis density decreases for the first 1 Myr and then stays the same until around 5 Myr where it then starts to increase again. For the supervirial simulations this initial decrease in the 6D phase space happens more rapidly than in the subvirial simulations. It also does not reach the low densities that the subvirial regions attain. In the supervirial regions, the stars will expand together in co-moving groups; therefore they will have similar positions but can still have velocities that exhibit kinematic substructure. In contrast, in the subvirial simulations the stars interact more and erase this substructure. The difference in the velocities explains why



(a) $D_f = 1.6$, subvirial, 1 pc

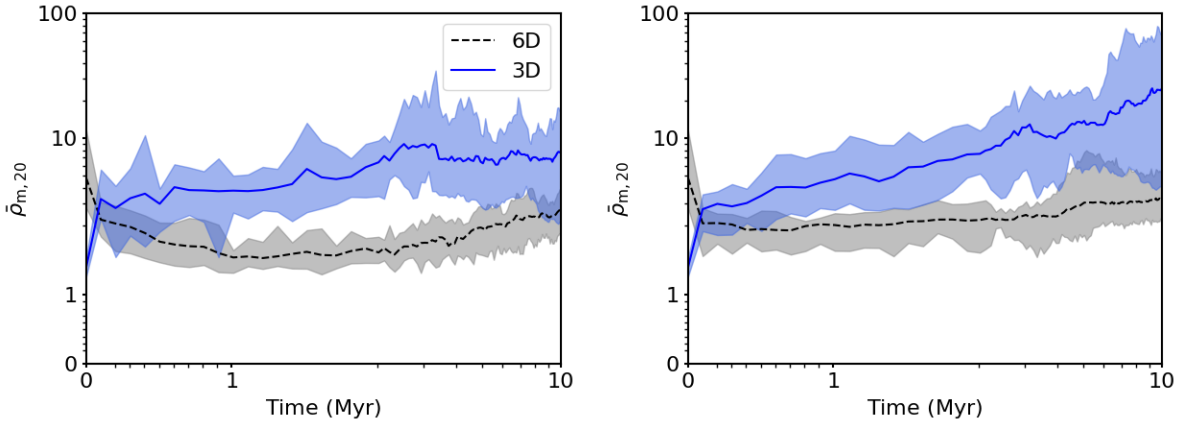
(b) $D_f = 1.6$, supervirial, 1 pc

Figure 4.2: Plots of the mean Mahalanobis distance calculated in both the 3D and 6D phase spaces against time for regions with both high initial volume densities and high degrees of substructure (i.e. fractal dimension $D_f = 1.6$ with radii of 1 pc) consisting of 1000 stars. The shaded areas show the range of mean Mahalanobis distances found across all 10 of the simulations at the current time. The solid lines show the mean of the mean Mahalanobis distances across all 10 simulations. The blue area and solid blue line shows the 3D phase space and the black dashed line and the grey area show the same but for the 6D phase space.

the 6D Mahalanobis density is much lower than the 3D density. Some of the supervirial regions attain higher 6D Mahalanobis densities compared to the subvirial regions at the end of the 10 Myr.

The Mahalanobis density and distance is also calculated for low density simulations for simulations with $D_f = 1.6$ and $D_f = 3.0$, where the initial radii regions are 5 pc (with a mean number density of around 3 stars pc^{-3} and a mean stellar mass density of around 1.6 $M_\odot$ pc^{-3}). The evolution of $\bar{M}_d$ is almost identical to the its evolution in the high density simulations, both in 3D and 6D phase spaces.

I show the evolution of the 3D and 6D Mahalanobis densities for these more diffuse simulations in Figure 4.4, which shows $\bar{\rho}_{m,20}$ plotted against time for the same substructured regions with fractal dimension $D_f = 1.6$ with initial radii of 5 pc. For the 3D phase space the same trends as in Figure 4.3 are seen but there is a slight difference for the 6D phase space. Now $\bar{\rho}_{m,20}$ decreases in the subvirial simulations as the regions evolve, and



(a) $D_f = 1.6$, subvirial, 1 pc

(b) $D_f = 1.6$, supervirial, 1 pc

Figure 4.3: The mean Mahalanobis density against time for each of the 10 subvirial (left-hand panel) and supervirial simulations (right-hand panel) with fractal dimension $D = 1.6$ and radii of 1 pc. The simulations consist of 1000 stars. The blue shaded area shows the minimum and maximum mean Mahalanobis density (in the 3D phase space) found across all 10 of the simulations at the current time. The solid blue line shows the mean of the means for the Mahalanobis density in the 3D phase space. The grey shaded area and the dashed black line shows the same but for the Mahalanobis density calculated in the 6D phase space. The mean number densities of the 10 simulations is around 314 stars pc^{-3} with a mean stellar mass density of around 201 $M_\odot$ pc^{-3}.

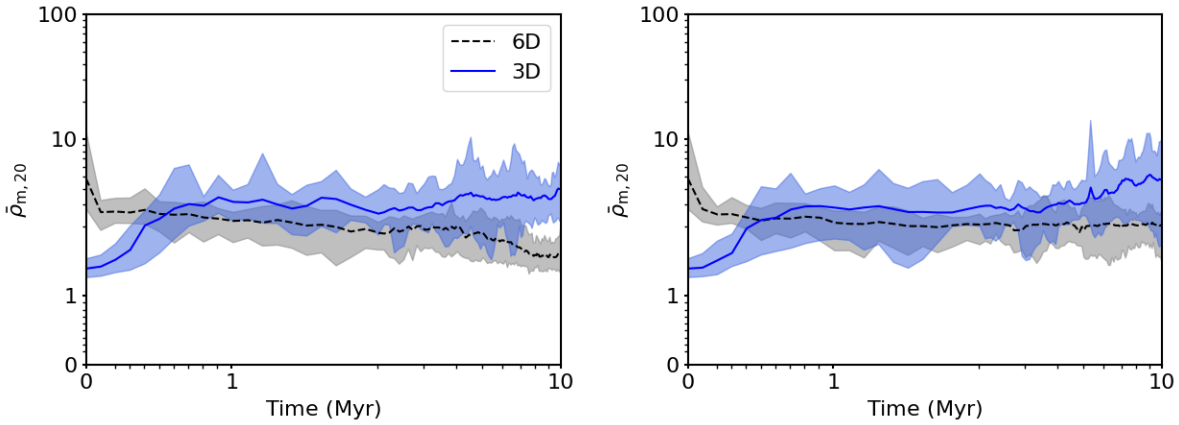
the supervirial simulations show a steady Mahalanobis density after 1 Myr.

I show the evolution of the Mahalanobis distance (measured between the points and the mean values of the phase in each snapshot), and Mahalanobis density in the simulations that have no primordial substructure (i.e. they are uniform spheres at $t = 0$ Myr).

Figure 4.5 shows the Mahalanobis distance over time for regions with a fractal dimension $D_{\rm f} = 3.0$ with radii of 1 pc. For the 3D phase space there is a decrease over time for the sub- and supervirial simulations.

Comparing these results to Fig. 4.2 it is seen that the $D_{\rm f} = 3.0$ regions' Mahalanobis distances decrease at a slower rate compared to regions with initially more substructure. This behaviour also results in a slightly greater 3D Mahalanobis distance being measured for regions with fractal dimension $D_{\rm f} = 3.0$ at 10 Myr. However, this is not seen in the 6D phase space Mahalanobis distances which show little change over the 10 Myr in the simulations. Also, very little difference is seen when comparing the 6D Mahalanobis distances between the sub- and supervirial simulations.

Figure 4.6 shows the 3D and 6D $\bar{\rho}_{\rm m,20}$ against time for the simulations with no primordial substructure (with an initial fractal dimension $D_{\rm f} = 3.0$ and radii 1 pc). The left-hand panel shows $\bar{\rho}_{\rm m,20}$ against time for the subvirial simulations. It shows the same increase in Mahalanobis density as the $D_{\rm f} = 1.6$ simulations but is lacking the initial decrease seen in the 6D $\bar{\rho}_{\rm m,20}$ for the $D_{\rm f} = 1.6$ simulations. At around 0.5 Myr a decrease in the 3D Mahalanobis density is seen, then a second period of increasing Mahalanobis density is seen around 0.9 Myr before attaining a steady Mahalanobis density for the rest of the simulations' run time. This initial increase is due to the region collapsing and stars moving closer to each other which raises the 3D Mahalanobis density. What stops it increasing further is likely the dynamical interactions causing stars to move further away from each other. Once this initial dynamical stage settles down the density can increase again due to stars being close to each other near the centre of the region.



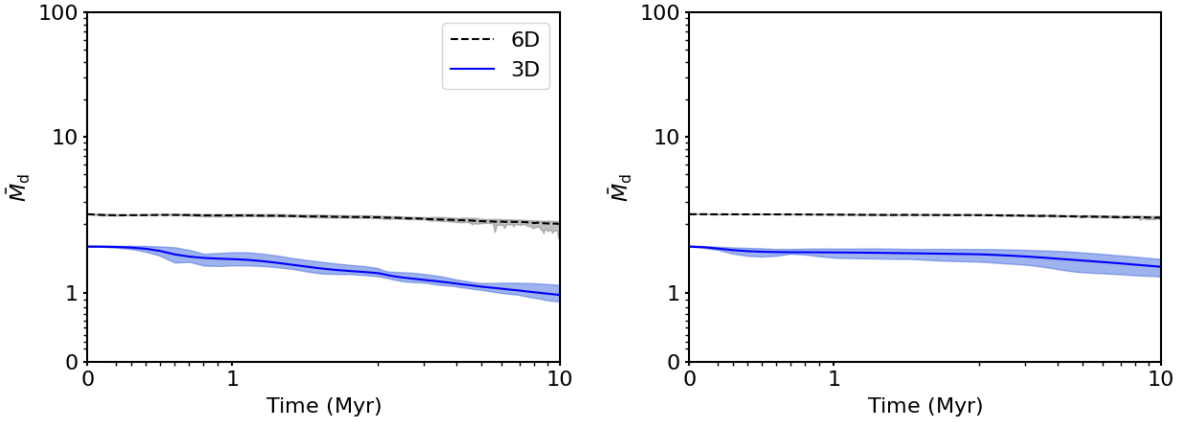
(a) $D_f = 1.6$, subvirial, 5 pc

(b) $D_f = 1.6$, supervirial, 5 pc

Figure 4.4: Plots showing the mean Mahalanobis density against time for each of the 10 subvirial (left-hand panels) and supervirial (right-hand panels) simulations with fractal dimension $D_f = 1.6$ and radii of 5 pc. These simulations have a low initial stellar number density with a mean around 3 stars pc^{-3} and a mean stellar mass density of around 1.6 $M_\odot$ pc^{-3}. The shaded blue area shows the minimum and maximum mean Mahalanobis density found across all 10 simulations in the 3D phase space. The solid blue line shows the mean of the means Mahalanobis density against time. The shaded grey area and the dashed black line show the same but for the 6D phase space.

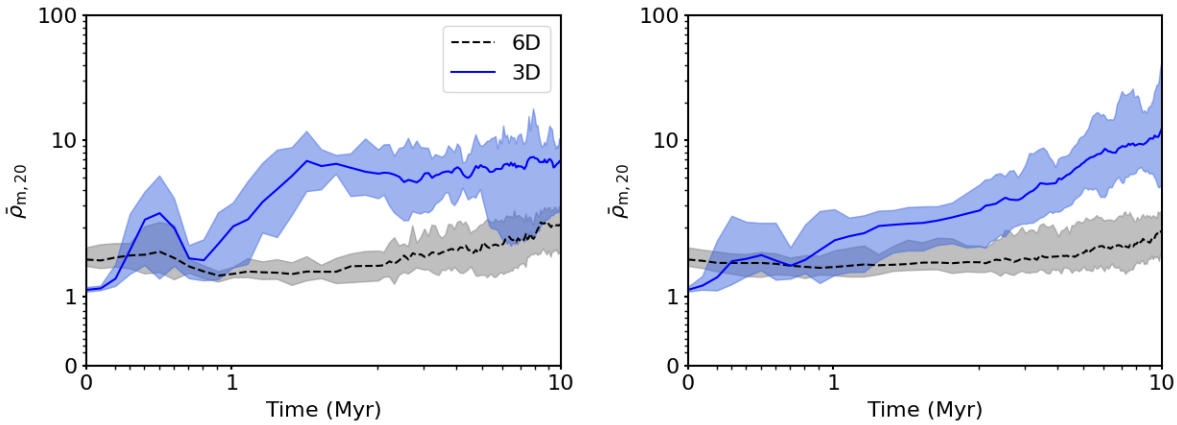
The right-hand panel of Figure 4.6 shows the 3D and 6D Mahalanobis density calculated for regions that are initially supervirial. The "bump"-like feature is much smaller for the 3D phase space calculations than in the initially subvirial simulations. The decrease in the bump compared to the subvirial simulations is due to the fact that the stars are constantly and continuously moving away from each other, meaning that the slight increase that is still present is due to small groupings of stars clumping together. Similar behaviour is seen for the 6D phase space in both the subvirial and supervirial simulations.

The 3D and 6D Mahalanobis densities for simulations with $D_f = 3.0$ and initial radii of 5 pc are shown in Figure 4.7. These low density simulations have mean number density of around 1 star pc^{-3}, or mean stellar mass density of around 0.7 $M_\odot$ pc^{-3}. I find similar results to the more dense subvirial simulations, where the 3D Mahalanobis density clearly traces the bump of the collapse, which then decreases as stars move apart.



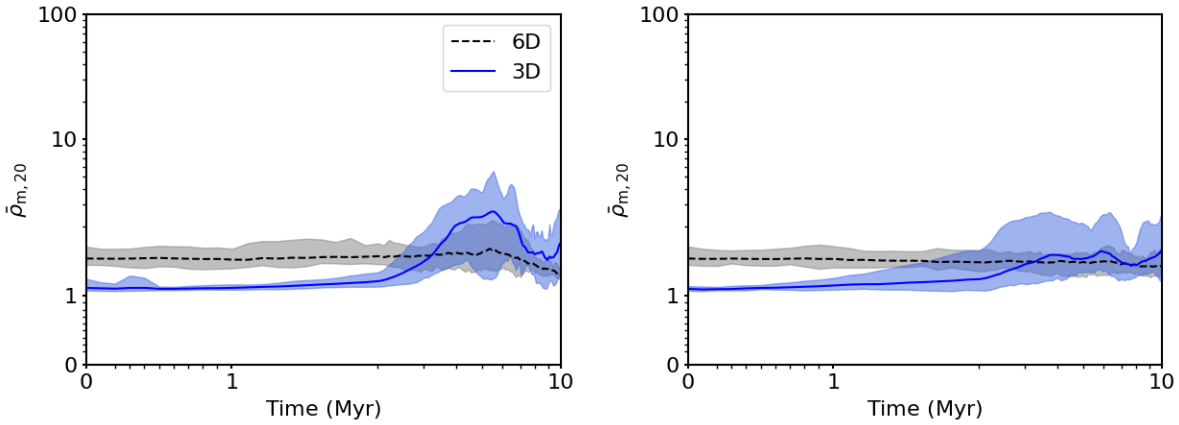
(a) $D_f = 3.0$, subvirial, 1 pc (b) $D_f = 3.0$, supervirial, 1 pc

Figure 4.5: Plots of the mean Mahalanobis distance from each star to the average in simulations without primordial substructure, i.e. a fractal dimension of $D_f = 3.0$, over time. The shaded blue area and solid blue line show the minimum and maximum mean Mahalanobis distance found across the 10 simulations and the solid blue line shows the mean of the means across all 10 simulations, respectively. The Mahalanobis distance is calculated in the 3D phase space for the blue area and line and calculated in the 6D phase space, shown by the grey shaded area and the black dashed line.



(a) $D_f = 3.0$, subvirial, 1 pc (b) $D_f = 3.0$, supervirial, 1 pc

Figure 4.6: The mean Mahalanobis density against time for each of the 10 subvirial (left-hand panel) and supervirial (right-hand panel) simulations without primordial substructure (fractal dimension $D_f = 3.0$) and radii of 1 pc. The shaded blue area and solid blue line show the range of mean Mahalanobis densities calculated in the 3D phase space and the mean of the means found across all 10 simulations, respectively. The grey shaded area and the dashed black line show the same but for the 6D phase space.



(a) $D_f = 3.0$, subvirial, 5 pc

(b) $D_f = 3.0$, supervirial, 5 pc

Figure 4.7: The mean Mahalanobis density against time for each of the 10 subvirial (left-hand panel) and supervirial (right-hand panel) simulations without primordial substructure (fractal dimension $D_f = 3.0$) with radii of 5 pc. The shaded blue area and solid blue line show the range of mean Mahalanobis densities calculated in the 3D phase space and the mean of the means found across all 10 simulations, respectively. The grey shaded area and the dashed black line show the same but for the 6D phase space.

The time the bump occurs is delayed by several Myr compared to the higher density regions, due to the longer dynamical time scales.

4.2.3 Comparison to other methods of quantifying structure

I now plot the 3D and 6D Mahalanobis densities against other measures of quantifying structure in star-forming regions. Figure 4.8 shows $\bar{\rho}_{m,20}$ plotted against the established methods of Λ_{MSR}, Q and Σ_{LDR} for the simulations with initial fractal dimension $D_f = 1.6$ and radius 1 pc. For the initially subvirial simulations the $\bar{\rho}_{m,20}$ values stay below 12 for the first 5 Myr whereas the supervirial simulations can achieve much higher values. This is due to the stars in supervirial regions forming small groupings as the region expands which causes an increase in the Mahalanobis density. In contrast, for the subvirial regions more stars are interacting with each other, which erases spatial and kinematic substructure and also ejects stars (Schoettler et al., 2020). As I am measuring the mean Mahalanobis density it is sensitive to a small number of stars being ejected which

manifests as a decrease in the mean Mahalanobis density for the subvirial simulations.

I first show the Mahalanobis density versus the amount of mass segregation as defined by $\Lambda_{\rm MSR}$ in Figure 4.8(a) and (b) (Allison et al., 2009). For the subvirial simulations mass segregation is detected for 6 of the 10 simulations at 1 Myr and only one simulation has mass segregation present at 5 Myr, with $\Lambda_{\rm MSR} > 2$. The reason for the dissipation in the amount of mass segregation is due to the ejection of massive stars from unstable Trapezium-like systems (Allison et al., 2010; Allison & Goodwin, 2011; Parker & Goodwin, 2015). In the supervirial simulations (see panel (b)) one region becomes mass segregated at 1 Myr and another at 5 Myr. If the cluster splits in two, with the most massive stars located in one of the halves then $\Lambda_{\rm MSR}$ can increase to the value shown in Figure 4.8(b) of around 5.5. As discussed in Parker et al. (2014), this is because the massive stars generally do not interact with each other as they do in the subvirial simulations where there is more mixing resulting in any structure in the phase spaces being erased.

The supervirial simulations display a wider spread in the Mahalanobis densities meaning that the plot of $\bar{\rho}_{\rm m,20}$ versus $\Lambda_{\rm MSR}$ can be used to distinguish between different initial virial states after at least 5 Myr of dynamical evolution.

The clearest distinction between different times in the simulations comes when $\bar{\rho}_{\rm m,20}$ is combined with the Q-parameter. Figure 4.8(c) and (d) show this clearly for both the subvirial simulations and the supervirial simulations. The plots also show that, as expected, the supervirial simulations maintain substructure for longer, with some regions maintaining traces of substructure for 5 Myr as measured using Q (i.e. $Q < 0.8$).

Panels (e) and (f) of Figure 4.8 show $\bar{\rho}_{\rm m,20}$ plotted against the relative local surface density ratio, $\Sigma_{\rm LDR}$. It finds an increase in the local surface density for the 10 most massive stars compared to all stars in the region for both sub- and supervirial simulations.

Interestingly the simulations with the highest local surface density around the 10 most massive stars do not necessarily have the highest Mahalanobis densities. This is

likely due to the local surface density being calculated on the plane of the sky whereas the Mahalanobis density is being calculated for the full 3D phase space.

The supervirial regions display high Mahalanobis densities at later times, it is this difference that can be used, and the different evolution of the Q-parameter and $\Lambda_{\rm MSR}$ to distinguish between initial conditions after several Myr of dynamical evolution.

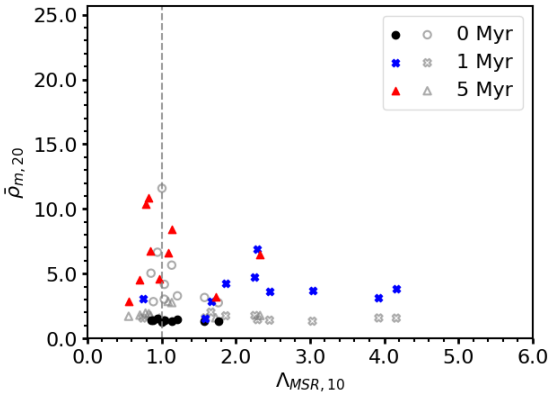
Using the 6D Mahalanobis density does not improve the diagnostic ability of the metric. I find that the range of $\bar{\rho}_{\rm m,20}$ decreases when calculated in 6D. Meaning that the $\bar{\rho}_{\rm m,20}$ values overlap making differentiating between different snapshots and virial states impractical.

I now show the same plots but for simulations with little to no primordial spatial or kinematic substructure. Figure 4.9 shows the mean 3D and 6D Mahalanobis densities plotted against the established methods for simulations that have an initial fractal dimension $D_{\rm f} = 3.0$ and radius 1 pc. The Mahalanobis densities for these simulations increase over time, with supervirial simulations having higher Mahalanobis densities compared to the subvirial simulations after 10 Myr of evolution.

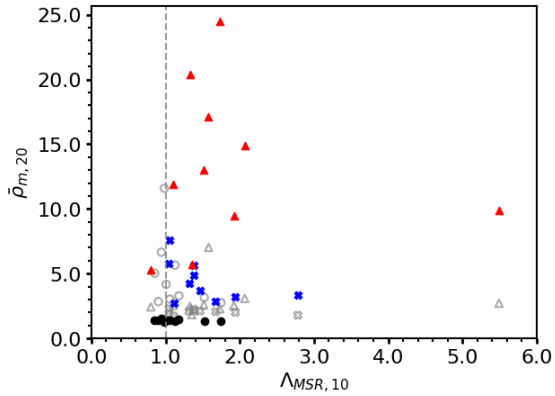
Panels (a) and (b) show $\bar{\rho}_{\rm m,20}$ plotted against $\Lambda_{\rm MSR}$ for the 10 simulations. For the subvirial simulations mass segregation detected in three of the 10 simulations, for the supervirial simulations mass segregation is detected in two of the 10 simulations (recall that the threshold for declaring mass segregation is $\Lambda_{\rm MSR} > 2$).

As for the highly substructured simulations ($D_{\rm f} = 1.6$), the plot of the mean Mahalanobis density when combined with the Q-parameter gives the clearest distinction between the different snapshots. Panels (e) and (f) show $\bar{\rho}_{\rm m,20}$ against $\Sigma_{\rm LDR}$. It can see that the 10 most massive stars can end up in a wide range of local surface density ratios The subvirial simulations have a wider range of values, with $\Sigma_{\rm LDR}$ between 0.1 and 10, whereas the supervirial simulations all finish with $\Sigma_{\rm LDR} > 1$.

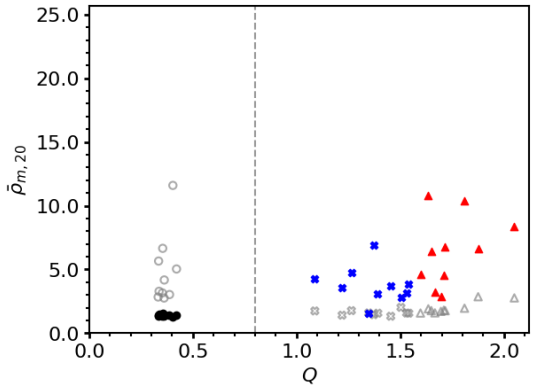
The grey markers in Figure 4.9 show the 6D Mahalanobis densities against the established methods. I find once again that the spread in the Mahalanobis densities has



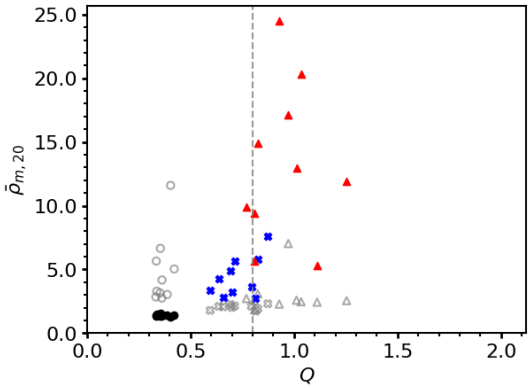
(a) $\bar{\rho}_{\mathrm{m},20}$ vs Λ_{MSR}, subvirial, 1 pc



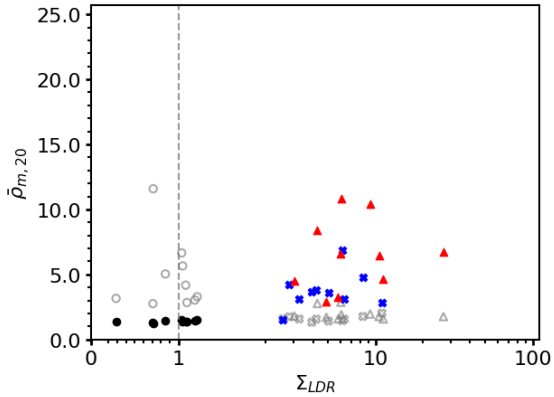
(b) $\bar{\rho}_{\mathrm{m},20}$ vs Λ_{MSR}, supervirial, 1 pc



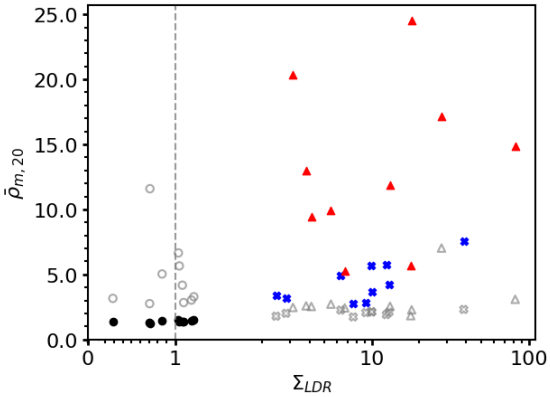
(c) $\bar{\rho}_{\mathrm{m},20}$ vs Q, subvirial, 1 pc



(d) $\bar{\rho}_{\mathrm{m},20}$ vs Q, supervirial, 1 pc

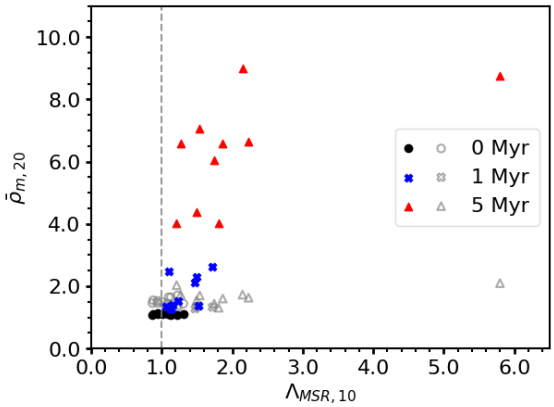


(e) $\bar{\rho}_{\mathrm{m},20}$ vs Σ_{LDR}, subvirial, 1 pc

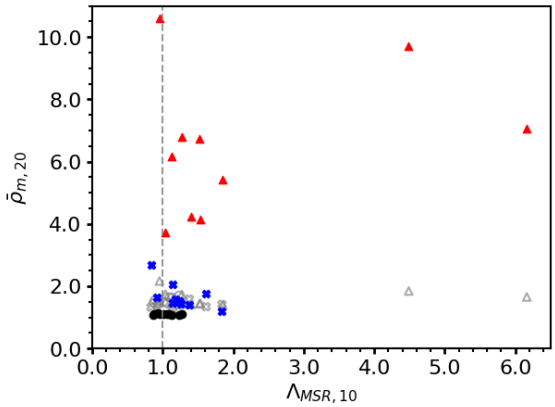


(f) $\bar{\rho}_{\mathrm{m},20}$ vs Σ_{LDR}, supervirial, 1 pc

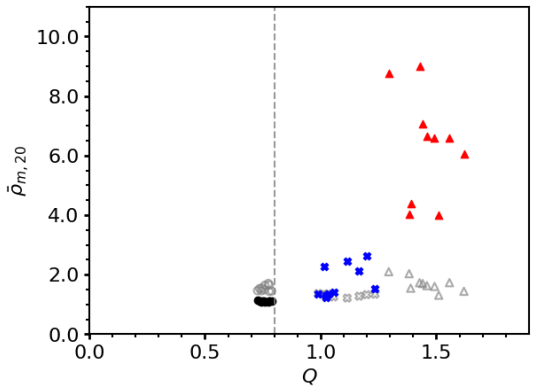
Figure 4.8: The mean Mahalanobis density calculated for 3D and 6D phase spaces plotted against other methods of quantifying structure for 10 subvirial and supervirial simulations which are initially substructured with fractal dimension $D_{\mathrm{f}} = 1.6$ and 1 pc radii. The left-hand panels show the results for the subvirial regions and the right-hand panels show the results for the supervirial regions. The initial values at 0 Myr are represented by the black circles, the blue crosses show 1 Myr and the red triangles show 5 Myr. The grey open circles show the comparison of the 6D Mahalanobis density at 0 Myr, the open grey crosses show it for 1 Myr and the open grey triangles for 5 Myr. From top to bottom the rows show the different methods which $\bar{\rho}_{\mathrm{m},20}$ is plotted against, with the top row showing Λ_{MSR}, second row showing Q and the bottom row showing Σ_{LDR}.



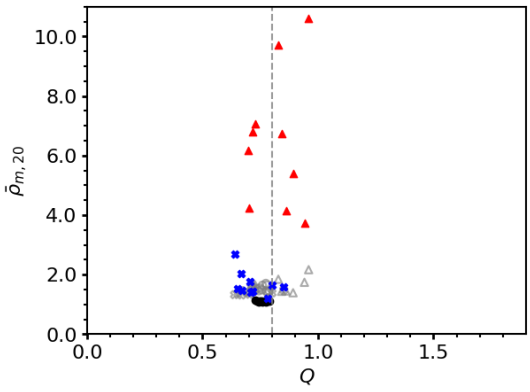
(a) $\bar{\rho}_{\mathrm{m},20}$ vs Λ_{MSR}, subvirial, 1 pc



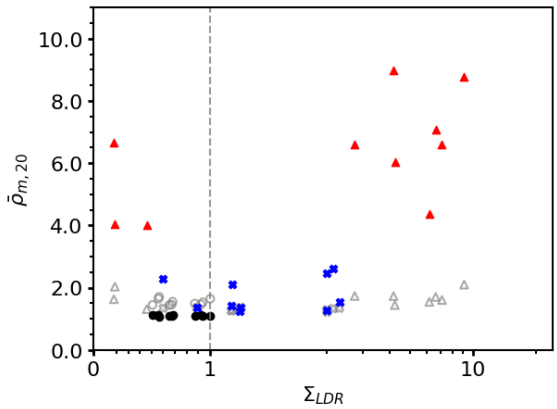
(b) $\bar{\rho}_{\mathrm{m},20}$ vs Λ_{MSR}, supervirial, 1 pc



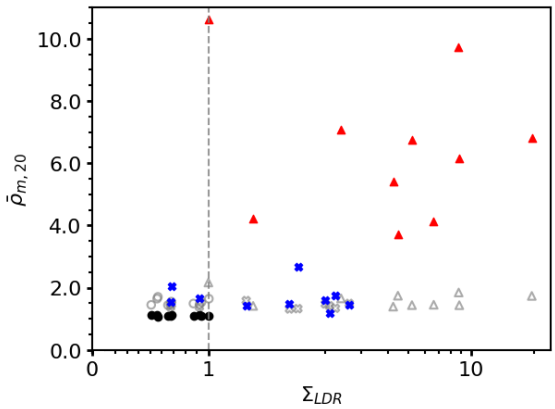
(c) $\bar{\rho}_{\mathrm{m},20}$ vs Q, subvirial, 1 pc



(d) $\bar{\rho}_{\mathrm{m},20}$ vs Q, supervirial, 1 pc

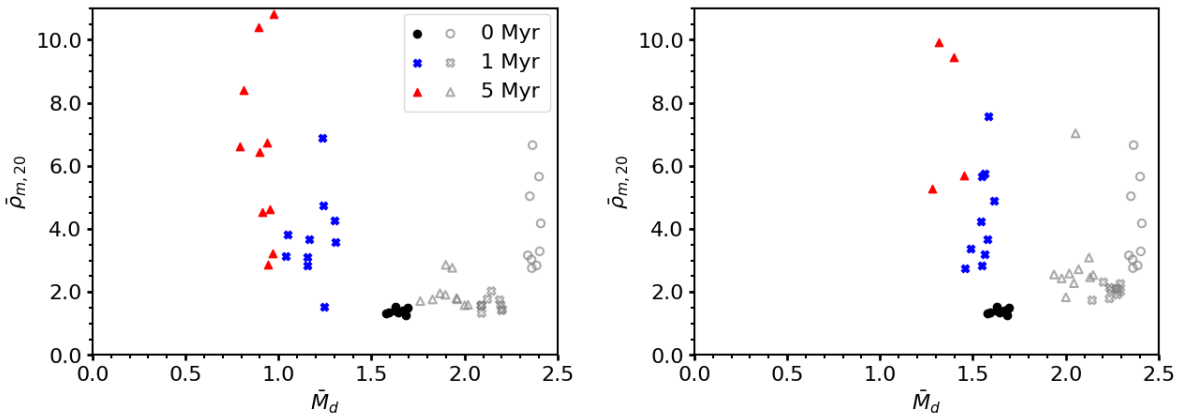


(e) $\bar{\rho}_{\mathrm{m},20}$ vs Σ_{LDR}, subvirial, 1 pc



(f) $\bar{\rho}_{\mathrm{m},20}$ vs Σ_{LDR}, supervirial, 1 pc

Figure 4.9: The mean Mahalanobis density calculated for the 3D and 6D phase spaces plotted against other methods of quantifying substructure for 10 subvirial and supervirial simulations which have little to no initial substructure with fractal dimension $D_{\mathrm{f}} = 3.0$ and 1 pc radii. The left-hand panels show the subvirial results and the right-hand panels show the supervirial regions. The initial values at 0 Myr are represented by the black circles, the blue pluses show 1 Myr and the red triangles show 5 Myr. The mean 6D Mahalanobis densities at 0 Myr, 1 Myr and 5 Myr are shown by the grey open circles, grey open crosses and grey open triangles, respectively. From top to bottom the rows show the different methods which $\bar{\rho}_{\mathrm{m},20}$ is plotted against, with the top row showing Λ_{MSR}, second row showing Q and the bottom row showing Σ_{LDR}.



(a) $\bar{\rho}_{\mathrm{m},20}$ vs $\bar{M}_{\mathrm{d}}$, subvirial, 1pc

(b) $\bar{\rho}_{\mathrm{m},20}$ vs $\bar{M}_{\mathrm{d}}$, supervirial, 1 pc

Figure 4.10: The mean Mahalanobis density ($\bar{\rho}_{\mathrm{m},20}$) plotted against the mean Mahalanobis distance ($\bar{M}_{\mathrm{d}}$) for highly substructured regions with fractal dimensions $D_{\mathrm{f}} = 1.6$ and initial radii of 1 pc. Each region contains 1000 stars. The black circles the values at 0 Myr, the blue plus signs are the values at 1 Myr and the red triangles are the values at 5 Myr. The grey open circles, crosses and triangles show the same information but for the Mahalanobis distance and density calculated in the 6D phase space.

decreased, making differentiating between different times or virial states impractical.

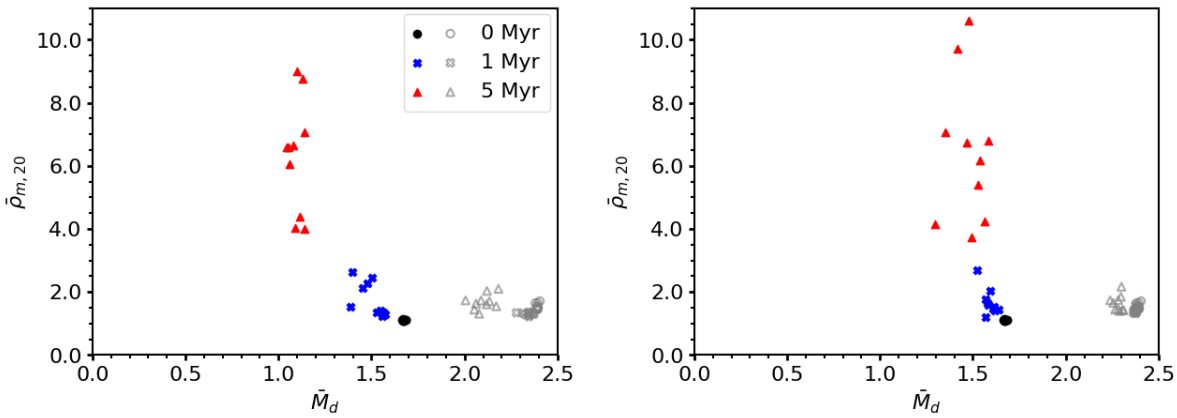
4.2.4 Mahalanobis Distance versus Mahalanobis Density

Figure 4.10 shows the relation between the Mahalanobis distance and density across the two different phase spaces investigated and the two different initial virial states.

In the positional phase space (3D, the coloured markers) there is significant overlap in both the Mahalanobis distance and density, making differentiating between different snapshots impractical. For the supervirial regions there is less overlap in the Mahalanobis density between the snapshots. However, there is significant overlap between the sub- and supervirial simulations, meaning that neither the Mahalanobis distance nor density can reliably distinguish between different virial states.

Both Figure 4.10 and Figure 4.11 shows the position-velocity phase space (6D) with the grey open markers.

Figure 4.11 shows the mean Mahalanobis distance plotted against the mean Ma-



(a) $\bar{\rho}_{\mathrm{m},20}$ vs $\bar{M}_{\mathrm{d}}$, subvirial, 1 pc

(b) $\bar{\rho}_{\mathrm{m},20}$ vs $\bar{M}_{\mathrm{d}}$, supervirial, 1 pc

Figure 4.11: The mean Mahalanobis density ($\bar{\rho}_{\mathrm{m},20}$) plotted against the mean Mahalanobis distance ($\bar{M}_{\mathrm{d}}$) for substructured regions with fractal dimensions $D_{\mathrm{f}} = 3.0$ and scales 1 pc for different snapshots. Each region contains 1000 stars. The black circles the values at 0 Myr, the blue crosses are the values at 1 Myr and the red triangles the values at 5 Myr. The grey open circles, crosses and triangles show the same information but for the Mahalanobis distance and density calculated in the 6D phase space.

halanobis density for high density (radii of 1 pc) region with little or no substructure ($D_{\mathrm{f}} = 3.0$).

4.3 Discussion

The work to test the Mahalanobis density in this chapter has been motivated due to its recent applications in quantifying the phase space of exoplanet host stars. In Winter et al. (2020) they propose that hot Jupiters are more likely to be found around host stars that are in high 6D phase space density, as measured using the Mahalanobis density. However, this was questioned by Mustill et al. (2022) who show the peculiar velocities introduce a bias, that once accounted for, results in no significant excess of hot Jupiters around host stars in high 6D phase space density. The aim of this chapter is not to make a scientific assessment of the Mahalanobis density in its use in planet formation specifically, but simply to see how it evolves when looking at simple N-body simulations

of SFRs to see what information, if any, it may be able to provide about the initial conditions (i.e. virial state, density and initial morphology).

Due to the simplicity of the simulations there are a number of important caveats that must be taken into account. First, there is no galactic potential or tidal force acting on our simulated SFRs. The presence of an external Galactic tidal field would likely increase the dissolution of the SFRs by causing outlying stars to become unbound, which would in turn increase the potency of the Galactic tidal field at later ages.

Two more important caveats are that there is no gas simulated, and therefore there is no gas potential and also that the systems are fully isolated. The most important caveat that disallows direct comparison to the works of Winter et al. (2020) and Mustill et al. (2022) is that there are no planets in the simulations and so how representative these simulations are of real SFRs with exoplanet host stars is uncertain. I investigate if there is a measurable link between a star's initial formation environment and any exoplanets they may have in Chapter 5.

In §4.2.3 I show that the 6D Mahalanobis density in isolation cannot be used to reliably infer the initial conditions of SFRs due to overlap in the sub- and supervirial values. However, when the Mahalanobis density in the 3D phase space is combined with either $\Lambda_{\rm MSR}$, $\Sigma_{\rm LDR}$ or Q then the initial virial conditions can be inferred. This is most clear to see in Figure 4.9.

The regions that are initially supervirial attain higher final phase space densities than subvirial regions. This is somewhat counter intuitive, but is due to the fact that as the region expands small groupings of stars can form which will have similar positions and therefore higher phase space densities. In the subvirial cases lower 3D phase space densities are found due to stars being ejected and ending up in relative isolation compared to the rest of the region. Because the mean Mahalanobis density is being used the results are sensitive to only a few stars being ejected.

The initial conditions of the simulations cannot be determined using the 6D Maha-

lanobis density. One would assume that the more data there is (and therefore dimensions in the phase space), the more clearly the distinction between sub- and supervirial simulations. However, somewhat counter intuitively, adding more dimensions to the phase space effectively dilutes out any information that would allow the initial conditions to be determined in the SFRs. For example, in the $D_{\rm f} = 1.6$ supervirial simulations, as the stars dynamically evolve they may get further apart spatially, but kinematically they may be quite similar. If two stars that are very far apart end up moving in the same general direction the velocity and positional phases spaces will effectively cancel each other out and therefore remove any information about the initial conditions of the region (i.e. different positions but similar velocities).

I find that the 6D Mahalanobis density for all simulations is similar at 10 Myr for both sub- and supervirial regions independent of the fractal dimension and the initial radii of the region. I therefore suggest that the Mahalanobis distance, and its associated density, are not suitable for quantifying the initial conditions of star formation, nor any subsequent dynamical evolution.

4.4 Conclusion

I have presented N-body simulations with different initial fractal dimensions, virial states and initial radii and quantify their resepective 3D and 6D phase space density using the Mahalanobis density. I compare the performance of the Mahalanobis density to more established methods for quantifying structure in SFRs, namely $\Lambda_{\rm MSR}$, $\Sigma_{\rm LDR}$ and Q. I also applied the Mahalanobis distance in 3D phase space to sets of static synthetic regions of different morphologies to test its ability to discriminate between different morphologies. The conclusions of the chapter are as follows:

1. The Mahalanobis distance in the 3D phase space is degenerate across a wide range of morphologies commonly observed in SFRs, associations and clusters, and so it

cannot be used to differentiate between different morphologies.

2. The 3D Mahalanobis densities, $\rho_{\mathrm{m},20}$ can be used to distinguish between the high and low stellar density regions with large amounts of substructure ($D_{\mathrm{f}} = 1.6$). The low stellar density regions show similar behaviour but delayed by around 0.6 Myr compared to the high volume density regions due to the dynamical timescales being longer. This effect is even more pronounced in the simulations with little to no initial substructure ($D_{\mathrm{f}} = 3.0$) where the bump occurs several Myr later than in the higher stellar density simulations corresponding to the subvirial collapse.

3. I show that the Mahalanobis density calculated in the 3D phase space can be used with the Q-parameter, Λ_{MSR} or Σ_{LDR} to infer information about a region's initial virial state.

4. When using the 6D Mahalanobis densities there are no significant differences between any of the simulations. Adding more parameters (adding more dimensions) to the phase space suppresses any changes in the Mahalanobis density over time.

It is therefore advised that using the Mahalanobis distance as a method to quantify the morphology of star-forming regions should be avoided due to its degeneracy across both substructured regions and smooth, centrally concentrated regions.

When applied to spatial and kinematic phase space (6D), all of its discriminatory power is washed out (similar to the issues encountered when applying the Q-parameter to kinematic data, Cartwright (2009)), and so I advocate using combinations of spatial and kinematic metrics instead.

In summary, I have found that the Mahalanobis density in 6D is similar for simulations with and without substructure. It is important to mention that the way the Mahalanobis density is applied in this Chapter is different to the way it is calculated in Winter et al. (2020), meaning that direct comparisons between these works should not

be made, instead what should be taken from this work is the challenge of physically interpreting the 6D Mahalanobis density, and the somewhat counterintuitive finding that the 6D Mahalanobis density is less reliable at discerning between the different initial conditions of the simulations.

5

Phase Space Densities of Exoplanet Host Stars

5.1 Summary

In Chapter 4 I showed that the 6D Mahalanobis density, calculated with respect to all other stars, cannot be used to reliably infer the initial conditions of simulated star-forming regions (SFRs). In this chapter I discuss my investigations into the Mahalanobis density of exoplanet host stars. There are two main tests I perform in this chapter. One is investigating how the 6D Mahalanobis density evolves for different subsets of stars (hosts, non-hosts and former hosts). The other is testing if the number of hosts in high phase space density regimes correlates to the initial density of the regions.

I will also discuss other works that have also investigated the Mahalanobis density such as Mustill et al. (2022), Adibekyan et al. (2021) and Kruijssen et al. (2021).

The main aim of this chapter is to understand the origin of the high Mahalanobis density regimes, in which Winter et al. (2020) finds an overabundance of hot Jupiter host stars. In Winter et al. (2020) they calculate the relative Mahalanobis density, different to how I calculate it in Chapter 4. In this chapter I recreate how Winter et al. (2020) calculate the relative Mahalanobis density and calculate it for N-body simulations of 1000 stars with ∼500 Jupiter mass planets in orbit either at 5 AU of 30 AU.

I find that the Mahalanobis density of former hosts does appear to evolve differently depending on the initial density of a region. In low mass density simulations the former hosts that initially had planets at 5 AU can have much lower Mahalanobis densities than in the high mass density simulations. The 5 AU former hosts in the high mass density simulations tend to finish with a minimum mean Mahalanobis density of ~ 1.5 and the 5 AU former hosts in the low mass density simulations finish with minimum mean Mahalanobis densities of ~ 1.

When splitting the host stars into different phase space density regimes I find that most host stars are most are found in high phase space density regimes. This is true regardless of the initial conditions of the simulations. Therefore, the high Mahalanobis

density regime does not correspond to the initial conditions of the simulations (i.e. density) but rather the kinematics of the host stars themselves.

I construct synthetic field distributions by randomly picking snapshots from simulations and superimposing them. This is done to mimic the field distribution, which contains SFRs of differing ages. This is to test if the results found for the simulations are also found in noisy crowded fields. I then compare the Mahalanobis densities of hosts in the synthetic fields to the work of Winter et al. (2020), as they propose that high and low density regimes correspond to the initial conditions, specifically the density, of SFRs which the host star formed in. I find an overabundance of host stars in high phase space density regimes in the synthetic field distributions, regardless of the initial mass density of the simulations I pick the snapshots from. There is no significant difference in the proportion of host stars in high phase space density regimes in the field distributions, regardless of the initial conditions of the simulation used to construct the fields.

I count how many perturbed hosts are found in high phase space density regimes and also find that out of three density regimes (low, ambiguous or high) most perturbed hosts are found in high phase space density regimes. The overall number of perturbed hosts is lower, which means the difference between the number of perturbed hosts in low, ambiguous and high phase space density regimes is smaller. The exception to this is in the low density simulations with planets initially at 30 AU, where the number of perturbed hosts is greater, and so the number of perturbed hosts in high phase space densities is also greater.

5.2 Introduction

Various simulations have shown that planet formation is affected by different phenomena within SFRs. The dominant effect on planet formation in SFRs is photoevaporation (see Chapter 1 §1.9.2) due to nearby massive stars. Photoevaporation is a significant factor

even at relatively low stellar mass densities of $10s\ M_{\odot}\ \mathrm{pc}^{-3}$ and will decrease the amount of volatiles in the disk (effectively quenching gas giant formation) (Nicholson et al., 2019; Concha-Ramírez et al., 2019; Parker et al., 2021).

The other major effect on planet formation is that of direct gravitational interactions between stars (see Chapter 1 § 1.9.1). The dynamical interactions of stars will lead to some planets becoming unbound from their parent star. The proportion of stars that lose a planet can be used to infer the initial densities of SFRs (Daffern-Powell et al., 2022). Investigating the effects dynamical interactions within SFRs have on planet formation, specifically how orbits are changed due to flybys, may reveal a new way to infer the initial conditions of SFRs based on the kinds of planets we observe around stars.

To investigate this, Winter et al. (2020) developed the Mahalanobis density (see Chapter 2 §2.3.5) (an N dimensional density metric that rescales and removes the units from the data allowing different parameters to be compared) to see if the location of Galactic exoplanet host stars' in the position-velocity phase space density depends on the type of planets observed around them.

Winter et al. (2020) focuses on hot Jupiters as their formation mechanism is still debated. There are several theories to explain the origin of hot Jupiters; they can form far away from their host star then migrate inwards (disk migration) or they may form far away from their host stars and through interactions with a third body have their eccentricity increased a significant amount. Then, through tidal interactions at the closest point in the orbit to the star the planet will lose energy, reducing its apoapsis and circularising its orbit (Dawson & Johnson, 2018). Winter et al. (2020) proposes that the location of hot Jupiter host stars primarily being in high phase space densities is due to the initial formation conditions of the stars, imprinted in both their spatial distributions and kinematics. This high phase space density is attributed to hot Jupiter host stars forming in dense regions, meaning they will share similar velocities (co-moving) even though they may be physically far apart. Winter et al. (2020) posits that it is due to

dynamical interactions in the dense regions between stars, which they hypothesise is the primary cause of hot Jupiters' having such small semi-major axes.

In Mustill et al. (2022) they recreate the work of Winter et al. (2020) and also find an overabundance of hot Jupiters in high phase space density regimes. However, this overabundance disappears when the ages of the host stars are taken into account. The reason for this is due to hot Jupiter host stars being relatively young and young stars will have low peculiar velocities - the velocity difference between the host star compared to an ideal circular orbit around the galactic centre at a distance of the Sun. Since the hot Jupiter host stars are generally young, they will have similar velocities and therefore they will have higher phase space densities.

Other works looking into the origin of this overabundance find other possible sources other than sharing the same birth cluster. In Kruijssen et al. (2021) they propose the overabundance is primarily caused by the large scale structure in the Milky Way, with some contribution from the initial conditions of the SFR.

The work of Adibekyan et al. (2021) uses data consisting of FGK dwarf stars that have short period exoplanets. In their investigation they find there is a difference in the phase space density and the period of the exoplanets. They find that host stars in the low phase space density regimes are significantly older than hosts in the high phase space density regimes. The reasons older stars appear in low phase space densities is likely due to the greater amount of dynamical interactions they have undergone in the Galaxy. This is supported by the findings of Mustill et al. (2022), where they find that generally hot Jupiter host stars are younger, meaning that they still have similar velocities to other stars from their natal group, which for the Mahalanobis density translates as hot Jupiter hosts being in high phase space density regimes.

Table 5.1: Initial conditions of the four sets of simulations. From left to right the columns the degree of substructure, the number of stars, the minimum and maximum number of planets in each set, the virial ratio, the initial semi-major axis of the planets and the stellar mass density (volume defined as a sphere with a radius equal to the distance from the centre of mass to the furthest star in a simulation) of the simulations.

Fractal Dimension	R (pc)	Range $N_{\rm p}$	$\alpha_{\rm vir}$	$a_{\rm i}$ (AU)	$\tilde{\rho}$ ($M_\odot\,{\rm pc}^{-3}$)
$D_{\rm f} = 1.6$	1	$462 - 530$	0.3	30	10^4
$D_{\rm f} = 1.6$	5	$465 - 530$	0.3	30	10^2
$D_{\rm f} = 1.6$	1	$480 - 530$	0.3	5	10^4
$D_{\rm f} = 1.6$	5	$464 - 519$	0.3	5	10^2

5.3 Simulations

I use simulations from Daffern-Powell et al. (2022), which were originally used to investigate how the dynamical interactions between stars in SFRs affect planetary orbits, either by changing their orbits or unbinding them from their host star. The simulations were run using the `Kira` integrator (see Chapter 2 §2.2). I am using four sets of simulations, each containing 20 different realisations of SFRs with the same statistical initial conditions. All sets are initially subvirial with a virial ratio of $\alpha_{\rm vir} = 0.3$ (different to the virial ratio used in the simulations in Chapter 4 where $\alpha_{\rm vir} = 0.1$), the simulations each have 1000 stars and a high initial degree of substructure, $D_{\rm f} = 1.6$. The regions are simulated for 10 Myr.

For all simulations $\sim$ 500 Jupiter mass planets are added in orbit around randomly chosen stars with masses $< 3M_\odot$. The sets either have planets with initial semi-major axes of 5 AU or 30 AU. Two of the sets have an initial median stellar mass density of $\sim 10^4\ M_\odot\,{\rm pc}^{-3}$, with the other two having a lower median stellar mass density of $\sim 10^2\ M_\odot\,{\rm pc}^{-3}$. The initial conditions of the simulations are summarised in Table 5.1.

5.3.1 The Neighbourhood Mahalanobis density

In this section I will describe the neighbourhood Mahalanobis density method that was used in Winter et al. (2020), where they assume that stars in their Gaia data belong to

either a low, ambiguous or high phase space density regime. I follow Winter et al. (2020) and split the phase space into a low, ambiguous and high density regime by assuming the Mahalanobis density distributions can be described using two Gaussians (though, see Mustill et al. (2022) where they use the Bayesian information criteria to determine the number of Gaussians that best describe the Mahalanobis density distributions and find they are best described using more than two Gaussians, something I also find when applied to my synthetic SFRs).

To calculate which density regime a host star belongs to I calculate the Mahalanobis density using the following neighbourhood method, with a large neighbourhood cut-off of 80 pc and local neighbourhood cut-off of 40 pc. These distance cut-offs are measured from the host star. The use of these neighbourhood cut-offs has been done to allow direct comparison with the findings of the original work, but due to the way the synthetic fields are constructed (made up of SFRs with ages ¡ 10 Myr) an 80 pc cut-off will cover the majority of stars within the synthetic field (Winter et al., 2020). This is also the case for the individual simulations. The reason for this cut-off in Winter et al. (2020) was to avoid picking up Galactic structure but in the simulations the regions are assumed to be in isolation with no galactic structure or galactic tidal field that will affect them. The method is as follows:

a) Find a host star.

b) Define a large neighbourhood (80 pc) and a local neighbourhood (40 pc).

c) Check that the chosen host star's local neighbourhood has 400-600 stars within it. If there are fewer than 400 stars skip this host star, if it is more than 600 randomly pick 600 stars in the local neighbourhood.

d) Calculate the Mahalanobis density of the host star with respect to all other stars in the local neighbourhood, not just the ones picked in the previous step.

e) Then for each of the chosen stars (the 400-600 picked in step c)) in the local neighbourhood of the host star, calculate their Mahalanobis density with respect to all stars in their own local neighbourhoods.

f) Once the Mahalanobis density has been calculated for the host star and the chosen stars within its local neighbourhood the phase space densities are then normalised to the median. This allows comparisons between the different regions around each of the host stars.

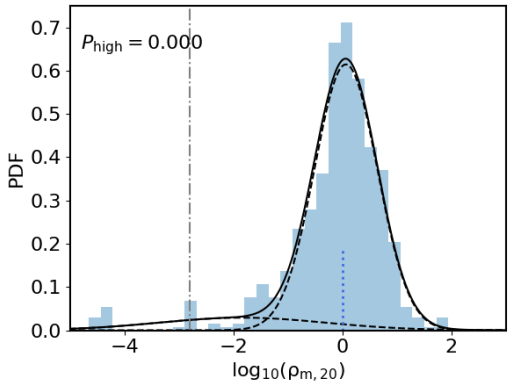
g) Go back to step a).

Once the Mahalanobis density has been calculated for all host stars and up to 600 other stars in their local neighbourhoods I then calculate the probability that a host star is found in a higher than average Mahalanobis density compared to the stars in its local neighbourhood. I follow Winter et al. (2020) and use a Gaussian mixture model with the number of Gaussians set to two. I show examples fitting the Gaussians in Figure 5.1 for 30 AU and 5 AU hosts in the top and bottom rows, respectively.

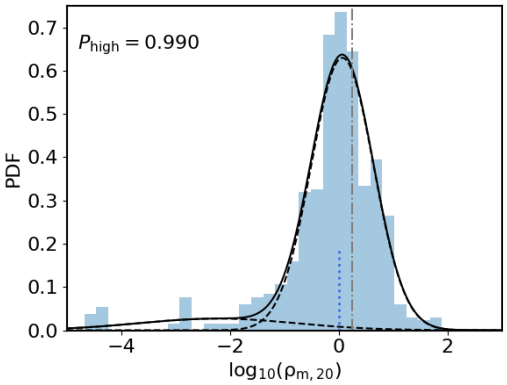
The criteria for a host star to be in the high phase space density regime is based on $P_{\rm high}$, the probability it was picked from the high phase space Gaussian. A probability $P_{\rm high} > 0.84$ is the criteria for the host to belong to the high phase space density regime and $P_{\rm high} < 0.16$ is the criteria for it to belong to the low phase space density regime. These values have been chosen so that the two thresholds are one standard deviation apart across the entire Mahalanobis density distribution. If $0.16 < P_{\rm high} < 0.86$ the host is said to be in an ambiguous phase space density regime.

5.3.2 Recreating the Field

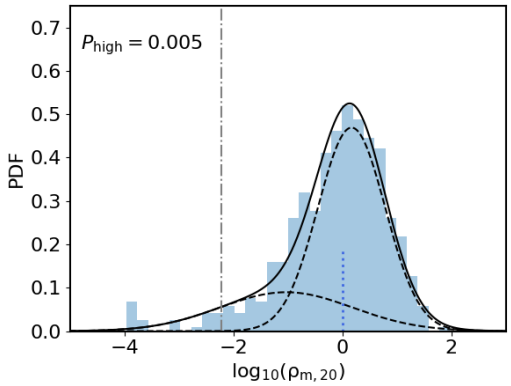
I recreate the Galactic field population by randomly picking single snapshots from each of the simulations in a set and then superimposing them to make fields of 20,000 stars. The field population of stars can be thought of as the summation of different SFRs



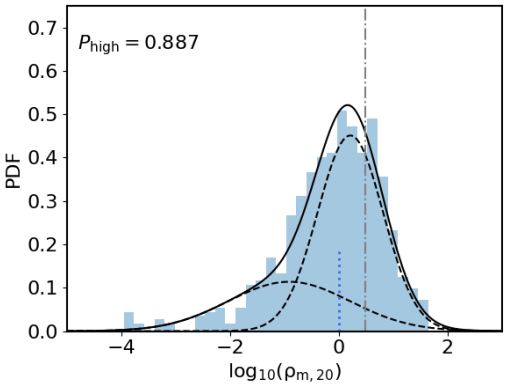
(a) Host star in low density regime. Planets at 5 AU.



(b) Host star in high density regime. Planets at 5 AU.

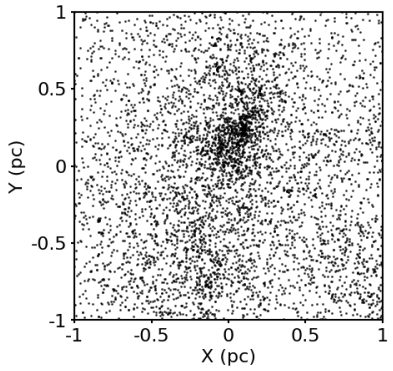


(c) Host star in low density regime. Planets at 30 AU.



(d) Host star in high density regime. Planets at 30 AU.

Figure 5.1: Example of host stars in high and low density regimes in initial snapshots from single simulations with planets either at 5 AU or 30 AU. P_{high} is the probability that the host star belongs to the high density regime. The vertical grey dash dotted line is the Mahalanobis density of the host star which the region is centred on. The solid black line shows the Gaussian mixture model, with its two components shown by the black dashed lines. The vertical blue dotted line shows the median Mahalanobis density of the region. The top row shows examples from a simulation with planets initially at 5 AU and the bottom row shows results for a simulation with planets initially at 30 AU. The left-hand column shows host stars in low phase space density regimes and the right-hand column shows host stars in high phase space density regimes.



(a) 4 pc^2 from the artificial field

Figure 5.2: Plot showing the distribution of stars in the synthetic field constructed from single random snapshots taken from each of the 20 simulations in a set. This plot shows the distribution of stars taken from the set of initially high mass density simulations with planets initially at 5 AU from their host stars.

at different times in their dynamical evolution. Some stars may still be in co-moving groups (high phase space density regime, generally younger stars) while others are no longer bound to their birth cluster and are classified as field stars (low phase space density regime, generally older stars).

There are several important caveats with the way the field is generated. First, in each of the simulations the galactic gravitational tidal field is not being simulated, therefore the cluster dissolution is only due to internal interactions within a cluster. Finally, each synthetic field is made up of SFRs that share the same statistical initial conditions; in reality SFRs may have a wide range of initial conditions.

The aim of testing the Mahalanobis density on synthetic fields is to test if the number of host stars in high phase space density regimes can be detected in a field of stars. An example $4\,pc^2$ section of a synthetic field is shown in Figure 5.2 made up from snapshots picked from initially high density simulations with planets initially at 5 AU.

5.4 Results

In this section I show the phase space density evolution for three different populations of stars in the sets of N-body simulations. The host-stars, the non-host stars and the former host stars. Whether a star is a host or not is dependent on if there is a planet gravitationally bound to it or not. Former hosts are determined by comparing non-host stars in subsequent snapshots to the initial snapshot of the simulation to determine which non-hosts initially had a planet. The Mahalanobis density is calculated for all stars, relative to all other stars in the current snapshot.

I then determine the number of host stars that are found in low, ambiguous and high phase space density regimes across the 4 sets of simulations for the initial and final snapshots. This is calculated using the neighbourhood Mahalanobis density method described in this chapter. The phase space density regimes of perturbed hosts is also investigated in the final snapshots of the simulations.

I calculate the neighbourhood Mahalanobis density for 500 host stars in the sets of synthetic fields and determine the number of hosts in low, ambiguous or high phase space density regimes.

5.4.1 Evolution of Host Star Phase Space Densities

The 6D Mahalanobis density is calculated the same way as in Blaylock-Squibbs & Parker (2023). It is calculated for stars with respect to all other stars. For hosts, non-hosts and former hosts the Mahalanobis densities are calculated, and the ranges of the mean Mahalanobis densities are found across the 20 simulations in each of the four sets (low and high mass density simulations with planets initially at 5 AU and low and high mass density simulations with planets initially at 30 AU). In Figure 5.3 I show the results for the four sets. Figure 5.3 (a) shows that the hosts, non-hosts and former hosts share a similar range of Mahalanobis densities. Of particular note is the extended range of

Mahalanobis density values found for the former host stars with planets initially at 5 AU in low mass density simulations, shown in Figure 5.3 (d). This is due to the amount of energy needed to unbind a planet at 5 AU being greater than for a planet at 30 AU, and as a consequence the former host stars' velocities will be changed by a greater amount. However, this difference is only detectable on the simulated data as I know exactly which stars have lost planets.

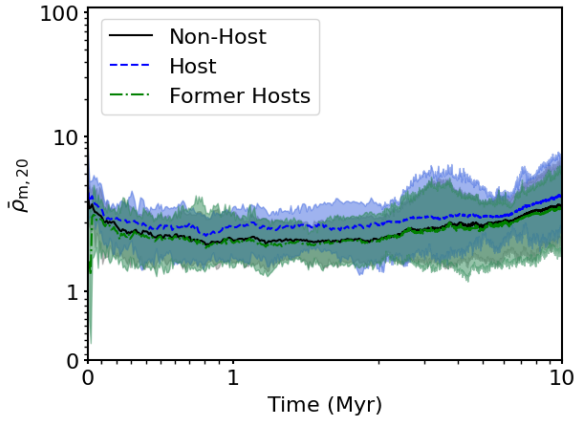
The minimum Mahalanobis density values found for the low mass density regions is lower than in the higher mass density regions, and also takes longer to reach similar mean values. This is in part due to low number statistics as early on in the simulation there will be only a few former hosts. It takes a few Myr in these simulations for a significant number of hosts to lose their planets and become former hosts.

The range is clearly less in Figure 5.3 (b) due to the fact that planets at 30 AU will be easier to lose compared to planets at 5 AU, meaning that the change to the 30 AU former host stars' velocities will be less.

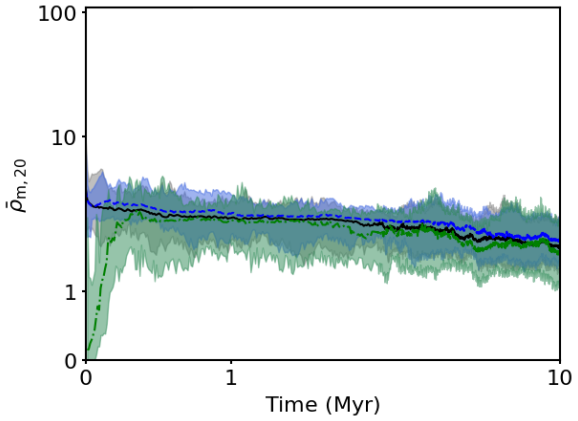
5.4.2 Comparing Density Regimes of Host Stars in the Initial and Final Snapshots

Here I present the number of times host stars are found in low, ambiguous and high phase space density regimes across each of the four sets of initial conditions (high and low mass density 5 AU and also high and low mass density 30 AU). Their Mahalanobis densities are calculated using the neighbourhood method. Figure 5.4 shows the number of times host stars are found in the different density regimes, with subfigure (a) showing the results in the initial snapshots ($t = 0.0\,\mathrm{Myr}$) and subfigure (b) showing the results for the final snapshots ($t = 10.0\,\mathrm{Myr}$).

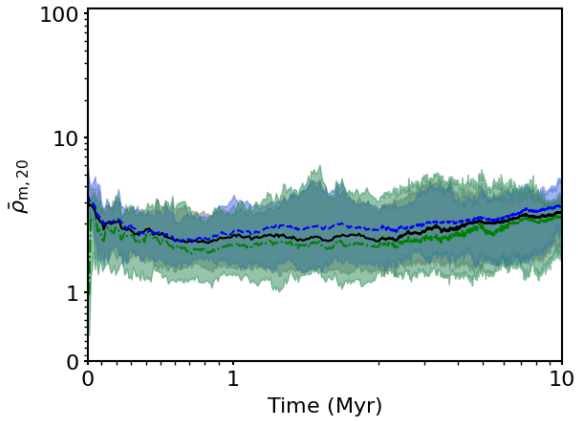
Initially most host stars are either in ambiguous phase space densities (i.e. $0.16 < P_{\mathrm{high}} < 0.86$) or high phase space densities (i.e. $P_{\mathrm{high}} > 0.86$). This is seen in both



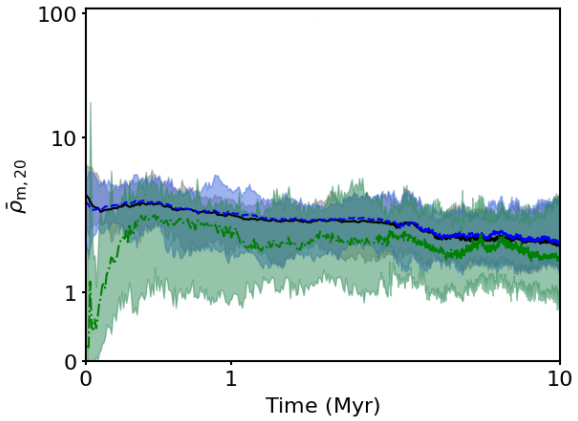
(a) High density SFR, planets initially at 30 AU



(b) Low density SFR, planets initially at 30 AU



(c) High density SFR, planets initially at 5 AU



(d) Low density SFR, planets initially at 5 AU

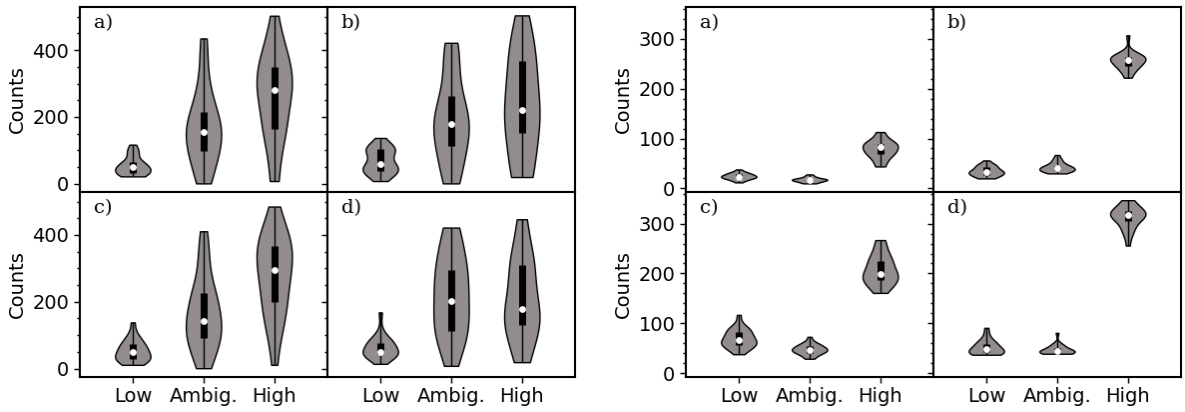
Figure 5.3: The shaded areas show the range of mean 6D Mahalanobis densities found across all 20 simulations for the different subsets. The grey, green and blue shaded areas show the range of mean Mahalanobis densities found across all 20 simulations for non-hosts, former hosts and hosts, respectively. The black, green and blue lines show the mean of the mean Mahalanobis densities found for each of the subsets. The top row from left to right shows the results for high and low spatial density simulations with planets initially at 30 AU. The bottom row from left to right shows the results for high and low spatial density simulations with planets initially at 5 AU.

high and low density simulations with planets at 30 AU shown in panels a) and b) of Figure 5.4. Panels c) and d) show the same for high and low density simulations, respectively, but with planets initially at 5 AU.

Figure 5.4(b) shows the number of times host stars are found in the different phase space density regimes in the final snapshots of the simulations. The spread of values for the sets has decreased, due to there being fewer host stars at the end of all sets of simulations and dynamical interactions changing positions and velocities of the host stars, moving them into high phase space density regimes. Across the four sets I find that most host stars end up in high phase space density regimes. Stars that remain host stars after all this time will be similar to other surviving hosts in the simulations. If a host star experiences an interaction significant enough to change its velocity, lowering its Mahalanobis density, it will also likely lose its planet. Evidence of this is shown in panel a) where I show the number of times 30 AU host stars are found in high phase space density regimes. Comparing the high mass density (panel a)) and low mass density simulations (panel b)) with planets at 30 AU initially it is clear to see that more hosts end up in high phase space densities in the initially less dense simulations, as more hosts are likely to survive till the end of the simulations.

I also calculate the phase space density regimes for perturbed host stars. A host star is considered perturbed if its semi-major axis in the final snapshot has changed by more than ten per cent of its original value (either 30 AU or 5 AU). The results of this are shown in Figure 5.5.

For all sets, perturbed hosts are found in predominantly high phase space density regimes. For simulations with planets at 30 AU with initially low mass densities the number of perturbed hosts is much greater. This is due to the fact that even in a low density environment the planets can have their orbits more easily perturbed compared to if the planets are initially at 5 AU, like in panel d) as they are less strongly bound. The number of perturbed 5 AU hosts in low mass density simulation that end up in



(a) Initial Snapshots

(b) Final Snapshots

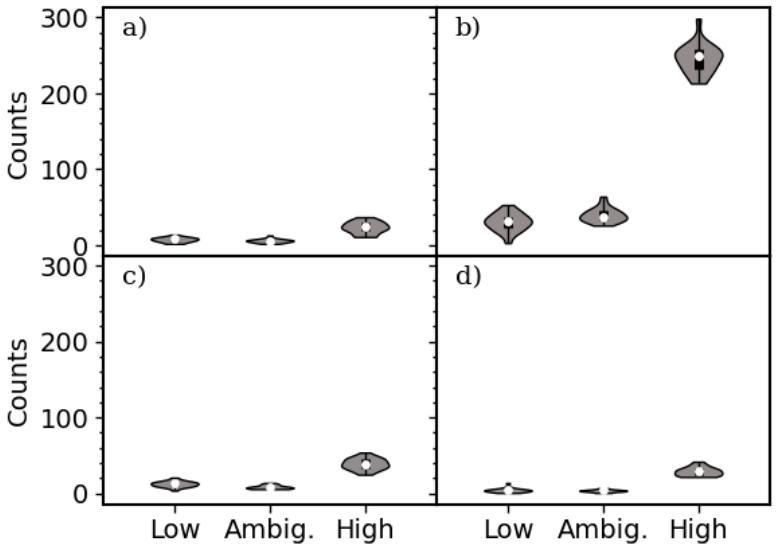
Figure 5.4: Violin plots of the number of times host stars are found in either low, ambiguous or high phase space density regimes across all 20 simulations in the four sets for the initial snapshot and the final snapshot. The median is shown by the white dot, the black bar shows the interquartile range and the line shows the full range of the data. The shaded area shows the probability of counting a particular value, where the width of the plot corresponding to the probability of measuring this value. Panels a), b), c) and d) in both subfigures correspond to the 30 AU high mass density, 30 AU low mass density, 5 AU high mass density and 5 AU low mass density sets, respectively.

high phase space density regimes is lower because it is much harder to perturb close in planets, unlike planets further away from their host stars.

The number of perturbed 30 AU hosts in the high mass density simulations that end up in high phase space densities is lower compared to the low mass density simulations. This is because in the high mass density simulations the 30 AU planets are more often going to become unbound, rather than just perturbed.

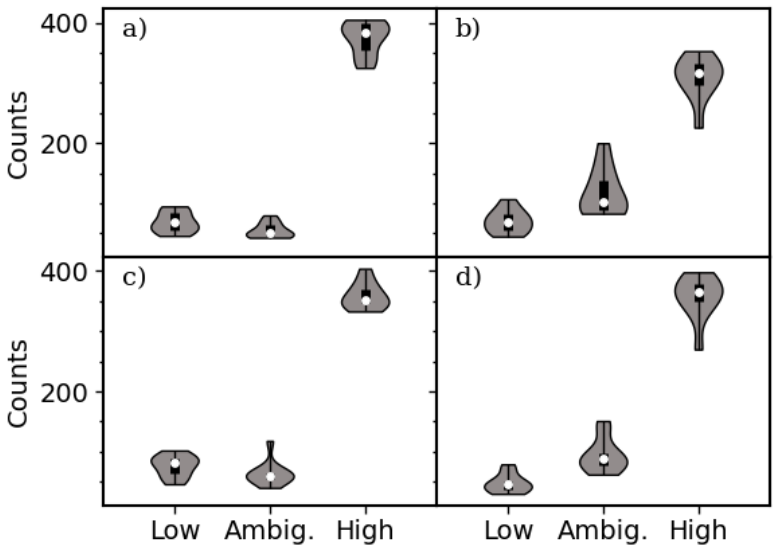
5.4.3 Density Regimes in Synthetic Fields

The Mahalanobis density is calculated using the neighbourhood method for 500 host stars in each of the four sets (each set contains 10 different synthetic fields). Figure 5.6 shows the number of times host stars are found in low, ambiguous or high phase space density regimes. I limit the analysis to 500 hosts in each of the synthetic fields to better match the initial sample size of host stars in the single simulations.



(a) Final Snapshots perturbed hosts

Figure 5.5: Violin plots of the number of times perturbed host star are found in either low, ambiguous or high phase space density regimes across all 20 simulations in the four sets of simulations for the initial snapshot and the final snapshot. The median is shown by the white dot, the thick black bar shows the interquartile range and the thin black line shows the full range of the data. The shaded area shows the probability of counting a particular value. Panels a), b), c) and d) correspond to the 30 AU high mass density, 30 AU low mass density, 5 AU high mass density and 5 AU low mass density sets, respectively.



(a) Violin plot showing the number of host stars in different phase space density regimes.

Figure 5.6: Violin plots of the number of times host stars are found in either low, ambiguous or high phase space density regimes across the four sets of 10 synthetic field distributions. The median number of hosts in a phase space density regime is shown by the white circle, the thick black bar shows the interquartile range, the thin black line shows the full range of values and the shaded area shows the probability of getting a particular value (wider areas meaning greater probability of getting that value). Panel a) shows the counts for the set of field distributions generated using random snapshots from simulations with planets initially at 30 AU, with high initial mass densities. Panel b) shows the same, but the field distributions are made from simulations with low initial mass densities. Panels c) shows the high mass density set and d) shows the low mass density set for field distributions made from simulations with planets initially at 5 AU.

All sets of synthetic fields have similar proportions of host stars in low, ambiguous and high phase space density regimes. This means that the number of host stars in high phase space densities is similar regardless of the initial densities of the star forming regions that make up the field distribution.

5.5 Discussion

Firstly, I will discuss the results of calculating the Mahalanobis density for three different populations (hosts, non-hosts and former hosts) in the N-body simulations over time. When calculated with respect to all other stars in a region the Mahalanobis density does evolve slightly differently depending on the initial conditions. The difference is most evident when comparing the phase space density of former hosts in high and low mass density simulations with planets initially at 5 AU. This difference is due to the severity of the interaction needed to remove a planet at 5 AU from its host star which changes the former host star's velocity by a greater amount. This difference in velocities between former hosts and the other two subsets manifests as a lower minimum mean Mahalanobis density (since in the 6D phase space the star has moved further away from most other stars due to its velocity being changed by a large amount).

In the simulations I have knowledge of which stars have lost planets due to dynamical interactions. While it would be worthwhile investigating if this is also seen in reality, it would be challenging as we do not know which stars are former hosts. A statistical approach would be needed involving estimating which stars in a sample have lost a planet in the past.

The main difference between Mahalanobis density evolution of the sets of simulations is due to the initial mass density of the regions. The high mass density regions finish with slightly higher Mahalanobis densities compared to the low mass density regions. The low mass density simulations typically finish with lower Mahalanobis densities than

they started with.

There is a difference in the number of host stars in high phase space densities when comparing the final snapshots of high and low spatial density simulations which is due to the fact that there will be more 5 AU host stars left at the end of simulations compared to 30 AU hosts.

Counting the number of host stars in high phase space density regimes in the field distributions shows no significant difference regardless of the initial conditions of the simulations that make the field distributions.

When comparing the number of hosts in different phase space density regimes there is still an overabundance of host stars in high phase space densities. This is seen in the final snapshots of the N-body simulations, in the synthetic fields and also when determining the phase space densities of perturbed host stars

5.6 Conclusion

In this chapter I have applied the Mahalanobis density in different ways to N-body simulations and large synthetic field distributions of stars. I have calculated the Mahalanobis density for every star with respect to all other stars in the four sets of N-body simulations and find that there is a difference between the minimum Mahalanobis densities of former hosts that had planets at 5 AU initially when comparing between high and low initial mass densities. Any interaction that is able to remove a Jupiter mass planet at 5 AU will change the former host stars velocity by a greater amount than for an interaction that removes a planet at 30 AU from a star. The greater the difference in former hosts' velocities to non-host stars the lower their phase space density will be.

When calculating the Mahalanobis density for the final snapshots of the simulations using the neighbourhood method I find that the proportion of host stars in high phase space density regimes does not significantly differ, regardless of the initial conditions.

This means that the Mahalanobis density cannot be used to infer the initial conditions of SFRs based on the number of host stars in low, ambiguous or high phase space density regimes.

In the synthetic fields there is no dependence on the initial conditions of the constituent snapshots on the proportion of host stars in high phase space density regimes.

In summary the Mahalanobis density cannot infer the initial conditions of the simulated SFRs based on overabundances of exoplanet host stars (regardless of their initial semi-major axis) in high phase space density regimes. The reason host stars are found more often in high phase space density regimes is due the kinematics of the host stars. Stars in the simulations will undergo dynamical processing, changing their velocities. Host stars that make it to the end of the simulations will not have undergone significant interactions with other stars. Therefore, host stars will have similar velocities to other host stars in the simulations compared to other stars in the simulations whose velocities can be very different from each other. This supports the findings of Adibekyan et al. (2021) and Mustill et al. (2022), that the phase space overdensities are due to the kinematics of host stars, not the initial mass density of the stars-forming regions. However, this does not mean the initial conditions of SFRs are not embedded in the exoplanetary system architecture, just that such information cannot be measured using the Mahalanobis density.

I find no significant difference in the proportion of perturbed hosts in high phase space density regimes between the sets with different initial conditions. There are significantly more perturbed hosts in high phase space density in the set of simulations with initial low mass density with planets initially at 30 AU, compared to low mass density simulations with planets initially at 5 AU. This is due to a preference for planets at 30 AU in low mass simulations to be perturbed rather than becoming unbound, like they do in the high mass density simulation set.

When comparing the evolution of host stars and non-host stars there is no signif-

icant difference in the evolution of their Mahalanobis densities. This further supports a kinematic origin for the host stars being in high phase space density regimes. Stars who remain hosts by the end of the simulation will have experienced fewer dynamical interactions increasing the Mahalanobis density.

While the Mahalanobis density of former-host stars does trace the dynamical evolution of different subsets to some degree. However, this is due to having all information about the stars within the simulations (i.e. which stars have lost planets).

6

Conclusions

6.1 Investigating INDICATE

I tested the ability of INDICATE to quantify the clustering of stars and investigated if there was a link between the degree of clustering and the morphology of the cluster. I found that INDICATE cannot be used to differentiate between different morphologies of synthetic star-forming regions like the Q-parameter can.

I found that INDICATE can be used to detect the traces of two commonly used definitions of mass segregation (i.e. the most massive stars are centrally located in star-forming regions or that the most massive stars are located in areas of the greatest surface density). When INDICATE is applied to just the 50 most massive stars it finds the first definition of mass segregation in most synthetic regions when the most massive stars are swapped with the 10 most central or clustered stars.

The majority of regions that have mass segregation, as measured using INDICATE, also have mass segregation when measured using $\Lambda_{\rm MSR}$.

I apply INDICATE to several observed star-forming regions, and find a wide range of clustering. When testing for classical mass segregation INDICATE gives results in agreement with other works. However, INDICATE finds traces of classical mass segregation in NGC 1333, but in Parker & Alves de Oliveira (2017) they find no traces of mass segregation using the $\Lambda_{\rm MSR}$ method.

My findings show that INDICATE is robust at quantifying the clustering of stars within synthetic and observed star-forming regions. Therefore, INDICATE can be used to infer the physics and star formation histories of star-forming regions.

6.2 Phase Space Densities of Star-Forming Regions

I tested the ability of the Mahalanobis density at both differentiating between star-forming regions with different morphologies and its ability to infer the initial conditions

of N-body simulations. This work was undertaken to test the Mahalanobis density metrics ability to differentiate between simulations with different initial conditions. The motivation for these tests was due to the conclusions in Winter et al. (2020), in which they report a measurable link between stars with hot Jupiters in orbit around them and the stars initial formation environment having a high mass density. This finding was also used to propose that hot Jupiters are likely formed via external dynamical interactions with other stars in high mass density star-forming regions.

I find that the 3D Mahalanobis distance (calculated from the mean position in the cluster to each star) in unable to differentiate between clusters with different morphologies, with very different spatial distributions producing similar results. This is because the Mahalanobis density variance normalises each of the parameters in the data.

When applied to N-body simulations with different initial degrees of substructure, mass density and virial states I find no clear differences in the evolution of the 6D Mahalanobis density.

The 3D Mahalanobis density can be used to differentiate between highly substructured regions with different initial mass densities. For low mass density regions that are initially subvirial the peak 3D Mahalanobis density is seen around 0.6 Myr later on in low mass density simulations compared to the high mass density ones, which is expected as it takes longer for the clumps of stars to move nearer to each other. The need for the full history of the region to make this distinction limits the observational use of the Mahalanobis density when used in the way.

I combine the 3D and 6D Mahalanobis density with other methods, such as the Q-parameter and $\Lambda_{\rm MSR}$. I find that the 3D Mahalanobis density and the Q-parameter most reliably infer the initial conditions of the simulated SFRs. However, this is mainly due to the discriminatory power of the Q-parameter.

6.3 Phase Space densities of Exoplanet Host Stars

The origins of hot Jupiter exoplanets is not fully understood. How do massive gas giants get so close to their host stars? The findings of Winter et al. (2020) would suggest that hot Jupiter host stars formed within high mass density regions, and within these dense regions the orbits of exoplanets can be perturbed. These perturbation leads to their orbital distance decreasing, which can lead to gas giants moving very closely to their host star. This implies that the origins of hot Jupiters is likely dynamical in nature, with interactions between the Jupiter mass gas giant and other stars causing the exoplanet to undergo orbital decay bringing it closer to its host star (Fabrycky & Tremaine, 2007; Dawson & Johnson, 2018). In the previous science chapter I applied the Mahalanobis density to N-body simulations of 1000 stars only.

This chapter continues the work of the previous, by applying the Mahalanobis density to simulations that contain planets, with the aim of investigating if there is any difference in the Mahalanobis density evolution for planet host stars, non-hosts and former hosts. In this chapter I also investigate the suitability of counting the number of exoplanet host stars in different density regimes in N-body simulations to determine their initial conditions.

I show that there is a slight difference in the Mahalanobis density for former host stars. This is most clearly seen for the 5 AU host stars initially in low mass density environments. This is due to any host star that loses a planet initially at 5 AU having its velocity changed a significant amount, therefore lowering its phase space density. The reason it is more evident in the low density environment is that other stars will have velocities that are more similar to their initial ones later on in time due to there being fewer dynamical interactions per star.

I also construct synthetic field distributions by picking random snapshots from simulations in each of the sets and find the same, that number of host stars in high phase

space density regimes cannot be used to infer the initial star formation conditions when looking at synthetic fields of star-forming regions of varying ages.

The Mahalanobis density is unable to differentiate between the initial formation conditions of stars based on the presence of an overabundance of host stars in a high phase space density regime. For all initial conditions tested there is an overabundance of host stars in high phase space density regimes.

I also see an overabundance of perturbed hosts in high phase space density regimes, which is present in all sets of simulations. There are far more perturbed hosts in the set of low mass density simulations with planets at 30 AU, as these planets will be much easier to perturb in low density environments compared to planets at 5 AU.

From this and Chapter 4 it is clear that the Mahalanobis density of stars cannot be used to reliably infer the initial conditions of star-forming regions. In Chapter 5 I show that the initial conditions of host stars cannot be inferred using the 6D Mahalanobis density, but I would stress that just because the Mahalanobis is not measuring a link does not mean there isn't one.

7

Future Work

7.1 Machine Learning Classifier

Machine learning has been used in a wide array of astronomical fields. In star formation it has been used to classify galaxies as star-forming or not, to infer the star-forming properties in galaxies and to identify star clusters (Zhang et al., 2019; Surana et al., 2020; Pérez et al., 2021)

Investigating the suitability of machine learning classifiers at identifying the initial conditions of simulated star-forming regions should be performed. There are two ways a machine learning classifier could be used to do this. Firstly, the current methods (see Chapter 2 §2.3) can all be applied to simulated regions, then a classifier can be trained on the results of these methods for large sets of synthetic star-forming regions with different initial conditions.

Secondly, the classifier could be trained directly on the positions of velocities of stars in the simulations. It would of course also have to be trained on a limited amount data to better match what observers have at their disposal (i.e. plane of sky position and proper motion).

The kinds of classifiers and how many parameters I use are both important factors. The use of binary classification is unlikely to be suitable when trying to determine the virial state of star-forming region (i.e. they can be subvirial, virial or supervirial). If the model is being trained on the outputs of the methods discussed in Chapter 2 then a random forest classifier could be used for more than two parameters, which has been previously used to infer the initial conditions of star-forming regions (i.e. see Figure 4.9 in Chapter 4). Some classes could be subvirial-high-density and supervirial-low-density as just two examples.

7.2 Determining the Best Metrics

Star-forming regions are complex systems where different initial conditions can produce similar looking regions. Methods that can be applied to higher dimensional data might be able to infer the initial conditions of similar looking regions by using both the spatial and kinematic information of the stars within them. However, as I have shown, one such method (the Mahalanobis density) cannot alone reliably infer the initial conditions of star-forming regions.

To improve current methods and aid in the development of new methods (or sets of methods) I plan to perform a systematic comparison of methods that quantify the different parameters of star-forming regions. By comparing all the metrics, the most effective method (or methods in combination) can be determined. A statistical analysis of the metrics would be performed. A combination of metrics could be analysed, and some probability would be given that this combination of values from the different metrics corresponds to a particular initial density, degree of substructure or virial state.

The statistical analysis of the methods would be Bayesian in nature as it would allow the inference of star-forming region initial conditions based on the outputs of the methods discussed in this work. Testing the feasibility of a Bayesian framework for inferring the initial conditions of star-forming regions is needed due to the limitation of combining the current methods, mainly that their results are qualitative rather than quantitative.

Additional Software Acknowledgements

Plots in this work have been produced using `Matplotlib` 3.3.4 (Hunter, 2007). Data analysis performed using `Numpy` 1.20.1, `scikit-learn` 1.30.1 and `sci-py` 1.9.0 (Harris et al., 2020; Pedregosa et al., 2011; Virtanen et al., 2020).

Bibliography

Aarseth S. J., 1985, in Goodman J., Hut P., eds, Dynamics of Star Clusters Vol. 113, Direct N-body calculations.. pp 251–258

Aarseth S. J., Henon M., Wielen R., 1974, A&A, 37, 183

Adibekyan V., Santos N. C., Demangeon O. D. S., Faria J. P., Barros S. C. C., Oshagh M., Figueira P., Delgado Mena E., Sousa S. G., Israelian G., Campante T., Hakobyan A. A., 2021, A&A, 649, A111

Alcock H. L., Parker R. J., 2019, Monthly Notices of the Royal Astronomical Society, 490, 350

Alfaro E. J., González M., 2016, Monthly Notices of the Royal Astronomical Society, 456, 2900

Allison R. J., Goodwin S. P., 2011, Monthly Notices of the Royal Astronomical Society, 415, 1967

Allison R. J., Goodwin S. P., Parker R. J., Zwart S. F. P., de Grijs R., 2010, Monthly Notices of the Royal Astronomical Society, 407, 1098

Allison R. J., Goodwin S. P., Parker R. J., Zwart S. F. P., de Grijs R., Kouwenhoven M. B. N., 2009, Monthly Notices of the Royal Astronomical Society, 395, 1449

ALMA Partnership et al. 2015, The Astrophysical Journal, 808, L3

Alves F. O., Cleeves L. I., Girart J. M., Zhu Z., Franco G. A. P., Zurlo A., Caselli P., 2020, The Astrophysical Journal, 904, L6

Alves J., Lombardi M., Lada C. J., 2007, Astronomy & Astrophysics, 462, L17

Anderson M., Peretto N., Ragan S. E., Rigby A. J., Avison A., Duarte-Cabral A., Fuller G. A., Shirley Y. L., Traficante A., Williams G. M., 2021, Monthly Notices of the Royal Astronomical Society, 508, 2964

Andrews S. M., Huang J., Pérez L. M., Isella A., Dullemond C. P., Kurtovic N. T., Guzmán V. V., Carpenter J. M., Wilner D. J., Zhang S., Zhu Z., Birnstiel T., Bai X.-N., Benisty M., Hughes A. M., Öberg K. I., Ricci L., 2018, The Astrophysical Journal, 869, L41

André P., 1994, Observations of protostars and protostellar stages

Arnold B., Wright N. J., Parker R. J., 2022, Monthly Notices of the Royal Astronomical Society, 515, 2266

Arzoumanian D., André P., Könyves V., Palmeirim P., Roy A., Schneider N., Benedettini M., Didelon P., Di Francesco J., Kirk J., Ladjelate B., 2019, Astronomy and Astrophysics, 621, A42

Bae J., Isella A., Zhu Z., Martin R., Okuzumi S., Suriano S., 2023, 534, 423

Barnes A. T., Liu J., Zhang Q., Tan J. C., Bigiel F., Caselli P., Cosentino G., Fontani F., Henshaw J. D., Jiménez-Serra I., Kalb D. S., Law C. Y., Longmore S. N., Parker R. J., Pineda J. E., Sánchez-Monge A., Lim W., Wang K., 2023, A&A, 675, A53

Barsony M., 1994, 65, 197

Bate M. R., 2012, MNRAS, 419, 3115

Bate M. R., 2014, MNRAS, 442, 285

Bate M. R., Clarke C. J., McCaughrean M. J., 1998, Monthly Notices of the Royal Astronomical Society, 297, 1163

Bell C. P. M., Naylor T., Mayne N. J., Jeffries R. D., Littlefair S. P., 2013, Monthly Notices of the Royal Astronomical Society, 434, 806

Biazzo K., Randich S., Palla F., 2011, Astronomy & Astrophysics, 525, A35

Blaauw A., 1961, Bulletin of the Astronomical Institutes of the Netherlands, 15, 265

Blaylock-Squibbs G. A., Parker R. J., 2023, Monthly Notices of the Royal Astronomical Society, 519, 3643

Blaylock-Squibbs G. A., Parker R. J., Buckner A. S. M., Güdel M., 2022, Monthly Notices of the Royal Astronomical Society, 510, 2864

Bodenheimer P., 1995, Annual Review of Astronomy and Astrophysics, 33, 199

Bohlin R. C., Savage B. D., Drake J. F., 1978, The Astrophysical Journal, 224, 132

Bonnell I. A., 2005, arXiv:astro-ph/0501258, 327, 425

Bonnell I. A., Bate M. R., 1994, Monthly Notices of the Royal Astronomical Society, 271, 999

Bonnell I. A., Davies M. B., 1998, Monthly Notices of the Royal Astronomical Society, 295, 691

Bonnell I. A., Smith R. J., Clark P. C., Bate M. R., 2011, Monthly Notices of the Royal Astronomical Society, 410, 2339

Bonnor W. B., 1956, Monthly Notices of the Royal Astronomical Society, 116, 351

Bontemps S., Andre P., Kaas A. A., Nordh L., Olofsson G., Huldtgren M., Abergel A., Blommaert J., Boulanger F., Burgdorf M., Cesarsky C. J., Cesarsky D., Copet E., Davies J., Falgarone E., Lagache G., Montmerle T., Perault M., Persi P., Prusti T., Puget J. L., Sibille F., 2001, arXiv:astro-ph/0103373

Boulanger F., 1999, in Taylor A. R., Landecker T. L., Joncas G., eds, New Perspectives on the Interstellar Medium Vol. 168 of Astronomical Society of the Pacific Conference Series, Dust Emission and ISM Components. p. 173

Bracco A., Jelić V., Marchal A., Turić L., Erceg A., Miville-Deschênes M.-A., Bellomi E., 2020, Astronomy & Astrophysics, 644, L3

Bravi L., Zari E., Sacco G. G., Randich S., Jeffries R. D., Jackson R. J., Franciosini E., Moraux E., López-Santiago J., Pancino E., Spina L., Wright N. J., Jiménez-Esteban F. M., Klutsch A., Roccatagliata V., Gilmore G., Bragaglia A., Flaccomio E., Francois P., Koposov S. E., Bayo A., Carraro G., Costado M. T., Damiani F., Frasca A., Hourihane A., Jofré P., Lardo C., Lewis J., Magrini L., Morbidelli L., Prisinzano L., Sousa S. G., Worley C. C., Zaggia S., 2018, åp, 615, A37

Bressert E., Bastian N., Evans C. J., Sana H., Hénault-Brunet V., Goodwin S. P., Parker R. J., Gieles M., Bestenlehner J. M., Vink J. S., Taylor W. D., Crowther P. A., Longmore S. N., Gräfener G., Maíz Apellániz J., de Koter A., Cantiello M., Kruijssen J. M. D., 2012, Astronomy and Astrophysics, 542, A49

Bressert E., Bastian N., Gutermuth R., Megeath S. T., Allen L., Evans II N. J., Rebull L. M., Hatchell J., Johnstone D., Bourke T. L., Cieza L. A., Harvey P. M., Merin B., Ray T. P., Tothill N. F. H., 2010, Monthly Notices of the Royal Astronomical Society: Letters, 409, L54

Buckner A. S. M., Khorrami Z., Gonzalez M., Lumsden S. L., Moraux E., Oudmaijer R. D., Clark P., Joncour I., Blanco J. M., de la Calle I., Hacar A., Herrera-Fernandez J. M., Motte F., Salgado J., Valero-Martin L., 2020, A&A, 636, A80

Buckner A. S. M., Khorrami Z., Khalaj P., Lumsden S. L., Joncour I., Moraux E., Clark P., Oudmaijer R. D., Blanco J. M., de la Calle I., Herrera-Fernandez J. M., Motte F., Salgado J. J., Valero-Martin L., 2019, A&A, 622, A184

Burbidge E. M., Burbidge G. R., Fowler W. A., Hoyle F., 1957, Reviews of Modern Physics, 29, 547

Cameron A. G. W., 1962, The Astronomical Journal, 67, 112

Carpenter J. M., 2000, The Astronomical Journal, 120, 3139

Cartwright A., 2009, Monthly Notices of the Royal Astronomical Society, 400, 1427

Cartwright A., Whitworth A. P., 2004, Monthly Notices of the Royal Astronomical Society, 348, 589

Caselli P., Myers P. C., 1995, The Astrophysical Journal, 446, 665

Chabrier G., Baraffe I., 1997, Astronomy and Astrophysics, 327, 1039

Chapman N. L., Goldsmith P. F., Pineda J. L., Clemens D. P., Li D., Krčo M., 2011, ApJ, 741, 21

Clarke C. J., Gendrin A., Sotomayor M., 2001, Monthly Notices of the Royal Astronomical Society, 328, 485

Clarke S. D., Whitworth A. P., Spowage R. L., Duarte-Cabral A., Suri S. T., Jaffa S. E., Walch S., Clark P. C., 2018, Monthly Notices of the Royal Astronomical Society, 479, 1722

Clausius R., 1870, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 40, 122

Combet C., Ferreira J., 2008, Astronomy and Astrophysics, 479, 481

Concha-Ramírez F., Wilhelm M. J. C., Portegies Zwart S., Haworth T. J., 2019, Monthly Notices of the Royal Astronomical Society, 490, 5678

Cottaar M., Covey K. R., Foster J. B., Meyer M. R., Tan J. C., Nidever D. L., Chojnowski S. D., da Rio N., Flaherty K. M., Frinchaboy P. M., Majewski S., Skrutskie M. F., Wilson J. C., Zasowski G., 2015, ApJ, 807, 27

Cuello N., Ménard F., Price D. J., 2023, The European Physical Journal Plus, 138, 11

Daffern-Powell E. C., Parker R. J., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4925

Daffern-Powell E. C., Parker R. J., 2022, Monthly Notices of the Royal Astronomical Society, 517, 2103

Daffern-Powell E. C., Parker R. J., Quanz S. P., 2022, Monthly Notices of the Royal Astronomical Society, 514, 920

Dawson R. I., Johnson J. A., 2018, Annual Review of Astronomy and Astrophysics, 56, 175

De Maesschalck R., Jouan-Rimbaud D., Massart D., 2000, Chemometrics and Intelligent Laboratory Systems, 50, 1

Dehnen W., Read J. I., 2011, European Physical Journal Plus, 126, 55

Dobashi K., Uehara H., Kandori R., Sakurai T., Kaiden M., Umemoto T., Sato F., 2005, Publications of the Astronomical Society of Japan, 57, S1

Domínguez R., Fellhauer M., Blaña M., Farias J. P., Dabringhausen J., 2017, Monthly Notices of the Royal Astronomical Society, 472, 465

Dopcke G., Glover S. C. O., Clark P. C., Klessen R. S., 2011, The Astrophysical Journal, 729, L3

D'Orazi V., Randich S., Flaccomio E., Palla F., Sacco G. G., Pallavicini R., 2009, A&A, 501, 973

Ebert R., 1955, Zeitschrift fur Astrophysik, 37, 217

Eggleton P. P., Tokovinin A. A., 2008, Monthly Notices of the Royal Astronomical Society, 389, 869

Ehlerová S., Palouš J., 2013, Astronomy & Astrophysics, 550, A23

Eislöffel J., 2000, A&A, 354, 236

Elmegreen B. G., 2000, The Astrophysical Journal, 530, 277

Ester M., Kriegel H.-P., Sander J., Xu X., 1996, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise

Ewen H. I., Purcell E. M., 1951, Nature, 168, 356

Fabrycky D., Tremaine S., 2007, ApJ, 669, 1298

Fellhauer M., Kroupa P., 2005, The Astrophysical Journal, 630, 879

Ferrière K. M., 2001, Rev. Mod. Phys., 73, 1031

Finkbeiner et al. 2004, AJ, 128, 2577

Foong S. K., 2008, European Journal of Physics, 29, 987

Foster J. B., Cottaar M., Covey K. R., Arce H. G., Meyer M. R., Nidever D. L., Stassun K. G., Tan J. C., Chojnowski S. D., da Rio N., Flaherty K. M., Rebull L., Frinchaboy P. M., Majewski S. R., Skrutskie M., Wilson J. C., Zasowski G., 2015, The Astrophysical Journal, 799, 136

François P., Matteucci F., Cayrel R., Spite M., Spite F., Chiappini C., 2004, Astronomy & Astrophysics, 421, 613

Galli D., Walmsley M., Gonçalves J., 2002, A&A, 394, 275

Gieles M., Portegies Zwart S. F., 2011, Monthly Notices of the Royal Astronomical Society: Letters, 410, L6

Gomez M., Hartmann L., Kenyon S. J., Hewett R., 1993, The Astronomical Journal, 105, 1927

González M., Alfaro E. J., 2017, Monthly Notices of the Royal Astronomical Society, 465, 1889

Goodwin S. P., 2009, Astrophysics and Space Science, 324, 259

Goodwin S. P., 2010, Philosophical Transactions of the Royal Society of London Series A, 368, 851

Goodwin S. P., Bastian N., 2006, MNRAS, 373, 752

Goodwin S. P., Whitworth A. P., 2004, Astronomy & Astrophysics, 413, 929

Gould R. J., 1964, The Astrophysical Journal, 140, 638

Gouliermis D. A., Hony S., Klessen R. S., 2014, Monthly Notices of the Royal Astronomical Society, 439, 3775

Greaves J. S., Rice W. K. M., 2010, Monthly Notices of the Royal Astronomical Society, 407, 1981

Grudić M. Y., Guszejnov D., Hopkins P. F., Lamberts A., Boylan-Kolchin M., Murray N., Schmitz D., 2018, Monthly Notices of the Royal Astronomical Society, 481, 688

Guszejnov D., Hopkins P. F., 2015, Monthly Notices of the Royal Astronomical Society, 450, 4137

Gutermuth R. A., Megeath S. T., Myers P. C., Allen L. E., Pipher J. L., Fazio G. G., 2009, The Astrophysical Journal Supplement Series, 184, 18

Hacar A., Clark S., Heitsch F., Kainulainen J., Panopoulou G., Seifried D., Smith R., 2022

Harris C. R., Millman K. J., van der Walt S. J., Gommers R., Virtanen P., Cournapeau D., Wieser E., Taylor J., Berg S., Smith N. J., Kern R., Picus M., Hoyer S., van Kerkwijk M. H., Brett M., Haldane A., del Río J. F., Wiebe M., Peterson P., Gérard-Marchant P., Sheppard K., Reddy T., Weckesser W., Abbasi H., Gohlke C., Oliphant T. E., 2020, Nature, 585, 357

Haverkorn M., Spangler S. R., 2013, Space Science Reviews, 178, 483

Hayashi C., 1961, Publications of the Astronomical Society of Japan, 13, 450

Heggie D. C., 1975, MNRAS, 173, 729

Heggie D. C., Mathieu R. D., 1986, in Hut P., McMillan S. L. W., eds, The Use of Supercomputers in Stellar Dynamics Standardised units and time scales. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 233–235

Henshaw J. D., Caselli P., Fontani F., Jiménez-Serra I., Tan J. C., Longmore S. N., Pineda J. E., Parker R. J., Barnes A. T., 2016, Monthly Notices of the Royal Astronomical Society, 463, 146

Henyey L. G., Lelevier R., Levée R. D., 1955, Publications of the Astronomical Society of the Pacific, 67, 154

Hillenbrand L. A., Hartmann L. W., 1998, ApJ, 492, 540

Hills J. G., 1975, AJ, 80, 1075

Ho P. T. P., Martin R. N., Myers P. C., Barrett A. H., 1977, The Astrophysical Journal, 215, L29

Hunt L. K., Belfiore F., Lelli F., Draine B. T., Marasco A., García-Burillo S., Venturi G., Combes F., Weiß A., Henkel C., Menten K. M., Annibali F., Casasola V., Cignoni M., McLeod A., Tosi M., Beltrán M., Concas A., Cresci G., Ginolfi M., Kumari N., Mannucci F., 2023, A&A, 675, A64

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Inoue T., Inutsuka S.-i., 2016, The Astrophysical Journal, 833, 10

Inutsuka S.-I., Miyama S. M., 1992, The Astrophysical Journal, 388, 392

Jaffa S. E., Whitworth A. P., Lomax O., 2017, Monthly Notices of the Royal Astronomical Society, 466, 1082

Jakimiec M., Paciorek J., Bartkowiak A., 1991, Acta Astronomica, 41, 61

Jeffries R. D., Littlefair S. P., Naylor T., Mayne N. J., 2011, Monthly Notices of the Royal Astronomical Society, 418, 1948

Johnstone D., Matsuyama I., McCarthy I. G., Font A. S., 2004, 22, 38

Joncour I., Duchêne G., Moraux E., 2017, Astronomy & Astrophysics, 599, A14

Joncour I., Duchêne G., Moraux E., Motte F., 2018, Astronomy & Astrophysics, 620, A27

Jones C. E., Basu S., Dubinski J., 2001, ApJ, 551, 387

Jørgensen J. K., Johnstone D., Kirk H., Myers P. C., Allen L. E., Shirley Y. L., 2008, The Astrophysical Journal, 683, 822

Kim J.-G., Ostriker E. C., Filippova N., 2021, ApJ, 911, 128

King R. R., Parker R. J., Patience J., Goodwin S. P., 2012, Monthly Notices of the Royal Astronomical Society, 421, 2025

Klein R. I., Inutsuka S. I., Padoan P., Tomisaka K., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V Current Advances in the Methodology and Computational Simulation of the Formation of Low-Mass Stars. p. 99

Klessen R. S., Kroupa P., 2001, Astronomy & Astrophysics, 372, 105

Kley W., Crida A., 2008, Astronomy & Astrophysics, 487, L9

Korntreff C., Kaczmarek T., Pfalzner S., 2012, Astronomy & Astrophysics, 543, A126

Kos J., Lin J., Zwitter T., Žerjal M., Sharma S., Bland-Hawthorn J., Asplund M., Casey A. R., De Silva G. M., Freeman K. C., Martell S. L., Simpson J. D., Schlesinger K. J., Zucker D., Anguiano B., Bacigalupo C., Bedding T. R., Betters C., Da Costa G., Duong L., Hyde E., Ireland M., Kafle P. R., Leon-Saval S., Lewis G. F., Munari U., Nataf D., Stello D., Tinney C. G., Traven G., Watson F., Wittenmyer R. A., 2017, Monthly Notices of the Royal Astronomical Society, 464, 1259

Kounkel M., Deng T., Stassun K. G., 2022, The Astronomical Journal, 164, 57

Kraus A. L., Hillenbrand L. A., 2008, The Astrophysical Journal, 686, L111

Kroupa P., 2001, Monthly Notices of the Royal Astronomical Society, 322, 231

Kroupa P., 2008, pp 181–259

Kruijssen J. M. D., 2012, arXiv:1208.2963 [astro-ph]

Kruijssen J. M. D., Longmore S. N., Chevance M., Laporte C. F. P., Motylinski M., Keller B. W., Henshaw J. D., 2021, arXiv e-prints, p. arXiv:2109.06182

Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, Annual Review of Astronomy and Astrophysics, 57, 227

Kuhn M. A., Feigelson E. D., Getman K. V., Baddeley A. J., Broos P. S., Sills A., Bate M. R., Povich M. S., Luhman K. L., Busk H. A., Naylor T., King R. R., 2014, The Astrophysical Journal, 787, 107

Kuhn M. A., Hillenbrand L. A., Sills A., Feigelson E. D., Getman K. V., 2019, The Astrophysical Journal, 870, 32

Kuiper G. P., 1938, The Astrophysical Journal, 88, 472

Kuznetsova A., Hartmann L., Ballesteros-Paredes J., 2015, \textbackslashapj, 815, 27

Könyves V., André P., Men'shchikov A., Palmeirim P., Arzoumanian D., Schneider N., Roy A., Didelon P., Maury A., Shimajiri Y., Di Francesco J., Bontemps S., Peretto N., Benedettini M., Bernard J. P., Elia D., Griffin M. J., Hill T., Kirk J., Ladjelate B., Marsh K., Martin P. G., Motte F., Nguyên Luong Q., Pezzuto S., Roussel H., Rygl K. L. J., Sadavoy S. I., Schisano E., Spinoglio L., Ward-Thompson D., White G. J., 2015, Astronomy and Astrophysics, 584, A91

Küpper A. H. W., Maschberger T., Kroupa P., Baumgardt H., 2011, Monthly Notices of the Royal Astronomical Society, 417, 2300

Lada C. J., Lada E. A., 2003, Annu. Rev. Astron. Astrophys., 41, 57

Lamb J. B., Oey M. S., Werk J. K., Ingleby L. D., 2010, The Astrophysical Journal, 725, 1886

Larson R. B., 1981, Monthly Notices of the Royal Astronomical Society, 194, 809

Larson R. B., 1995, Monthly Notices of the Royal Astronomical Society, 272, 213

Larson R. B., 1998, Monthly Notices of the Royal Astronomical Society, 301, 569

Limongi M., Chieffi A., 2018, The Astrophysical Journal Supplement Series, 237, 13

Lomax O., Bates M. L., Whitworth A. P., 2018, MNRAS, 480, 371

Longmore S. N., Kruijssen J. M. D., Bastian N., Bally J., Rathborne J., Testi L., Stolte A., Dale J., Bressert E., Alves J., 2014, in , Protostars and Planets VI. University of Arizona Press

Low C., Lynden-Bell D., 1976, Monthly Notices of the Royal Astronomical Society, 176, 367

Luhman K. L., Mamajek E. E., Shukla S. J., Loutrel N. P., 2017, AJ, 153, 46

Lustig-Yaeger J., Fu G., May E. M., Ortiz Ceballos K. N., Moran S. E., Peacock S., Stevenson K. B., López-Morales M., MacDonald R. J., Mayorga L. C., Sing D. K., Sotzen K. S., Valenti J. A., Adams J., Alam M. K., Batalha N. E., Bennett K. A., Gonzalez-Quiles J., Kirk J., Kruse E., Lothringer J. D., Rustamkulov Z., Wakeford H. R., , 2023, A JWST transmission spectrum of a nearby Earth-sized exoplanet

Mac Low M.-M., Klessen R. S., 2004, Reviews of Modern Physics, 76, 125

Machida M. N., Inutsuka S.-i., Matsumoto T., 2010, The Astrophysical Journal, 724, 1006

Machida M. N., Inutsuka S.-i., Matsumoto T., 2011, The Astrophysical Journal, 729, 42

McMillan S. L. W., Vesperini E., Portegies Zwart S. F., 2007, The Astrophysical Journal, 655, L45

Mahalanobis P. C., 1936 On the generalized distance in statistics. National Institute of Science of India

Makino J., 1991, The Astrophysical Journal, 369, 200

Makino J., Aarseth S. J., 1992, Publications of the Astronomical Society of Japan, 44, 141

Mamajek E. E., 2009, 1158, 3

Marks M., Kroupa P., 2011, Monthly Notices of the Royal Astronomical Society, 417, 1702

Maschberger T., 2013, Monthly Notices of the Royal Astronomical Society, 429, 1725

Maschberger T., Clarke C. J., 2011, Monthly Notices of the Royal Astronomical Society, pp no–no

Mikkola S., Aarseth S. J., 1998, New Astronomy, 3, 309

Moeckel N., Bonnell I. A., 2009, Monthly Notices of the Royal Astronomical Society, 396, 1864

Moeckel N., Holland C., Clarke C. J., Bonnell I. A., 2012, Monthly Notices of the Royal Astronomical Society, 425, 450

Mouschovias T. C., Paleologou E. V., 1979, The Astrophysical Journal, 230, 204

Mouschovias T. C., Tassis K., Kunz M. W., 2006, The Astrophysical Journal, 646, 1043

Murray N., 2011, ApJ, 729, 133

Mustill A. J., Lambrechts M., Davies M. B., 2022, A&A, 658, A199

Myers P. C., Dame T. M., Thaddeus P., Cohen R. S., Silverberg R. F., Dwek E., Hauser M. G., 1986, The Astrophysical Journal, 301, 398

Nelson R. P., 2018

Nicholson R. B., Parker R. J., Church R. P., Davies M. B., Fearon N. M., Walton S. R. J., 2019, Monthly Notices of the Royal Astronomical Society

Nony T., Robitaille J.-F., Motte F., Gonzalez M., Joncour I., Moraux E., Men'shchikov A., Didelon P., Louvet F., Buckner A. S. M., Schneider N., Lumsden S. L., Bontemps S., Pouteau Y., Cunningham N., Fiorellino E., Oudmaijer R., André P., Thomasson B., 2021, Astronomy & Astrophysics, 645, A94

Oh S., Kroupa P., 2016, Astronomy and Astrophysics, 590, A107

Onishi T., Mizuno A., Kawamura A., Tachihara K., Fukui Y., 2002, ApJ, 575, 950

Ostriker J., 1964, The Astrophysical Journal, 140, 1056

Padoan P., Nordlund A., Jones B. J. T., 1997, Monthly Notices of the Royal Astronomical Society, 288, 145

Parker R. J., 2014, Monthly Notices of the Royal Astronomical Society, 445, 4037

Parker R. J., 2018, Monthly Notices of the Royal Astronomical Society, 476, 617

Parker R. J., 2020, Royal Society Open Science, 7, 201271

Parker R. J., Alves de Oliveira C., 2017, Monthly Notices of the Royal Astronomical Society, 468, 4340

Parker R. J., Bouvier J., Goodwin S. P., Moraux E., Allison R. J., Guieu S., Güdel M., 2011, Monthly Notices of the Royal Astronomical Society, 412, 2489

Parker R. J., Dale J. E., 2017, Monthly Notices of the Royal Astronomical Society, 470, 390

Parker R. J., Goodwin S. P., 2015, arXiv:1503.02692 [astro-ph]

Parker R. J., Goodwin S. P., Kroupa P., Kouwenhoven M. B. N., 2009, Monthly Notices of the Royal Astronomical Society, 397, 1577

Parker R. J., Maschberger T., Alves de Oliveira C., 2012, Monthly Notices of the Royal Astronomical Society, 426, 3079

Parker R. J., Nicholson R. B., Alcock H. L., 2021, MNRAS, 502, 2665

Parker R. J., Wright N. J., 2018, Monthly Notices of the Royal Astronomical Society, 481, 1679

Parker R. J., Wright N. J., Goodwin S. P., Meyer M. R., 2013, arXiv:1311.3639 [astro-ph]

Parker R. J., Wright N. J., Goodwin S. P., Meyer M. R., 2014, MNRAS, 438, 620

Pattle K., Fissel L., Tahani M., Liu T., Ntormousi E., 2023, in Inutsuka S., Aikawa Y., Muto T., Tomida K., Tamura M., eds, Astronomical Society of the Pacific Conference Series Vol. 534 of Astronomical Society of the Pacific Conference Series, Magnetic Fields in Star Formation: from Clouds to Cores. p. 193

Pavlidou T., Scholz A., Teixeira P. S., 2021, Monthly Notices of the Royal Astronomical Society, 503, 3232

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., Duchesnay E., 2011, Journal of Machine Learning Research, 12, 2825

Pérez G., Messa M., Calzetti D., Maji S., Jung D. E., Adamo A., Sirressi M., 2021, ApJ, 907, 100

Pflamm-Altenburg J., Kroupa P., 2006, MNRAS, 373, 295

Pilbratt G. L., Riedinger J. R., Passvogel T., Crone G., Doyle D., Gageur U., Heras A. M., Jewell C., Metcalfe L., Ott S., Schmidt M., 2010, Astronomy and Astrophysics, 518, L1

Pineda J. E., Arzoumanian D., André P., Friesen R. K., Zavagno A., Clarke S. D., Inoue T., Chen C.-Y., Lee Y.-N., Soler J. D., Kuffmeier M., 2022

Plummer H. C., 1911, Monthly Notices of the Royal Astronomical Society, 71, 460

Portegies Zwart S. F., 2016, MNRAS, 457, 313

Portegies Zwart S. F., McMillan S. L. W., Hut P., Makino J., 2001, Monthly Notices of the Royal Astronomical Society, 321, 199

Prentice A. J. R., Ter Haar D., 1971, Monthly Notices of the Royal Astronomical Society, 151, 177

Raghavan D., McAlister H. A., Henry T. J., Latham D. W., Marcy G. W., Mason B. D., Gies D. R., White R. J., ten Brummelaar T. A., 2010, The Astrophysical Journal Supplement Series, 190, 1

Rathborne J. M., Jackson J. M., Chambers E. T., Simon R., Shipman R., Frieswijk W., 2005, The Astrophysical Journal, 630, L181

Rathborne J. M., Jackson J. M., Simon R., 2006, The Astrophysical Journal, 641, 389

Ray T. P., Ferreira J., 2021, New Astronomy Reviews, 93, 101615

Redfield S., Falcon R. E., 2008, ApJ, 683, 207

Reggiani M., Robberto M., Da Rio N., Meyer M. R., Soderblom D. R., Ricci L., 2011, Astronomy & Astrophysics, 534, A83

Reiter M., Parker R. J., 2022, European Physical Journal Plus, 137, 1071

Richert A. J. W., Getman K. V., Feigelson E. D., Kuhn M. A., Broos P. S., Povich M. S., Bate M. R., Garmire G. P., 2018, Monthly Notices of the Royal Astronomical Society, 477, 5191

Salpeter E. E., 1955, The Astrophysical Journal, 121, 161

Scally A., Clarke C., 2001, Monthly Notices of the Royal Astronomical Society, 325, 449

Scally A., Clarke C., 2002, Monthly Notices of the Royal Astronomical Society, 334, 156

Schmeja S., Klessen R. S., 2006, Astronomy & Astrophysics, 449, 151

Schoettler C., de Bruijne J., Vaher E., Parker R. J., 2020, Monthly Notices of the Royal Astronomical Society, 495, 3104

Schoettler C., Parker R. J., Arnold B., Grimmett L. P., de Bruijne J., Wright N. J., 2019, Monthly Notices of the Royal Astronomical Society, 487, 4615

Segura-Cox D. M., Schmiedeke A., Pineda J. E., Stephens I. W., Fernández-López M., Looney L. W., Caselli P., Li Z.-Y., Mundy L. G., Kwon W., Harris R. J., 2020, Nature, 586, 228

Shirley Y. L., 2015, PASP, 127, 299

Shukirgaliyev B., Parmentier G., Berczik P., Just A., 2017, A&A, 605, A119

Siegal B. S., Griffiths J. C., 1974, The Moon, 9, 397

Simon M., 1997, The Astrophysical Journal, 482, L81

Smith R. C., 1983, The Observatory, 103, 29

Solomon P. M., Rivolo A. R., Barrett J., Yahil A., 1987, The Astrophysical Journal, 319, 730

Spitzer L., 1998, Physical Processes in the Interstellar Medium, 1 edn. Wiley

Surana S., Wadadekar Y., Bait O., Bhosale H., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4808

Sánchez N., Alfaro E. J., 2009, The Astrophysical Journal, 696, 2086

Takahashi S., Ho P. T. P., Teixeira P. S., Zapata L. A., Su Y.-N., 2013, The Astrophysical Journal, 763, 57

Tobin J. J., Hartmann L., Furesz G., Mateo M., Megeath S. T., 2009, ApJ, 697, 1103

Tomisaka K., Habe A., Ikeuchi S., 1981, Astrophysics and Space Science, 78, 273

Tomisaka K., Ikeuchi S., 1983, Astronomical Society of Japan, 35, 187

Utomo D., Sun J., Leroy A. K., Kruijssen J. M. D., Schinnerer E., Schruba A., Bigiel F., Blanc G. A., Chevance M., Emsellem E., Herrera C., Hygate A. P. S., Kreckel K., Ostriker E. C., Pety J., Querejeta M., Rosolowsky E., Sandstrom K. M., Usero A., 2018, ApJL, 861, L18

Vincke K., Pfalzner S., 2016, The Astrophysical Journal, 828, 48

Virtanen et al. 2020, Nature Methods, 17, 261

Vorobyov E. I., 2011, The Astrophysical Journal, 729, 146

Wareing C. J., Falle S. A. E. G., Pittard J. M., 2019, Monthly Notices of the Royal Astronomical Society, 485, 4686

Whitworth A. P., Ward-Thompson D., 2001, The Astrophysical Journal, 547, 317

Williams B. A., Walker D. L., Longmore S. N., Barnes A. T., Battersby C., Garay G., Ginsburg A., Gomez L., Henshaw J. D., Ho L. C., Kruijssen J. M. D., Lu X., Mills E. A. C., Petkova M. A., Zhang Q., 2022, Monthly Notices of the Royal Astronomical Society, 514, 578

Wilson C. D., Walker C. E., Thornley M. D., 1997, The Astrophysical Journal, 483, 210

Winter A. J., Kruijssen J. M. D., Longmore S. N., Chevance M., 2020, Nature, 586, 528

Wright N. J., Kounkel M., Zari E., Goodwin S., Jeffries R. D., 2023, 534, 129

Xu F.-W., Wang K., Liu T., Goldsmith P. F., Zhang Q., Juvela M., Liu H.-L., Qin S.-L., Li G.-X., Tej A., Garay G., Bronfman L., Li S., Wu Y.-F., Gómez G. C., Vázquez-Semadeni E., Tatematsu K., Ren Z., Zhang Y., Toth L. V., Liu X., Yue N., Zhang S., Baug T., Issac N., Stutz A. M., Liu M., Fuller G. A., Tang M., Zhang C., Dewangan L., Lee C. W., Zhou J., Xie J., Jiao W., Wang C., Liu R., Luo Q., Soam A., Eswaraiah C., 2023, Monthly Notices of the Royal Astronomical Society, 520, 3259

Zhang G.-Y., André P., Men'shchikov A., Wang K., 2020, Astronomy & Astrophysics, 642, A76

Zhang K., Schlegel D. J., Andrews B. H., Comparat J., Schäfer C., Vazquez Mata J. A., Kneib J.-P., Yan R., 2019, ApJ, 883, 63

Zilinskas M., Miguel Y., Mollière P., Tsai S.-M., 2020, Monthly Notices of the Royal Astronomical Society, 494, 1490

Zinchenko I. I., , 2022, Observational studies of high-mass star formation

Zinnecker H., Yorke H. W., 2007, Annual Review of Astronomy and Astrophysics, 45, 481

Zwart S. P., Makino J., McMillan S. L. W., Hut P., 1999, Technical Report arXiv:astro-ph/9812006, Star cluster ecology III: Runaway collisions in young compact star clusters, `http://arxiv.org/abs/astro-ph/9812006`. arXiv