Article:

The evolution of phase space densities in star-forming regions

George A. Blaylock-Squibbs ⭐ and Richard J. Parker †
Department of Physics and Astronomy, The University of Sheffield, Hounsfield Road, Sheffield S3 7RH, UK

ABSTRACT

The multidimensional phase space density (both position and velocity) of star-forming regions may encode information on the initial conditions of star and planet formation. Recently, a new metric based on the Mahalanobis distance has been used to show that hot Jupiters are more likely to be found around exoplanet host stars in high six-dimensional 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. We test the ability of the Mahalanobis distance and density to differentiate more generally between star-forming regions with different morphologies by applying it to static regions that are either substructured or smooth and centrally concentrated. We find that the Mahalanobis distance is unable to distinguish between different morphologies, and that the initial conditions of the *N*-body simulations cannot be constrained using only the Mahalanobis distance or density. Furthermore, we find that the more dimensions in the phase space, the less effective the Mahalanobis density is at distinguishing between different initial conditions. We 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. We therefore recommend continued use of multiple diagnostics for determining the initial conditions of star-forming regions, rather than relying on a single multidimensional metric.

Key words: galaxies: star formation – methods: statistical – methods: numerical.

1 INTRODUCTION

Star formation is observed to take place along filaments within giant molecular clouds (GMCs) (Palmeirim et al. 2013; Schisano et al. 2014; André 2017). The initial formation and distribution of these filaments is likely due to supersonic turbulence within GMCs (Larson 1981). It is along these filaments that cores can form, with further fragmentation of these cores leading to stars forming in groups containing tens to thousands of members (Lada & Lada 2003; Bastian et al. 2009).

One of the foundational questions of star formation is whether star formation is a universal process or not. Are the initial conditions of star-forming regions dependent on the environment, where differences in the stellar density, initial mass function (IMF), and stellar multiplicity are due to the initial conditions of the star-forming region? Or does star formation happen in a similar way everywhere, and any differences we observe in these regions are stochastic in nature?

There are two main proposed modes of star formation, monolithic and hierarchical. In monolithic formation modes, the gas is already contained within the final volume of the region before stars begin to form, whereas in hierarchical formation the gas extends beyond the final volume of the region (Longmore et al. 2014; Williams et al. 2022). In the hierarchical mode, stars are forming while at the same time the gas is collapsing.

The kinds of star-forming regions these modes produce are of interest not only in the context of star formation but also in the way in which the final star-forming regions that form may influence the architecture of the planetary systems that are produced within them (Adams 2010; Parker 2020).

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 while 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 star-forming region (Parker et al. 2014). To enable comparisons of observations and simulations, we need to be able to quantify parameters of the star formation regions, such as the degree of substructure and mass segregation (Cartwright & Whitworth 2004; Allison et al. 2009; Sánchez & Alfaro 2009; Gouliermis, Hony & Klessen 2014; Kuhn et al. 2014; Alfaro & González 2016; González & Alfaro 2017; Jaffa, Whitworth & Lomax 2017; Joncour et al. 2018; Buckner et al. 2019; Arnold, Wright & Parker 2022).

Early methods such as the autocorrelation 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 were used extensively to determine

---

the degree of substructure, with early work suggesting that breaks in the two-point correlation function corresponded to the Jeans length (Simon 1997) (though see Bate, Clarke & McCaughrean 1998) and the size of the widest stellar binaries in the regions in question (Kraus & Hillenbrand 2008; Joncour, Duchêne & Moraux 2017).

Subsequent work made extensive use of minimum spanning trees (MSTs) to quantify structures in star-forming regions. Cartwright & Whitworth (2004) introduced the Q-parameter to quantify spatial substructure, and Allison et al. (2009) introduced the Λ_{MSR} method to quantify mass segregation.

Parker et al. (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 (Küpper et al. 2011; Maschberger & Clarke 2011).

However, the majority of the above methods are designed to operate on two- or three-dimensional (3D) spatial data, whereas recent observational data (e.g. from *Gaia* and associated ground-based surveys) have provided high-resolution spatial and kinematic data (six-dimensional, 6D).

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, Paciorek & Bartkowiak (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 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 ($M > 50$ M$_\oplus$ and $a < 0.2$ au) around them compared to the lower phase space densities. However, Mustill, Lambrechts & Davies (2022) show 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 paper we aim to test this metric when applied to both synthetic static regions and assess its performance in quantifying phase space structures of N-body simulations of star-forming regions.

The paper is structured as follows. In Section 2, we present the methods used. In Section 3.1, we present the results of testing the Mahalanobis distance's ability to differentiate between different morphologies. In Section 3.2, we show how the Mahalanobis distance and density change with time in N-body simulations. In Section 3.3, we compare the Mahalanobis distance and density to other methods for quantifying structure in star-forming regions. In Section 4, we present a discussion of our results and we conclude in Section 5.

2 METHODS

In this section, we describe the set-up of the N-body simulations, including the initial spatial and kinematic distributions of stars. We
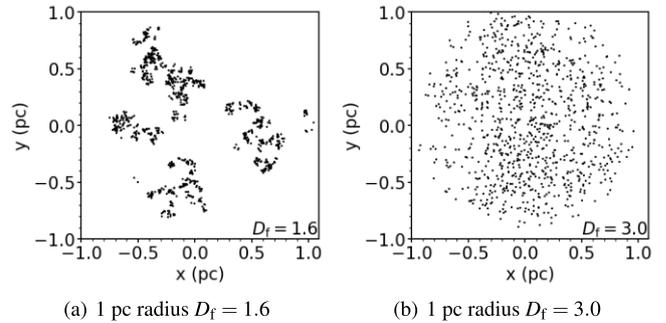
Figure 1. Examples of fractal regions with 1000 stars (stars are shown as black dots): on the left-hand side is a highly substructure region of fractal dimension $D_f = 1.6$ and on the right-hand side is a region with far less substructure with fractal dimension $D_f = 3.0$.

then describe the methods used to quantify the spatial and kinematic distributions in our simulated star-forming regions.

2.1 *N*-body simulations

The *N*-body simulations are run using the KIRA integrator, part of the STARLAB[1] package (Portegies Zwart et al. 1999, 2001). Each simulation has a population of 1000 stars and is run for a simulated time of 10 Myr. Our choice of 1000 stars comes from Lada & Lada (2003) where they find the following relation:

$$N_{cl} \propto M_{cl}^{-2}, \tag{1}$$

where N_{cl} is the number of clusters and M_{cl} is the mass of the cluster. This power law is obeyed for star clusters between masses of $10 < M_{cl}/M_\odot < 10^5$, and our choice of 1000 stars puts our simulations close to the middle of this distribution.

Observations of star-forming regions show that within these complexes there are filaments of denser gas, inside of which pre-stellar cores are observed (Myers 2009; André 2017). These filaments are thought to be caused by supersonic turbulence, which is likely to be responsible for the substructure we observe in these regions (Larson 1981; Kraljic et al. 2014). To mimic this substructure, we make use of a box-fractal generator to initialize our simulations with fractal dimensions of $D_f = 1.6$ and $D_f = 3.0$ (Goodwin & Whitworth 2004; Daffern-Powell & Parker 2020).

We run subvirial simulations to match observations that indicate pre-stellar cores may be subvirial with respect to one another and may then virialize and form bound smooth, centrally concentrated star clusters (Foster et al. 2015; Kuznetsova, Hartmann & Ballesteros-Paredes 2015; Parker et al. 2016). We also run a set of simulations that are initially supervirial. We do this to mimic the observations that some young star-forming regions ($\sim$1$-$5 Myr) are supervirial and therefore expanding (Bravi et al. 2018; Kuhn et al. 2019; Kounkel, Deng & Stassun 2022).

Fig. 1 shows example clusters with the fractal dimensions $D_f = 1.6$ and $D_f = 3.0$.

We use the following to define the virial ratio:

$$\alpha_{vir} = \frac{T}{|\Omega|}, \tag{2}$$

where T is the total kinetic energy of the region and Ω is the total potential energy of the region. By using this equation, we can scale

---

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

Table 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)
$D_f = 1.6$	1000	0.1, 0.9	1, 5
$D_f = 3.0$	1000	0.1, 0.9	1, 5

the initial velocities to our desired virial ratio either $\alpha_{vir} = 0.1$ for subvirial regions or $\alpha_{vir} = 0.9$ for supervirial regions. The initial conditions for the simulations are summarized in Table 1.

We assign masses using the Maschberger IMF with a lower mass limit of 0.01 $M_\odot$, upper mass limit of 50.0 $M_\odot$, and a mean mass of 0.2 $M_\odot$ (Maschberger 2013).

The probability distribution function of the Maschberger IMF is

$$p(m) \propto \left(\frac{m}{\mu}\right)^{-\alpha} \left(1 + \left(\frac{m}{\mu}\right)^{1-\alpha}\right)^{-\beta}, \tag{3}$$

where μ is the mean stellar mass, $\alpha = 2.3$ is the high-mass exponent, and $\beta = 1.4$ is the low-mass exponent (Salpeter 1955).

2.2 Smooth centrally concentrated regions

We generate smooth, centrally concentrated regions with radial density profiles in which the stars are randomly distributed using the following relation,

$$n \propto r^{-\alpha}, \tag{4}$$

where r is the distance from the centre of the region and α is the radial density index and has values $\alpha = 0.0, 1.0, 2.0$, and 2.5. We also generate Plummer spheres, regions with a 3D density distribution of the form

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

where M_0 is the total mass of the region, r is the distance from the centre of the region, and a is the Plummer radius (Plummer 1911; Kroupa 2008).

2.3 Generating fractal regions

We follow Goodwin & Whitworth (2004) and Cartwright & Whitworth (2004) to generate substructured regions using the box-fractal method. Other examples of this method can be found in Allison et al. (2009), Parker & Goodwin (2015), and Daffern-Powell & Parker (2020).

The method proceeds as follows. A single star is placed at the centre of a cube whose side length is chosen to be N_{Div}. This cube is then subdivided down into N_{Div}^3 subcubes. A star is then placed at the centre of each of the subcubes. Each of these subcubes then has a probability of being subdivided again given by $N_{Div}^{(3-D_f)}$, where D_f is the fractal dimension. Cubes that are not subdivided have their stars removed along with any previous generations of stars that came before them. A small amount of noise is added to each of the stars to prevent them having a regular looking appearance.

These steps are repeated until the desired number of stars is reached or exceeded in the latest generation. Once this condition is met all previous generations of stars are removed, then the remaining stars

are randomly removed until the desired number of stars is reached. By removing the stars in this manner, we end up with stars distributed inside of a spherical volume.

The velocities of the first generation of stars are picked from a Gaussian with mean zero, with each subsequent generation of stars inheriting the velocity of the previous generation plus a random component. This results in stars that are close to each other having similar velocities and stars far apart from one another having very different velocities. The velocities are related to the length-scale of the region with the following relation $V(L) \propto L^{3-D_f}$ where L is the length-scale of interest and D_f is the desired fractal dimension (Parker & Wright 2018).

2.4 The Mahalanobis distance

The Mahalanobis distance is a metric that measures distances between points to the average in the distribution in N-dimensional phase spaces (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 the phase space; this is the Mahalanobis distance, M_d.

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

$$\vec{x} = (x_1, x_2, x_3, \ldots, x_N)^T, \tag{6}$$

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

$$M_d(\vec{x}, \vec{\mu}) = \sqrt{(\vec{x} - \vec{\mu})^T \, \mathbf{S}^{-1} \, (\vec{x} - \vec{\mu})}, \tag{7}$$

where the $\vec{x}$ is the point vector, $\vec{\mu}$ 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.

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

To calculate the Mahalanobis density, we first must define the Mahalanobis distance between points in the phase space (i.e. distance between $\vec{x}$ and $\vec{y}$). We follow Winter et al. (2020) and use

$$m_d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \, \mathbf{S}^{-1} \, (\vec{x} - \vec{y})}, \tag{8}$$

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 of interest.

The calculation of the Mahalanobis density proceeds as follows. First, we find the Mahalanobis distance to the Nth nearest neighbour, then we divide the nearest neighbour number by the volume whose

---

[2]We have used different notation for the Mahalanobis distance to avoid confusion with the fractal dimension and also the number of dimensions. We instead use M_d instead of D as used in Winter et al. (2020) to avoid confusion with the fractal dimension. We also change the number of dimensions in the phase space from D to D_p again to avoid confusion with the fractal dimension, which we represent as D_f.

side length is defined as the Mahalanobis distance to the Nth nearest neighbour. The Mahalanobis density is then defined as

$$\rho_{\mathrm{m,N}} = N m_{\mathrm{d,N}}^{-D_{\mathrm{p}}}, \tag{9}$$

where $\rho_{\mathrm{m},N}$ is the Mahalanobis density, N is the nearest neighbour number, $m_{\mathrm{d},N}$ is the Mahalanobis distance to the Nth nearest neighbour, and D_{p} is the number of dimensions in the phase space (Winter et al. 2020). The Mahalanobis densities are then normalized so that the median Mahalanobis density is unity.

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

2.5 Local surface density ratio

The local surface density ratio Σ_{LDR} was introduced in Maschberger & Clarke (2011) to quantify the differences between the surface densities of subsets of stars within their host regions and for this work we choose the 10 most massive stars as the subset of interest. The algorithm proceeds as follows. For each star we find the distance to its Nth nearest neighbour, then we calculate the circular area whose radius is the distance to the Nth nearest neighbour. To find the surface density of the stars, we divide the nearest neighbour number by this area. We use a nearest neighbour number of 5 for this work.

The ratio is defined as

$$\Sigma_{\mathrm{LDR}} = \frac{\tilde{\Sigma}_{\mathrm{subset}}}{\tilde{\Sigma}_{\mathrm{all}}}, \tag{10}$$

where $\tilde{\Sigma}_{\mathrm{subset}}$ is the median surface density found for the 10 most massive stars and $\tilde{\Sigma}_{\mathrm{all}}$ is the median surface density found for the entire region. Therefore, if $\Sigma_{\mathrm{LDR}} > 1$ the 10 most massive stars are found in areas of higher than average stellar surface density, and conversely, if $\Sigma_{\mathrm{LDR}} < 1$ 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 $p < <0.01$, we reject the null hypothesis that the 10 most massive stars share the same underlying distribution of surface densities compared to the entire region.

2.6 Mass segregation ratio

The mass segregation ratio Λ_{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. The method makes use of MSTs which are graphs of points connected to each other in such a way that the total length of the tree is minimized and that all points are connected to at least one other point with no closed loops.

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

$$\Lambda_{\mathrm{MSR}} = \frac{\langle l_{\mathrm{average}} \rangle}{l_{10}} {}^{+\sigma_{5/6}/l_{10}}_{-\sigma_{1/6}/l_{10}}, \tag{11}$$

where $<l_{\mathrm{average}}>$ 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 we construct 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, we mark the value 1 to show the boundary between mass segregation and inverse mass segregation. However, we follow Parker & Goodwin (2015) and only take ratio values above 2 to be signs of mass segregation, to avoid false positives.

We follow Parker (2018) and calculate the uncertainty using the randomly constructed MSTs. First, we 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 us values which correspond to a 66 per cent deviation from the median MST length found.

2.7 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. First, the normalized 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 normalize 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 normalized by diving it by $\frac{N_{\mathrm{total}} A}{N_{\mathrm{total}} - 1}$, where N_{total} is the number of stars in the region and A is the area of the region. We use the circular area [see Schmeja & Klessen (2006); Parker (2018) for a discussion on normalization 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

$$Q = \frac{\bar{m}}{\bar{s}}, \tag{12}$$

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

3 RESULTS

We show the results of the Mahalanobis distance applied to both static and N-body simulations of star-forming regions with various initial conditions. We 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 star-forming regions.

3.1 Static regions

We first find the Mahalanobis distances between stars and the average point in the region, $\vec{\mu}$, and calculate the average Mahalanobis distance in their respective regions ($\bar{M}_{\mathrm{d}}$) for sets of synthetic and static star clusters, with each set having a different structural parameter. Each region in the set consists of 1000 stars. We calculate $\bar{M}_{\mathrm{d}}$ for substructured regions with fractal dimensions $D_{\mathrm{f}} = 1.6, 2.0, 2.6, 3.0$ and clusters with radial density profile indexes, $\alpha = 0.0, 1.0, 2.0,$ 2.5. We also show the results of a set of 100 Plummer spheres which have a radial density profile described by equation (5) with a radial density index of 2.5 (see Section 2.2).
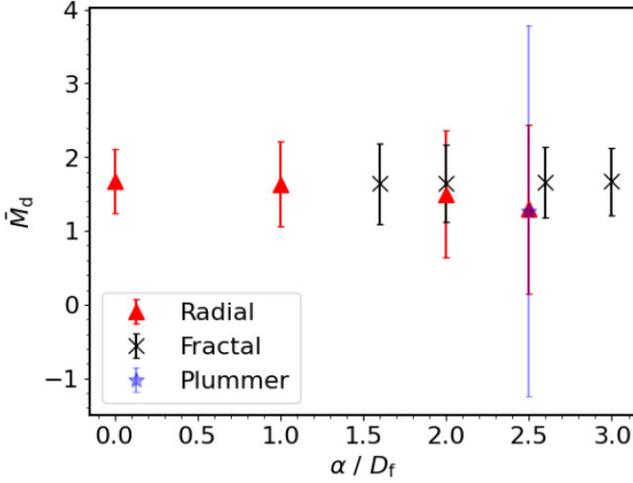
Figure 2. Mean of the mean Mahalanobis distances calculated in the 3D phase space for sets of 100 different star-forming regions plotted against the structural parameter used to make the sets. The red triangles are the smooth, centrally concentrated radial regions, the purple star (on top of the red 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.

Fig. 2 shows the mean of the means for the M_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. We first calculate the mean Mahalanobis distance in each of the 100 regions in the set and then we 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. Fig. 2 clearly shows that $\bar{M}_d$ is degenerate across a wide range of morphologies and we are therefore unable to use $\bar{M}_d$ calculated in the 3D phase space to differentiate between the different morphologies. There is much more spread in the values for the Plummer sphere compared to the radial and fractal models, with the fractals having the smallest spread of $\bar{M}_d$ 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 normalize over very distant stars.

3.2 *N*-body results

Fig. 3 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 $D_f = 1.6$ and 1 pc radii. The left-hand panel shows the results for subvirial (collapsing) regions, 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 for 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 we see a less pronounced decrease in $\bar{M}_d$ compared to the initially subvirial regions. The $\bar{M}_d$ calculated in the 6D phase space shows more modest decrease over time for both initially sub- and supervirial simulations.

Fig. 4 shows the mean Mahalanobis density ($\bar{\rho}_{m,20}$) 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 $D_f = 1.6$ and initial radii of 1 pc. The highest Mahalanobis densities are calculated in the 3D phase space 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 the Appendix, we show the relationship between the Mahalanobis distance and density for the high density simulations with and without substructure. We 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

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

(b) $D_f = 1.6$, supervirial, 1 pc
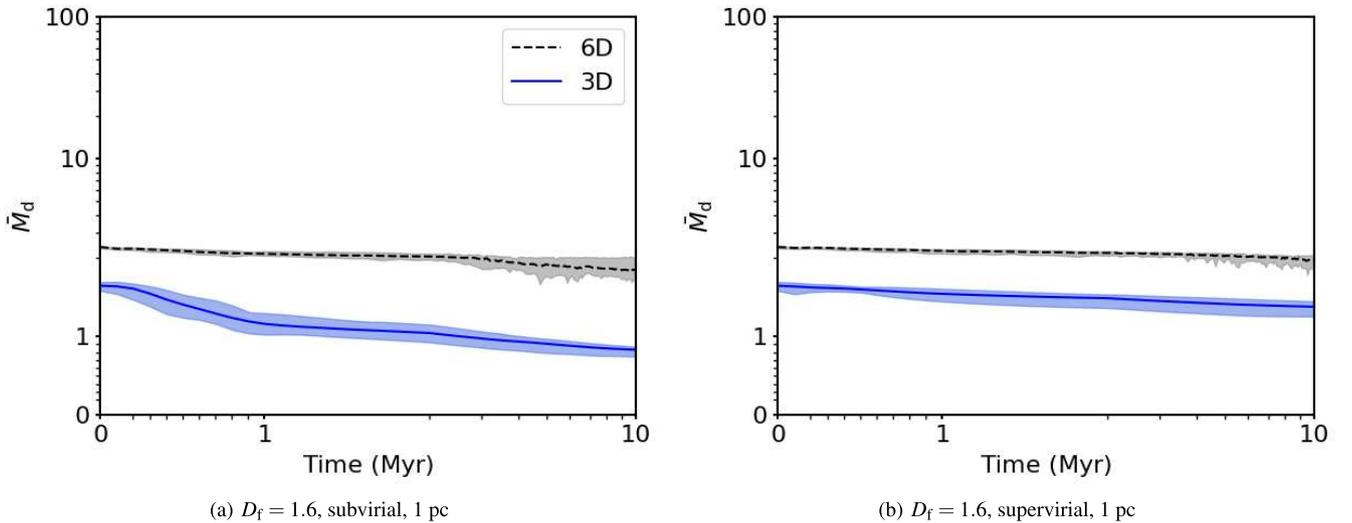
Figure 3. Plots of the mean Mahalanobis distance calculated in both the 3D and 6D phase spaces against time for regions with both high initial volume densities and high degrees of substructure (i.e. fractal dimension $D_f = 1.6$ with radii of 1 pc) consisting of 1000 stars. The shaded areas show the 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.

(a) $D_f = 1.6$, subvirial, 1 pc   (b) $D_f = 1.6$, supervirial, 1 pc
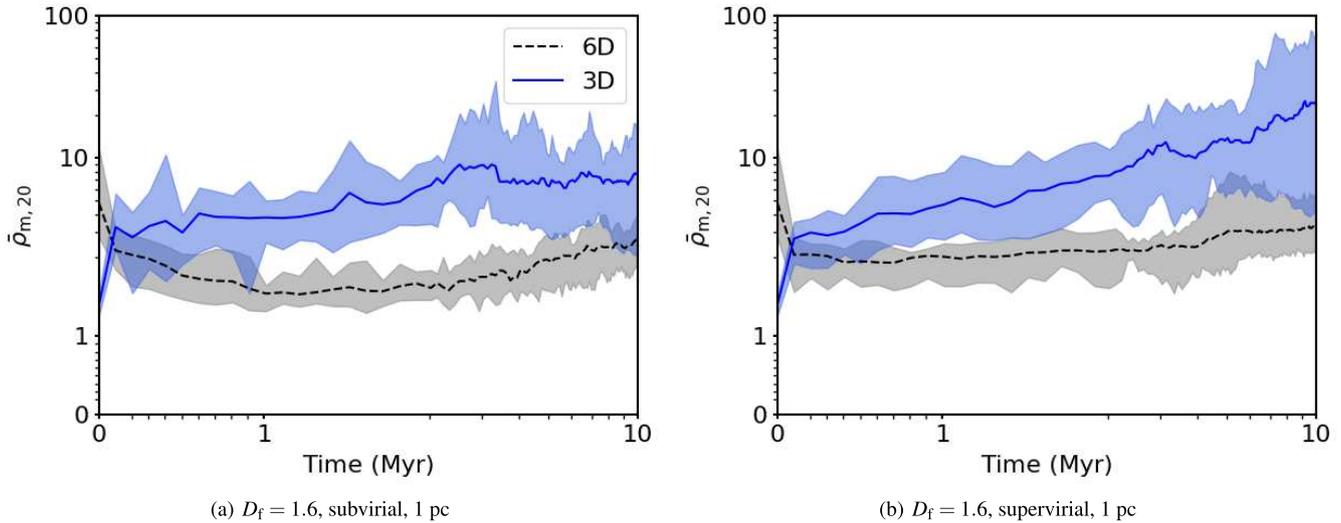
Figure 4. The mean Mahalanobis density against time for each of the 10 subvirial (left-hand panel) and supervirial simulations (right-hand panel) with fractal dimension $D = 1.6$ and radii of 1 pc. The simulations consist of 1000 stars. The blue shaded area shows the minimum and maximum mean Mahalanobis 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 show 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^{-1} with a mean stellar mass density of around 201 $M_\odot \, pc^{-3}$.

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 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.

We also calculate the Mahalanobis density and distance 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 $M_\odot \, pc^{-3}$). We find that 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.

We show the evolution of the 3D and 6D Mahalanobis densities for these more diffuse simulations in Fig. 5, 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, we see the same trends as in Fig. 4 but a slight difference for the 6D phase space. Now $\bar{\rho}_{m,20}$ decreases in the subvirial simulations as the regions evolve, and the supervirial simulations show a steady Mahalanobis density after 1 Myr.

We 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).

Fig. 6 shows the Mahalanobis distance over time for regions with a fractal dimension $D_f = 3.0$ 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. 3, we see that the $D_f = 3.0$ 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 $D_f = 3.0$ at 10 Myr. However, this is not seen

in the 6D phase space Mahalanobis distances that 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.

Fig. 7 shows the 3D and 6D $\bar{\rho}_{m,20}$ against time for the simulations with no primordial substructure (with an initial fractal dimension $D_f = 3.0$ and radii 1 pc). The left-hand panel shows $\bar{\rho}_{m,20}$ against time for the subvirial simulations. It shows the same increase in Mahalanobis density as the $D_f = 1.6$ simulations, but we do not see the same initial decrease in the 6D $\bar{\rho}_{m,20}$ that we see in the $D_f = 1.6$ 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.

The right-hand panel of Fig. 7 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. We see similar behaviour for the 6D phase space in both the subvirial and supervirial simulations.

We show the 3D and 6D Mahalanobis densities for simulations with $D_f = 3.0$ and initial radii of 5 pc in Fig. 8. These low-density simulations have mean number density of around 1 star pc^{-3}, or mean stellar mass density of around 0.7 $M_\odot \, pc^{-3}$. We find similar results to the more dense subvirial simulations simulations, where the 3D Mahalanobis density clearly traces the 'pump' of the collapse, which then decreases as stars move apart. The time the 'bump' occurs is
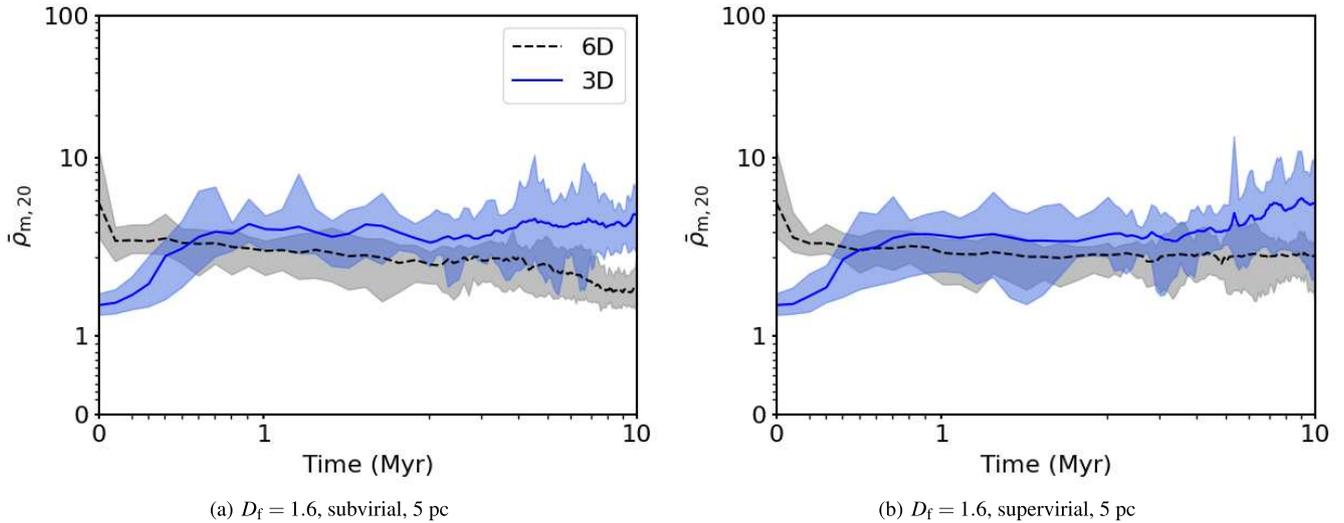
Figure 5. Plots showing the mean Mahalanobis density against time for each of the 10 subvirial (left-hand panels) and supervirial (right-hand panels) simulations with fractal dimension $D_f = 1.6$ and radii of 5 pc. These simulations have a low initial stellar number density with a mean around 3 stars 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.
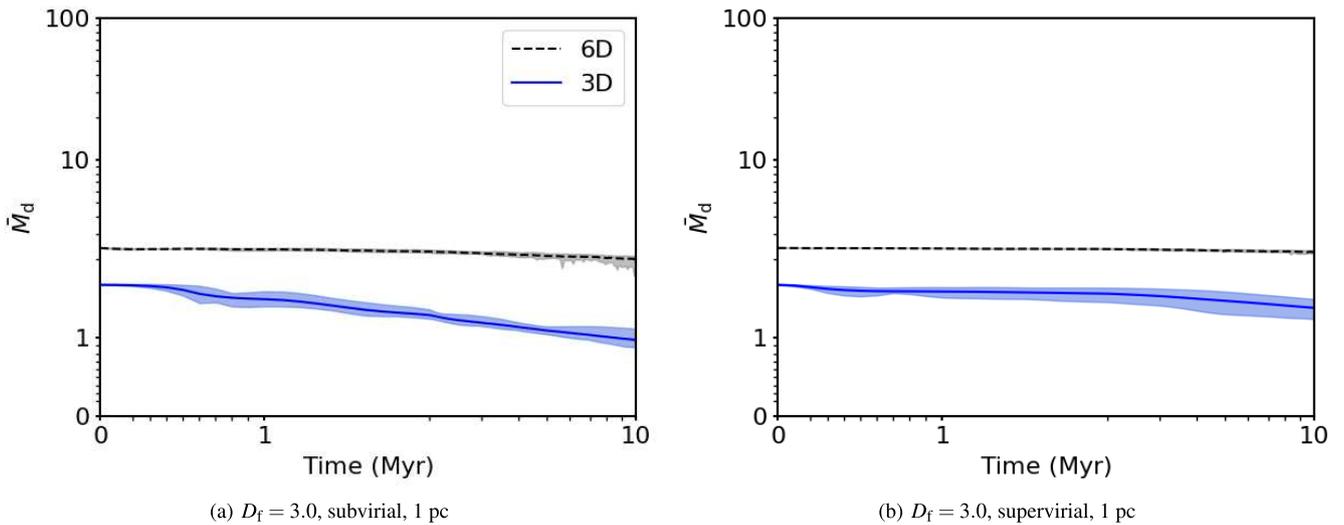
Figure 6. Plots of the mean Mahalanobis distance from each star to the average in simulations without primordial substructure, i.e. a fractal dimension of $D_f = 3.0$, over time. The shaded blue area and solid blue line show the minimum and maximum mean Mahalanobis distance found across the 10 simulations 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.

delayed by several Myr compared to the higher density regions, due to the longer dynamical time-scales.

3.3 Comparison to other methods of quantifying structure

We now plot the 3D and 6D Mahalanobis densities against other measures of quantifying structure in star-forming regions. Fig. 9 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 we are measuring the mean Mahalanobis density, we are sensitive to a small number of stars being ejected which we see as a decrease in the mean Mahalanobis density for the subvirial simulations.

We first show the Mahalanobis density versus the amount of mass segregation as defined by Λ_{MSR} Allison et al. (2009) in Fig. 9(a) and (b). 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 $\Lambda_{MSR} > 2$. The reason for the dissipation in the amount of mass segregation is due to the ejection

(a) $D_\mathrm{f} = 3.0$, subvirial, 1 pc

(b) $D_\mathrm{f} = 3.0$, supervirial, 1 pc
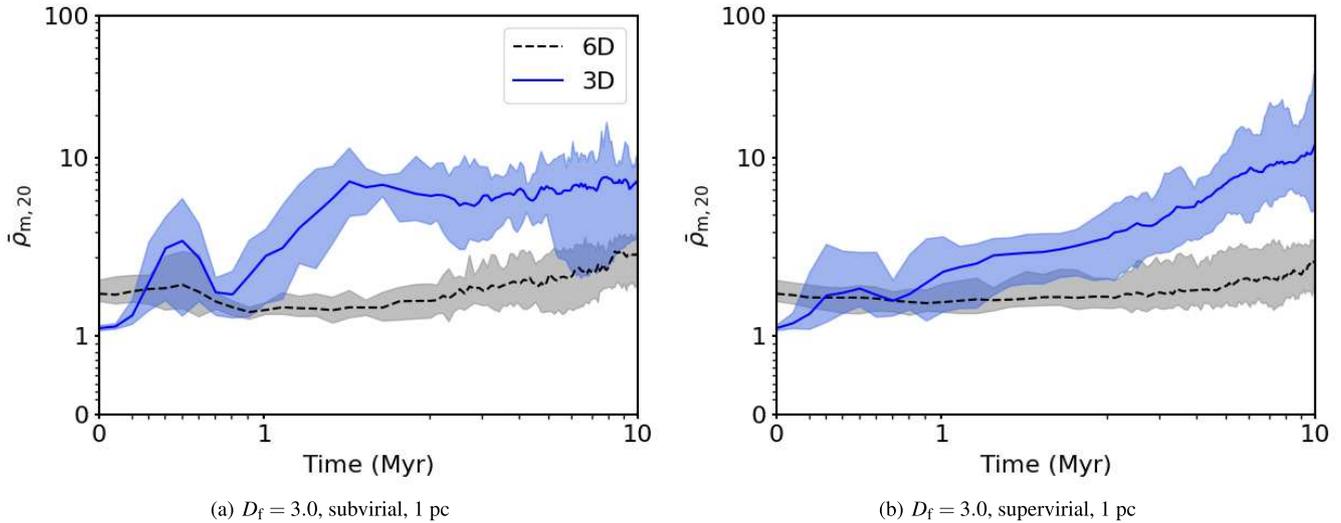
Figure 7. The mean Mahalanobis density against time for each of the 10 subvirial (left-hand panel) and supervirial (right-hand panel) simulations without primordial substructure (fractal dimension $D_\mathrm{f} = 3.0$) and radii of 1 pc. The shaded blue area and solid blue line show the range of mean 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) $D_\mathrm{f} = 3.0$, subvirial, 5 pc

(b) $D_\mathrm{f} = 3.0$, supervirial, 5 pc
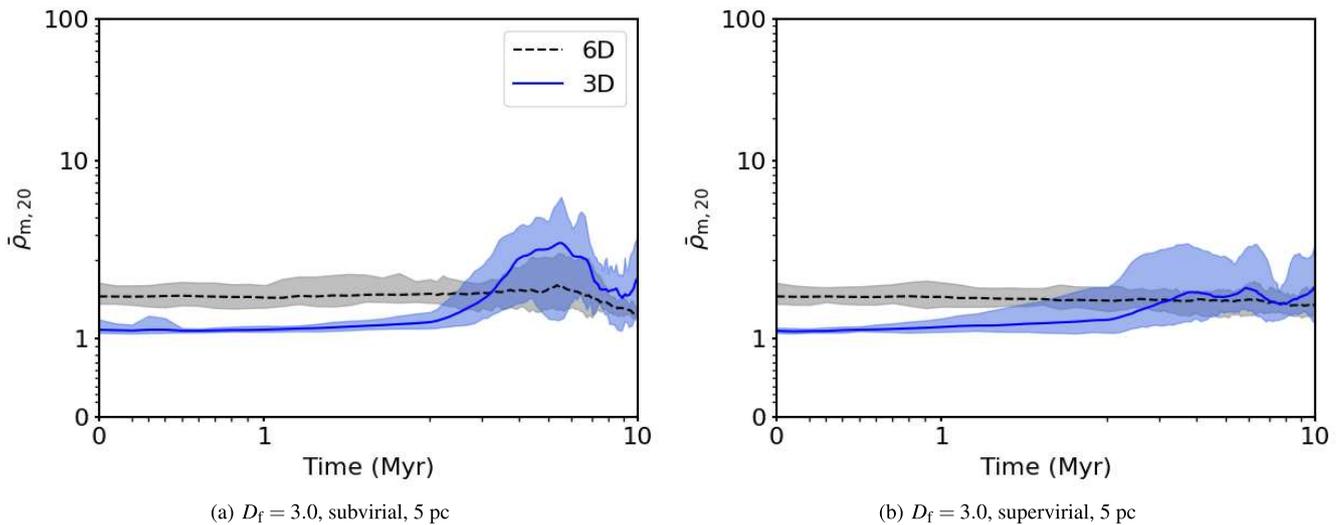
Figure 8. The mean Mahalanobis density against time for each of the 10 subvirial (left-hand panel) and supervirial (right-hand panel) simulations without primordial substructure (fractal dimension $D_\mathrm{f} = 3.0$) with radii of 5 pc. The shaded blue area and solid blue line show the range of mean 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.

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 Λ_MSR can increase to the value we see in Fig. 9 (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 $\bar{\rho}_\mathrm{m,20}$ versus Λ_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 $\bar{\rho}_\mathrm{m,20}$ is combined with the Q-parameter. Fig. 9(c) and (d) shows 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 of the regions maintaining traces of substructure for 5 Myr as measured using Q (i.e. $Q < 0.8$).

Fig. 9 (e) and (f) shows $\bar{\rho}_\mathrm{m,20}$ plotted against the relative local surface density ratio, Σ_LDR. We find that for both subvirial and supervirial simulations there is an increase in the local surface density of the 10 most massive stars compared to all stars in the region. 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
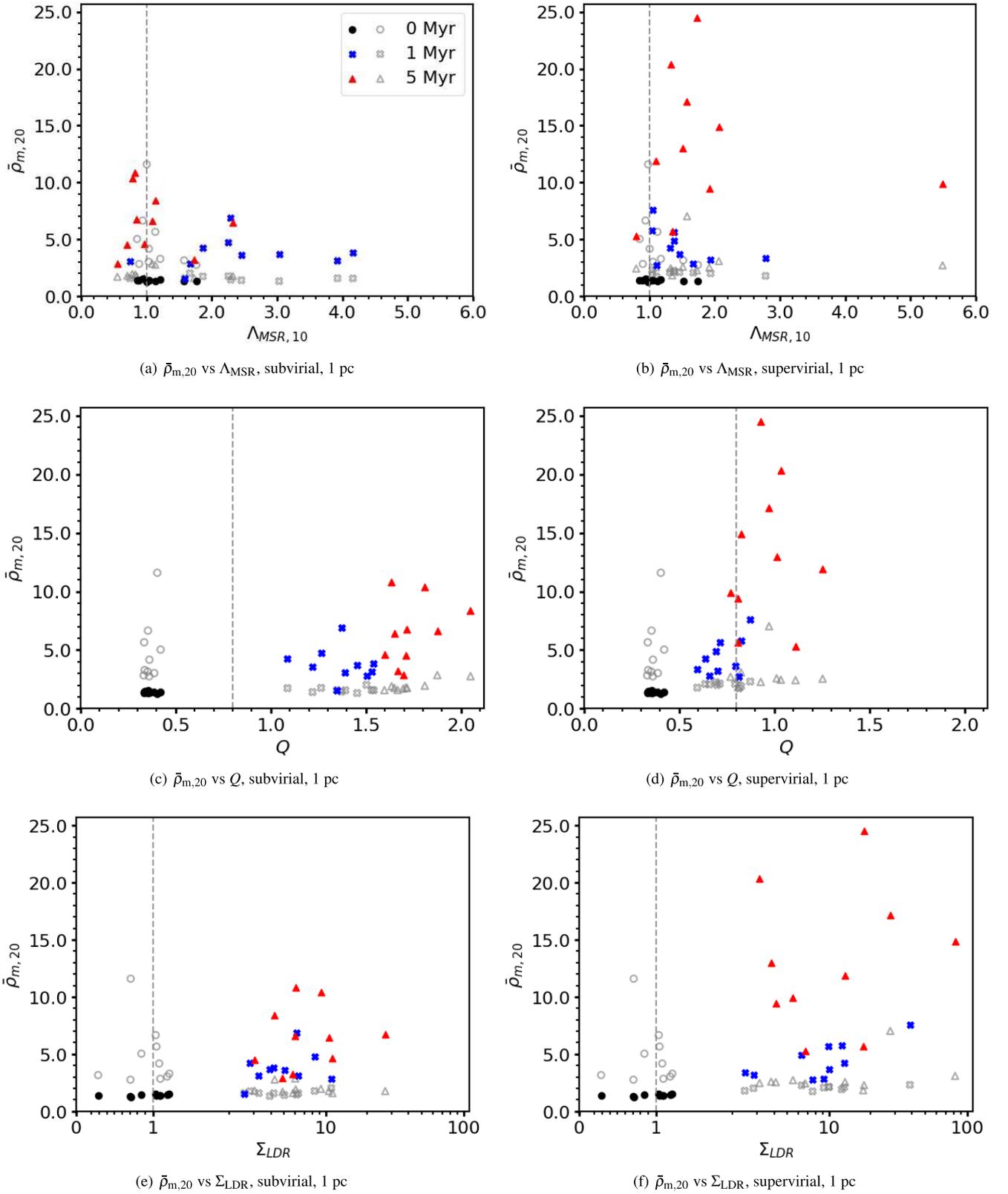
Figure 9. The mean Mahalanobis density calculated for 3D and 6D phase spaces plotted against other methods of quantifying structure for 10 subvirial and supervirial simulations that are initially substructured with fractal dimension $D_{\rm f} = 1.6$ and 1 pc radii. The left-hand panels show the results 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 $\bar{\rho}_{\rm m,20}$ is plotted against, with the top row showing $\Lambda_{\rm MSR}$, second row showing Q, and the bottom row showing $\Sigma_{\rm LDR}$.

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, and we can use this, and the different evolution of the Q-parameter and Λ_{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. We find that the range of $\bar{\rho}_{m,20}$ decreases when calculated in 6D. Meaning, we see that the $\bar{\rho}_{m,20}$ values overlap making differentiating between different snapshots and virial states impractical.

We now show the same plots but for simulations with little to no primordial spatial or kinematic substructure. Fig. 10 shows the mean 3D and 6D Mahalanobis densities plotted against the established methods for simulations that have an initial fractal dimension $D_f = 3.0$ 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 $\bar{\rho}_{m,20}$ plotted against Λ_{MSR} for the 10 simulations. For the subvirial simulations, we see mass segregation detected in 3 of the 10 simulations; for the supervirial simulations, we detect mass segregation in 2 of the 10 simulations (recall that our threshold for declaring mass segregation is $\Lambda_{MSR} > 2$.)

As for the highly substructured simulations ($D_f = 1.6$), the plot of the mean Mahalanobis density when combined with the Q-parameter gives the clearest distinction between the different snapshots. For panels (e) and (f), we show $\bar{\rho}_{m,20}$ against Σ_{LDR}. We 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 Σ_{LDR} between 0.1 and 10, whereas the supervirial simulations all finish with $\Sigma_{LDR} > 1$.

The grey markers in Fig. 10 show the 6D Mahalanobis densities against the established methods. We find once again that the spread in the Mahalanobis densities has decreased, making differentiating between different times or virial states impractical.

4 DISCUSSION

We have been motivated to test the Mahalanobis density 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 6D 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 work is not to make a scientific assessment of the Mahalanobis density in its use in planet formation specifically but simply to see how it changes over time when looking at simple *N*-body simulations of star-forming regions to see what information, if any, it may be able to give us about the initial conditions (i.e. virial state, density, and initial morphology).

Due to the simplicity of our 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 star-forming regions. The presence of an external Galactic tidal field would likely increase the dissolution of the star-forming region 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 we do not simulate any gas, and therefore, there is no gas potential and also that our 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 we do not simulate planets in our simulations and so how representative our simulations are of real regions with exoplanet host stars is uncertain.

In Section 3.3, we show that the 6D Mahalanobis density in isolation cannot be used to reliably infer the initial conditions of star-forming regions due to overlap in the sub- and supervirial values. However, when the Mahalanobis density in the 3D phase space is combined with Λ_{MSR}, Σ_{LDR}, or Q, then the initial virial conditions can be inferred. This is most clear to see in Fig. 10.

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 that will have similar positions and therefore higher phase space densities. The Mahalanobis distances between these groupings is reduced when we multiply by the inverse of the covariance matrix. In the subvirial cases, we see lower 3D phase space densities due to stars being ejected and ending up in relative isolation compared to the rest of the region. As we are using the mean Mahalanobis density, we are sensitive to only a few stars being ejected.

We cannot determine the initial conditions using the 6D Mahalanobis density. One would assume that the more data we have (and therefore dimensions in the phase space), the more clearly we would see the distinction between sub- and supervirial simulations. However, somewhat counter-intuitively, adding more dimensions to the phase space effectively 'washes' out any information that would allow us to determine the initial conditions of our star-forming regions. For example, in our $D_f = 1.6$ 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 removing any information about the initial conditions of the region (i.e. different positions but similar velocities).

We 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 D_f and the initial radii of the region. We 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.

5 CONCLUSION

We present *N*-body simulations with different initial fractal dimensions, virial states, and initial radii, and quantify the 3D and 6D phase space densities using the Mahalanobis distance. We compare the performance of the Mahalanobis density to more established methods for quantifying structure in star-forming regions, namely Λ_{MSR}, Σ_{LDR}, and Q. We 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. Our conclusions are as follows:

(i) The Mahalanobis distance in the 3D phase space is degenerate across a wide range of morphologies commonly observed in star-forming regions, associations and clusters, and so it cannot be used to differentiate between different morphologies.

(ii) The 3D Mahalanobis densities, $\rho_{m,20}$, can be used to distinguish between the high and low stellar density regions with large amounts of substructure ($D_f = 1.6$). The low stellar density regions

(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
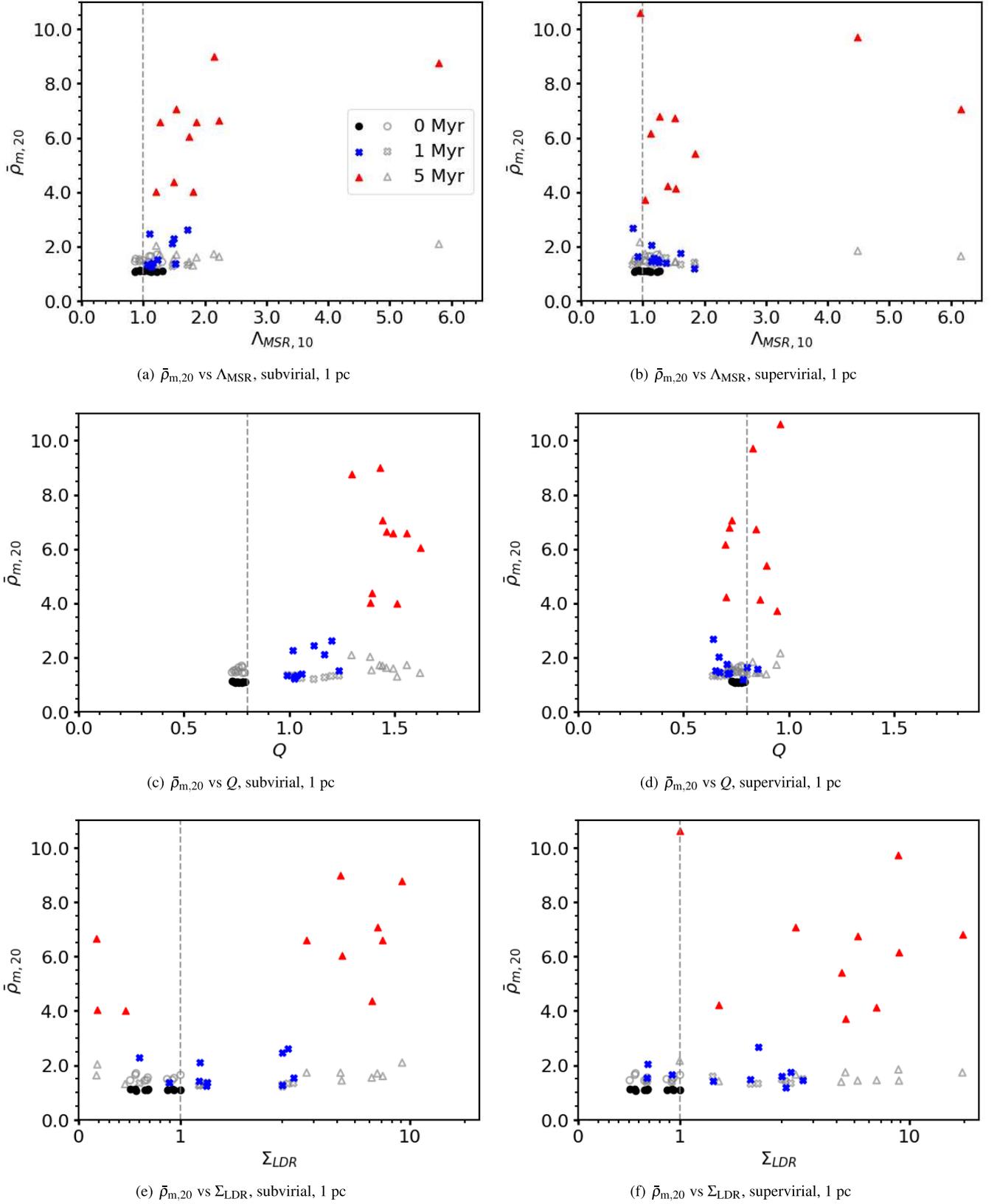
Figure 10. The mean Mahalanobis density calculated for the 3D and 6D phase spaces plotted against other methods of quantifying substructure for 10 subvirial and supervirial simulations which have little to no initial substructured with fractal dimension $D_{\mathrm{f}} = 3.0$ and 1 pc radii. The left-hand 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. We show the mean 6D Mahalanobis densities at 0, 1, and 5 Myr with 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}.

show similar behaviour but delayed by around 0.6 Myr compared to the high volume density regions due to the dynamical time-scales being longer. This effect is even more pronounced in the simulations with little to no initial substructure ($D_f = 3.0$), where we see the 'bump' occurring several Myr later than in the higher stellar density simulations corresponding to the subvirial collapse.

(iii) We 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.

(iv) When using the 6D Mahalanobis densities, we see 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.

We therefore advise against using the Mahalanobis distance as a method to quantify the morphology of star-forming regions due to its degeneracy across both substructured regions and also 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 we advocate using combinations of spatial and kinematic metrics instead.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

Data are available on reasonable request to the corresponding author.

REFERENCES

Adams F. C., 2010, Ann. Rev. Astron. Astrophys., 48, 47
Alcock H. L., Parker R. J., 2019, MNRAS, 490, 350
Alfaro E. J., González M., 2016, MNRAS, 456, 2900
Allison R. J., Goodwin S. P., 2011, MNRAS, 415, 1967
Allison R. J., Goodwin S. P., Parker R. J., Zwart S. F. P., de Grijs R., Kouwenhoven M. B. N., 2009, MNRAS, 395, 1449
Allison R. J., Goodwin S. P., Parker R. J., Zwart S. F. P., de Grijs R., 2010, MNRAS, 407, 1098
André P., 2017, Comptes Rendus Geosci., 349, 187
Arnold B., Wright N. J., Parker R. J., 2022, MNRAS, 515, 2266
Bastian N., Gieles M., Ercolano B., Gutermuth R., 2009, MNRAS, 392, 868
Bate M. R., Clarke C. J., McCaughrean M. J., 1998, MNRAS, 297, 1163
Bravi L. et al., 2018, A&A, 615, A37
Buckner A. S. M. et al., 2019, A&A, 622, A184
Cartwright A., 2009, MNRAS, 400, 1427
Cartwright A., Whitworth A. P., 2004, MNRAS, 348, 589
Daffern-Powell E. C., Parker R. J., 2020, MNRAS, 493, 4925
Elmegreen B. G., 2000, ApJ, 530, 277
Foster J. B. et al., 2015, ApJ, 799, 136
Gomez M., Hartmann L., Kenyon S. J., Hewett R., 1993, AJ, 105, 1927
González M., Alfaro E. J., 2017, MNRAS, 465, 1889
Goodwin S. P., Whitworth A. P., 2004, A&A, 413, 929
Gouliermis D. A., Hony S., Klessen R. S., 2014, MNRAS, 439, 3775

Harris C. R. et al., 2020, Nature, 585, 357
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jaffa S. E., Whitworth A. P., Lomax O., 2017, MNRAS, 466, 1082
Jakimiec M., Paciorek J., Bartkowiak A., 1991, Acta Astron., 41, 61
Joncour I., Duchêne G., Moraux E., 2017, A&A, 599, A14
Joncour I., Duchêne G., Moraux E., Motte F., 2018, A&A, 620, A27
Klessen R. S., Kroupa P., 2001, A&A, 372, 105
Kounkel M., Deng T., Stassun K. G., 2022, AJ, 164, 57
Kraljic K., Renaud F., Bournaud F., Combes F., Elmegreen B., Emsellem E., Teyssier R., 2014, ApJ, 784, 112
Kraus A. L., Hillenbrand L. A., 2008, ApJ, 686, L111
Kroupa P., 2008, The Cambridge N-Body Lectures, 760, 181
Kuhn M. A. et al., 2014, ApJ, 787, 107
Kuhn M. A., Hillenbrand L. A., Sills A., Feigelson E. D., Getman K. V., 2019, ApJ, 870, 32
Küpper A. H. W., Maschberger T., Kroupa P., Baumgardt H., 2011, MNRAS, 417, 2300
Kuznetsova A., Hartmann L., Ballesteros-Paredes J., 2015, ApJ, 815, 27
Lada C. J., Lada E. A., 2003, Annu. Rev. Astron. Astrophys., 41, 57
Larson R. B., 1981, MNRAS, 194, 809
Larson R. B., 1995, MNRAS, 272, 213
Longmore S. N. et al., 2014, in Protostars and Planets VI. Univ. Arizona Press, USA
Reprint of: Mahalanobis P. C., 1936, Sankhya A, 80, 7
Maschberger T., 2013, MNRAS, 429, 1725
Maschberger T., Clarke C. J., 2011, MNRAS, 416, 541
Mustill A. J., Lambrechts M., Davies M. B., 2022, A&A, 658, A199
Myers P. C., 2009, ApJ, 700, 1609
Palmeirim P. et al., 2013, A&A, 550, A38
Parker R. J., 2018, MNRAS, 476, 617
Parker R. J., 2020, R. Soc. Open Sci., 7, 201271
Parker R. J., Goodwin S. P., 2015, MNRAS, 449, 3381
Parker R. J., Wright N. J., 2018, MNRAS, 481, 1679
Parker R. J., Wright N. J., Goodwin S. P., Meyer M. R., 2014, MNRAS, 438, 620
Parker R. J., Goodwin S. P., Wright N. J., Meyer M. R., Quanz S. P., 2016, MNRAS, 459, L119
Plummer H. C., 1911, MNRAS, 71, 460
Portegies Zwart S., Makino J., McMillan S. L. W., Hut P., 1999, Star Cluster Ecology III: Runaway Collisions in Young Compact Star Clusters, 348, 117
Portegies Zwart S. F., McMillan S. L. W., Hut P., Makino J., 2001, MNRAS, 321, 199
Salpeter E. E., 1955, ApJ, 121, 161
Sánchez N., Alfaro E. J., 2009, ApJ, 696, 2086
Schisano E. et al., 2014, ApJ, 791, 27
Schmeja S., Klessen R. S., 2006, A&A, 449, 151
Schoettler C., Parker R. J., Arnold B., Grimmett L. P., de Bruijne J., Wright N. J., 2019, MNRAS, 487, 4615
Schoettler C., de Bruijne J., Vaher E., Parker R. J., 2020, MNRAS, 495, 3104
Siegal B. S., Griffiths J. C., 1974, Moon, 9, 397
Simon M., 1997, ApJ, 482, L81
Virtanen P. et al., 2020, Nature Methods, 17, 261
Williams B. A. et al., 2022, MNRAS, 514, 578
Winter A. J., Kruijssen J. M. D., Longmore S. N., Chevance M., 2020, Nature, 586, 528

APPENDIX: MAHALANOBIS DISTANCE VERSUS MAHALANOBIS DENSITY

In Fig. A1, we show the relation between the Mahalanobis distance and density across the two different phase spaces investigated and the two different initial virial states.

We see that 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

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

(b) $\bar{\rho}_{m,20}$ vs $\bar{M}_d$, supervirial, 1 pc
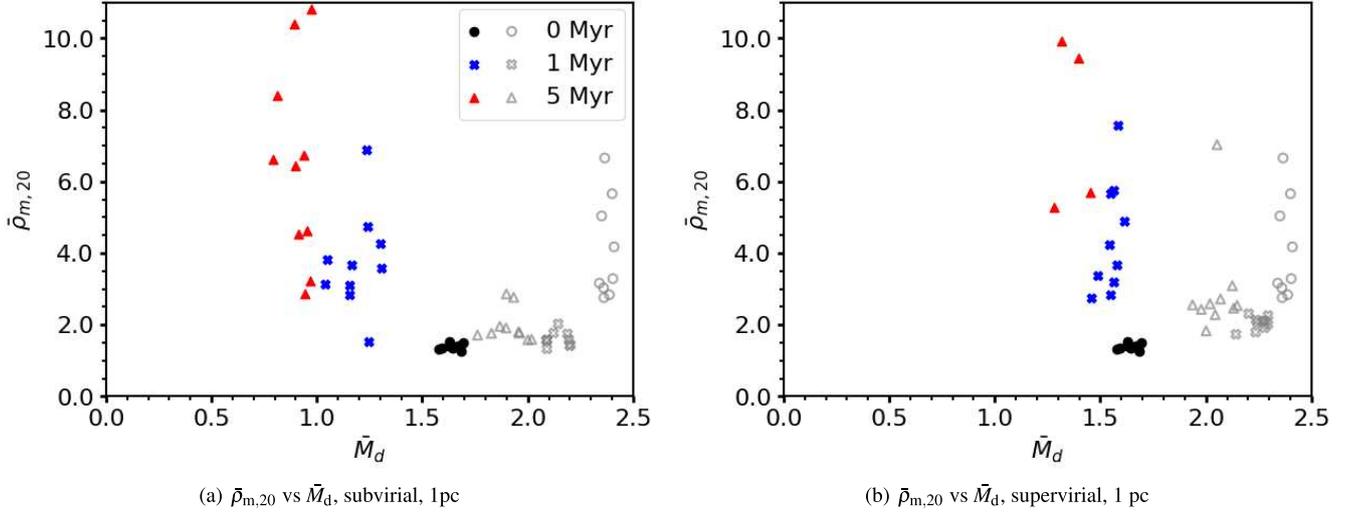
Figure A1. The mean Mahalanobis density ($\bar{\rho}_{m,20}$) plotted against the mean Mahalanobis distance ($\bar{M}_d$) for highly substructured regions with fractal dimensions $D_f = 1.6$ and initial radii of 1 pc. Each region contains 1000 stars. The black circles the values at 0 Myr, the blue plus 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.

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

(b) $\bar{\rho}_{m,20}$ vs $\bar{M}_d$, supervirial, 1 pc
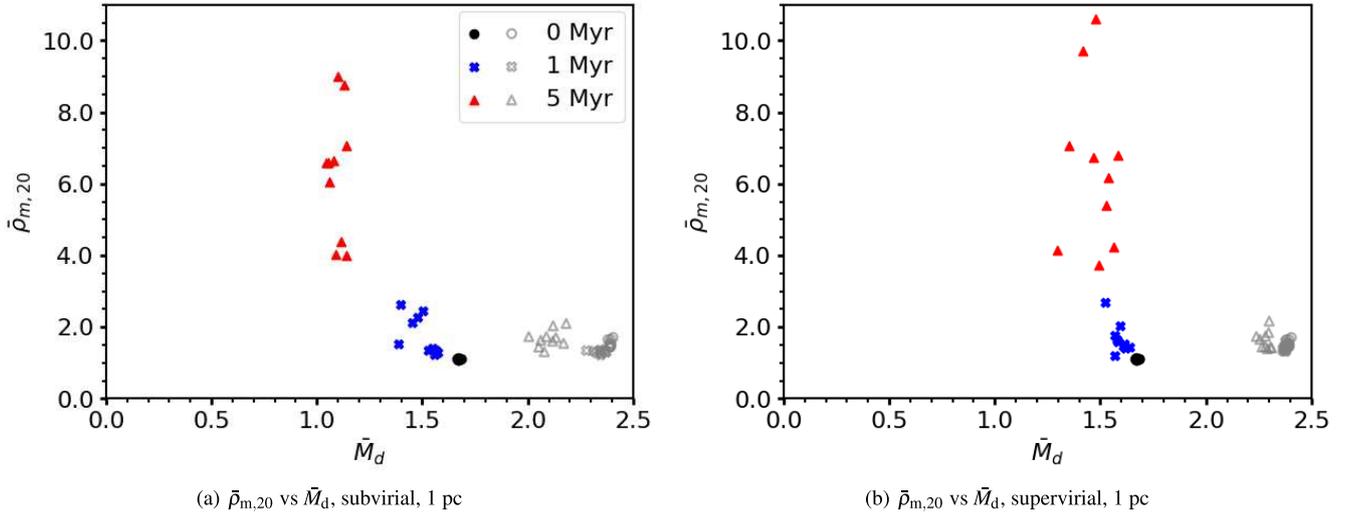
Figure A2. The mean Mahalanobis density ($\bar{\rho}_{m,20}$) plotted against the mean Mahalanobis distance ($\bar{M}_d$) for substructured regions with fractal dimensions $D_f = 3.0$ and scales 1 pc for different snapshots. Each region contains 1000 stars. The black circles the values at 0 Myr, the 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.

impractical. For the supervirial regions, we see 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.

In Figs A1 and A2, we show the position–velocity phase space (6D) with the grey open markers.

Fig. A2 shows the mean Mahalanobis distance plotted against the mean Mahalanobis density for high density (radii of 1 pc) region with little or no substructure ($D_f = 3.0$).

This paper has been typeset from a TeX/LaTeX file prepared by the author.