Statistical descriptors of clouds and clusters

Sarah Elizabeth Jaffa

A Thesis submitted to
Cardiff University
for the degree of
Doctor of Philosophy

March 2018

‘Many people are aware of the Weak and Strong Anthropic Principles. The Weak One says, basically, that it was jolly amazing of the universe to be constructed in such a way that humans could evolve to a point where they make a living in, for example, universities, while the Strong One says that, on the contrary, the whole point of the universe was that humans should not only work in universities but also write for huge sums books with words like “Cosmic” and “Chaos” in the titles.’
Terry Pratchett, Hogfather

ACKNOWLEDGEMENTS

Firstly I would like to thank my excellent supervisor, Ant Whitworth, without whom I would never have started down this path. Thanks, Ant, for helping and encouraging me as well as challenging me, and always being in my corner. Secondly, thanks to Matt Griffin and Nick Wright for making my viva not just bearable but actively enjoyable.

I had the privilege to work with the best group of office mates, lunch crew and tea buddies I could have wished for in a research group. Particular thanks go to Seamus and Scott, for teaching me how to be a PhD student and to Olly for teaching me how to be Bayesian; to Rachael, Gwen and Camilo for being the best thesis support group ever; to Alex and Liz for reminding me why I started this when I most needed it.

I'm so lucky to have the wonderfully supportive and silly family that I do. Lots of love to Mum, Dad and Fran for helping me see the big picture, asking me strange astronomy questions, proof-reading, and reminding me that the wider world exists.

Thanks to Hannah and Matt for dragging me out of the academic bubble, going on adventures and being patient with me when work took over my life. Thanks to Simon, Tom and the rest of the Real Ale crew for taking me to beer festivals, book festivals and ceilidh dances. Thanks to Sam for making me see how small my problems are compared to the stars. Thanks to Dan for helping me learn to walk in someone else's shoes. Thanks to Chris, Tom, Phil and everyone who came to tea, ate cake, and went to the Flute on Fridays or the Blackweir on rugby days. Thanks to everyone who has climbed, picnicked and danced with me.

And finally, thanks to Jess and her iron-fisted Pomodoro timing, without which this thesis would never have been finished.

ABSTRACT

Star formation is a chaotic process, involving the evolution and interaction of a wide variety of structures. The interstellar medium exhibits substructure over a range of scales, and the clusters which form from the densest parts of this material may be imprinted with this clumpy distribution. In this thesis, we describe and evaluate statistical tools for quantifying structures that are important to the star formation process, in order to constrain the underlying physics and robustly compare observations, simulations and synthetic observations.

We describe the basic theory and some common applications of fractal theory in astronomy. We show that some common measures of fractal structure are inconsistent and that comparing values derived from different data types (e.g. continuum data of molecular cloud maps and discrete data of star distributions) can lead to confusion.

We introduce the $\mathcal{Q}^+$ algorithm which quantifies the substructure in star clusters in terms of a fractal distribution. We describe the derivation and validation of this method and apply it to observed and simulated data sets. We examine the possibility of applying this same analysis to continuum data by converting the greyscale image into a statistically representative distribution of points.

We introduce the J plots algorithm which uses the principal moments of inertia of a two-dimensional pixelated structure to quantify its shape. We show that this can be used to identify the shapes of structures extracted from astrophysical images using dendrograms. We apply this method (i) to data from the Hi-GAL survey to demonstrate the identification of ring-like shapes, and (ii) to simulations of filament formation to quantify the differences in structure resulting from the nature of turbulence in the accreting material.

Contents

CHAPTER 1

INTRODUCTION

A philosopher once asked, "Are we human because we gaze at the stars, or do we gaze at them because we are human?" Pointless, really… "Do the stars gaze back?" Now, that's a question.

-Neil Gaiman, Stardust

1.1 AN OVERVIEW OF STAR FORMATION

The study of star formation is crucial to our understanding of the Universe. Stars create most of the heavy elements necessary for life, either by nuclear fusion or in supernovae. Interaction of newly forming stars, in dense clusters or in stellar multiple systems, change the properties of disks around those stars, and therefore influence the formation of planets. Massive stars and clusters can impact the mass budgets and luminosities of entire galaxies. We therefore need to study how and where stars are formed, how they evolve, and how they interact with their environment. Star formation is a chaotic process spanning orders of magnitude in scale, density and temperature. The interplay of physical and chemical processes is complex, non-linear, and in most cases impossible to replicate in a controlled laboratory environment. We therefore rely on the interpretation of observations and the predictive power of simulations and analytic models to constrain our understanding. A very simple one-dimensional model was described over a hundred years ago by Jeans (1902). By examining the dispersion relation of a wave in an isothermal, self-gravitating gas we can find a critical wavelength, λ_J, above which perturbations have exponentially increasing amplitudes. This defines a critical length at which gas of a set density, ρ, and temperature, T, will collapse. This can be extended to three dimensions to define the critical mass, M_J, at which a sphere of ideal gas with constant volume density ρ and mean gas particle mass μ is confined by the balance between internal and external pressure and self-gravity:

$$M_J = \frac{a_0^3}{G^{3/2}\rho^{1/2}}, \tag{1.1}$$

where $a_0 = \sqrt{(k_B T/\mu)}$. Masses above this limit are therefore gravitationally unstable and will contract, while masses below this limit are supported against self-gravity and will not condense out of the medium.

This leads to a simple picture of star formation in which nature only needs to assemble the correct amount of mass within a small enough radius for it to contract and form a star. However, this leads to further questions: how is mass assembled and compressed to reach the required radius? Why do stars form with different masses? Why do stars form more readily in some areas than others? In the past century we have progressed some way towards answering these questions, through the advance of technology bringing higher resolution observations in many different wavelengths, through ever increasing computing power enabling larger and more complex simulations, and through the comparison of these using synthetic observations and data analysis routines. However, there are still many contentious issues in star formation, such as the role of magnetic fields, whether feedback from newly formed massive stars has a positive or negative effect on the star formation rate, and even how to strictly define such simple terms as 'cloud' and 'cluster'.

Recently, the advent of large scale high resolution surveys has greatly increased the amount of raw data available, and enabled the analysis of large numbers of objects as distinct from in-depth studies of single objects. This 'big data' revolution comes with its own challenges, including the reduction, storage and sharing of data as well as robust, speedy and meaningful methods of analysis. Since the processes involved in star formation are highly non-linear, we cannot usually directly compare individual objects or regions, whether observed or simulated. Therefore, we require effective metrics to compare objects in a statistical sense. There are the additional complications of projection, confusion, and optical depth effects which make the definition and fidelity of such metrics particularly difficult. In this thesis I will present some statistical methods for the analysis of astronomical data sets, with the aim of constraining the underlying physics that controls star formation.

1.1.1 THE INTERSTELLAR MEDIUM AND MOLECULAR CLOUDS

The diffuse material from which molecular clouds condense and eventually form stars is called the interstellar medium (ISM). The ISM consists predominantly of hydrogen gas, along with some helium and heavy elements. About 1% by mass is dust and an even smaller percentage is high-energy particles. All of this is bathed in radiation of different intensities and wavelengths, and interacts with the magnetic field and gravitational potential.

The simplest model of the ISM invokes a mixture of three components in approximate pressure equilibrium: cold, dense neutral medium (CNM); warm, relatively dense neutral medium (WNM); and hot rarefied ionised medium (HIM) (McKee & Ostriker, 1977). Table 1.1 gives typical values for the density and temperature of each of these three components, along with their approximate filling factor. These three phases interact dynamically and chemically with one another, and with the interstellar radiation field.

	Temperature (K)	Density (10^{-24} g/cm^{-3})	Filling factor
CNM	100	25	~ 0.03
WNM	8000	0.175	~ 0.2
HIM	500,000	0.01	~ 0.75

Table 1.1. The approximate temperature and density of the three phases in the simple model of the ISM

To observe the ISM we need tracers that cover this wide range of physical conditions. Unfortunately there is no single atom or molecule that is observable in all these environments. The 21cm line of atomic hydrogen traces the bulk of the neutral ISM in diffuse HI clouds, and the warm inter-cloud medium. Both neutral and ionised interstellar matter can be traced by optical and UV absorption against bright background sources, but this is limited to lines of sight with a suitably characterised background source. Ionised matter in HII regions can also be traced by Hα and other recombination lines, as well as free-free emission.

In the dynamic ISM, shocks from supernova or colliding flows compress the gas. Where gas becomes dense enough for self-shielding to take effect the interstellar radiation field is no longer able to heat the gas and it can start to condense and form molecules. We call these regions of higher density Molecular Clouds. They typically have densities in the range 10^{-22} g cm^{-3} $\leq \bar{\rho} \leq 10^{-20}$ g cm^{-3} and masses of 10^2 to 10^6 M$_\odot$. In these environments there is a range of molecules we can observe to probe the different conditions.

Each molecular transition has a different critical density, defined as the density at which the distribution of level populations is dominated by collisions. This is usually the typical density at which we observe a particular molecular line, although other factors such as the optical depth can change the critical density. If the line is optically thin, we can convert the detected intensity into an estimate of column density. If it is optically thick we will not be probing the full depth of the cloud, and so cannot make the same approximation. We therefore need to observe a range of different molecular transitions in order to fully map the structure of the cloud at different densities. Molecular line observations also have the advantage of providing information about the radial velocity of the emitting gas through Doppler shifting, as well as the velocity dispersion and temperature through broadening of the lines.

Dust blocks our view of certain astrophysical phenomena, absorbing or scattering light at short wavelengths so we cannot see, for example, the light produced by young stars that are deeply embedded in dusty clouds. However, this energy is re-emitted in the infrared and can be observed to trace the density and temperature of the dust, and thereby infer the possible sources of heating. Elongated dust grains produce polarised light. Polarisation is more challenging to measure than intensity, but can provide vital information on the strength and direction of the magnetic field, as elongated grains will preferentially align perpendicular to the field.

Various algorithms have been developed to identify molecular cloud structures in astronomical images. SCIMES (Colombo et al., 2015) uses spectral clustering to identify regions with similar properties (such as volume or integrated luminosity) in

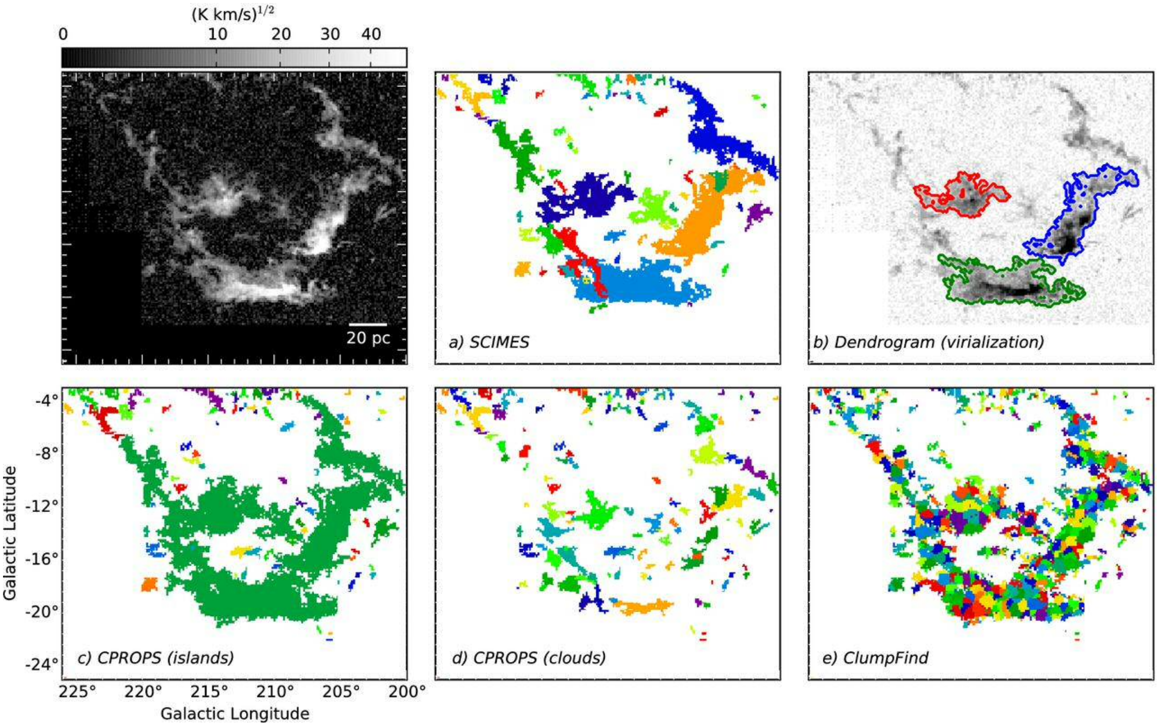

Figure 1.1. An illustration of some different methods for identifying Giant Molecular Clouds (GMCs) applied to the Orion-Monoceros region, taken from Colombo et al. (2015). The top left panel shows the integrated intensity map of CO(1-0). Each subsequent panel shows the projection of GMCs identified in the position-position-velocity data cube by the following algorithms: (a) Scimes (Colombo et al., 2015), (b) Dendrograms (Rosolowsky et al., 2008), (c) and (d) CPROPS using two different settings (Rosolowsky & Leroy, 2006), and (e) ClumpFind (Williams, de Geus & Blitz, 1994).

order to find Giant Molecular Clouds (molecular clouds with mass greater than 10^5 $M_\odot$) in molecular line data cubes. GaussClumps (Stutzki & Güsten, 1990) uses a least-squares fitting procedure to recursively fit the data with Gaussian profiles. ClumpFind (Williams, de Geus & Blitz, 1994) identifies peaks in PPV data cubes and follows them to lower intensities taking steps of some user-defined multiple of the rms noise. This does not assume a set shape for the clumps, but has been noted to be very sensitive to the contour interval chosen (Elia et al., 2007, and references therein). FellWalker (Berry, 2015) is a watershed algorithm, taking all emission above a background threshold and tracing gradients upwards from that level to split the 1, 2 or 3-dimensional data into regions associated with individual significant peaks. This has the advantage of being less sensitive to user-defined parameters and also not assuming a shape for the cloud. Figure 1.1 demonstrates the significant differences in structures identified with different algorithms when applied to the same data set.

1.1.2 THE FORMATION OF STARS WITHIN MOLECULAR CLOUDS

Molecular clouds can have sizes between 1 and 100 pc, and are not homogeneous. They have complex internal structure and the increasing resolution of new observational facilities, from *Herschel* to ALMA, is probing substructure within clouds down to scales of 0.01 pc (André et al., 2010; Hacar et al., 2013). The statistical similarity of this substructure over a range of scales is understood to be a product of turbulence, although the exact nature of turbulence and its injection and dissipation scales are still a matter of debate (Bodenheimer, 2011). The sub-structure within molecular clouds can be quantified in terms of a fractal dimension, which describes how clumpy or smooth the cloud is. However, there are many methods of estimating a fractal dimension and their definitions may not be consistent, so comparisons of these values are difficult (we discuss this in detail in Chapter 2).

Herschel observations have shown that filaments are common in molecular clouds, both in low- and high-mass star forming regions, and in regions that are not actively forming stars (Molinari et al., 2010; André et al., 2010; Hennemann et al., 2012; Palmeirim et al., 2013). Further studies have estimated that 60% to 75% of bound prestellar cores lie along filaments whose line mass is greater than the critical value estimated for the onset of fragmentation (André et al., 2010; Könyves et al., 2015; Marsh et al., 2016). Recent work examining the velocity-coherent structures within filamentary regions suggests that these filaments may be made up of smaller elongated structures, termed 'fibers'.

Finding elongated structures such as filaments in images is a common task in computer vision. Several algorithms have been developed specifically for astrophysics, often borrowing techniques from computer vision or from other astronomical disciplines. The most commonly used method, DISPERSE (Sousbie, 2011), was originally designed for identifying filaments of the cosmic web in cosmological simulations. It has been widely used in star formation studies, but has been criticised for being too dependent on user defined parameters. FILFINDER (Koch & Rosolowsky, 2015) uses an adaptive threshold, allowing the recovery of filamentary structures with a range of brightnesses within the same image. The Hessian matrix method (Schisano et al., 2014; Salji et al., 2015) uses local orthogonal gradients to identify pixels on the edge of elongated structures and then aggregates these pixels into continuous filaments.

Molecular clouds are supported against gravitational collapse by thermal pressure, magnetic fields and turbulence. The relative importance of these three forces is still debated and may vary with the larger scale environment. The dense condensations within molecular clouds are referred to as clumps or cores. Cores are often

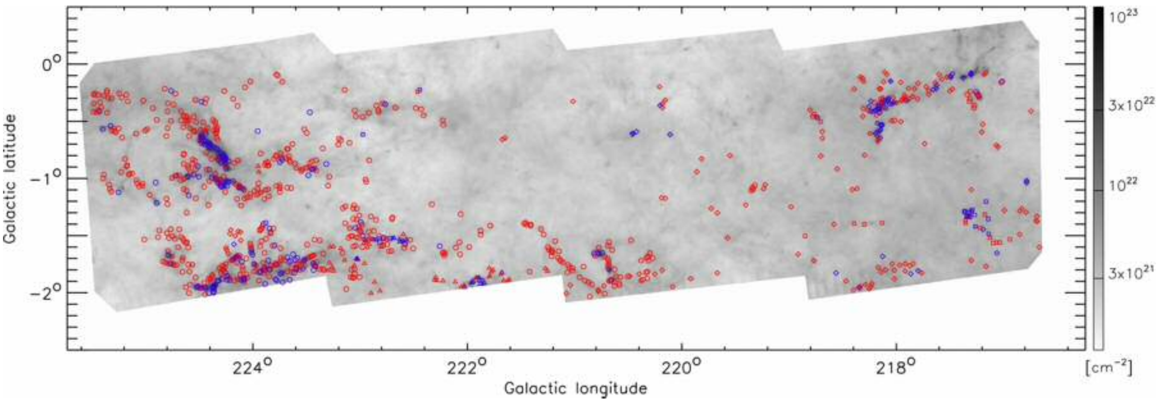

Figure 1.2. The distribution of compact sources across four tiles of the Hi-GAL survey, taken form Elia et al. (2013). Blue symbols represent protostellar sources and red symbols represent starless (prestellar or unbound) sources, overlaid on the column density map. The different symbols (circles, triangles, diamonds and squares) indicate to which of four CO(1-0) components the source has been associated to give an estimate of radial velocity.

separated into protostellar, prestellar, and unbound starless cores by examining their masses and spectral energy distributions (SEDs) (e.g. Elia et al., 2013). Prestellar and unbound cores will both have SEDs well fit by a simple black body, implying that they are composed of cool material. Sources that show an excess of emission at 70 μm are classified as protostellar, as this suggests the clump is internally heated by a deeply embedded protostar. Prestellar and unbound sources are identified by estimating their masses and radii. Elia et al. (2013) use the Bonnor-Ebert mass (M_{BE}) as the critical mass above which cores are gravitationally bound.

$$M_{BE} = 2.4 R_{BE}\ a^2/G, \tag{1.2}$$

where R_{BE} is the Bonner-Ebert radius, estimated as the observed radius of the core at a particular wavelength, a is the sound speed (which depends on the source temperature), and G is the gravitational constant. Objects that have masses greater than this limit are classified as prestellar, and those with lower masses are labelled as unbound starless cores.

Figure 1.2 shows the distribution of compact sources across four tiles of the Hi-GAL survey in the outer Galaxy, taken from Elia et al. (2013). The authors remark that the compact sources are spatially correlated with the dense filamentary structures seen in the column density map, in agreement with the other papers discussed above.

1.1.3 STAR CLUSTERS AND FEEDBACK

Star clusters form from dense cores within Giant Molecular Clouds, and it is often inferred that most stars form in embedded clusters (Lada & Lada, 2003). These clusters come in many sizes and shapes (Megeath et al., 2016). Young stellar clusters are important for the study of star formation, because they present several key observables that can be used to test theories and simulations. Figure 1.3 shows two clusters of Young Stellar Objects (YSOs), separated into protostars and pre-main sequence stars, as well as the dense gas traced by 850 μm emission. We can see that YSOs of different evolutionary stages have different spatial distributions, with protostars more closely associated with the dense gas and pre-main sequence stars showing a more extended distribution.

Large young clusters are presumed to have formed from a single cloud and within a small time frame, so their masses should fully populate the IMF. It is still debated whether this is universally constant or varies with environment, but it is a useful metric against which to test simulations and theories of star formation.

The formation of massive stars is another contentious point - how can a large enough amount of mass be assembled before feedback clears material away? Different theories of massive star formation, for example, monolithic collapse or competitive accretion, should lead to differences in where massive stars are born within clusters. If massive stars are only born in the denser centres of clusters then we should observe that young clusters are mass segregated. Some observed young clusters do show mass segregation but it has recently been suggested that these measures are unreliable, as the blending of sources in very dense regions will result in overestimation of the mass of point sources (see Khorrami et al. (2016) and references therein).

This discussion is further complicated by different definitions of mass segregation. It can be defined as massive stars being closer to each other on average than the lower mass stars, or as the massive stars being in regions of locally higher density (Parker et al., 2014). Unfortunately, clusters can also become dynamically mass segregated as low mass stars are thrown out of the cluster and more massive ones sink towards the centre, so mass segregation at a later stage of cluster evolution may not be able to test the formation scenarios of massive stars. However, if mass segregation is not imprinted on clusters at birth then it can be used as a measure of the age of the cluster. N-body simulations have shown that using the combination of segregation and substructure as a diagnostic of evolution allows us to constrain the initial conditions of older clusters (Wright et al., 2014). For example, studies of Cyg OB2 show that it is substructured but has no mass segregation, which implies that

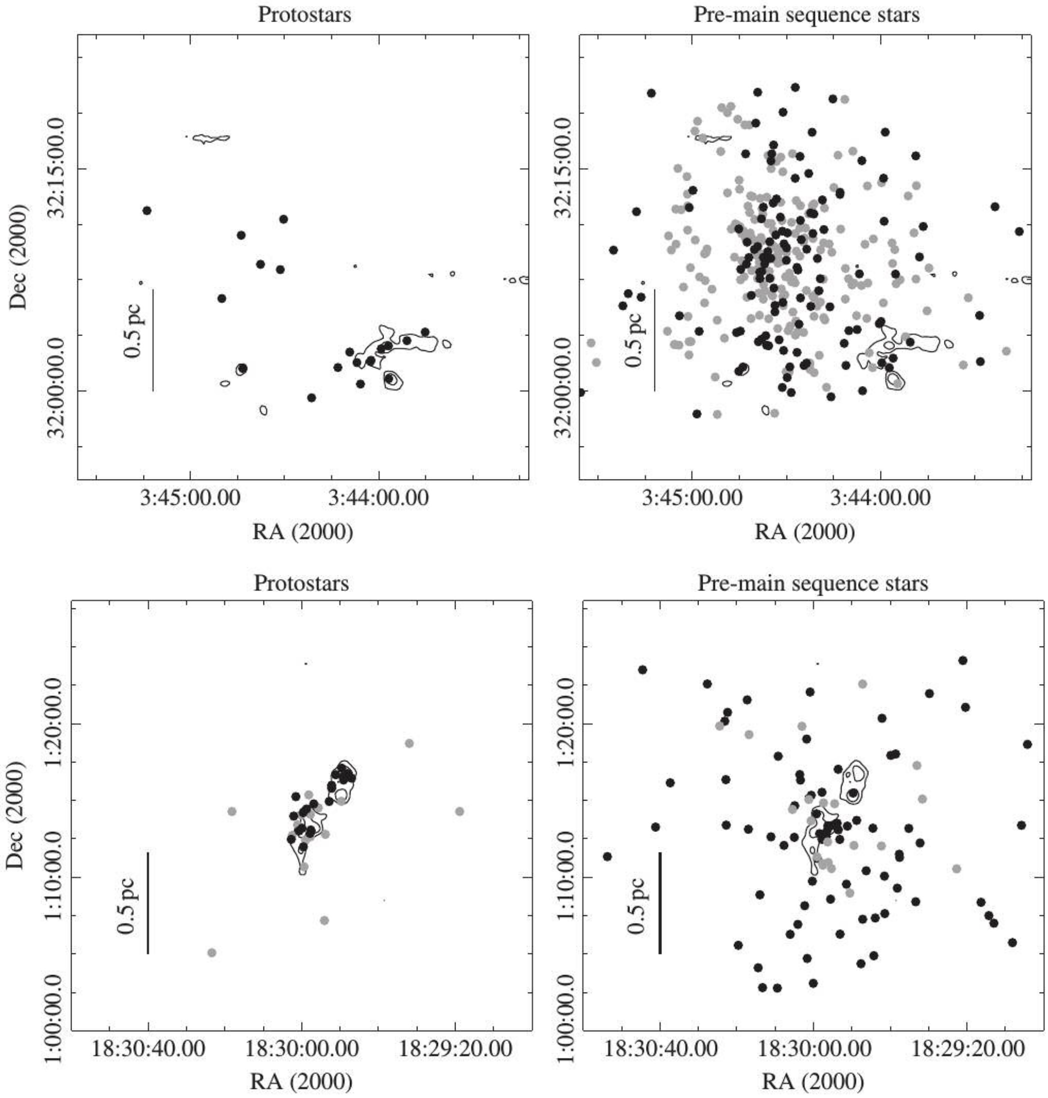

Figure 1.3. Figures taken from Chabrier (2009). The top left panel shows the distribution of protostars in IC 348, and the top right panel shows the pre-main sequence (PMS) stars in IC 348 identified by Muench et al. (2007). The black points are PMS stars with disks, and the grey points are those without disks. The bottom left panel shows the protostars in the Serpens cloud core cluster, classified as Class I/0 and flat spectrum (black and grey markers, respectively). The bottom right panel shows the PMS stars of this cluster, again with and without disks (black and grey markers, respectively). All Serpens data was taken from Winston et al. (2007).The contours show the dense gas detected in an 850 m SCUBA map (IC 348 data from J. DiFrancesco, personal communication. Serpens data from Davis et al. (2000).)

it is dynamically young and was not born as a denser cluster (Wright et al., 2014).

Simulations of star formation often involve a box or sphere of turbulent gas evolving under self-gravity, with some density threshold set for the onset of star formation. For a long time a persistent problem in star formation has been the efficiency of converting molecular clouds into stars. Observations indicate that in most environments this value is very low, on the order of 10% (Lada & Lada, 2003; McKee & Ostriker, 2007). Simulations involving a range of physical effects have shown that this is highly dependent on the local environment (Vázquez-Semadeni, Kim & Ballesteros-Paredes, 2005; Geen, Soler & Hennebelle, 2017; Howard, Pudritz & Harris, 2017)

Another element of the star formation process that could reduce the SFE is feedback. Even at the protostellar stage, new stars give off powerful jets which disrupt the local cloud and drive small scale turbulence. When many stars have formed the cloud will be dispersed by radiation pressure and stellar winds, and the cluster will no longer be embedded and therefore be easier to observe. Massive stars have very strong stellar winds and emit copious energetic radiation. This ionises the material in a large volume producing a HII region which expands and sweeps up a dense shell. Later, supernovae are even more violent and can disrupt entire giant molecular clouds and affect star formation on a galactic scale. We can relate the amount of gas residing within a cluster to its evolutionary stage. For example, Figure 1.3 shows the IC 348 cluster, which has very little dense gas and a centrally concentrated distribution of stars. These factors indicate that it may be more evolved than the Serpens cluster as this still has gas residing at the centre of the cluster.

On the scale of individual clusters, ionising radiation is currently considered to be the most important form of feedback in sculpting or dispersing the molecular cloud, and therefore impacting the rate of star formation. Simulations have shown that the impact of feedback on the clouds varies greatly with the mass and/or density of the cloud (Dale, 2015; Geen, Soler & Hennebelle, 2017).

It is important to note that feedback has both a positive and negative effect on the star formation efficiency. As material is pushed away from the region already forming stars, thereby reducing the local efficiency, it may be pushed into other material, thereby increasing the density and possibly triggering another wave of star formation offset from the first. Rigorously identifying sites of triggered star formation is difficult (e.g. Dale, Haworth & Bressert, 2015) and therefore the net effect of feedback is not well understood.

Stars are born in the densest parts of molecular clouds and therefore young clusters should reflect this structure. The motions and gravitational interactions of newly formed stars will quickly alter their distribution, and so the morphology of the molecular cloud structures from which they formed may be lost in as little as one Myr. The expulsion of molecular gas form a cluster can also affect it's evolution. The removal of a large amount of mass changes the potential energy of the cluster and can result in it becoming unbound. Simulations of young clusters often do not consider the interaction of the stars with the potential of the gas reservoir, or make simplified assumptions about gas removal. Hopefully with the increases in computing power and advanced numerical techniques more simulations will be performed with coupled hydrodynamical and N-body codes (e.g. Sills et al., 2018) to realistically model the expulsion of residual gas from young clusters.

1.2 Thesis outline

In Chapter 2 we review the use of fractal dimensions to quantify the chaotic structures of molecular clouds and young star clusters, and discuss issues relating to comparing different fractal dimensions. In Chapter 3 we introduce a new algorithm designed to quantify the structure of star clusters in terms of three parameters: the fractal dimension, the number of levels in the fractal hierarchy, and the density scaling between levels. In Chapter 4 we apply this algorithm to real observations of Gould Belt clusters, to simulations of clusters formed by cloud-cloud collisions, and to simulations of the evolution of massive clusters. In Chapter 5 we discuss how this algorithm can be applied to molecular cloud data to enable a like-for-like comparison with the clusters they will eventually form, and demonstrate its application to column density maps of the Galactic plane. In Chapter 6 we introduce a second new algorithm which aims to classify and quantify the shapes of structures in greyscale images by calculation of their principal moments of inertia, and we show that this is able to identify centrally concentrated structures, elongated filament-like structures, and hollow ring-like structures. In Chapter 7 we summarise the main conclusions of this thesis and discuss future directions for this work.

The appendices contain detailed mathematical definitions which underpin some areas of this thesis. Appendix A details the construction of Complete Graphs and Minimum Spanning Trees, which are used in Chapter 3. Appendix B gives a formal definition of the Bayesian framework also applied in Chapter 3. Appendix C defines some basic terms in matrix algebra, which are in turn used in Appendix D to describe Principal Component Analysis. This is applied in Chapter 3. Appendix E details the mathematical identification of the principal axes of a shape, which is applied in Chapter 6.

CHAPTER 2

FRACTAL DIMENSIONS IN ASTROPHYSICS

The concept of a fractal dimension is explained and its application to astrophysical images, as a way to quantify substructure, is reviewed. The many methods used to estimate a fractal dimension for an astrophysical object are described, and the problem of comparing values obtained with different methods is discussed.

2.1 INTRODUCTION

Both star clusters and molecular clouds were first defined simply as over-dense regions with respect to some background or detection limit. Early astronomers even referred to the entire Milky Way galaxy as a stellar cluster (Mädler, 1847). With increasing resolution and wavelength range one observes more complex structures, both in the gas and dust of the ISM, and in the compact sources we associate with star formation. Many efforts have been made to quantify the morphology and spatial properties of such structures, and the methods used depend critically on the properties of the data and the questions being asked of it.

Continuum images such as single waveband intensity images or derived column density plots can be analysed statistically to describe some aspect of their structure, for example by using a fractal dimension to describe the degree of substructure. Other statistics such as the velocity power spectrum or spectral correlation function have been used to infer the presence of turbulence and estimate its characteristics. A variety of algorithms have also been designed to find certain shapes in images, such as filaments or compact objects, which can then be analysed themselves.

The distribution of compact sources (be they clumps of condensed gas, prestellar cores, protostellar objects or stars themselves) has been extensively characterised. Clusters can be extracted from the multitude of sources using k-means or Minimum Spanning Trees. Their substructure can also be described by a fractal dimension, and other elements of their distribution quantified by the two point correlation function, angular dispersion parameter or nearest neighbour distance.

Henceforth, we will use the term ‘cluster’ to describe any collection of discrete points representing stars or compact objects, and ‘cloud’ to refer to any greyscale image of molecular clouds or star forming regions. Fractal dimensions can be applied to both data types, and therefore they are particularly useful for comparing how structure changes over the star formation cycle and how substructure in molecular clouds is transferred to, or affected by, the stars that form in these clouds. In Section 2.2 we describe methods of creating artificial fractal clouds and clusters. In Section 2.3 we present some measures of fractal dimension used in astrophysics, and in Section 2.4 we discuss the difficulties in comparing values given by these methods, with reference to their different definitions, the different data types they apply to, and the distinction between measures performed on two dimensional and three dimensional data sets.

As I will be discussing many definitions of fractal dimensions in this Chapter I will use a different subscript to indicate each method. A summary is given in Table 2.1. It is worth noting that this is in no way a complete discussion of fractal dimensions, or even of those applied to astronomical images, as a great number of methods are used for different purposes. Here I aim only to give an overview of the topic and highlight some important issues related to the use of fractal dimensions in astrophysics. I will describe a few of the multitude of fractal dimensions which I deem to be important to this discussion due to their prevalence, originality or mathematical neatness.

2.2 Creating fractal distributions

There are as many ways of creating fractal distributions as there are ways of measuring fractal dimensions. Here I will limit myself to two of the simplest and most widely used methods. (i) Three-dimensional artificial fractal star clusters can be created using the so called 'box fractal' method introduced by Goodwin & Whitworth (2004), which has been the underpinning of the $\mathcal{Q}$ (Cartwright & Whitworth, 2004) and (in an adapted form) $\mathcal{Q}^+$ (Jaffa, Whitworth & Lomax, 2017) methods of analysing clusters. (ii) The fBm method, based initially on simple Brownian motion, creates realistic looking clouds with a range of fractal properties in two and three dimensions.

Since many methods of analysing fractal distributions apply only to clouds (or other continuum data sets) or only to clusters (or other discrete point data sets), I will describe simple mathematical techniques to convert point data into greyscale maps and vice versa. This allows the direct comparison of different fractal parameters and analysis tools.

2.2.1 Box-fractals

The algorithm to construct a synthetic three-dimensional fractal star cluster presented in Goodwin & Whitworth (2004) has two parameters: the fractal dimension, $\mathcal{D}_{BF}$, and the number of stars in the cluster, $\mathcal{N}$. It uses recursive octal partitioning, starting with a root-cube of side $L_0 = 2$ centred on the origin, i.e. $-1 < x, y, z < +1$. The root-cube constitutes level $\ell = 0$, and it is *de facto* 'fertile', meaning it will be partitioned into 8 sub-cubes, which constitute the next level. The likelihood of a sub-cube being fertile is controlled by the fractal dimension, $\mathcal{D}_{BF}$. Cubes that are not fertile are not partitioned into smaller structures and contribute no mass to the cluster, as shown in the two-dimensional example in Figure 2.1. The partitioning continues until the number of sub-sub-..cubes on the last level exceeds the number of stars in the cluster. One star is then randomly placed in each of the fertile cubes on the last level.

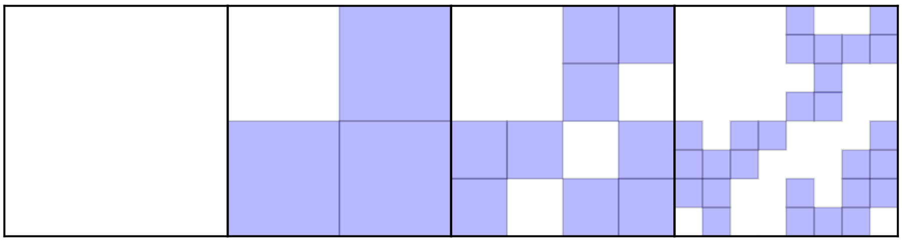

Figure 2.1. A two-dimensional illustration of the algorithm for creating box-fractal clusters. This example has a two-dimensional fractal dimension of 1.58, so that 3 out of 4 sub-cubes are randomly chosen to be fertile at each partitioning.

The probability that a given cube is fertile is given by

$$\mathcal{P}_{\text{FERTILE}} = 2^{(\mathcal{D}_{BF}-3)}\,, \tag{2.1}$$

so reducing $\mathcal{D}_{BF}$ decreases the probability of a cube being fertile. It follows that, since there are 8 sub-cubes in each fertile parent cube, the average number of fertile sub-cubes per parent cube is given by

$$\mathcal{F} = 2^{\mathcal{D}_{BF}} \tag{2.2}$$

The best-behaved results are obtained when $\mathcal{F}$ is an integer, as this results in randomly selecting exactly $\mathcal{F}$ fertile sub-cubes at each subdivision. When $\mathcal{F}$ is not an integer we must randomly select either $\lceil\mathcal{F}\rceil$ or $\lfloor\mathcal{F}\rfloor$ sub-cubes at each subdivision, keeping a cumulative remainder, so that over the whole structure we have selected an average of $\mathcal{F}$ fertile sub-cubes per subdivision. This causes problems because variation in the number of fertile cubes selected at the first subdivision (level 1) can have a large effect on the resultant structure and the measured $\mathcal{D}$. We therefore limit the artificial clusters generated to integer values of $\mathcal{F}$ and assume that interpolation between these values is valid.

2.2.2 Fractional Brownian Motion

Fractional Brownian motion (sometimes called fractal Brownian motion), as presented in detail in Stutzki et al. (1998), can be used to create one, two, or three dimensional distributions characterised by a power spectrum exponent, β. This can be analytically related to a fractal dimension $\mathcal{D}_{fBm}$ (see Section 2.3.3) by

$$\mathcal{D}_{fBm} = (3E + 2 - \beta)/2, \tag{2.3}$$

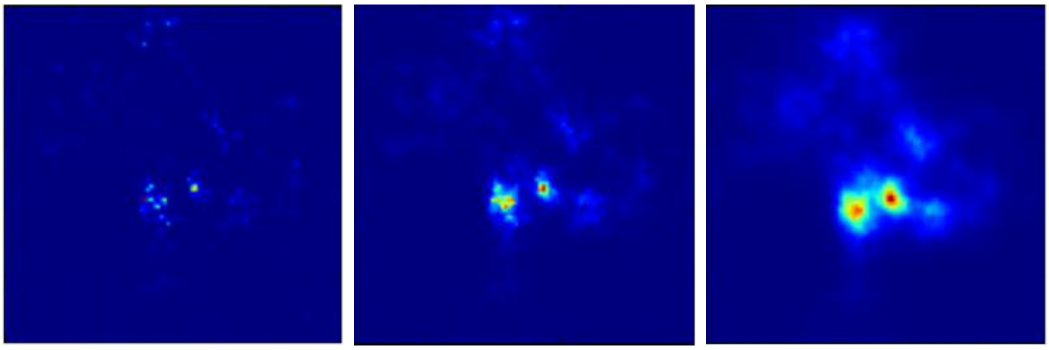

Figure 2.2. Two-dimensional (E=2) fBm images created with $\beta = 2, 3$ and 4, all with 100×100 pixels and $\sigma = 1$. These values correspond to fractal dimensions $\mathcal{D}_{fBm} = 3, 2.5, 2$.

where E is the Euclidean dimension.

Having specified a power law index, we can generate the corresponding amplitudes and a set of random phases. Fourier transformation of this spectrum then creates an image, as shown in the two-dimensional examples in Figure 2.2. Some book-keeping needs to be done to ensure that the values are real and positive, and this is discussed in more detail in Stutzki et al. (1998).

The code used here, written by O. Lomax, exponentiates the values and scales the image to a preset standard deviation, σ. The resulting image is therefore fully described by three user defined parameters: the power law index, β; the scaling factor, σ; and the range of scales or number of pixels in the image.

2.2.3 Clouds to points

Creating a set of points with a statistically similar distribution to a greyscale image is analogous to randomly drawing points from a two-dimensional probability density function (PDF). We have a two dimensional PDF (image), $p(x, y)$, bounded by limits $x_{min}, x_{max}, y_{min}, y_{max}$, and normalised so that

$$\int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} p(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1\,. \tag{2.4}$$

We wish to randomly sample coordinates (X, Y).

First we find X. Here, we calculate the PDF of x by marginalising (integrating out) y,

$$p(x) = \int_{x_{min}}^{x_{max}} p(x, y)\,\mathrm{d}y\,. \tag{2.5}$$

We draw a random number $\mathcal{R}_x$ from the uniform distribution in the interval $[0, 1]$ and solve

$$\int\limits_{x_{min}}^{X} p(x)\,\mathrm{d}x = \mathcal{R}_x \tag{2.6}$$

for X. In practice this is achieved by tabulating

$$P(X) = \int\limits_{x_{min}}^{X} p(x)\,\mathrm{d}x \tag{2.7}$$

and looking up X for $P(X) = \mathcal{R}_x$.

In order to find Y, we calculate the PDF of y, given X,

$$p(y|X) = \frac{p(X, y)}{\int_{y_{min}}^{y_{max}} p(X, y)\,\mathrm{d}y}\,. \tag{2.8}$$

As before, we draw another random number $\mathcal{R}_y$ and solve

$$\int\limits_{y_{min}}^{Y} p(y|X)\,\mathrm{d}y = \mathcal{R}_y \tag{2.9}$$

for Y. Again, we tabulate $P(y|X)$ and lookup $P(Y|X) = \mathcal{R}_y$.

When performing these calculations numerically, $p(x, y)$ is stored in a discrete grid. Integration is performed using the trapezoidal rule and linear interpolation is used to find $p(x, y)$ between grid points.

This method can be extended to three dimensions (and beyond) by marginalising z from $p(x, y, z)$ and finding X and Y using the above method. Z may then be sampled from the distribution $p(z|X, Y)$.

2.2.4 Points to clouds

The inverse of the above process, converting a set of points into a greyscale image, can be accomplished in a variety of ways. After choosing a suitable grid size in two or three dimensions, the most simple method would be to simply increment each cell that contains a point by the 'weight' of that point, where the weight may be one for every point, or may be the mass of a star or some other important property. This produces reasonable results when the pixel size is larger than the separation between points, therefore ensuring that each point is not isolated in a single pixel. However, if the points are sparsely distributed in some areas or the pixel size is small, we need to spread the 'weight' of each point over a patch of pixels to create a smoother distribution.

The most common shape of kernel is a truncated Gaussian. The size of the kernel, σ_{smooth}, can be chosen by eye or for a specific purpose, or various equations and rules-of-thumb can suggest an appropriate size (Bashtannyk & Hyndman, 2001). The most commonly used rules for choosing a suitable Gaussian kernel size are those suggested by Silverman (1986) and Scott (1992). In two dimensions these are equal and depend only on the number of points:

$$\sigma_{smooth} = \mathcal{N}_{pts}^{-1/6} \tag{2.10}$$

2.3 MEASURING FRACTAL DIMENSIONS

2.3.1 PERIMETER-RULER: THE FIRST FRACTAL DIMENSION

The original concept of a 'fractional dimension' is often erroneously attributed to Mandelbrot (1967) in his description of coastlines. He was indeed the first to coin the term (later changed to 'fractal dimension') and a pioneer in bringing the abstract mathematical ideas to bear on many real-world problems, but the characteristic equation of fractals (Equation 2.11) was first published six years earlier by Richardson (1961) as an empirically derived "characteristic of a frontier ... expected to have some positive correlation with one's immediate visual perception of the irregularity of the frontier." The equation given in this first application of a fractal dimension is

$$L(G) \propto G^{(1-\mathcal{D}_{PR})} \tag{2.11}$$

where G is the length of an ungraduated stick used to measure an irregular coastline, $L(G)$ is the length that is measured with that particular stick (and therefore an integer multiple of G), and $\mathcal{D}_{PR}$ is the non-integer dimension inferred from a series of such measurements with sticks of different length. Smaller sticks will be able to follow more of the small deviations of the coastline, and therefore measure a longer total length L. This is increasingly true as the complexity of the coastline is increased, and this is the property that the fractal dimension aims to capture. In this case the value of $\mathcal{D}_{PR}$ is limited to the range $1 < \mathcal{D}_{PR} \leq 2$, because this application is in two dimensions. A lower fractal dimension indicates a more irregular shape and a higher fractal dimension indicates a smoother shape, up to the limit of the Euclidean dimension of the shape. This approach is very simplistic and has not been widely used in astrophysics, but was originally applied to the shape of coastlines (Richardson, 1961).

2.3.2 PERIMETER-AREA

The perimeter-area relation has been used occasionally in an astrophysical context (see Sánchez, Alfaro & Pérez (2005) and references therein), as it is easily applied to molecular cloud structures defined by isocontours. It is defined in Mandelbrot (1983) as

$$\begin{aligned} P(G)^{(1/\mathcal{D}_{PA})} &\propto A(G)^{1/2} \\ P(G) &\propto A(G)^{(\mathcal{D}_{PA}/2)} \end{aligned} \tag{2.12}$$

where $P(G)$ is the perimeter of a two dimensional shape when measured with an ungraduated stick of length G and $A(G)$ is the surface area in units G^2. In practice this method is often applied, not by using different measurement lengths, G, but by assuming a set of objects will conform to a single fractal dimension and therefore measuring all their perimeters with a single G and fitting a straight line to the resulting log-log relation.

This relation can also be extended to three dimensions, as described in Voss (1988):

$$S(G) \propto V(G)^{(\mathcal{D}_{AV}/3)} \tag{2.13}$$

where $S(G)$ is the surface area of a three dimensional shape measured in units G^2 and $V(G)$ is the volume measured in units G^3. However, as pointed out in Elia et al. (2014), this describes only the contour or perimeter of the shape and does not reflect the internal structure of the cloud. The relation between the fractal dimension of the isocontour and that of the image is unclear in many cases.

2.3.3 BOX COUNTING

The box counting dimension, $\mathcal{D}_{BC}$, is popular in many branches of astrophysics, likely due to its simplicity of application and intuitive analysis. Again, it was originally described by Mandelbrot (1983), but the details of the procedure applied may vary between publications.

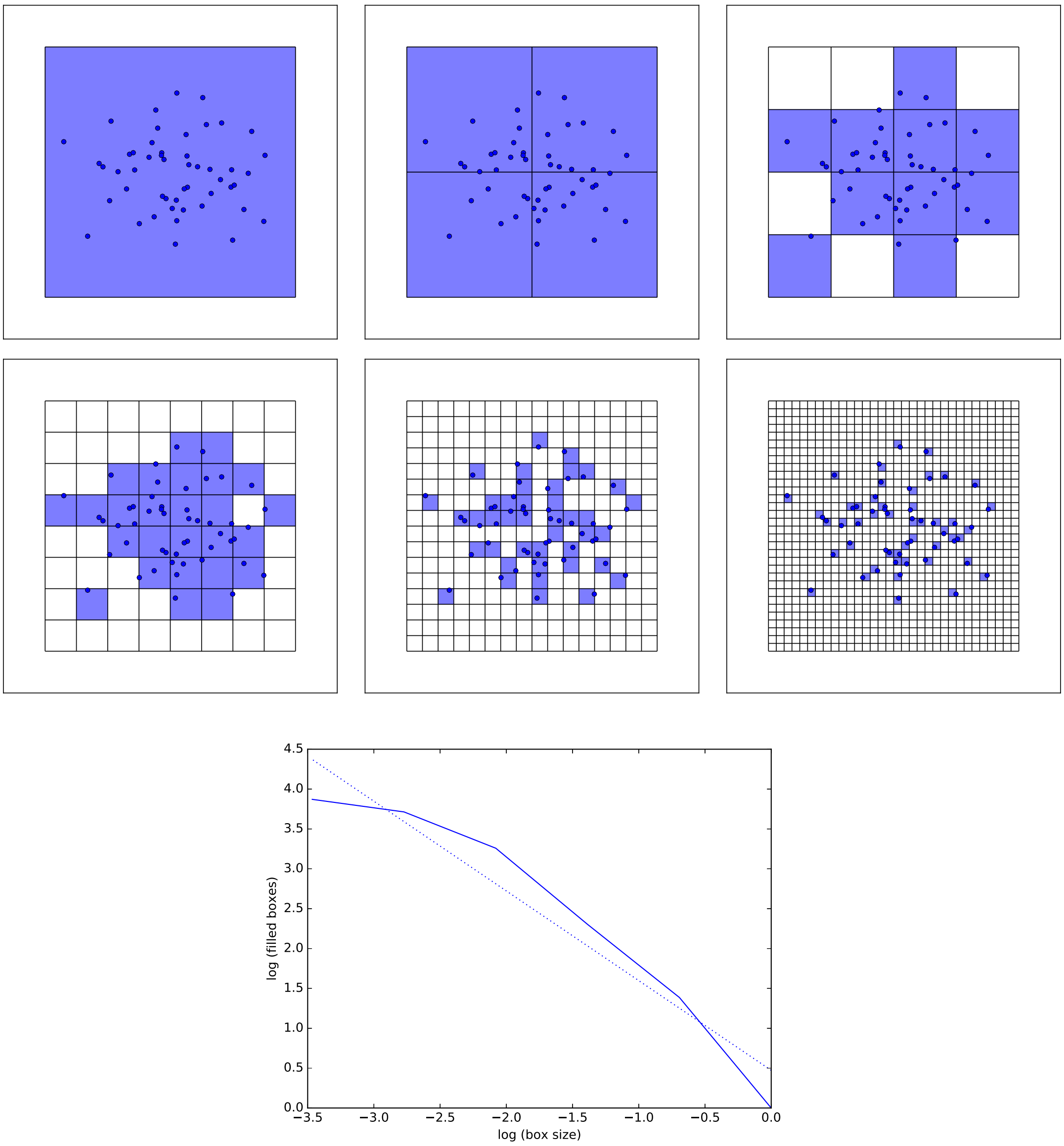

Figure 2.3. Representation of the box-counting method for estimating a fractal dimension. Grids of varying size are overlaid on the distribution of points. The number of occupied cells is counted for each size of grid. In the bottom figure, a log-log plot of the box size versus the number of occupied boxes (solid line) can be fit (dashed line) to yield a gradient which is used to estimate $\mathcal{D}_{BC}$.

The procedure for calculating the box-counting fractal dimension of a set of points, illustrated in Figure 2.3, involves overlaying the points with grids of various box sizes, L. One then finds how many of the boxes are occupied by the points, $O(L)$. A distribution that is smooth will fill more of the space, and therefore more of the boxes, while a very porous and clumpy (highly fractal) distribution of points will not fill as many boxes. If one plots the box size against the number of occupied boxes on a log-log scale the gradient of the line will reflect the nature of the distribution, and is therefore used as a fractal dimension:

$$O(L) \propto L^{-\mathcal{D}_{BC}} \tag{2.14}$$

Some authors suggest that the results may not be as readily comparable as expected due to sensitivity to the exact details of application (Sánchez, Alfaro & Pérez, 2005), such as the initial placement and limits of the bounding box.

2.3.4 Δ-VARIANCE

The Δ-variance was developed by Stutzki et al. (1998), derived from the Allan variance (Allan, 1966) which was originally intended to characterise the properties of noise in atomic clocks. On a two-dimensional field A(x,y) it is defined as:

$$\sigma_{\Delta}^{2}(L) = \frac{1}{2\pi} \langle (A * \odot_L)^2 \rangle_{x,y} \tag{2.15}$$

where $\odot_L$ represents a filter function that is convolved with the map. Stutzki et al. (1998) originally used the so-called *French hat function* (see Stutzki et al. (1998) Appendix B), while later improvement of the Δ-variance method by Ossenkopf, Krips & Stutzki (2008) used the *Mexican hat function*:

$$\odot_L\,(r) = \frac{4}{\pi L^2}\, e^{\frac{r^2}{(L/2)^2}} - \frac{4}{\pi L^2 (v^2-1)} [e^{\frac{r^2}{(vL/2)^2}} - e^{\frac{r^2}{(L/2)^2}}] \tag{2.16}$$

Plotting the logarithm of $\odot_L(r)$ against the logarithm of L will display a linear behaviour over a certain range of scales. Fitting the slope of this linear regime gives a measure of the power spectrum exponent, β, which, by comparison with fractal Brownian motion (see Section 2.2.2), can then be related to a fractal dimension,

$$\mathcal{D}_{\Delta \mathrm{v}} = \frac{3E + 2 - \beta}{2} \tag{2.17}$$

where E is the Euclidean dimension. It is important to note that β can take values between two and four, resulting in $\mathcal{D}_{\Delta \mathrm{v}}$ having a range between 3 and 2 with 2 being smoother and 3 being clumpier, which is in the opposite sense to other measures we have discussed.

2.3.5 $\mathcal{Q}$-parameter

The $\mathcal{Q}$-parameter (Cartwright & Whitworth 2004; hereafter CW04) uses the Complete Graph and Minimum Spanning Tree (described in more detail in Appendix A) to quantify the structure of a star cluster, and can separate sub-structured distributions from those with a radial density gradient. It has frequently been used to characterise the three-dimensional structure of observed and simulated star clusters. It is an empirical measure calibrated against a set of artificial three-dimensional clusters created using the box-fractal method described in Section 2.2.1. These clusters are then projected at a random viewing angle to create a two dimensional image as would be observed on the sky, and the three-dimensional fractal dimension is inferred from the properties of this two-dimensional set.

$\mathcal{Q}$ is evaluated by first constructing the Complete Graph of a two-dimensional set of points (e.g. the projected positions of stars in a cluster) and computing the mean length, $\bar{s}$, of all the edges on the Complete Graph (i.e. all the straight lines connecting each point to all the other points). Next, one constructs the Minimum Spanning Tree (MST, the set of edges which connects all the points together with the shortest possible total length) of the points and computes the normalised mean length, $\bar{m}$, of the edges on the MST. Finally, one computes $\mathcal{Q} = \bar{m}/\bar{s}$. Values of $\mathcal{Q} < 0.8$ can be translated into a notional fractal dimension, $\mathcal{D}_{\mathcal{Q}}$, for star clusters with substructure, and values of $\mathcal{Q} > 0.8$ can be translated into a notional radial density exponent, $\alpha = -\mathrm{d}\ln[n]/\mathrm{d}\ln[r]$ for spherically symmetric star clusters (here n is the mean volume-density of stars at distance r from the centre of the cluster). The data presented in CW04 are displayed in Figure 2.4, showing how the measured value of $\mathcal{Q}$ can be used to estimate either the fractal dimension or radial density exponent.

In CW04, the number of stars in an artificial cluster was held approximately constant at ≈ 300 while the fractal dimension, $\mathcal{D}_{BF}$, was varied. For a lower fractal dimension, this necessitates splitting the root cube more times to produce the required number of sub-sub-..cubes on the last level. The resulting changes in structure were attributed to the change in $\mathcal{D}_{BF}$ only. In CW04, one star was placed in every fertile cube on the final level. This constrains the artificial clusters to having a bimodal density distribution, varying only slightly with the random placement of stars within their cubes.

An improved method for creating box-fractal clusters which takes account of

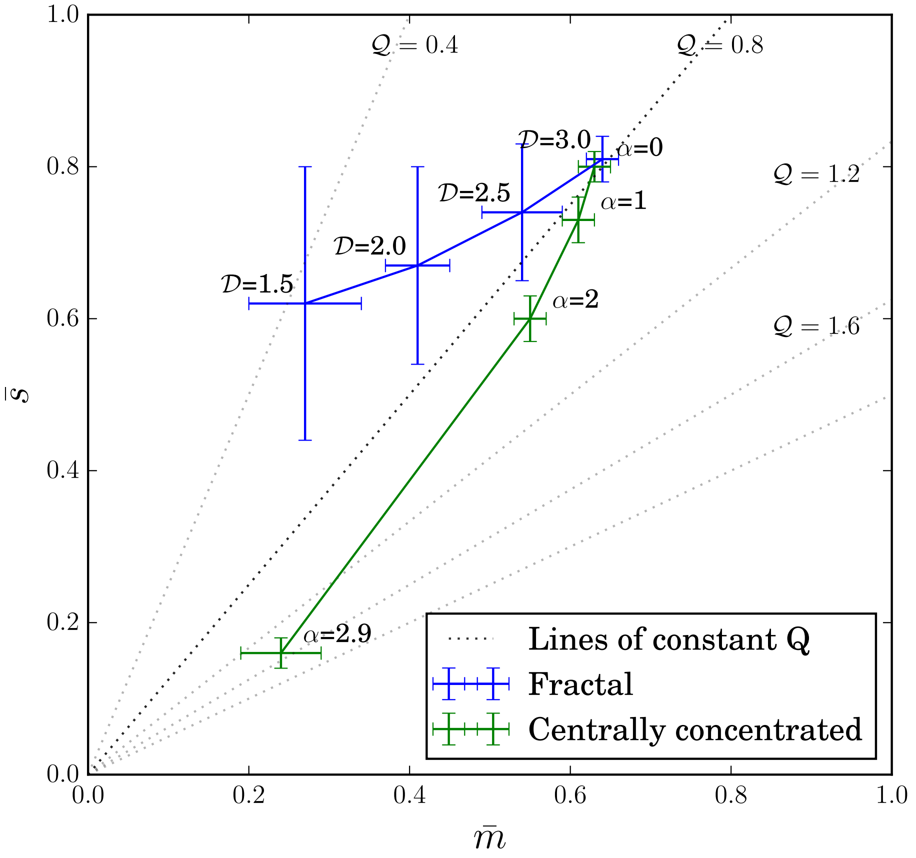

Figure 2.4. Results presented in CW04, showing how the measured value of $\mathcal{Q}$ can be used to estimate either the fractal dimension or radial density exponent. The blue points are artificial fractal clusters created using the 'box-fractal' method of Goodwin & Whitworth (2004), green points are centrally concentrated clusters created using a simple algorithm described in CW04, and the dashed lines are lines of constant $\mathcal{Q}$.

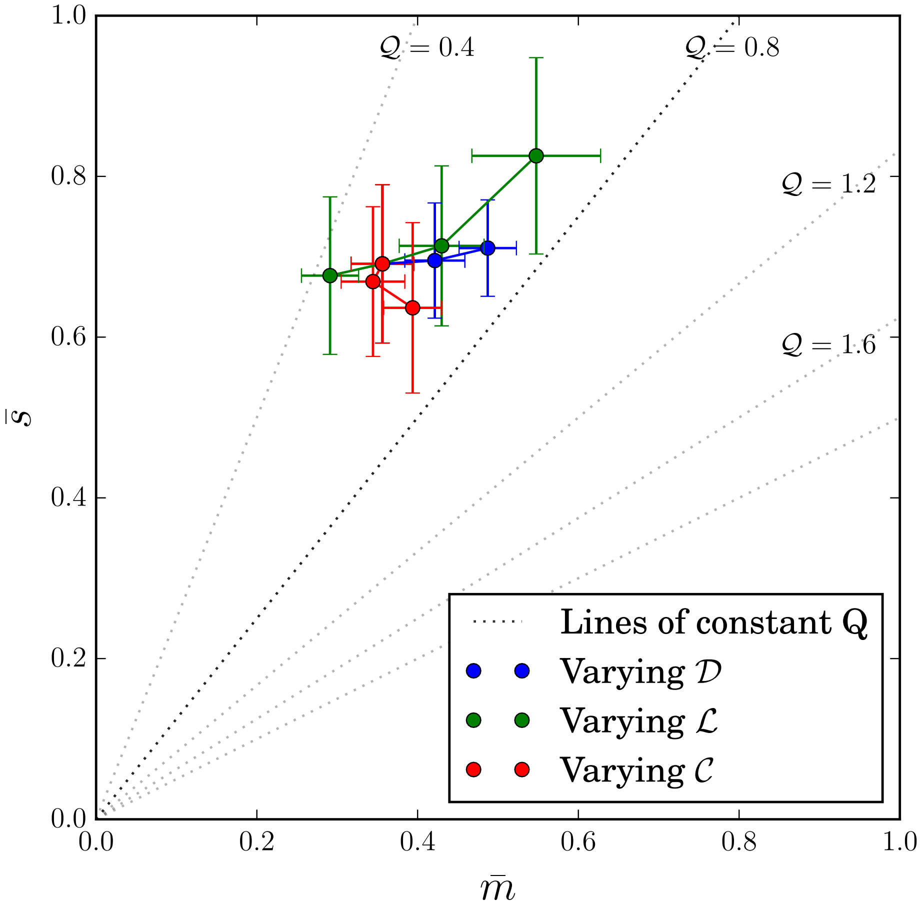

Figure 2.5. The mean edge length of the MST and complete graph for artificial clusters using the improved box-fractal method. The $\mathcal{Q}$-parameter does not distinguish between changes in each of the three parameters that describe the structure. The parameters $\mathcal{D}, \mathcal{L}, \mathcal{C}$ are defined explicitly in Section 3.2

changes in both the number of levels in the hierarchy and the distribution of mass on different levels will be described in Section 3.2. When all three parameters are controlled, we find that the $\mathcal{Q}$-parameter fails to distinguish between the changes in fractal structure (see Figure 2.5).

2.4 Comparing fractal dimensions in two dimensions

Name	Designation	Reference	Data types	Dimensions	Smooth	Irregular
Box-fractal	$\mathcal{D}_{BF}$	Goodwin & Whitworth (2004)	Points	2,3	3	1
fBm	$\mathcal{D}_{fBm}$	Stutzki et al. (1998)	Greyscale	1,2,3	2	3
Perimeter-ruler	$\mathcal{D}_{PR}$	Richardson (1961)	Isocontour	2	2	1
Perimeter-area	$\mathcal{D}_{PA}$	Mandelbrot (1983)	Isocontour	2	2	1
Area-volume	$\mathcal{D}_{AV}$	Mandelbrot (1983)	Isocontour	3	3	1
Box counting	$\mathcal{D}_{BC}$	Mandelbrot (1983)	Points	1,2,3	3	1
Δ-variance	$\mathcal{D}_{\Delta\mathrm{v}}$	Stutzki et al. (1998)	Greyscale	2	2	3
$\mathcal{Q}$ parameter	$\mathcal{D}_{\mathcal{Q}}$	Cartwright & Whitworth (2004)	Points	2	3	0.58

Table 2.1. Details of the fractal dimensions discussed here, including the number of dimensions in which it is applied, the type of shape (data type) it measures, and the limiting values of smooth and irregular shapes.

Many studies of fractal dimensions of star clusters have found values in the range $1 \leq \mathcal{D} \leq 2$(e.g. Sánchez, Alfaro & Pérez, 2005, and references therein). We therefore explore artificial clouds and star clusters with fractal dimensions in this range. We create a set of three-dimensional box-fractal star clusters with known $\mathcal{D}_{BF}$ and all other parameters kept constant, project them at a random angle to create a two-dimensional cluster, and then measure the box counting fractal dimension.

Intuitively, the measured box counting dimension should be equivalent to the input box-fractal dimension, as these both use repeated partitioning of space. However, Figure 2.6 shows that the box counting method gives consistently lower results than the input box-fractal dimension. Furthermore, when the number of levels in the hierarchy is varied, the offset changes even though the fractal dimension remains constant. The box counting measure is more accurate for clusters with a higher number of levels.

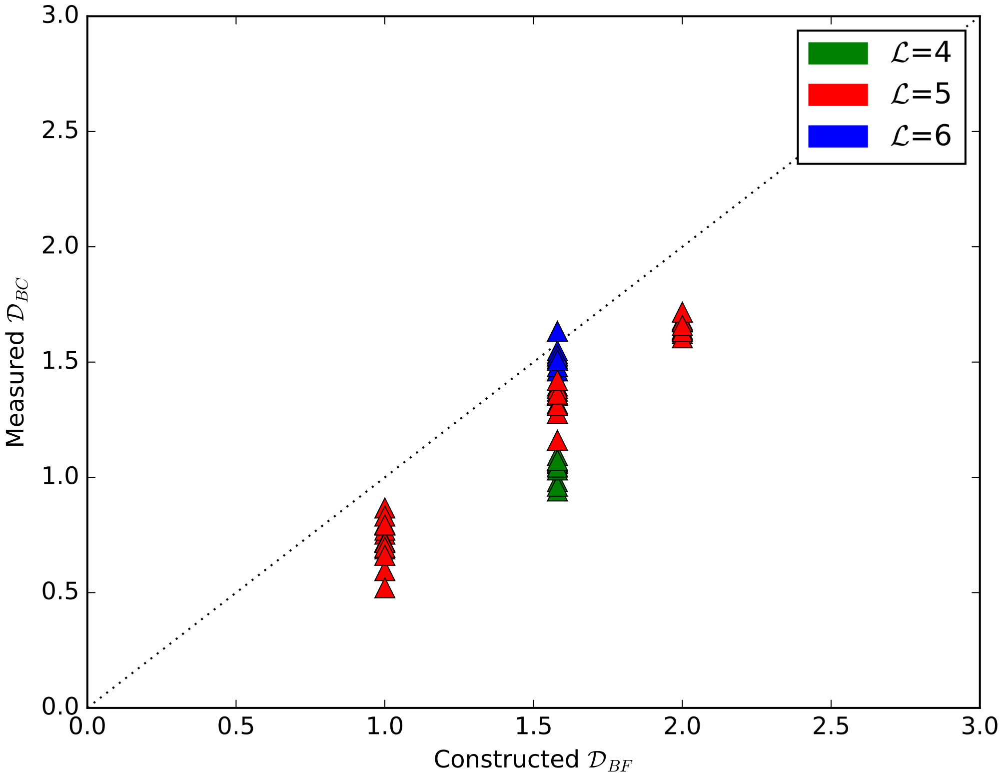

Figure 2.6. The box counting fractal dimension measured on a set of box-fractal star clusters. The dotted line indicates equality.

The measured Δ-variance dimension should accurately measure the fractal

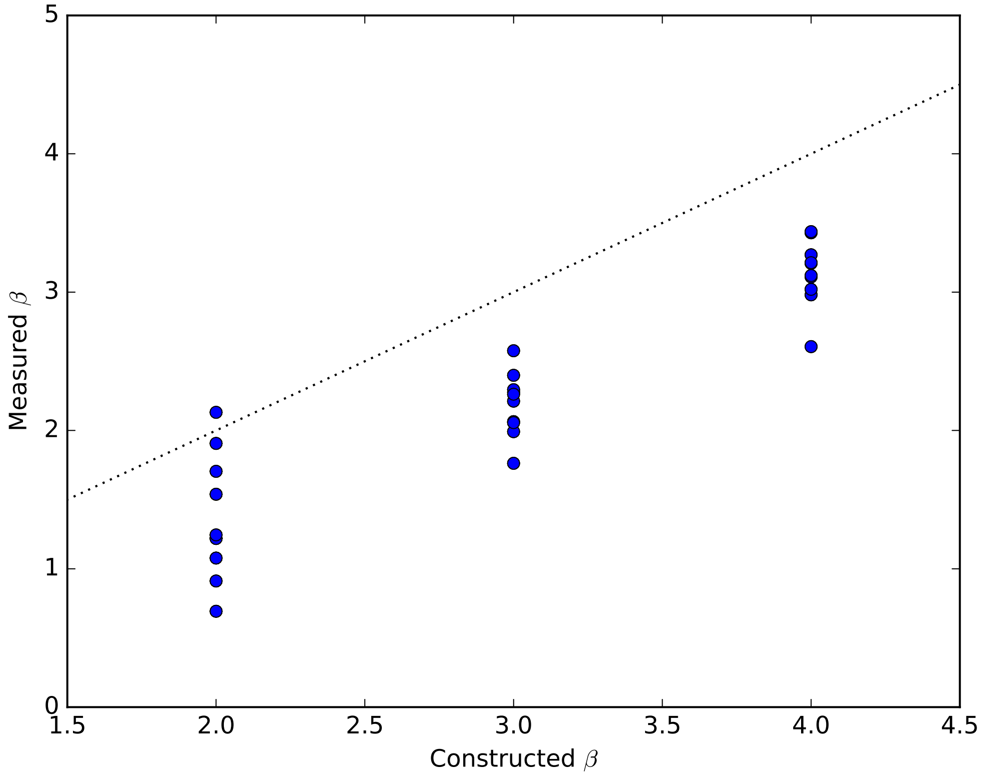

Figure 2.7. The value of β measured using the Δ-variance on a set of fBm fractal clouds. The dotted line indicates equality.

Brownian motion dimension*. Both of these fractal dimensions are more commonly quantified in terms of the power spectrum exponent β which is related to the fractal dimension through Equation 2.17. However, we find that this measure also consistently underestimates the value of β. Also, the range of measured values is very large when β is small, indicating that the smoother structure is not well constrained.

We can also convert our distributions of points to greyscale images and vice versa. Firstly, we can convert fractal structures created as greyscale images into a set of points. Figure 2.8 shows the box counting fractal dimension measured for a set of fBm fractal structures that have been converted into a distribution of points. We do not expect a direct correlation in these two measures as, according to their mathematical definition, they have different ranges and vary in the opposite direction in terms of clumpy or smooth structure. However, our results do not show an inverse correlation either. These two descriptions of fractal structure seem to describe a fundamentally different aspect of the structure as the measured $\mathcal{D}_{BC}$ shows no trend

---

[*] To calculate the Δ-variance we use the IDL script DELTAVARWIDGET described in Ossenkopf, Krips & Stutzki (2008)

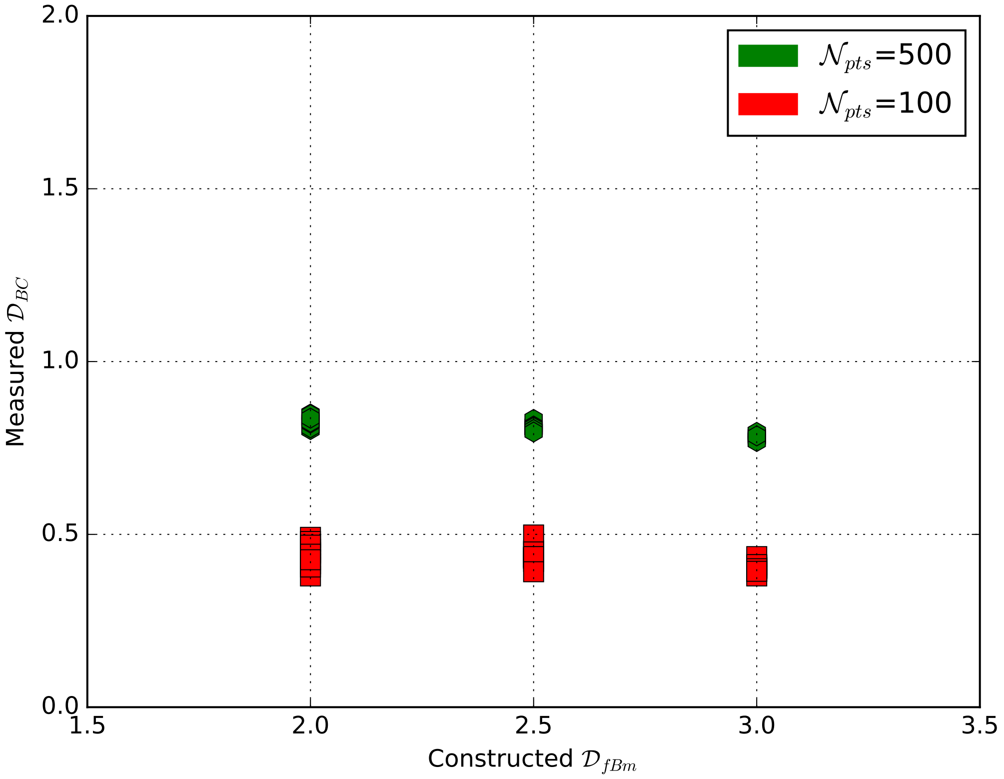

Figure 2.8. The box counting fractal dimension measured on a set of fBm fractal structures which have been converted into a distribution of points.

with $\mathcal{D}_{fBm}$. The number of points in the cluster does have an effect on this measure but does not change the lack of correlation with $\mathcal{D}_{fBm}$.

Secondly, we can convert the box-fractal star clusters to greyscale images by Gaussian smoothing. Measuring the Δ-variance of these images returns an estimate of the power law slope β which can be related to a fractal dimension. Figure 2.9 shows that this also does not give the expected inverse correlation. We would expect that box-fractal clusters with a lower $\mathcal{D}_{BF}$ would produce a higher $\mathcal{D}_{\Delta\mathrm{v}}$ i.e. a negative correlation, but the results show a small positive correlation. The box-fractal clusters used here have varying numbers of points. They therefore have different sizes of smoothing kernel, σ_{smooth}, because this is calculated from the number of points. As was pointed out by Turlach (1993), the size of the kernel used can have a large influence on the resulting structure, even more than the shape of the kernel itself. The estimation based on the number of points may be over-smoothing in dense areas of the more substructured distributions.

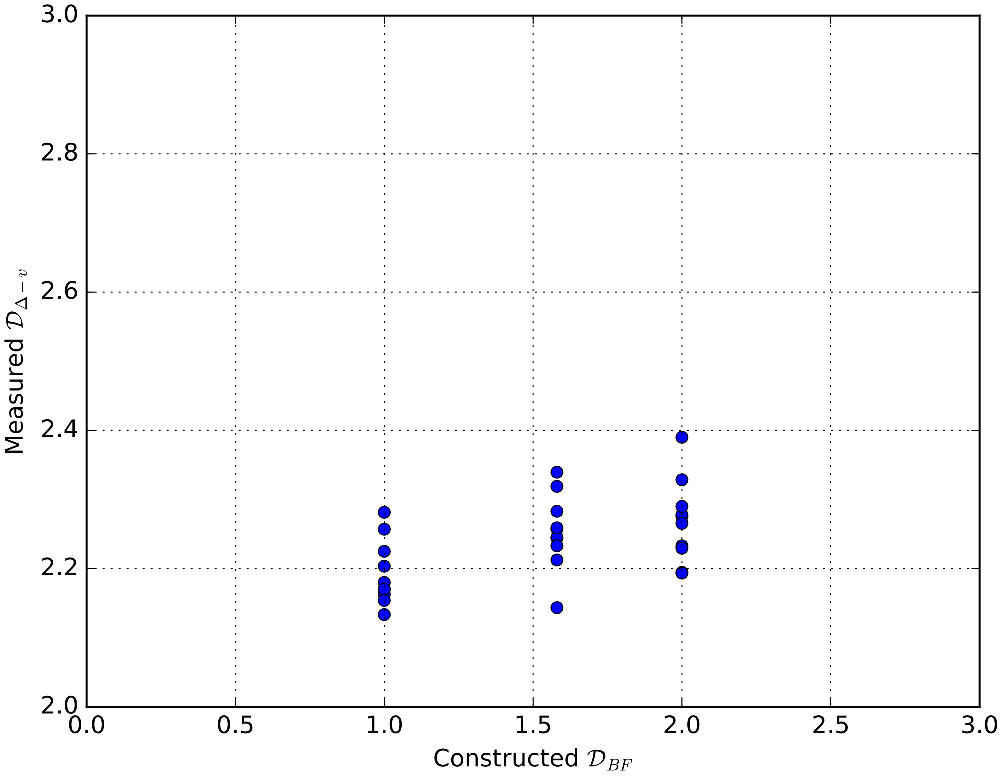

Figure 2.9. The Δ-variance fractal dimension measured on a set of box fractal structures which have been converted into greyscale images.

2.5 Two- and three-dimensional fractals

As summarised in Table 2.1, some of the fractal dimensions described above are designed to be applied to two dimensional data sets, and some to three dimensional data sets. The $\mathcal{Q}$-parameter is alone among the measures discussed here in attempting to directly constrain the three-dimensional fractal dimension from the two-dimensional image, which is a fundamental problem faced in observational astronomy[†]. Many authors have discussed this issue, from both a mathematical and empirical standpoint.

One of the oldest and most commonly used relations is that described by Mandelbrot (1983) in relation to the surface area-volume and perimeter-area fractal dimensions:

$$\mathcal{D}_{PA} = \mathcal{D}_{AV} - 1 \tag{2.18}$$

However, this describes a slice through a three-dimensional object rather than a projection onto a plane. Falconer (1990) derived the following relation for the projection of Borel sets in any number of dimensions:

$$D^{k-1} = min\{2, D^k\} \tag{2.19}$$

Beech (1992) tested the relations derived in Falconer (1990) with a simple experiment involving the shadows of scrunched up paper balls: measuring the three-dimensional fractal dimension of the paper balls by the mass-size relation, holding them in front of a lamp to trace their shadow and calculating the two-dimensional fractal dimension of the shadow using the perimeter-area relation.

Fractal Brownian motion images (see Section 2.2.2) are mathematically defined such that projection into two dimensions does not change the power law slope, resulting in a change in the fractal dimension of

$$D_{fBm,2} = D_{fBm,3} - 3/2 \tag{2.20}$$

Gouliermis, Hony & Klessen (2014) empirically measured the fractal dimension of box-fractal clusters in three dimensions and in projection into two dimensions, and found the relation:

[†] Simulated molecular clouds or star clusters may be viewed in three dimensions or projected into two dimensions to make them directly comparable to observations.

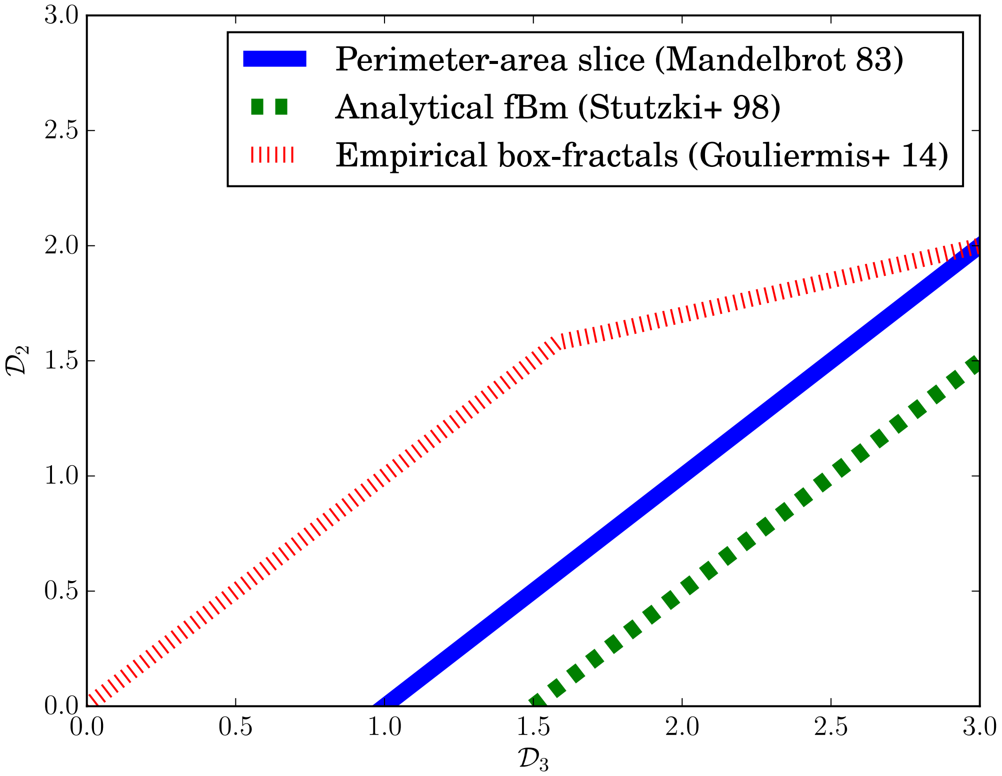

Figure 2.10. Various relations in the literature for converting two-dimensional fractal dimensions to the three-dimensional fractal dimensions.

$$\mathcal{D}_{BF,2} = \begin{cases} \mathcal{D}_{BF,3} & \mathcal{D}_{BF,3} \leq 1.6 \\ 1.1 + 0.3\mathcal{D}_{BF,3} & \mathcal{D}_{BF,3} \geq 1.6 \end{cases} \tag{2.21}$$

Figure 2.10 compares these three relations to illustrate the large differences. It is also important to note that for the perimeter-area and box-fractal relations, a lower number indicates a more clumpy distribution and a higher number indicates a smoother distribution, while the fBm fractal dimension varies in the opposite direction.

2.6 CONCLUSIONS

In his seminal 1967 work on fractal geometry, Benoit Mandelbrot said, *"The concept of 'dimension' is elusive and very complex, ...Different definitions frequently* *yield different results, and the field abounds in paradoxes."* Here we have demonstrated that even the simplest definitions of a fractal dimension may be incompatible, and the notion that all the statistically important structure in a cloud or cluster can be constrained by a single number is an oversimplification. However, each of the tested methods of quantifying fractal structure, when applied correctly to a suitable data set, does yield sensible results and enables a quantitative and statistical discussion of some aspects of the internal structure of clouds or clusters.

The box counting method is able to constrain the input fractal dimension of box-fractal star clusters, and the Δ-variance method can likewise constrain the input dimension of fBm fractal clouds. However, converting data sets from points to clouds or vice versa indicates that these two dimensions are not compatible and do not accurately measure the input dimension of data sets constructed using the other method, although the effects of the conversion are hard to measure. When converting a greyscale image into a set of points, the number of points is a free parameter that may have a large effect on the resultant fractal structure, and when converting a set of points into a greyscale image the size of the smoothing kernel is similarly important.

Real astrophysical structures are three-dimensional and appear projected on the plane of the sky. We can create artificial data sets in three dimensions, either analytically or through simulations and project them at a random angle to test how the measurable structure changes, but we can only make statistical arguments about the real three dimensional structures if we can find a well understood relation. The various equations in the literature produce disparate results and may apply in different situations.

We conclude that, while the measurement of a fractal dimension is undoubtedly a valuable tool for quantifying the chaotic structures inherent in the star formation process, the comparison of measurements from different methods is fraught with issues which have yet to be fully explored. This makes the case for clear communication of the exact analysis method used in any published work, and the increase in open source software being freely shared with the scientific community is strongly encouraged in this case.

Chapter 3

$\mathcal{Q}^{+}$: AN ALGORITHM FOR QUANTIFYING STRUCTURE IN STAR CLUSTERS [*]

Fractal distributions need three parameters to fully describe their structure. In the model that we use, these are: the fractal dimension, $\mathcal{D}$*, the number of levels in the fractal hierarchy,* $\mathcal{L}$*, and the density scaling exponent,* $\mathcal{C}$*. These parameters are used to construct three-dimensional artificial star clusters, which are then projected onto an arbitrary plane so that they may be compared to two-dimensional observed star clusters. A set of statistical measures derived from the Complete Graph and Minimum Spanning Tree, when combined using Principal Component Analysis, are shown to distinguish between the changes in structure caused by varying* $\mathcal{D}$ *and* $\mathcal{L}$*.* $\mathcal{C}$ *has a more subtle effect and so this is considered separately using Bayesian inference. A robust algorithm for classifying the structure of two-dimensional star clusters is tested using artificial clusters is shown to reliably constrain the input parameters.*

---

[*] The work presented here is published in Jaffa, Whitworth & Lomax, 2017, MNRAS, 446, 1082

3.1 Introduction

Many previous authors have utilised the theory of fractal geometry to quantify the structure of star clusters by defining a fractal dimension which describes the clumpiness or substructure of the distribution as against a smooth distribution, as discussed in Chapter 2. However, even if star clusters are fractally sub-structured, the fractal dimension, $\mathcal{D}$, is not enough to fully describe the structure.

Firstly, this simple analysis assumes that fractals have an infinite range of scales over which the self-similarity persists. This is not the case in the study of star clusters as the range is limited - at its maximum by the size of the parent cloud or cloud complex, and at its minimum by the spatial scale of individual stars or stellar systems (see Section 3.4.1 for a discussion of binary and multiple systems in relation to fractal clustering). Therefore we must specify the range of scales over which the structure can be self-similar, i.e. the ratio between the overall linear size of the cluster and the size of the smallest separation between stellar systems. We quantify this by the number of levels in the fractal hierarchy, $\mathcal{L}$.

Secondly, much of the theory of fractal structures is built on so-called 'binary fractal' structures, in which space can only be occupied or unoccupied. While this can be analogous to star clusters in terms of individual stars or systems being treated as discrete objects, so a volume of space may have a single star in it or not, this is only useful at scales comparable to the average separation between stars. The simplest definition of a star cluster is a local overdensity of stars compared to the background of field stars. Therefore we are often interested in the local density of stars, which means we need to describe the distribution of density in the fractal structure. We quantify this with a volume-density scaling exponent, $\mathcal{C}$, which dictates how the mass of the structure is distributed on the different scales of the hierarchy.

It follows that more sophisticated tools are required to constrain these extra parameters, and we present such a tool here.

This chapter is organised as follows. In Section 3.2, we present a procedure for generating synthetic fractal star clusters characterised by three parameters, $\mathcal{D}$, $\mathcal{L}$ and $\mathcal{C}$, and illustrate how their properties are influenced by varying these parameters. In Section 3.3, we explain how Complete Graphs and Minimum Spanning Trees are constructed, and define the discriminating measures that can be derived from them. In Section 3.4, we demonstrate how these measures can be combined to express the maximum variation with $\mathcal{D}$ and $\mathcal{L}$ by using Principal Component Analysis, and how $\mathcal{C}$ can be found once these have been estimated. In Section 3.5, we apply the algorithm to synthetic data to evaluate its reliability. In Section 3.6, we summarise our main conclusions.

For mathematical convenience, we define a 3-component statistical state vector for a fractal star cluster,
$\mathbf{Y} \equiv (\mathcal{D}, \mathcal{L}, \mathcal{C})$.

3.2 CONSTRUCTING SYNTHETIC THREE-DIMENSIONAL FRACTAL STAR CLUSTERS

The algorithm to construct a synthetic three-dimensional fractal star cluster with three statistical parameters is adapted from Goodwin & Whitworth (2004). It uses recursive octal partitioning, starting with a root-cube of side $L_0 = 2$ centred on the origin, i.e. $-1 < x, y, z < +1$. The root-cube constitutes level $\ell = 0$, and it is *de facto* 'fertile', meaning it will be partitioned into 8 sub-cubes, which constitute the next level. Cubes that are not fertile are not partitioned into smaller structures and contribute no mass to the cluster. The partitioning continues until the required number of levels, $\mathcal{L}$, has been reached. The parameter $\mathcal{L}$ therefore controls the range of scales in the cluster. The likelihood of a sub-cube being fertile is controlled by the fractal dimension, $\mathcal{D}$, and the density of stars in the fertile cubes on different levels is controlled by the density scaling exponent, $\mathcal{C}$.

In this section, we illustrate three-dimensional clusters generated using the algorithm summarised in Section 3.2.5 and projected onto a randomly oriented plane, in order to demonstrate the effect of varying the underlying parameters, $\mathcal{D}$, $\mathcal{L}$ and $\mathcal{C}$, on the structure that would be seen in observations in the plane of the sky. For reference we define a fiducial cluster with $\mathcal{D} = 1.58$, $\mathcal{L} = 5$, and $\mathcal{C} = 3$. Figures 3.1, 3.2 and 3.3 show representative randomly generated clusters that have not been rotated. Viewing along the Cartesian axes we can more clearly identify the structural influence of each parameter.

3.2.1 THE FRACTAL DIMENSION, $\mathcal{D}$

The fractal dimension, $\mathcal{D}$, controls what proportion of sub-cubes are fertile. This is the same on every level, as the structure is self-similar. The probability that a given cube is fertile is given by

$$\mathcal{P}_{\text{FERTILE}} = 2^{(\mathcal{D}-3)}\,, \tag{3.1}$$

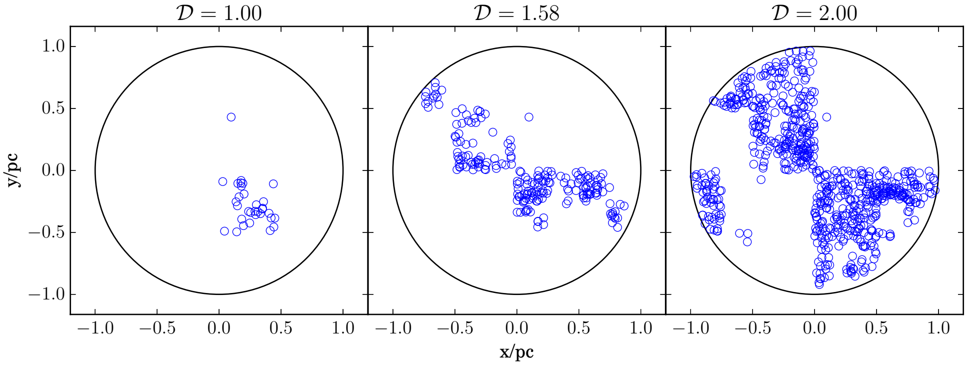

Figure 3.1. 2D projections of 3D fractal star clusters, all with the same $\mathcal{L} = 5$ and $\mathcal{C} = 3$, but different $\mathcal{D}$. (a) $\mathcal{D} = 1.00$; (b) $\mathcal{D} = 1.58$ (the fiducial case); (c) $\mathcal{D} = 2.00$. Increasing $\mathcal{D}$ reduces the amount of empty space in the cluster and increases the number of stars.

so reducing $\mathcal{D}$ decreases the probability of a cube being fertile. It follows that, since there are 8 sub-cubes in each fertile parent cube, the average number of fertile sub-cubes per parent cube is given by

$$\mathcal{F} = 2^{\mathcal{D}} \tag{3.2}$$

The best-behaved results are obtained when $\mathcal{F}$ is an integer, as this results in randomly selecting exactly $\mathcal{F}$ fertile subcubes at each subdivision. When $\mathcal{F}$ is not an integer we must randomly select either $\lceil \mathcal{F} \rceil$ or $\lfloor \mathcal{F} \rfloor$ subcubes at each subdivision, keeping a cumulative remainder, so that over the whole structure we have selected an average of $\mathcal{F}$ fertile subcubes per subdivision. This causes problems because variation in the number of fertile cubes selected at the first subdivision (level 1) can have a large effect on the resultant structure and the measured $\mathcal{D}$. We therefore limit the artificial clusters generated to integer values of $\mathcal{F}$ and assume that interpolation between these values is valid.

The fractal dimension, $\mathcal{D}$, is the only parameter explicitly investigated in CW04, but the effect it has on the $\mathcal{Q}$-parameter is similar to that of varying $\mathcal{L}$, as shown in Figure 2.5.

Figure 3.1 shows clusters with three different values of $\mathcal{D}$, but the same $\mathcal{L} = 5$ and $\mathcal{C} = 3$. The left-hand image shows a cluster with $\mathcal{D} = 1.00$; in this case, the fractal dimension is low, and each parent cube has only 2 fertile child-cubes, so the cluster is very sparse. The middle image shows a cluster with $\mathcal{D} = 1.58$; this is the fiducial cluster with a middling fractal dimension, and each parent-cube has 3 fertile child-cubes, so the cluster is more uniformly populated. The right-hand image shows a cluster with $\mathcal{D} = 2.00$; this is a higher fractal dimension, and each parent-cube

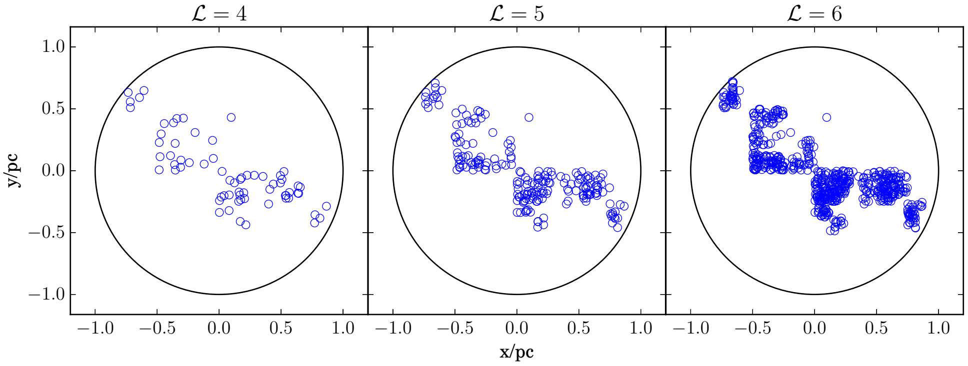

Figure 3.2. 2D projections of 3D fractal star clusters, all with the same $\mathcal{D} = 1.58$ and $\mathcal{C} = 3$, but different $\mathcal{L}$. (a) $\mathcal{L} = 4$; (b) $\mathcal{L} = 5$ (the fiducial case); (c) $\mathcal{L} = 6$. Increasing $\mathcal{L}$ decreases the size of the smallest separations compared to the overall size of the cluster and increases the number of stars.

has 4 fertile child-cubes and 4 sterile child-cubes, so the cluster is populated more uniformly. Thus the effect of increasing $\mathcal{D}$ is to increase the volume-filling factor on every level and increase the number of stars.

3.2.2 THE NUMBER OF LEVELS, $\mathcal{L}$

The root-cube is labelled as level 0, and at each splitting the fertile parent cubes are split into 8 children. The recursive division is terminated at level $\mathcal{L}$, as soon as we have created a level of sub-sub-...-sub-cubes which are smaller than the root-cube by a factor R, so

$$\mathcal{L} = \log_2(R) \tag{3.3}$$

The root-cube has side 2, and the cubes on level ℓ have side $2^{1-\ell}$, so the range of sizes is $2^{\mathcal{L}} = R$.

In CW04, the number of stars in an artificial cluster was held approximately constant at ≈ 300. For a lower fractal dimension, this necessitates populating more levels and therefore $\mathcal{L}$ was varied at the same time as $\mathcal{D}$. The resulting changes in structure were attributed to the change in $\mathcal{D}$ only. In this work, we accept that a set number of stars can only populate a certain fractal dimension to a certain number of levels, and therefore as these parameters vary the number of stars in the artificial cluster will also vary.

Figure 3.2 shows clusters with three different values of $\mathcal{L}$, but the same $\mathcal{D} = 1.58$ and $\mathcal{C} = 3$. The left hand image shows a cluster with $\mathcal{L} = 3$. The middle image again shows the fiducial cluster with $\mathcal{L} = 5$. The right hand image shows a cluster with $\mathcal{L} = 7$. The broad structures seen in the higher-$\mathcal{L}$ cases are visible, but less

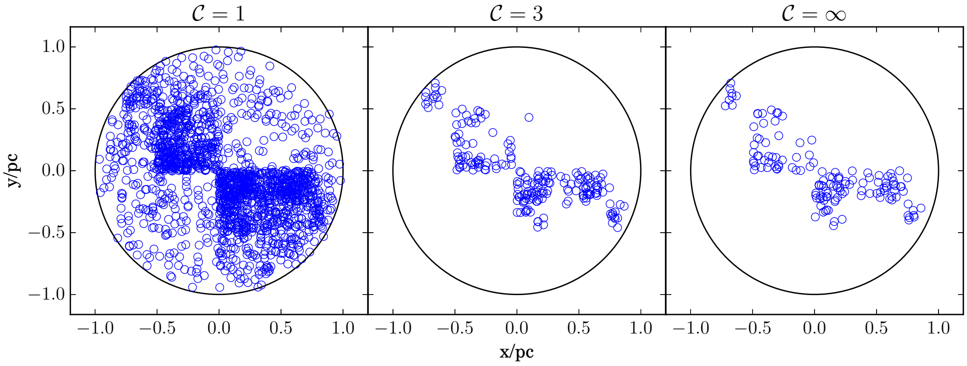

Figure 3.3. 2D projections of 3D fractal star clusters, all with the same $\mathcal{D} = 1.58$ and $\mathcal{L} = 5$, but different $\mathcal{C}$. (a) $\mathcal{C} = 1$; (b) $\mathcal{C} = 3$; (c) $\mathcal{C} = \infty$. Increasing $\mathcal{C}$ concentrates the star more on the lower levels (smaller structures) of the hierarchy and decreases the number of stars.

clearly defined, in the lower-$\mathcal{L}$ case.

3.2.3 THE VOLUME-DENSITY SCALING EXPONENT, $\mathcal{C}$

Once the overall structure of fertile cubes has been specified by $\mathcal{D}$ and $\mathcal{L}$, we then have to decide how the mass of the star cluster will be allocated to this structure. In CW04, one star was placed in every fertile cube on the final level. This constrains the artificial clusters to having only a single volume density, varying only slightly with the random placement of stars within their cubes. In this work we populate all levels of the hierarchy, assuming that the smaller structures will be of equal or higher density than the larger structures they are nested inside. We therefore define the volume-density scaling exponent, $\mathcal{C}$, which specifies by how much the extra density in a fertile child-cube will exceed the density of its parent.

We define the volume-density initially assigned to the root-cube as n_0. The additional volume-density assigned to the fertile cubes on level $\ell = 1$ is then $\delta n_1 = n_0 2^{\mathcal{C}}$. The additional volume-density assigned to the fertile sub-cubes on level $\ell = 2$ is $\delta n_2 = \delta n_1 2^{\mathcal{C}} = n_0 2^{2\mathcal{C}}$. Sub-sub-...-sub-cubes on level ℓ are assigned additional volume-density $\delta n_\ell = n_0 2^{\ell\mathcal{C}}$. We define our clusters such that each of the smallest fertile cubes on the lowest level contains one star and scale the densities of the larger cubes on higher levels accordingly.

Figure 3.3 shows clusters with three different values of $\mathcal{C}$, but the same $\mathcal{D} = 1.58$ and $\mathcal{L} = 5$. The left hand image shows a cluster with $\mathcal{C} = 1$; this is a small scaling exponent, which means that the excess volume-density in a child-cube is not much greater than that of its parent-cube (i.e. the substructure is not very well defined). The middle image again shows the fiducial cluster with $\mathcal{C}=3$, with child-cubes slightly denser than their parent-cubes. The right hand image shows a cluster with $\mathcal{C}=\infty$, where all the stars are located in the cubes on the final level. When $\mathcal{C}=1$ the many stars on the first level swamp any substructure on the lower levels. We therefore concentrate on clusters with $\mathcal{C}>1$.

3.2.4 The number of stars, $\mathcal{N}_\star$

The volume of space occupied by *all* the fertile cubes on level ℓ is $V_g = 8\mathcal{P}^{\ell}_{\text{FERTILE}}$, and hence the total number of stars in the root-cube (including stars assigned to the smaller cubes that are its descendants in the hierarchy) is

$$\begin{aligned}
\mathcal{N}_{\text{ROOT CUBE}} &= V_0 n_0 + \sum_{\ell=1}^{\mathcal{L}} V_\ell \delta n_\ell \\
&= 8n_0 \sum_{\ell=0}^{\mathcal{L}} \mathcal{B}^\ell \\
&= \frac{8n_0\left(\mathcal{B}^{(\mathcal{L}+1)}-1\right)}{(\mathcal{B}-1)}\,, && (3.4)\\
\mathcal{B} &= 2^{(\mathcal{C}+\mathcal{D}-3)}\,. && (3.5)
\end{aligned}$$

The additional number of stars in a single fertile cube on the last level is $n_0 2^{3+\mathcal{L}(\mathcal{C}-3)}$, and we constrain this to be unity, to be compatible with the original 'box-fractal' algorithm of Goodwin & Whitworth (2004). Therefore,

$$\begin{aligned}
n_0 &= 2^{(\mathcal{L}(3-\mathcal{C})-3)}\,, && (3.6)\\
\mathcal{N}_{\text{ROOT CUBE}} &= \frac{2^{(\mathcal{L}(3-\mathcal{C}))}\left(\mathcal{B}^{(\mathcal{L}+1)}-1\right)}{(\mathcal{B}-1)}\,. && (3.7)
\end{aligned}$$

Each fertile cube on level ℓ is therefore allocated

$$\delta\mathcal{N}_\ell = 2^{(\ell-\mathcal{L})(\mathcal{C}-3)} \qquad (3.8)$$

stars, which are positioned randomly within the cube. Non-integer numbers of stars can occur when $\mathcal{C}<\infty$ and are accommodated with a cumulative remainder.

The cluster is then pruned to fit within a sphere of radius unity, which decreases the number of stars by a factor of approximately $\pi/6$. The total number of stars in

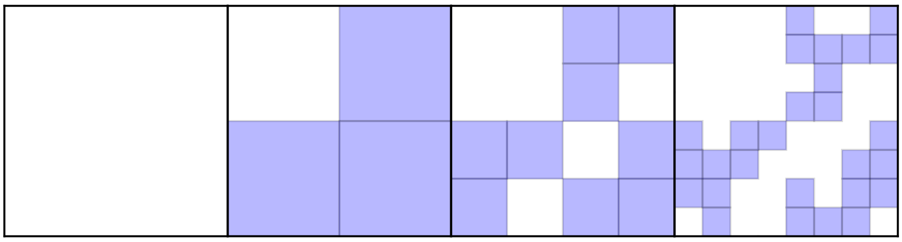

Figure 3.4. A two-dimensional illustration of the repeated partitioning described in Section 3.2.5. This example has a two-dimensional fractal dimension of 1.58, so that 3 out of 4 sub-cubes are randomly chosen to be fertile at each subdivision.

the cluster is therefore

$$\mathcal{N}_{\star} \simeq \frac{\pi}{6} \frac{2^{(\mathcal{L}(3-\mathcal{C}))} \left(\mathcal{B}^{(\mathcal{L}+1)} - 1\right)}{(\mathcal{B}-1)} \, . \tag{3.9}$$

The number of stars in a cluster increases with increasing $\mathcal{D}$, increasing $\mathcal{L}$, and decreasing $\mathcal{C}$.

3.2.5 SUMMARY OF ALGORITHM

1. Start with a fertile cube of side 2 (level 0).
2. Split this cube into 8 sub-cubes and randomly select $\mathcal{F}$ fertile sub-cubes (level 1).
3. Split each fertile sub-cube into 8 new sub-cubes and randomly select $\mathcal{F}$ to be fertile (level 2).
4. Repeat step 3 until $\mathcal{L}$ levels have been generated.
5. Calculate the number of stars to be placed in each fertile cube in each level, $\delta\mathcal{N}_{\ell}$, where $\delta\mathcal{N}_{\mathcal{L}} \equiv 1$.
6. Randomly position each star within its cube to reduce the imposition of a regularly gridded structure.
7. Finally, prune the root-cube to a sphere[†] with radius $r = 1$, and project it into two dimensions from a random viewing angle.

[†] We do not expect real star clusters to be exactly spherical. However, most cluster finding algorithms will search for spherical (or circular) overdensities.

3.3 The $\mathcal{Q}^+$ algorithm

Given a 2D image of a star cluster containing $\mathcal{N}_\star$ stars, we seek to constrain the parameters, $\mathbf{Y}$, describing its intrinsic 3D structure. Implicitly, we assume that the intrinsic 3D structure conforms to the fractal model described in Section 3.2.

We consider star clusters with the following properties,

$$\begin{array}{rllllll} \mathcal{D} = & 1.00, & 1.58, & 2.00, & 2.32; & & \\ \mathcal{L} = & 3, & 4, & 5, & 6, & 7, & 8; \\ \mathcal{C} = & 1, & 2, & 3, & \infty. & & \end{array}$$

There are 96 combinations in total. For each tabulated $\mathbf{Y} = (\mathcal{D}, \mathcal{L}, \mathcal{C})$, we calculate the expected number of stars (Equation 3.9). We exclude clusters whose $\mathcal{N}_\star$ would give too few stars for statistical significance or too many for computational efficiency, leaving 65 $\mathbf{Y}$ states with $20 \leqslant \mathcal{N}_\star \leqslant 10000$. For each of these $\mathbf{Y}$ states, we generate 100 independent star clusters and compute a variety of measures that might help to constrain the parameters of the underlying cluster structure.

3.3.1 Measures derived from Complete Graphs and Minimum Spanning Trees

We define a set of discriminating measures that distinguish the 3 parameters of an artificial fractal model based on the Minimum Spanning Tree and the Complete Graph (see Appendix A for a full description of these graphs and their derivation). $\mathcal{D}$ and $\mathcal{L}$ influence many of these measures so we combine the 6 best measures using Principle Component Analysis. However, $\mathcal{C}$ has a more subtle effect on the structure and does not strongly influence most of the measures. We therefore treat this parameter separately once $\mathcal{D}$ and $\mathcal{L}$ have been estimated.

We first construct the Complete Graph, i.e. the collection of $\mathcal{N}_\star(\mathcal{N}_\star - 1)/2$ edges (straight lines) connecting each star with every other star. The length of the edge joining stars i and j on the Complete Graph is termed s_{ij}.

Next we construct the Minimum Spanning Tree (MST), i.e. the collection of $\mathcal{N}_\star - 1$ edges that connects every star directly to at least one other star with no closed loops, and has the minimum total length. The length of the k^{th} shortest edge on the MST is termed m_k.

We assume that the cluster is spherical, and therefore its projection is circular, with radius r. We do not introduce the notion of a convex hull (c.f. Schmeja & Klessen, 2006), since a cluster generated on the assumption of spherical symmetry, but with a low fractal dimension, $\mathcal{D}$, and/or a high volume-density scaling exponent, $\mathcal{C}$, can have an extremely elongated convex hull (see Figures 3.1, 3.2 and 3.3); it still belongs to a family of clusters built upon the assumption of spherical symmetry.

The 7 statistical measures that are most useful in classifying the structure are the following:

1. the logarithm of the number of stars,
$$\log(\mathcal{N}_\star); \tag{3.10}$$

2. the logarithm of the range of edges on the Complete Graph, $\log(\mathcal{R})$, which is given by
$$\log(\mathcal{R}) = \frac{s_{\text{MAX}}}{s_5}, \tag{3.11}$$
where s_{MAX} is the largest edge on the Complete Graph, and s_5 is the fifth smallest;[‡]

3. the normalised mean edge length on the Minimum Spanning Tree, $\bar{m}$, which is given by
$$\bar{m} = \frac{(\mathcal{N}_\star - 1)}{(\pi\mathcal{N}_\star)^{1/2} r} \sum_{k=1}^{k=\mathcal{N}_\star - 1} \{m_k\}\,; \tag{3.12}$$

4. the normalised mean edge length on the Complete Graph, $\bar{s}$, which is given by
$$\bar{s} = \frac{2}{\mathcal{N}_\star(\mathcal{N}_\star - 1) r} \sum_{i=1}^{i=\mathcal{N}_\star - 1} \sum_{j=i+1}^{j=\mathcal{N}_\star} \{s_{ij}\}\ ; \tag{3.13}$$

5. the mean of the edge lengths on the Minimum Spanning Tree, μ_m, which is given by
$$\mu_m = \frac{1}{\mathcal{N}_\star - 1} \sum_{k=1}^{k=\mathcal{N}_\star - 1} \{m_k\}\ ; \tag{3.14}$$

6. the standard deviation of the edge lengths on the Minimum Spanning Tree, σ_m,

---

[‡] Whereas the largest edge on the Complete Graph, s_{MAX}, is relatively robust – in the sense that, if $(\mathcal{D}, \mathcal{L}, \mathcal{C})$ are held constant, it varies very little from one realisation to another – the smallest edge, s_{MIN}, is not. s_{MIN} has a large variance because it is usually determined by one chance alignment, and therefore can be arbitrarily small. We mitigate this problem by using the fifth smallest edge, s_5. In the same spirit, Larsen (2009) used the 5th brightest cluster in a galaxy as representative of the absolute magnitude, and this practice is often used in extragalactic statistics.

which is given by

$$\sigma_m^2 = \frac{1}{\mathcal{N}_\star - 1} \sum_{k=1}^{k=\mathcal{N}_\star - 1} \left\{ (m_k - \bar{m})^2 \right\} \; ; \tag{3.15}$$

7. the area above the cumulative distribution of MST edges, A, normalised by the number of stars and the size of the cluster. This is given by

$$A = 1 - \frac{m_0 + m_{\mathcal{N}_\star - 1} + 2 \sum_{k=1}^{\mathcal{N}_\star - 2} m_k}{2 s_{max} (\mathcal{N}_\star - 1)} \, , \tag{3.16}$$

   and reflects the proportion of very short edges on the MST (see Section 3.3.3).

Figure 3.5 shows how the means and standard deviations of the 7 statistical measures $(\log(\mathcal{N}_\star), \log(\mathcal{R}), \bar{m}, \bar{s}, \mu_m, \sigma_m, A)$ vary when each of the parameters defining the statistical state, $\mathbf{Y}$, is varied, with the other two held constant at their fiducial values, $\mathcal{D} = 1.58$, $\mathcal{L} = 5$, and $\mathcal{C} = 3$. The frames in the left hand column show what happens when $\mathcal{D}$ is varied. The frames in the middle column show what happens when $\mathcal{L}$ is varied. And the frames in the right hand column show what happens when $\mathcal{C}$ is varied. The plotted points are the means, and the error bars represent the standard deviations. The seventh measure, A, is only used to find $\mathcal{C}$ and therefore not plotted against $\mathcal{D}$ and $\mathcal{L}$.

3.3.2 ESTIMATING $\mathcal{D}$ AND $\mathcal{L}$

We use Principal Component Analysis (PCA, described in detail in Appendix D) to combine the measures described in Section 3.3.1 to reduce the number of dimensions.

In this work we have 6 variables, $\mathbf{Z} = (\log(\mathcal{N}_\star), \log(\mathcal{R}), \bar{m}, \bar{s}, \mu_m, \sigma_m)$, and 6500 data points (100 randomly generated clusters for each of the 65 $\mathbf{Y}$-states). Most of the measures can be reasonably approximated as linear except for μ_m and σ_m (see Figure 3.5). The ranges of the measures vary considerably, which would suggest that the correlation matrix should be used in this case. However, after analysing the effectiveness of this algorithm when using the correlation and covariance matrices, we find that using the covariance matrix results in a better separation of $\mathbf{Y}$-states. Examination of the relative weights of the measures in the two selected PCs reveals that when the covariance matrix is used, the weight of the non-linear measures is

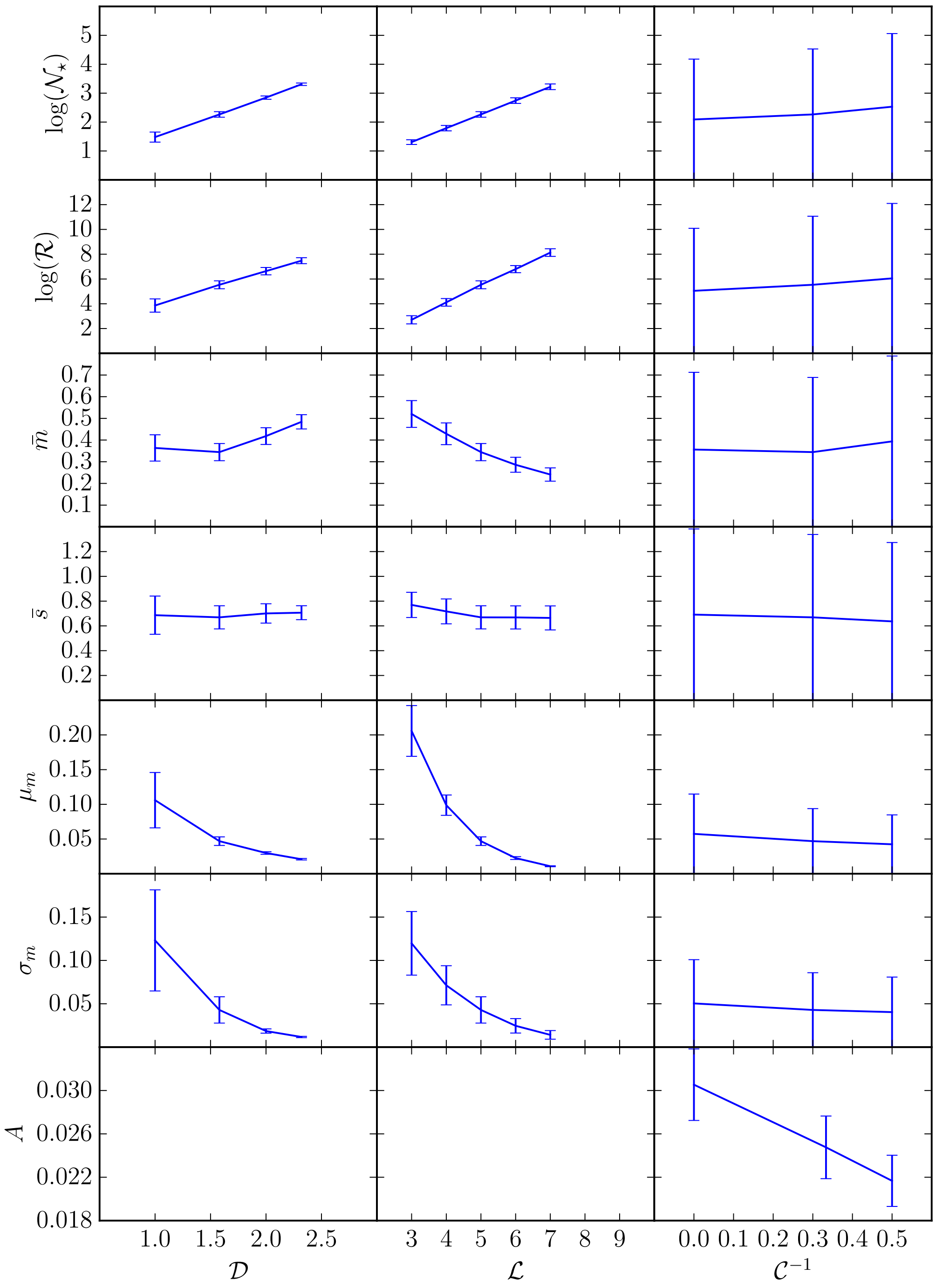

Figure 3.5. The variation of the statistical measures, (log($\mathcal{N}_\star$) (first/top row), log($\mathcal{R}$) (second row), $\bar{m}$ (third row), $\bar{s}$ (fourth row), μ_m (fifth row), σ_m (sixth row), and A (seventh/bottom row)) with the parameters defining the statistical state, ($\mathcal{D}$ (left hand column), $\mathcal{L}$ (middle column), and $\mathcal{C}^{-1}$ (right hand column)). The line gives the mean and the error bar gives the standard deviation. Whichever statistical-state parameter is being varied, the other two are held constant at their fiducial values, viz. $\mathcal{D} = 1.58$, $\mathcal{L} = 5$, and $\mathcal{C} = 3$.

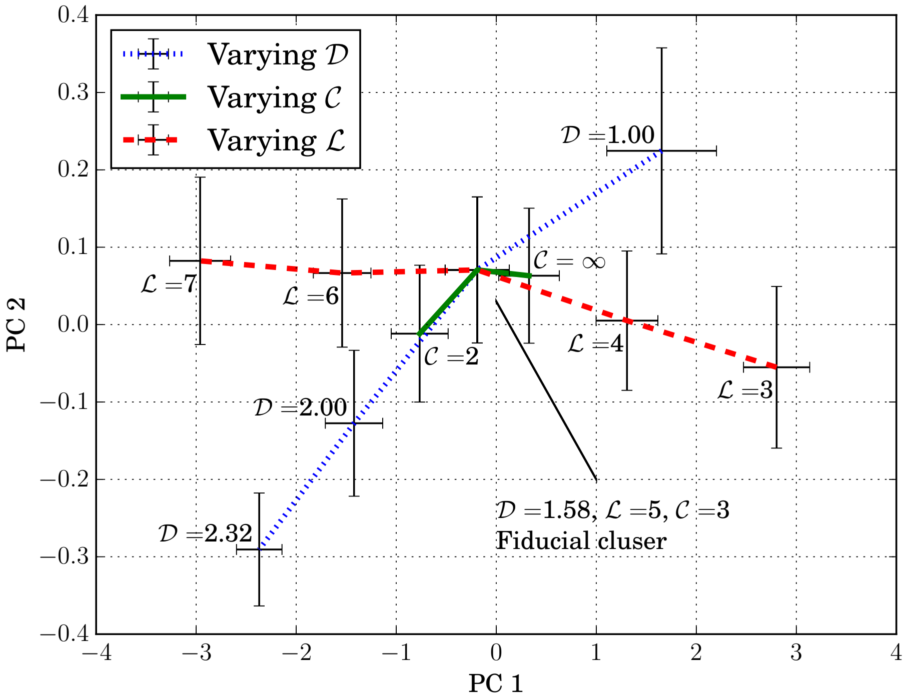

Figure 3.6. The variation of the principal components with the parameters defining the statistical state, ($\mathcal{D}$, $\mathcal{L}$, $\mathcal{C}$). The line gives the mean and the error bar gives one standard deviation. Whichever statistical-state parameter is being varied, the other two are held constant at their fiducial values, viz. $\mathcal{D} = 1.58$, $\mathcal{L} = 5$, and $\mathcal{C} = 3$. $\mathcal{D}$ and $\mathcal{L}$ vary almost orthogonally in the PC space, while the variation in $\mathcal{C}$ overlaps with $\mathcal{D}$ and is therefore more difficult to separate.

suppressed because of their much smaller ranges. We therefore use the covariance matrix in calculating the Principal Components.

We chose to keep only the two most significant PCs (henceforth referred to as PC1 and PC2), where

$$\begin{pmatrix} PC1 \\ PC2 \end{pmatrix} = \begin{pmatrix} -0.354 & -0.832 \\ -0.934 & 0.300 \\ 0.026 & -0.417 \\ 0.022 & -0.193 \\ 0.031 & -0.019 \\ 0.020 & 0.082 \end{pmatrix}^T \times [\left(\mathbf{Z}\right) - \left(\bar{\mathbf{Z}}\right)]^T \tag{3.17}$$

as they encapsulate most of the variance of the data set and make the system very easy to visualise. Figure 3.6 shows how the values of PC1 and PC2 change as the structure of the clusters is changed around the fiducial values.

Each of the 6500 clusters (100 for each of the 65 **Y** states) is transformed into

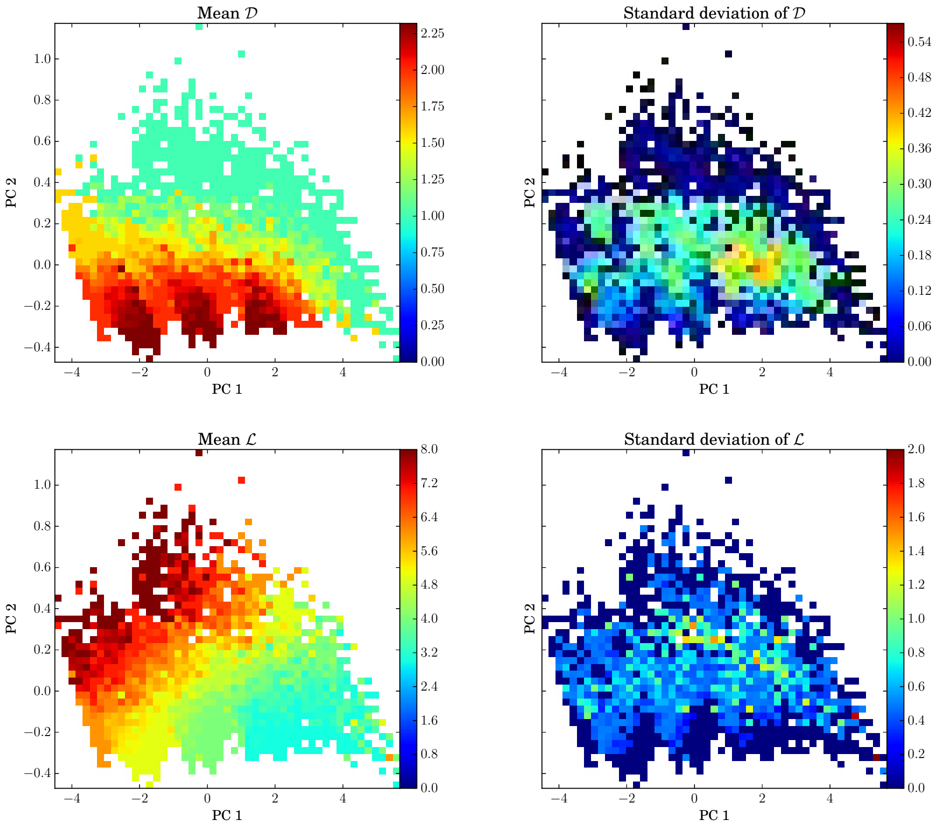

Figure 3.7. The variation of the principal components with $\mathcal{D}$ (top row) and $\mathcal{L}$ (bottom row). The colour corresponds to the mean value (left hand plots) and standard deviation (right hand plots) of each parameter for each square in a 50x50 grid covering the full parameter space. Areas where there are no data points are white.

PC space using Equation 3.17. In order to estimate the parameters of a test (real or artificial) cluster and quantify the uncertainty on this measurement, we grid all the 6500 clusters in PC space on a 50 x 50 regular grid covering the full extent of the data. In each grid square we find all clusters falling in that area, and calculate the mean and standard deviation of their $\mathcal{D}$ and $\mathcal{L}$ values (see Figure 3.7).

The fractal dimension increases somewhat with decreasing PC 1, and increases strongly with decreasing PC 2 (see Figure 3.7, top row). Areas of low $\mathcal{D}$ have very low $\sigma_{\mathcal{D}}$ (in many cases zero) because in this area a small change in parameters makes a large difference in detectable structure, so the different **Y** values are quite well separated. Middling fractal dimensions ($\mathcal{D} \approxeq 1.58$) have higher errors ($\sigma_{\mathcal{D}} > 0.3$) because the **Y** states are less well separated; however even the maximum standard deviation from the mean would only encompass the immediate neighbours in **Y**-space ($\mathcal{D} = 1.00$ or 2.00). The areas of highest $\mathcal{D}$ have again lower errors in $\mathcal{D}$; even though clusters in this area are not well separated, they all have high $\mathcal{D}$ as it is the edge of the parameter space explored. We find that any cluster with $\mathcal{D} \geq 2.32$ becomes indistinguishable from smooth structure. In three dimensions, clusters with $\mathcal{D} = 2.32$ occupy $\approx 60\%$ of the volume of their containing sphere. However, when projected into two dimensions this will fill $\geq 90\%$ of the area of the containing circle, so they appear smooth. We therefore classify all clusters with $\mathcal{D} \geq 2.32$ as not having distinguishable fractal structure. It should be noted that clusters with very low $\mathcal{C}$ also fall in this region, regardless of their $\mathcal{D}$ values, since the larger structures swamp the substructure on lower levels and erase the signs of substructure.

The number of levels increases with decreasing PC 1 and increasing PC 2 (see Figure 3.7, bottom row). The errors on $\mathcal{L}$ are low in most areas, particularly on the edges of parameter space where they drop to zero. The high values of $\sigma_{\mathcal{L}}$ occurring for PC $1 > 4$ and PC $2 \approx 0.2$ are caused by small number statistics at the very edge of the parameter space. We consider that clusters with $\mathcal{L} < 4$ will not have enough of a nested hierarchy to be described in terms of fractal structure. Clusters falling in this area are therefore classified as not fractal.

$\mathcal{D}$ and $\mathcal{L}$ both vary systematically across the range of the PCs (see Figure 3.6). $\mathcal{C}$, on the other hand, varies only locally around a particular $\mathbf{Y}$-state, and in the same direction as the variation in $\mathcal{D}$. Another method is therefore needed to estimate $\mathcal{C}$ once $\mathcal{D}$ and $\mathcal{L}$ have been estimated from the principal components.

3.3.3 ESTIMATING $\mathcal{C}$

Figure 3.8 shows how the cumulative distribution of MST edges (normalised by the size of the cluster, s_{max}, and the total number of edges, $\mathcal{N}_{\star} - 1$) varies with $\mathcal{C}$ for set values of $\mathcal{D}$ and $\mathcal{L}$. A cluster with higher $\mathcal{C}$ will have a greater proportion of short edges, shifting this curve to the right. The area above this curve, A, therefore increases with increasing $\mathcal{C}$, but this also varies with $\mathcal{D}$ and $\mathcal{L}$. Therefore, for each combination of $\mathcal{D}$ and $\mathcal{L}$ we measure A for all artificial clusters and find how it changes with $\mathcal{C}$.

For a test cluster, once $\mathcal{D}$ and $\mathcal{L}$ have been estimated using the PCs, we use Bayes' theorem to infer the posterior probability of $\mathcal{C}$, given A.

$$P(\mathcal{C}|A) = \frac{P(A|\mathcal{C})P(\mathcal{C})}{P(A)} \tag{3.18}$$

This equation is explained in greater detail in Appendix B. The likelihood of a particular A, given $\mathcal{C}$ ($P(A|\mathcal{C})$) is calculated from the mean and standard deviation of A ($\mu_{A_{\mathcal{C}}}, \sigma_{A_{\mathcal{C}}}$) over the 100 realisations at each $\mathbf{Y}$-state. We calculate the probability

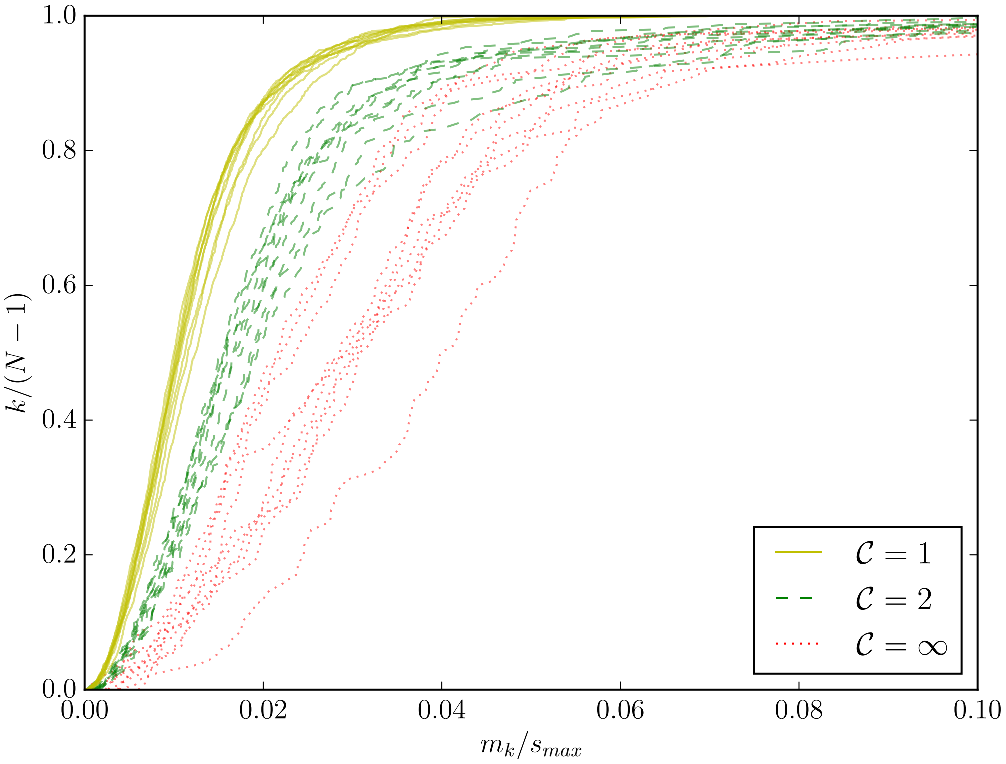

Figure 3.8. The cumulative distribution of minimum spanning tree edges for different density scaling exponents, and the fiducial $\mathcal{D} = 1.58$, $\mathcal{L} = 5$.

of the test A value belonging to each **Y**-state,

$$P(A|\mathcal{C}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(A-\mu_{A_\mathcal{C}})^2}{2\,\sigma^2_{A_\mathcal{C}}} \right\} \tag{3.19}$$

$P(\mathcal{C})$ is a weight given to each value of $\mathcal{C}$ based on prior knowledge of the distribution. In this case, each value of $\mathcal{C}$ is given equal weight. $P(A)$ is a normalisation constant to ensure that the probabilities add up to unity. We calculate the expectation value of $\mathcal{C}$ and its standard deviation from the posterior probabilities, i.e.

$$E_\mathcal{C} = \frac{\sum P(\mathcal{C}|A)\mathcal{C}}{\sum P(\mathcal{C}|A)} \tag{3.20}$$

$$\sigma^2_\mathcal{C} = \frac{\sum P(\mathcal{C}|A)(\mathcal{C}-E_\mathcal{C})^2}{\sum P(\mathcal{C}|A)} \tag{3.21}$$

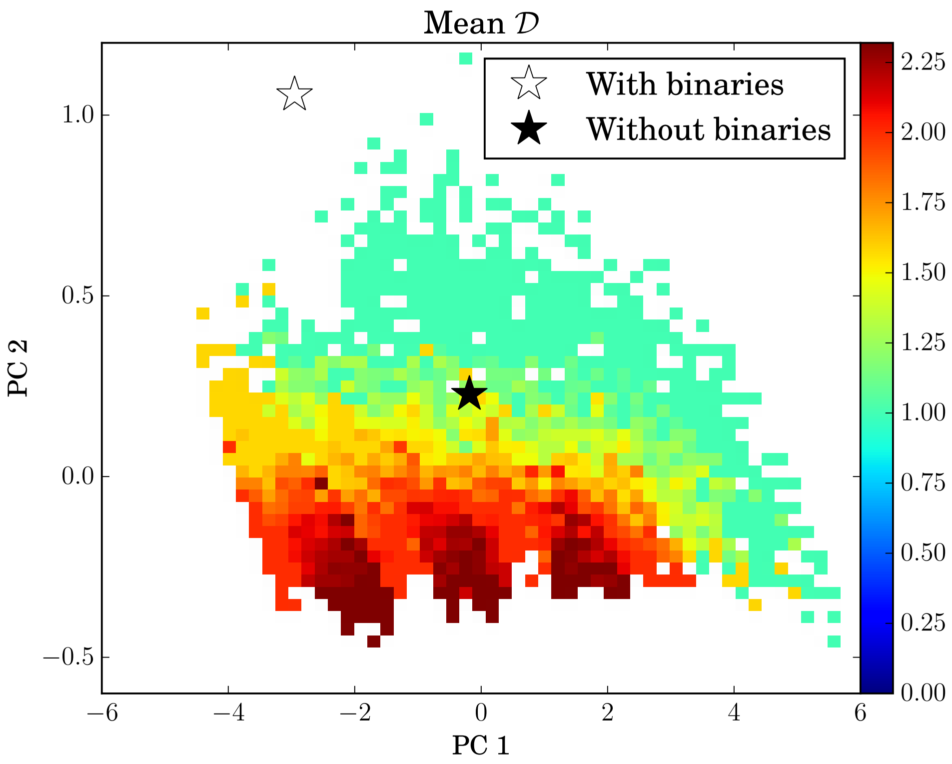

Figure 3.9. The Chamaeleon I cluster with (open star) and without (filled star) binaries superimposed on the $\mathcal{L}$ grid. It lies outside the parameter space if binaries are left in, but moves into the area of fractal clusters when binaries are removed. The colour gives the average value of $\mathcal{D}$ in each pixel, as in Figure 3.7.

3.4 USAGE

The only input to this algorithm is the two-dimensional positions of the stars in the real, synthetic or simulated cluster.

3.4.1 BINARIES AND MULTIPLE SYSTEMS

The artificial clusters generated for this analysis model hierarchical nested clustering, which provides a suitable way to quantify the sub-structure of young star clusters. However, one major feature of the structure of real star clusters that is not modelled in this work is binary or higher order multiple systems. If these are present in a cluster, they will produce many of the shortest edges in both the Complete Graph and the Minimum Spanning Tree, and will therefore significantly skew some of the measures.

Larson (1995) discussed the relation between hierarchical clustering and multiple systems and found that they showed a distinctly different distribution of separations. Gladwin et al. (1999) examined this in several clusters and found a characteristic separation of ≈ 0.03 pc that distinguishes hierarchical clustering from the regime of binary and multiple systems. We use this to remove the effect of binaries from real clusters before analysis.

Pairs of stars separated by less than 0.03 pc are classified as binaries. However, some pairs of stars with small enough separations to be defined as "binary" will be generated when an artificial cluster is viewed in two dimensions, due to chance alignments of stars which are well separated in the third dimension. Numerical experiments with artificial clusters across the parameter space show that this is simply dependent on average surface density:

$$\log_{10}(N_{bin}) = 1.86 \log_{10}(\mathcal{N}_\star/\pi R^2_{cluster}) - 0.56 \tag{3.22}$$

where N_{bin} is the number of chance alignments smaller than 0.03 pc and $R_{cluster}$ is the maximum distance in parsecs of any star from the mean position of all the stars. This can be easily calculated for a real cluster. When small separations are found in a real cluster they are randomly removed until the number does not exceed the predicted N_{bin}, thereby removing the effect of binaries on the statistical measures but without removing the proportion of small separations expected from projection. In the case of higher order multiples, this will result in the whole system being replaced by a single star. In effect, we are studying the hierarchical distribution of systems, rather than of stars.

In the 4 real clusters analysed in Section 4.1 the effect of removing binaries decreases $\log(\mathcal{R})$ by $\approx 30\%$ and increases μ_m by $\approx 20\%$.

Figure 3.9 shows the effect of binary systems in a real cluster on its position in Principle Component space. Using the original data of Chamaeleon I from Kirk & Myers (2011) as an example, the result lies well outside the parameter space of artificial fractal clusters. Once binaries have been removed, the cluster falls well within the parameter space explored. This procedure is therefore applied to all real clusters before the $\mathcal{Q}^+$ analysis.

3.4.2 RUNNING AND RESULTS

We first build the CG and MST of these points, and compute the six measures, $\mathbf{Z} = (\log(\mathcal{N}_\star), \log(\mathcal{R}), \bar{s}, \bar{m}, \mu_m, \sigma_m)$. These are transformed into principal components via Equation 3.17. The values of the PCs identify which grid square the test cluster falls into (see Figure 3.7) and the mean and standard deviation of artificial clusters in that grid square give an estimate and uncertainty for the $\mathcal{D}$ and $\mathcal{L}$ values of the test cluster. We then calculate the value of A for the test cluster and compare to the analytic clusters with the same $\mathcal{D}$ and $\mathcal{L}$ using a Bayesian approach to estimate $\mathcal{C}$ and its error.

3.4.3 Observational effects

Our methods of identifying and delineating star clusters are not perfect. When this algorithm is applied to observations of real star clusters there are several observational effects that need to be considered.

Stars in the background or foreground may be erroneously identified as cluster members. Assuming this population is more evenly distributed than the substructured young cluster being analysed, this should make the cluster appear smoother, with a higher $\mathcal{D}$. The addition of more evenly distributed stars would decrease the average edge length of the Minimum Spanning Tree as well as increasing $\log(\mathcal{N}_\star)$. This would lead to a small decrease in PC1 and a larger decrease in PC2, which in turn would result in a small overestimate of the fractal dimension and number of levels. This would also affect the estimate of $\mathcal{C}$, as the addition of more short edges to the MST would lead to a lower $\mathcal{C}$.

Crowding and incompleteness are very different issues observationally, but will have similar effects on the $\mathcal{Q}^+$ algorithm. Both result in a reduced number of stars - crowding in the centre of dense clusters causes multiple stars to blur together and be detected as one bright source, while incomplete observations will neglect a population of lower mass stars. Incompleteness only affects the lower mass stars, while crowding affects stars of any mass but only in the densest areas, but the $\mathcal{Q}^+$ algorithm does not take into account the mass of the stars so the effects will be similar. Both will reduce the number of stars and increase the average edge length of the MST. They may remove many of the shortest edges of the MST, therefore increasing $\log(\mathcal{R})$ which has a large effect on PC1. This will reduce the estimate of the number of levels. However, the $\mathcal{Q}^+$ algorithm does not examine the separation of binary and multiple systems, as discussed in Section 3.4.1. This puts a limit on the scales we need to resolve in order for the results to describe the true fractal structure.

Finally, observations are limited in terms of their field of view. In the fractal model described in Section 3.2 we account for this effect by imposing a spherical outline on the clusters. This will remove some of the outlying stars, but this effect should be the same in observed clusters as we tend to focus our observations on the bright centre of clusters and therefore may miss outlying members. If a small section of a cluster is analysed we would expect the fractal dimension of the cluster to still be recovered, as the nature of fractals implies similarity over a range of scales. However, this is only the case if there are enough levels in the hierarchy to still give meaningful results. If the full cluster has a high number of levels, we would expect the analysis of a part of that cluster to result in the same $\mathcal{D}$ and $\mathcal{C}$ but a lower $\mathcal{L}$.

3.5 EVALUATION WITH SYNTHETIC STAR CLUSTERS

In order to test the reliability of this algorithm, we apply it to artificial clusters with known structure and test the accuracy of the $\mathcal{Q}^{+}$ method in recovering the input parameters. We focus our attention on clusters with $\mathcal{D} = 1.00,\ 1.58$ and 2.00, $\mathcal{L} = 4,\ 5$ and 6 and $\mathcal{C} = 2,\ 3$ and ∞. For each of these 27 $\mathbf{Y}$-states, we have created 10 synthetic fractal star clusters, projected them at a random angle, and analysed them using the algorithm described in Section 3.3. We find the means and standard deviations of $\mathcal{D}_{\rm OUT}/\mathcal{D}_{\rm IN}$, $\mathcal{L}_{\rm OUT}/\mathcal{L}_{\rm IN}$, $\mathcal{C}_{\rm OUT}/\mathcal{C}_{\rm IN}$, where $Y_{i,\,\rm IN}$ is the value of Y_i that went into the construction of a synthetic three-dimensional star cluster, and $Y_{i,\,\rm OUT}$ is the value of Y_i estimated from the projected two-dimensional image of this cluster. These values are given in Table 3.1.

We can see that $\mathcal{D}$ and $\mathcal{L}$ are reliably estimated across the parameter space, although a low $\mathcal{D}$ is often overestimated when $\mathcal{L}$ is low. $\mathcal{C}$ is not as well constrained, particularly when $\mathcal{C} = \infty$. This could be improved by creating more analytic clusters with different values of $\mathcal{C}$ to give more a priori information for the Bayesian analysis, but as this is fairly computationally intensive we leave this to future work. Once use of this algorithm has revealed the areas of parameter space populated by real and simulated clusters, we can focus our attention on improvements in these regions.

$\mathcal{D}_{\rm IN} =$	1.00								
$\mathcal{L}_{\rm IN} =$	4			5			6		
$\mathcal{C}_{\rm IN} =$	2	3	∞	2	3	∞	2	3	∞
$\mathcal{D}_{\rm OUT}/\mathcal{D}_{\rm IN} =$	1.37±0.19	1.18±0.15	1.00±0.01	1.19±0.13	1.11±0.09	1.07±0.09	1.25±0.13	1.11±0.19	1.23±0.23
$\mathcal{C}_{\rm OUT}/\mathcal{C}_{\rm IN} =$	0.47±3.52	0.99±2.26	(2.53±0.55)	0.56±3.03	1.05±2.16	(5.82±2.30)	0.66±2.48	0.80±3.29	(7.06±2.47)
$\mathcal{L}_{\rm OUT}/\mathcal{L}_{\rm IN} =$	0.99±0.10	0.90±0.09	0.90±0.07	1.05±0.10	0.98±0.11	0.96±0.15	0.96±0.08	0.98±0.15	0.79±0.11
$\mathcal{D}_{\rm IN} =$	1.58								
$\mathcal{L}_{\rm IN} =$	4			5			6		
$\mathcal{C}_{\rm IN} =$	2	3	∞	2	3	∞	2	3	∞
$\mathcal{D}_{\rm OUT}/\mathcal{D}_{\rm IN} =$	1.07±0.15	1.08±0.17	0.89±0.15	1.00±0.15	0.96±0.11	0.93±0.15	1.01±0.09	1.00±0.11	0.98±0.11
$\mathcal{C}_{\rm OUT}/\mathcal{C}_{\rm IN} =$	0.29±3.63	1.02±2.34	(6.20±2.45)	0.68±3.33	1.11±2.39	(4.93±3.44)	0.63±3.18	1.11±2.23	(4.25±3.15)
$\mathcal{L}_{\rm OUT}/\mathcal{L}_{\rm IN} =$	1.07±0.12	0.92±0.13	1.03±0.15	1.06±0.13	1.02±0.10	0.99±0.13	1.07±0.07	0.99±0.06	0.94±0.09
$\mathcal{D}_{\rm IN} =$	2.00								
$\mathcal{L}_{\rm IN} =$	4			5			6		
$\mathcal{C}_{\rm IN} =$	2	3	∞	2	3	∞	2	3	∞
$\mathcal{D}_{\rm OUT}/\mathcal{D}_{\rm IN} =$	0.94±0.08	0.93±0.10	0.91±0.12	0.96±0.08	0.92±0.09	1.00±0.10	0.96±0.06	1.03±0.08	1.06±0.08
$\mathcal{C}_{\rm OUT}/\mathcal{C}_{\rm IN} =$	0.83±2.48	0.83±2.60	(3.30±3.54)	0.83±1.75	0.72±3.53	(7.48±3.70)	0.58±2.23	0.85±2.41	(6.05±3.61)
$\mathcal{L}_{\rm OUT}/\mathcal{L}_{\rm IN} =$	1.10±0.08	1.07±0.09	1.05±0.09	1.06±0.10	1.07±0.08	0.97±0.09	1.04±0.08	0.97±0.08	0.93±0.08

Table 3.1. Means and standard deviations for the ratios between the statistical-state parameters used in the construction of synthetic 3D fractal star clusters, $Y_{i,\rm IN}$, and the statistical-state parameters, $Y_{i,\rm OUT}$, derived from 2D projections of these clusters. In cases where $\mathcal{C} = \infty$ this ratio is meaningless, but we use a very high number in place of infinity so these results are presented for completeness.

3.6 CONCLUSIONS

We present a new algorithm for quantifying the fractal nature of young star clusters in terms of three parameters:

- the fractal dimension, $\mathcal{D}$ - a measure of the clumpiness or smoothness of the distribution;
- the number of levels, $\mathcal{L}$ - a measure of the range of scales of substructure within the overall cluster;
- the density scaling exponent, $\mathcal{C}$ - a measure of the relative distribution of mass on the different scales.

The algorithm is able to reliably classify the internal structure of young stellar clusters within a parameter space bounded by the following limitations:

1. $\mathcal{L} \leq 3$ will not have enough substructure to be detectably fractal.
2. $\mathcal{D} \geq 2.32$ will fill most of the area when projected into 2D, and therefore appear smooth.
3. $\mathcal{C} \leq 1$ will overpopulate the higher levels and swamp any substructure on smaller scales.

We anticipate that this method will be useful for:

- quantitative and objective analysis of large numbers of structures in the huge data sets available from modern observing methods;
- unbiased analysis of the results of simulations in 2 or 3 dimensions;
- analytical comparison of observational and simulated data sets to validate results and inform inferences about the similarity of observed regions to simulated regions.

The $\mathcal{Q}^+$ algorithm is available at `https://github.com/SJaffa/Q_plus`.

CHAPTER 4

$\mathcal{Q}^+$ APPLIED TO OBSERVED AND SIMULATED CLUSTERS [*]

The $\mathcal{Q}^+$ algorithm is applied to observed positions of YSOs in some well known clusters, and to data sets from SPH simulations of colliding clouds. The resulting estimates for $(\mathcal{D}, \mathcal{L}, \mathcal{C})$ are discussed and explained in terms of the visually apparent variation in structure. The evolution of fractal substructure is discussed and simulations of high mass cluster evolution are analysed.

---

[*] The work presented here is published in Jaffa, Whitworth & Lomax, 2017, MNRAS, 446, 1082, and in Sills, Rieder, Scora, McCloskey & Jaffa, 2018, MNRAS, *in press*

4.1 APPLICATION TO OBSERVED STAR CLUSTERS

4.1.1 DATA

Kirk & Myers (2011) have used *Spitzer* (Werner et al., 2004) near-IR data, in combination with ground based spectral surveys, to examine the clustering of Young Stellar Objects (YSOs) in four nearby star forming regions. They restrict their analysis to stars which are less than several Myr old, so that they are unlikely to have migrated far from their birthplace, but are old enough to have a reliable estimate of spectral type (Class I/II). For full details of the source catalogues, see their Appendix A. We use their data (plotted in Figures 4.1, 4.2, 4.3, and 4.4 and summarised in Table 4.1) to show the capabilities of the $\mathcal{Q}^+$ algorithm to quantify structure in observations.

Region:	Taurus	ChaI	Lupus 3	IC348
Number of sources:	352	237	70	363
Distance adopted (pc):	140	160	200	300

Table 4.1. Properties of the YSO groups collected by Kirk & Myers (2011)

4.1.2 RESULTS

Figure 4.5 shows the placement of the 4 real clusters in Principal Component space after binaries have been removed. The estimated parameters of the 4 clusters are given in Table 4.2 along with previous estimates of their structure.

Both Taurus and Chamaeleon I show clear sub-structure, but Taurus appears to be more clumpy, which is reflected in the lower $\mathcal{D}$. Taurus is the largest of these 4 regions, so the sub-structure has a greater range of scales. This is quantified by the high value of $\mathcal{L}$. The estimated parameters are compatible with a visual inspection of the data (see Figures 4.1, 4.2, 4.3, and 4.4) whilst having the merit of being objective and quantitative, whereas visual inspection is subjective and qualitative.

The estimated fractal dimensions are lower than those obtained using the original $\mathcal{Q}$ parameter, but the trend is the same. IC 348 was classified by CW04 as not having sub-structure but being centrally concentrated. Our analysis agrees with this, although we are not able to quantify the structure, only to say that it is not measurably sub-structured. Lupus 3 was not analysed in CW04, but our analysis

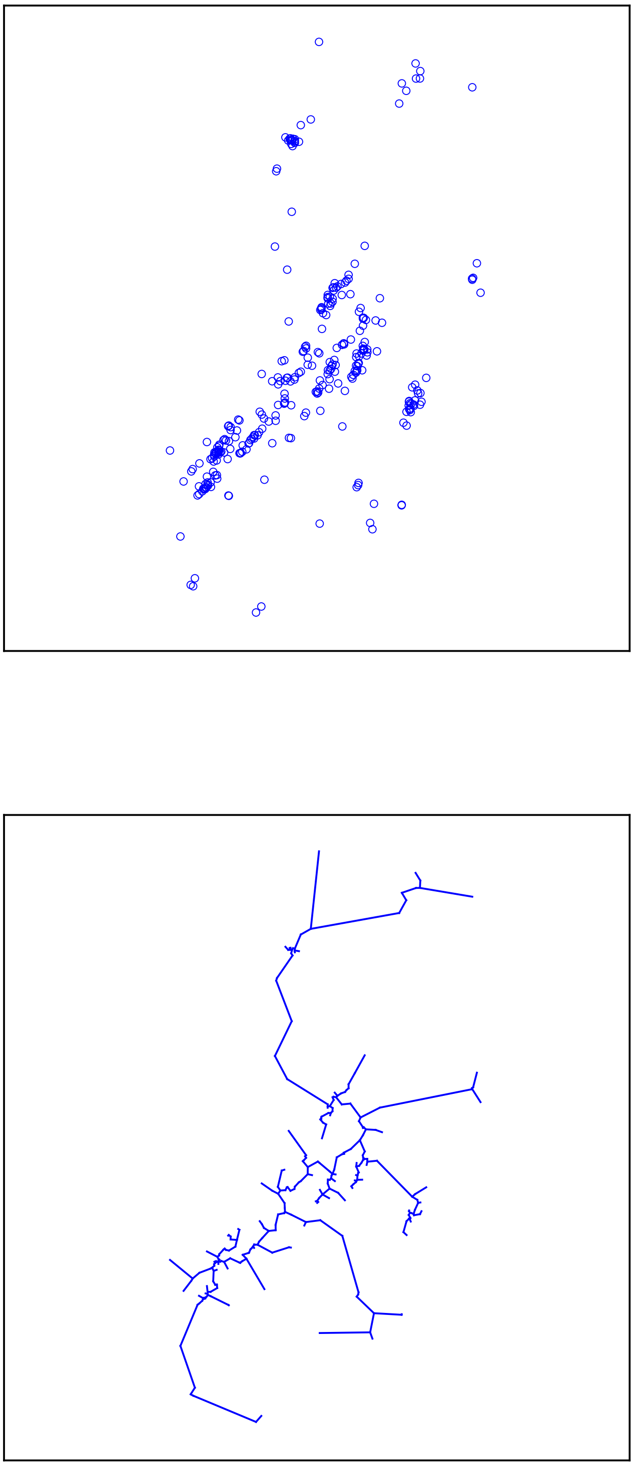

Figure 4.1. Young stellar objects in Taurus and the corresponding MST.

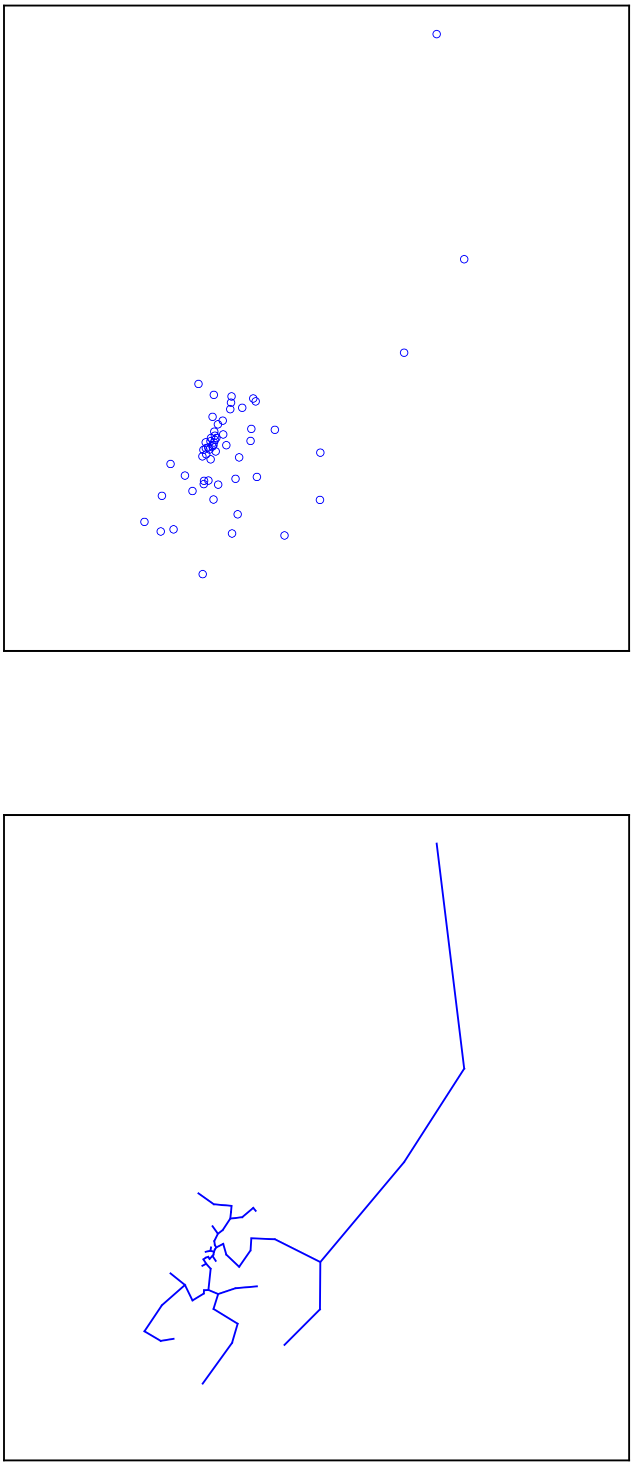

Figure 4.2. Young stellar objects in Lupus 3 the corresponding MST.

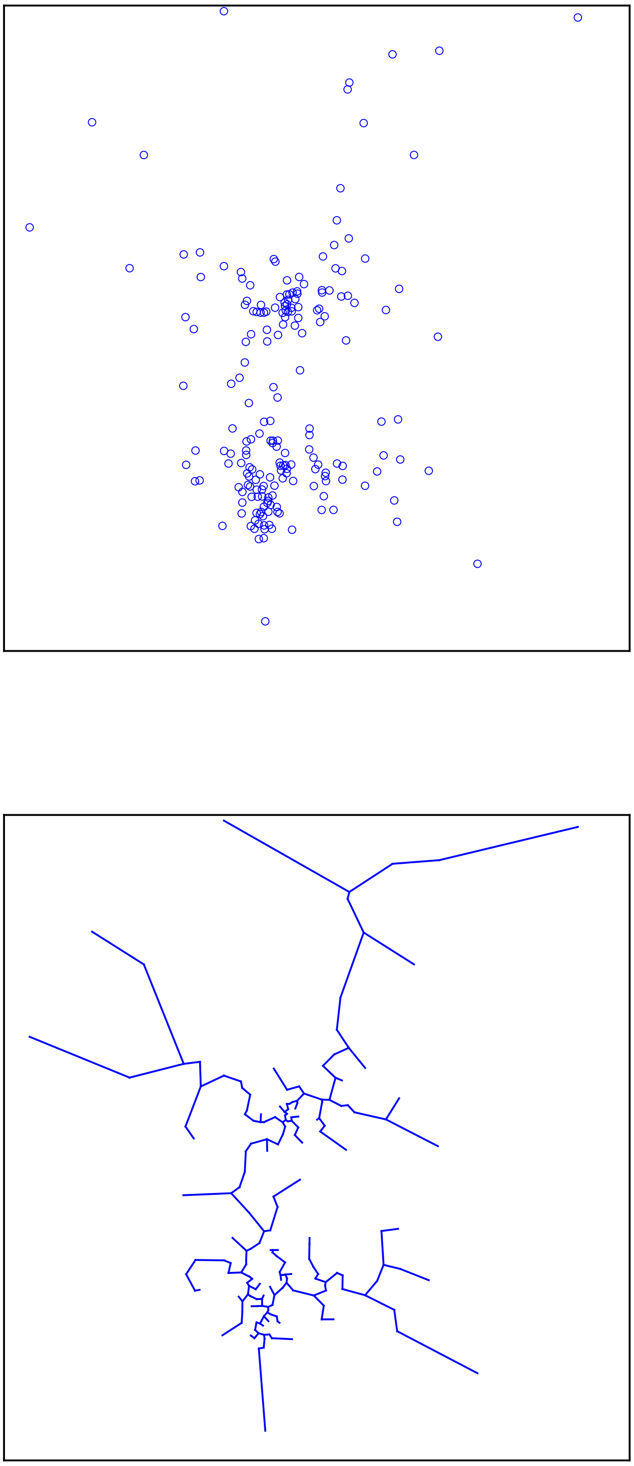

Figure 4.3. Young stellar objects in Chamaeleon I and the corresponding MST

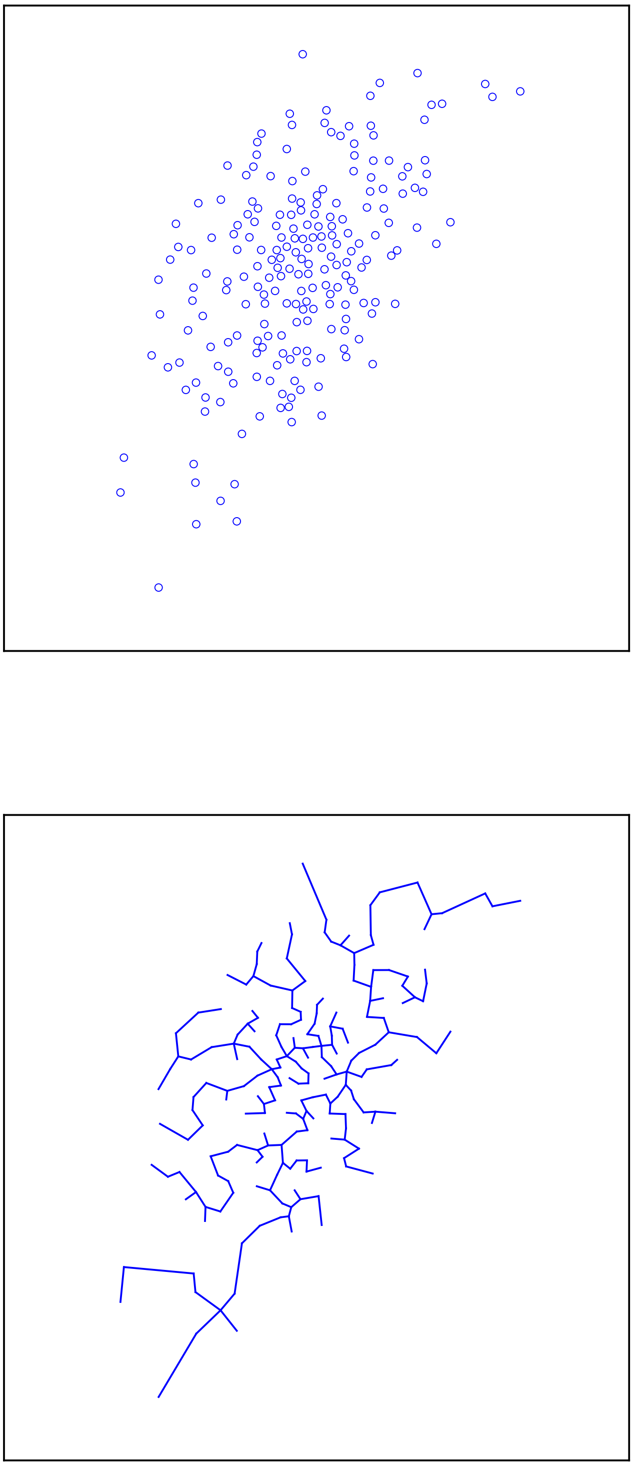

Figure 4.4. Young stellar objects in IC 348 and the corresponding MST.

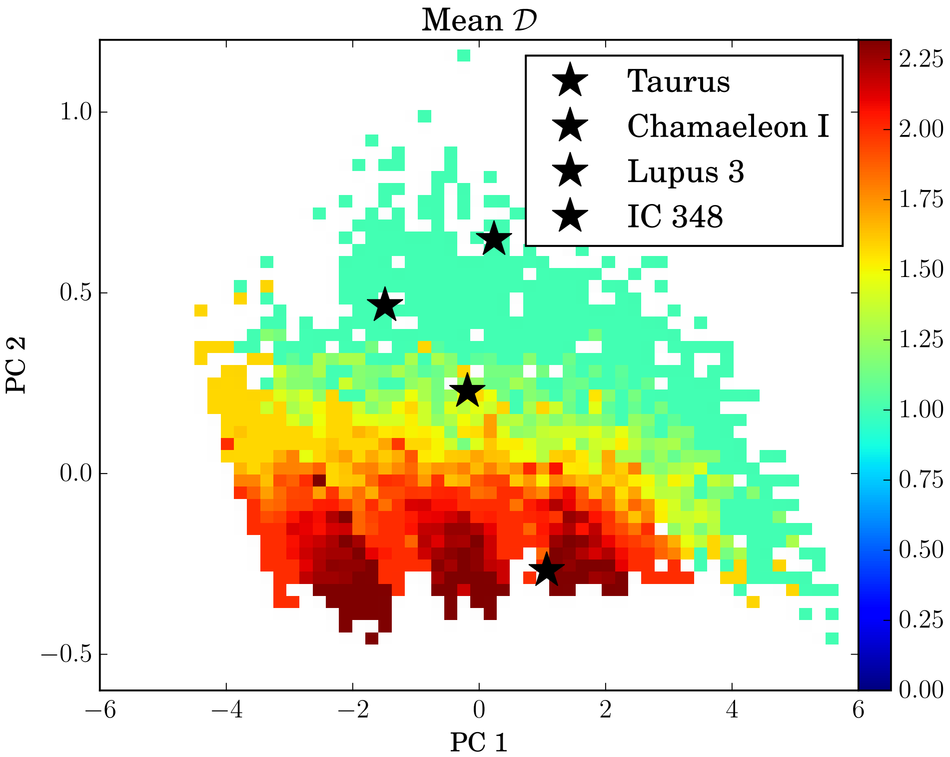

Figure 4.5. Real clusters superimposed on the background of mean $\mathcal{D}$.

shows it to be highly sub-structured, but over a smaller range of scales than Taurus (lower $\mathcal{L}$). It also has a higher $\mathcal{C}$, which reflects the fact that there are very few stars outside the main dense clump, while Taurus and Chamaeleon I have more widely distributed stars outside the densest regions.

The distribution of young stars is expected to follow that of the molecular clouds from which they are born (Elmegreen & Falgarone, 1996). Most fractal measures of molecular clouds have found values in the region of 2.3, which suggests a smoother distribution than that found in these clusters. However, combining fractal dimensions measured with different methods may be misleading (see Chapter 2). Furthermore, previous methods have often considered the fractal dimension to be the only parameter of import. In Chapter 5 we discuss the application of the $\mathcal{Q}^+$ algorithm to continuum images such as molecular cloud maps, which will enable a more meaningful like-for-like comparison.

Cluster:	Lupus 3	Chamaeleon I	Taurus	IC 348
$\mathcal{D} =$	1.0 ± 0.0	1.5 ± 0.2	1.0 ± 0.0	NOT FRAC.
$\mathcal{L} =$	7.0 ± 0.3	5.2 ± 0.4	8.0 ± 0.0	–
$\mathcal{C}^{-1} =$	0.1 ± 0.0	0.4 ± 0.0	0.4 ± 0.0	–
CW04	N/A	$\mathcal{D} = 2.25$	$\mathcal{D} = 1.5$	$\alpha = 2.2$

Table 4.2. Parameters estimated for the four real clusters after the removal of binary or multiple systems. The fourth row gives the results quoted in CW04 using the original $\mathcal{Q}$ parameter method, where $\mathcal{D}$ is the fractal dimension and α is the radial density exponent.

4.2 APPLICATION TO SIMULATIONS OF CLOUD-CLOUD COLLISIONS

4.2.1 DATA

Balfour et al. (2015) conducted Smooth Particle Hydrodynamics (SPH) simulations of colliding clouds to investigate how the collision speed influenced the resultant star formation. Figure 4.6 shows representative images from their simulations; the background image represents the column density of the cloud and the black dots are sink particles representing protostars. Each of these simulations is viewed along the collision axis. The slower collisions produced a 'hub and spoke' pattern of radial filaments in the compressed layer, with most star formation taking place in the centre. The higher speed collision produced a more complex 'spider's web' structure, where protostars are formed at the junctions between filaments.

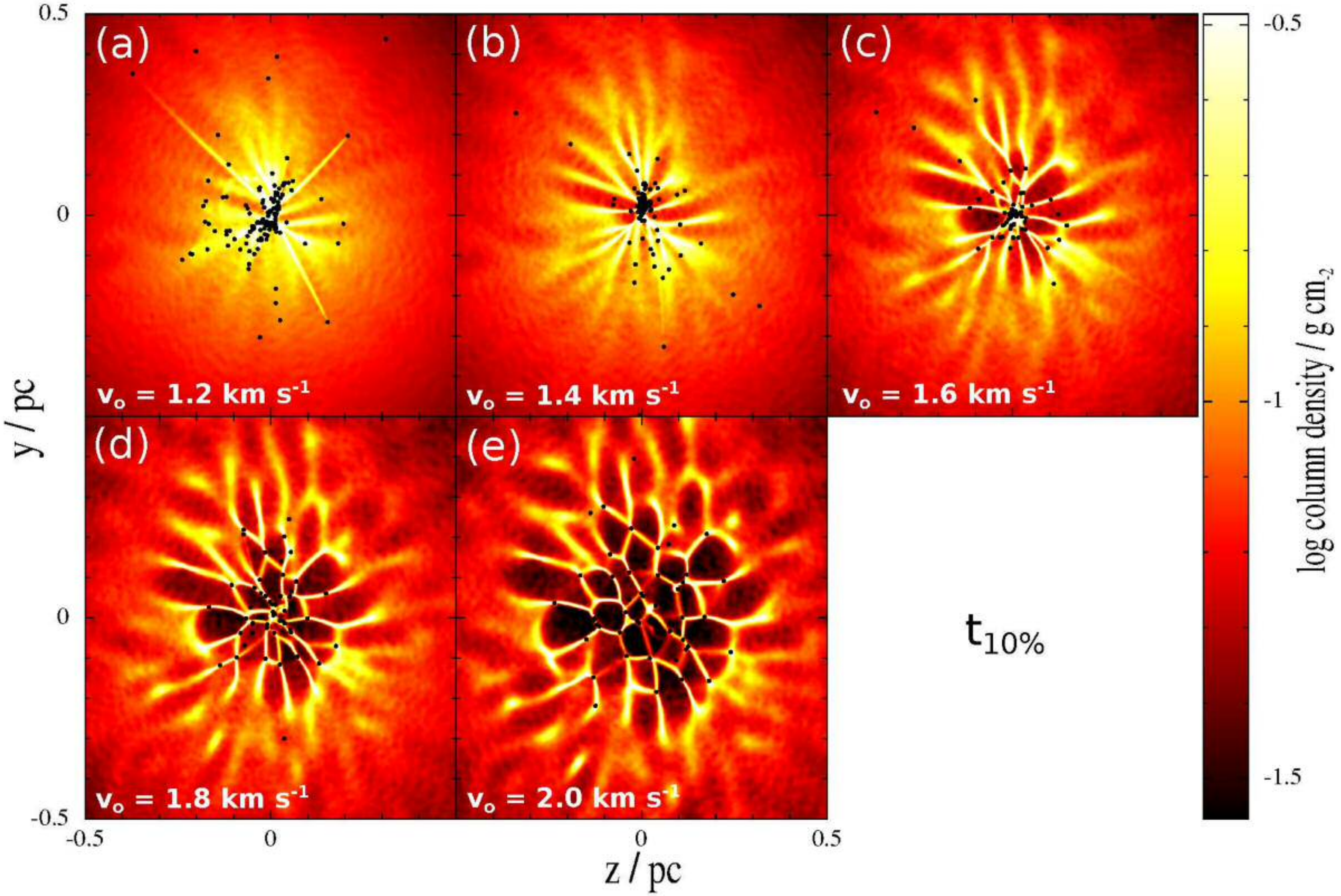

Figure 4.6. Images of simulations showing how the structure of gas and sink particles changes in collisions with different relative velocities, taken from Balfour et al. (2015). The simulations are terminated when 10% of the initial gas mass has been accreted into sink particles.

v_0:	1.2	1.4	1.6	1.8	2.0
$\mathcal{D} =$	1.0±0.0	1.1±0.2	1.2±0.3	1.0±0.1	NOT FRAC.
$\mathcal{L} =$	6.3±0.4	4.8±0.6	4.8±0.7	5.1±0.6	–
$\mathcal{C} =$	1.0±1.7	0.0±0.9	0.0±0.9	1.3±3.3	–

Table 4.3. Parameters estimated for the sink particles from the simulations of Balfour et al. of clouds colliding at different speeds, v_0, given in km s^{-1}.

4.2.2 RESULTS

Figure 4.7 and Table 4.3 show the results of analysis of these clusters. At the slowest speed, the compressed layer between the two clouds begins to collapse radially, resulting in a centrally concentrated layer. Star formation predominantly takes place in the centre of this layer resulting in a very tight cluster of stars with a low fractal dimension and a high number of levels. The results for the intermediate speeds are very similar as the gas distribution is still dominated by radial filaments but the central clump is not as tight. This is characterised by a lower $\mathcal{L}$ than the slowest collisions. These intermediate collisions also have a higher $\mathcal{C}$ than the slowest collisions. This indicates that more of the stars are concentrated in the small dense cluster. In the case of the slowest collision we believe that many of the stars that are outside the central clump were not formed there but ejected by dynamical interactions in the dense central clump. This therefore does not reflect the position at which the stars were born, but this dynamical evolution is very rapid in small, dense clusters so the chance of observing a cluster before this occurs is very small. The fastest collisions produce an extended network of filaments, with star formation taking place at the junctions between filaments. This produces a more dispersed cluster that is classified as not sub-structured.

The estimated parameters given for the slowest collision simulation are similar to the observational results for Lupus 3. Figure 4.8 shows the positions of YSOs in Lupus 3 and the sink particles in the simulation, for comparison. This similarity suggests that Lupus 3 may have formed following a slow collision like that simulated by Balfour et al. (2015).

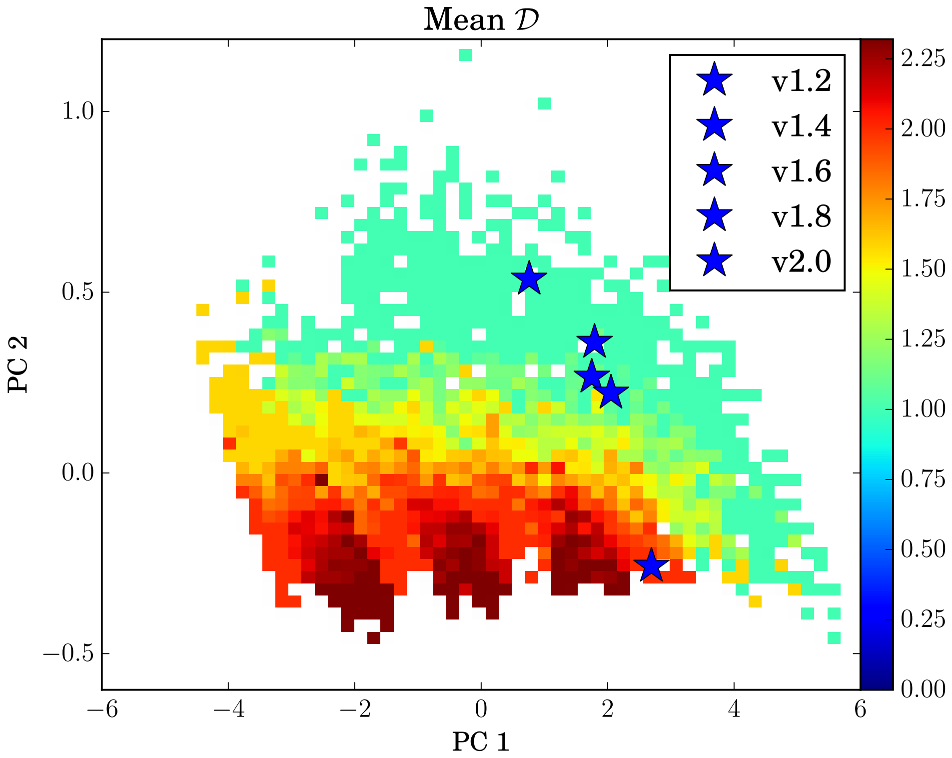

Figure 4.7. Results of analysis of sink particles produced by colliding clouds with different relative velocities superimposed on the background of mean $\mathcal{D}$.

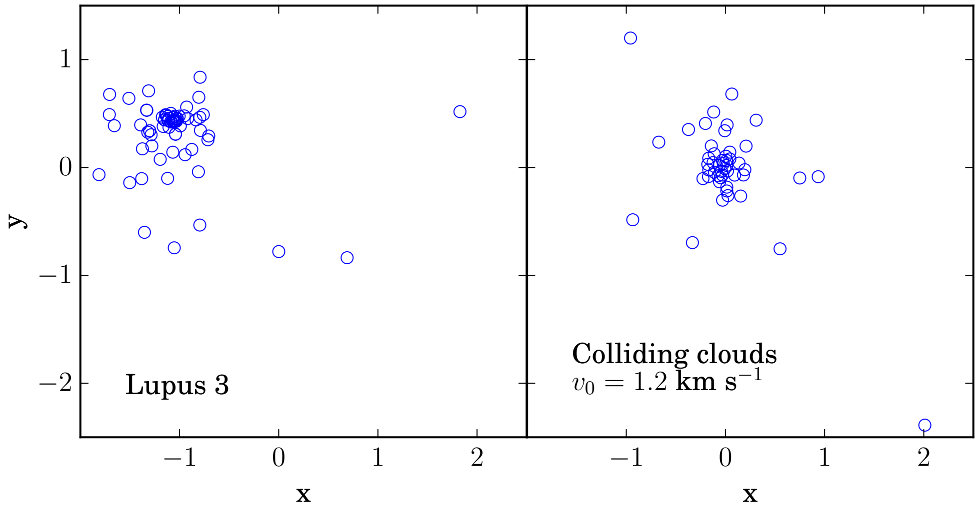

Figure 4.8. Comparison of the Lupus 3 cluster and simulation results for slowly colliding clouds.

4.3 APPLICATION TO SIMULATIONS OF EVOLVING YOUNG EMBEDDED CLUSTERS

4.3.1 DATA

Sills et al. (2018) have performed simulations of the evolution of massive young clusters in a gas envelope. Using initial conditions drawn from the MYStIX Project (Feigelson et al., 2013) and supplemented by a low mass stellar population (to extend the IMF from 0.83 down to 0.08 $M_\odot$) and an extrapolated gas envelope, they couple a collisional N-body code with a hydrodynamics code to follow the interaction of stars and gas. They study the evolution of three initially sub-structured massive clusters: DR 21, NGC 6357, and M17. The structure of each of these has been analysed by Kuhn et al. (2014) to identify sub-clusters, and the gas envelope and low mass stars have been constructed to match this substructure.They perform a range of simulations varying the initial velocities of the stars and gas, the gas mass fraction and the galactic position of each cluster. We concentrate here on the fiducial simulations which are initialised in virial equilibrium and with no relative velocity between the sub-clusters.

4.3.2 RESULTS

All of the clusters were initially analysed with the $\mathcal{Q}^+$ algorithm, to constrain their fractal structure. When the clusters evolved to become centrally concentrated we then used the original $\mathcal{Q}$ parameter to constrain their radial density exponent. The fractal dimension of the DR21 fiducial simulation was consistently low (quite clumpy) until the sub-clusters merged at around 0.5 Myr. However, during that time, the number of levels and the density scaling exponent both decrease, suggesting that the stars from the smallest sub-clump(s) are spreading out to fill the larger hierarchy. The fractal parameters are given in Table 4.4. After the clumps merge at ~ 0.5 Myr, the region completely loses its fractal structure and is consistent with a centrally concentrated system. The central stellar density increases with time, as does the radial density gradient.

If we limit our analysis to only the high mass stars ($M > 0.83 M_\odot$) that were taken from the MYStIX observations, and exclude the low mass stars added for completeness, we can compare the structure that would theoretically be seen by an observer to the full structure in the simulations. These results are shown in Figure 4.9. For the initial conditions of DR21, this mass limited analysis gave a lower fractal dimension and density scaling exponent than analysis of the full cluster, but a similar

Table 4.4. Fractal parameters recovered for the simulation of the evolution of DR 21. After 0.5 Myr the fractal structure is lost and the system becomes centrally concentrated.

Timestep (Myr)	$\mathcal{D}$	$\mathcal{L}$	$\mathcal{C}$
0.0	1.6 ± 0.0	8.0 ± 0.0	20 ± 0
0.1	1.6 ± 0.0	8.0 ± 0.0	20 ± 0
0.2	1.6 ± 0.0	8.0 ± 1.0	20 ± 0
0.3	1.6 ± 0.0	7.7 ± 0.0	20 ± 0
0.4	1.6 ± 0.0	7.0 ± 0.0	3 ± 10
0.5	1.6 ± 0.0	7.0 ± 0.0	3 ± 10

number of levels. This suggests that the observable higher mass stars populate a more clumpy distribution, but are not confined to the smallest sub-clumps. Their evolution showed the same trend as the full cluster. The high mass stars in M17 and NGC 6357 exhibit similar behaviour; the initial conditions have a low fractal dimension and high number of levels, but the substructure is quickly erased (within 1-2 Myr) and the cluster becomes centrally concentrated. We note that when the full clusters are analysed, M17 would not be classified as fractally sub-structured and NGC 6357 only just so. It is only when we analyse the high mass component of these two clusters that the sub-structure is reliably detected.

4.4 CONCLUSIONS

The analysis of observed clusters shows that real star clusters have structure across a large range of the parameter space modelled in this work, from highly sub-structured to smooth or centrally concentrated, and with a large difference in scales. Analysis of the four clusters studied in CW04 has confirmed that the $\mathcal{Q}^+$ method detects the same trends in structure as previously detected, although the estimates of fractal dimension are somewhat lower. Further investigations of observed data sets may reveal interesting similarities and differences between clusters in different environments that could constrain the underlying physical processes.

The quantitative analysis of simulation results enables a robust measurement of the effects of changing conditions as well as direct comparison to observations. The analysis of clusters produced in slow cloud-cloud collisions are fractally substructured, but faster collisions produce centrally concentrated clusters. While synthetic observations are invaluable in connecting simulation results to observable effects, the analysis of raw simulation data is also a useful tool for model testing and deeper understanding of the complex star formation process.

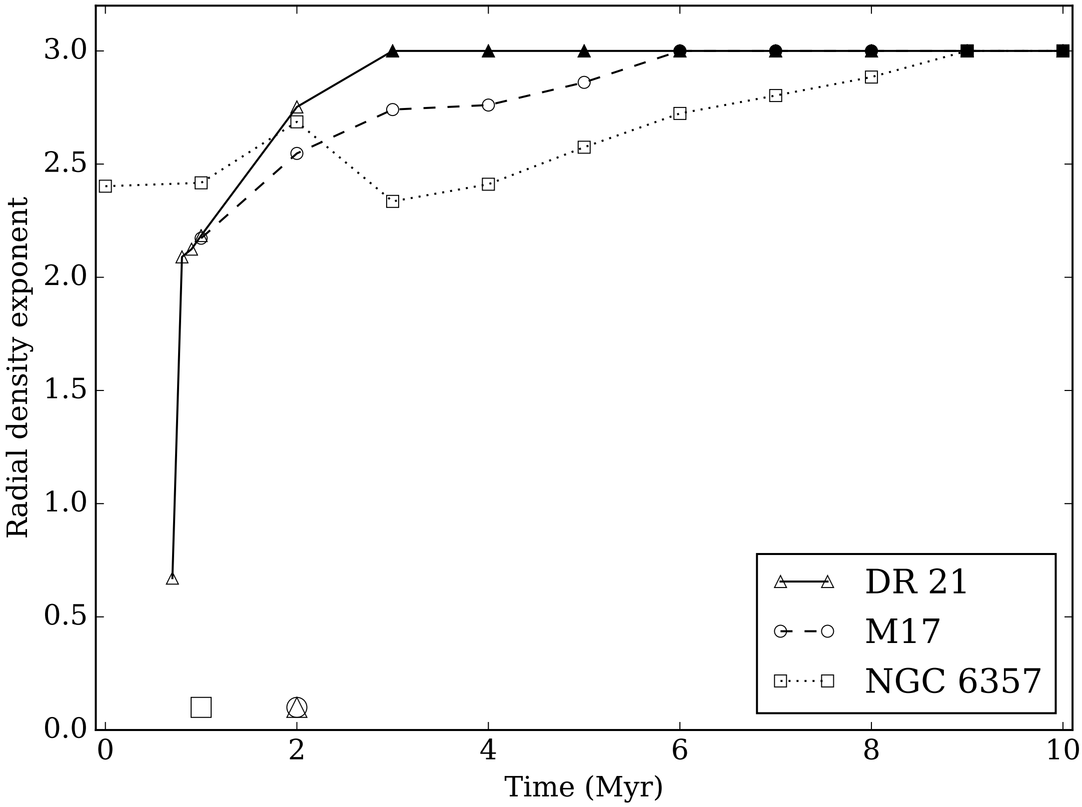

Figure 4.9. Figure showing the evolution of the structure over time in simulations of DR 21, M17 and NGC 6357. We analyse only the massive stars taken from the MYStIX observations, as this should be similar to what observers would be able to detect. Filled symbols indicate lower limits ($\alpha \geq 3$) as the analysis we perform cannot estimate radial density exponents greater than 3. The larger symbols on the x-axis indicate the time at which each cluster has lost its fractal substructure.

The evolution of the fractal parameters of young embedded clusters in simulations shows that the substructure with which clusters are born is very quickly dynamically erased, usually within 1 Myr. They then become increasingly centrally concentrated. This fits with the observationally informed picture that younger clusters are more substructured than older clusters, but investigation of a broader range of initial conditions would be needed to investigate how and when substructure can survive dynamical evolution.

CHAPTER 5

$\mathcal{Q}^{+}$ ADAPTED TO CONTINUUM IMAGES

The adaptation of the $\mathcal{Q}^{+}$ algorithm to analyse continuum images is discussed and tested. Column density maps of star forming regions are transformed into a statistically similar distribution of points and analysed to measure the fractal structure of star forming regions.

5.1 Introduction

The $\mathcal{Q}^+$ algorithm is designed to statistically constrain the 3D structure of a set of points from a 2D projection, and quantify the presence of sub-structure in terms of three parameters:

- the fractal dimension, $\mathcal{D}$ - a measure of the clumpiness or smoothness of the distribution;
- the number of levels, $\mathcal{L}$ - a measure of the range of scales of substructure;
- the density scaling exponent, $\mathcal{C}$ - a measure of the relative distribution of mass on the different scales.

This allows us to reliably constrain and compare structures in observed and simulated star clusters. However, the molecular clouds and cloud complexes in which star formation takes place also exhibit substructure, famously characterised by Larson's scaling relation (Larson, 1981),

$$M \propto R^2. \tag{5.1}$$

We assume that the hierarchical structure inherent in molecular clouds can also be described in terms of fractal geometry, and could therefore be analysed using $\mathcal{Q}^+$. The Larson scaling relation would be analogous, although not strictly identical, to a density scaling exponent of $\mathcal{C} = 1$.

Using this method to analyse both the molecular cloud structure and the clusters that form within it would enable unbiased comparison of structure between these two main phases of star formation. In Section 5.2 we describe the method to convert an image into a set of points, and in Section 5.2.1 we discuss how many points should be used to adequately represent the full cloud structure. Finally, in Section 5.3 we show how the $\mathcal{Q}^+$ algorithm can be applied to continuum data to analyse the structure in observations of the Galactic plane.

5.2 Turning clouds into points

The $\mathcal{Q}^+$ method is calibrated against star clusters created using the 'box fractal' method described in Section 3.2. Synthetic clouds created by simple Gaussian smoothing of these clusters should therefore give the most reliably similar cloud structure. We have tested this by smoothing the clusters to create a cloud and then converting the cloud back to a set of points (using the method discussed in Section 2.2.3), and reanalysing the results with the $\mathcal{Q}^+$ algorithm, to show that the statistically significant structure is maintained.

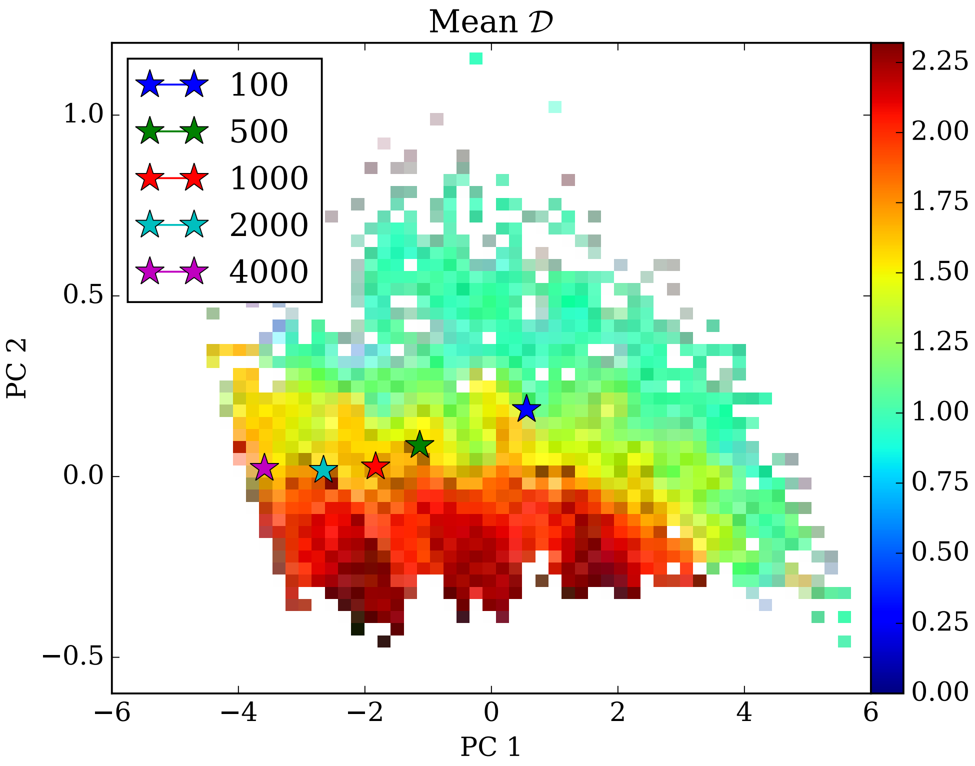

Figure 5.1. One fractal cloud decomposed into distributions with different N_{pts}, resulting in different locations in Principal Component space.

5.2.1 The number of points, N_{pts}

In the method explained is Section 2.2.3, the choice of how many points to place within the distribution is a free parameter. However, when analysing the data using the $\mathcal{Q}^+$ algorithm, the number of points can have a large effect on the parameters recovered. For example, Figure 5.1 shows the effect of analysing the same cloud $(\mathcal{D} = 1.58, \mathcal{L} = 6, \mathcal{C} = \infty)$ having converted it into different numbers of points. Lomax, Whitworth & Cartwright (2011) used an alternative method of decomposing an image into a set of points and advocated using $N_{pts} = N_{pixels}$. However, in the box fractal method of creating artificial star clusters, on which the $\mathcal{Q}^+$ algorithm is based, each particular **Y**-state $(\mathcal{D}, \mathcal{L}, \mathcal{C})$ is populated by a set number of points, as we scale all densities to the smallest scale which is constrained to have one star per fertile leaf cube. There is a small variation due to imposing a spherical envelope onto the cube structure, which slightly varies the final number of stars.

To test this we create synthetic fractal clouds with varied $\mathcal{D}$, $\mathcal{L}$, and $\mathcal{C}$ around the fiducial values $(\mathcal{D}, \mathcal{L}, \mathcal{C}) = (1.58,\ 5,\ 2)$. In this case we use a lower fiducial density scaling exponent than when this method is applied to clusters, because clouds are

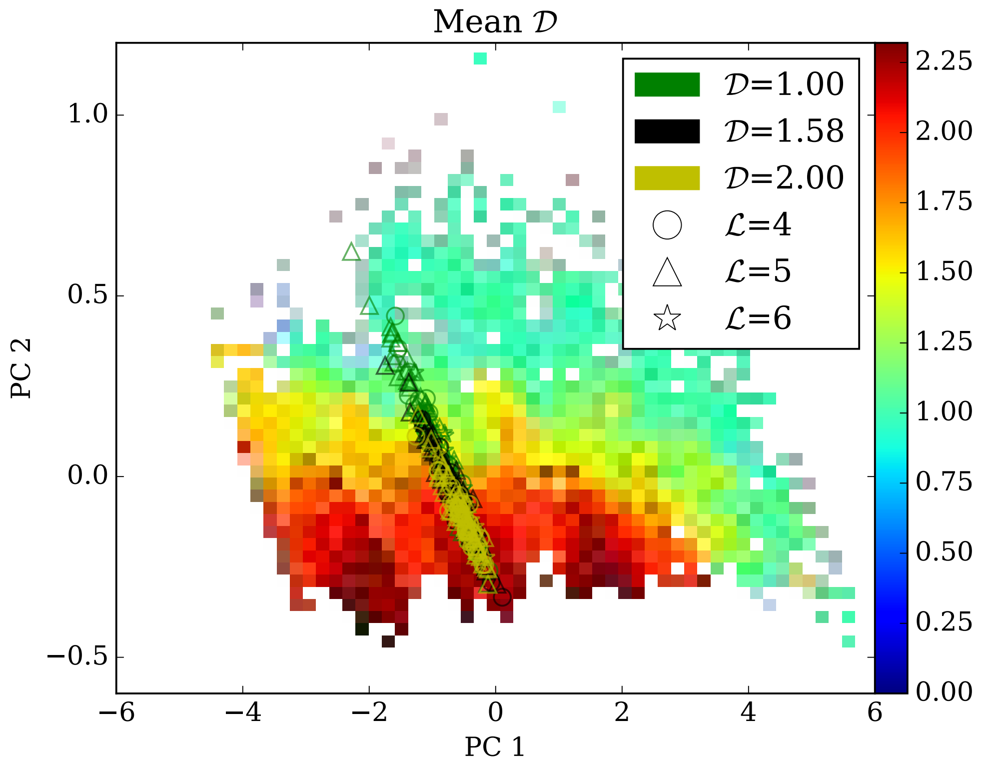

Figure 5.2. The range of fractal clouds with varying parameters when populated with 400 points in Principal Component space. The different symbols denote the number of levels of the input distribution and the colour denotes the fractal dimension. The scattered points also include varied density scaling exponent but this is not shown as it displays no clear trend.

expected to have $\mathcal{C} = 1$ from Larson's scaling law (Larson, 1981). We create 2D images each with 20 x 20 pixels and then convert each of these into points distributions with $N_{pts} = N_{pixels} = 400$. We can see in Figure 5.2 that the recovered parameters are confined to a thin strip of Principal Component space, corresponding to combinations of $\mathcal{D}$ and $\mathcal{L}$ that give $N \approx 400$. This severely limits the parameters that can be recovered and is therefore not suitable to recover the true fractal structure of the clouds.

An alternative approach is to scale the number of points proportional to the mass of the cloud. For example, the synthetic clouds constructed using the 'box fractal' method have masses in the range $10^2 - 10^4$ $\mathrm{M}_{\odot}$. The artificial clusters used to calibrate the $\mathcal{Q}^{+}$ method contained $10^1 - 10^3$ points. We therefore assume each point represents 10 $\mathrm{M}_{\odot}$, to produce a number of points that fits within the expected distribution. This method does not confine the results to a subset of $(\mathcal{D}, \mathcal{L}, \mathcal{C})$ values, and gives a better estimate of the input parameters of the artificial cloud. This is demonstrated in Table 5.1, which shows the ratio between the parameters used to construct the cloud and the parameters recovered by the $\mathcal{Q}^+$ algorithm. This method tends to overestimate the density scaling parameter $\mathcal{C}$, but gives a reliable estimate of the fractal dimension, $\mathcal{D}$, and number of levels, $\mathcal{L}$.

$\mathcal{D}_{\text{IN}} =$	1.00								
$\mathcal{L}_{\text{IN}} =$	4			5			6		
$\mathcal{C}_{\text{IN}} =$	1	2	3	1	2	3	1	2	3
$\mathcal{D}_{\text{OUT}}/\mathcal{D}_{\text{IN}} =$	1.91±0.23	1.29±0.15	1.04±0.08	2.16±0.10	1.45±0.12	1.09±0.08	Not populated	1.42±0.24	1.14±0.10
$\mathcal{L}_{\text{OUT}}/\mathcal{L}_{\text{IN}} =$	1.01±0.13	0.86±0.10	1.00±0.14	0.90±0.06	0.77±0.08	0.76±0.07	Not populated	0.79±0.14	0.65±0.09
$\mathcal{C}_{\text{OUT}}/\mathcal{C}_{\text{IN}} =$	11.11±8.90	4.14±2.47	2.28±1.84	7.50±8.19	5.04±3.51	1.56±0.69	Not populated	3.98±3.94	1.90±1.49
$\mathcal{D}_{\text{IN}} =$	1.58								
$\mathcal{L}_{\text{IN}} =$	4			5			6		
$\mathcal{C}_{\text{IN}} =$	1	2	3	1	2	3	1	2	3
$\mathcal{D}_{\text{OUT}}/\mathcal{D}_{\text{IN}} =$	1.25±0.07	0.99±0.13	0.87±0.16	1.24±0.07	1.18±0.08	0.96±0.13	Not populated	1.20±0.15	1.11±0.17
$\mathcal{L}_{\text{OUT}}/\mathcal{L}_{\text{IN}} =$	1.03±0.05	0.88±0.09	0.86±0.12	1.03±0.07	0.84±0.06	0.82±0.13	Not populated	0.84±0.09	0.76±0.10
$\mathcal{C}_{\text{OUT}}/\mathcal{C}_{\text{IN}} =$	5.84±5.42	2.48±1.77	2.10±1.89	4.80±5.19	4.53±3.97	2.80±2.23	Not populated	5.21±3.99	2.08±2.11
$\mathcal{D}_{\text{IN}} =$	2.00								
$\mathcal{L}_{\text{IN}} =$	4			5			6		
$\mathcal{C}_{\text{IN}} =$	1	2	3	1	2	3	1	2	3
$\mathcal{D}_{\text{OUT}}/\mathcal{D}_{\text{IN}} =$	1.12±0.04	0.95±0.10	0.96±0.11	1.10±0.07	1.03±0.06	1.06±0.09	Not populated	1.10±0.04	1.01±0.13
$\mathcal{L}_{\text{OUT}}/\mathcal{L}_{\text{IN}} =$	1.04±0.05	0.98±0.09	0.91±0.09	1.08±0.08	0.95±0.07	0.85±0.05	Not populated	0.89±0.05	0.91±0.11
$\mathcal{C}_{\text{OUT}}/\mathcal{C}_{\text{IN}} =$	7.96±6.25	6.14±4.11	2.78±2.57	8.75±6.75	7.33±3.78	2.65±2.31	Not populated	2.84±2.06	2.24±2.26

Table 5.1. Means and standard deviations for the ratios between the statistical-state parameters used in the construction of synthetic fractal clouds, $Y_{i\,\text{IN}}$, and the statistical-state parameters, $Y_{i\,\text{OUT}}$, derived from the representative points distribution with $N_{pts} = 0.1M_{TOT}$. States with $\mathcal{L} = 6$ and $\mathcal{C} = 1$ are not populated as these clusters have were too computationally expensive to produce at this stage.

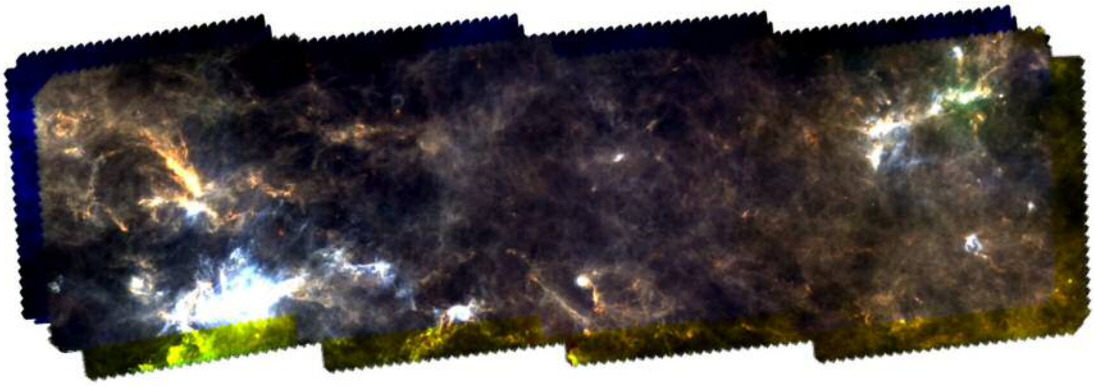

Figure 5.3. HiGAL tiles in the third galactic quadrant, taken from Elia et al. (2014).

5.3 APPLYING $\mathcal{Q}^{+}$ TO OBSERVATIONS

The Hi-GAL survey provides detailed images of the Galactic plane in 5 wavelength bands. We select 4 tiles in the third galactic quadrant, with longitudes between 217° and 224°, shown in Figure 5.3. Table 5.2 gives the results of the $\mathcal{Q}^{+}$ algorithm on these column density maps. The number of levels is reasonably consistent across the tiles. The central two tiles (ℓ222 and ℓ220) have lower fractal dimensions, indicating a more clumpy structure than the outer tiles (ℓ224 and ℓ217). At first glance this seems contrary to a visual inspection of the maps, as our eye is drawn to the bright substructures prominent in the outer tiles. However, the density scaling exponent of the outer tiles is very high, indicating that the mass is concentrated in the smallest levels of the hierarchy, while in the inner two tiles the density scaling exponent is lower, indicating that the mass is more evenly distributed on all levels.

Elia et al. (2014) (hereafter E+14) used these observations to examine the structure of the ISM. E+14 used the Δ-variance method (see Section 2.3.4) to infer a fractal dimension, $\mathcal{D}_{\Delta \mathrm{v}}$, for each of the 4 emission maps at different wavelengths (160, 250, 350, and 500 μm) for the 4 tiles across this longitude range, as well as the column density maps. Their results for the column density maps are also presented in Table 5.2. It should be noted once again that the fractal dimension measured by $\mathcal{Q}^{+}$ varies from 1 (clumpy) to 3 (smooth), while $\mathcal{D}_{fBm}$ varies from 3 (clumpy) to 2 (smooth). We therefore expect an inverse correlation. However, we do not find any clear correlation between the fractal dimensions measured using the two methods, as shown in Figure 5.4.

Tile :	$\ell = 224$	$\ell = 222$	$\ell = 220$	$\ell = 217$
$\mathcal{D} =$	2.2 ± 0.2	1.8 ± 0.2	2.1 ± 0.2	2.3 ± 0.0
$\mathcal{L} =$	5.5 ± 0.5	5.4 ± 0.5	5.0 ± 0.0	5.0 ± 0.0
$\mathcal{C} =$	10.0 ± 0.0	2.0 ± 0.0	2.0 ± 0.0	3.3 ± 0.0
$\mathcal{D}_{\Delta\mathrm{v}}$	2.61	2.64	2.83	2.93

Table 5.2. Parameters estimated for the HiGAL column density maps. The fourth row gives the results quoted in E+14, using the Δ-variance method. This measure of fractal dimension ranges from 2 (smooth) to 3 (clumpy), i.e. in the opposite direction to the $\mathcal{Q}^+$ definition of $\mathcal{D}_{\mathcal{Q}^+}$.

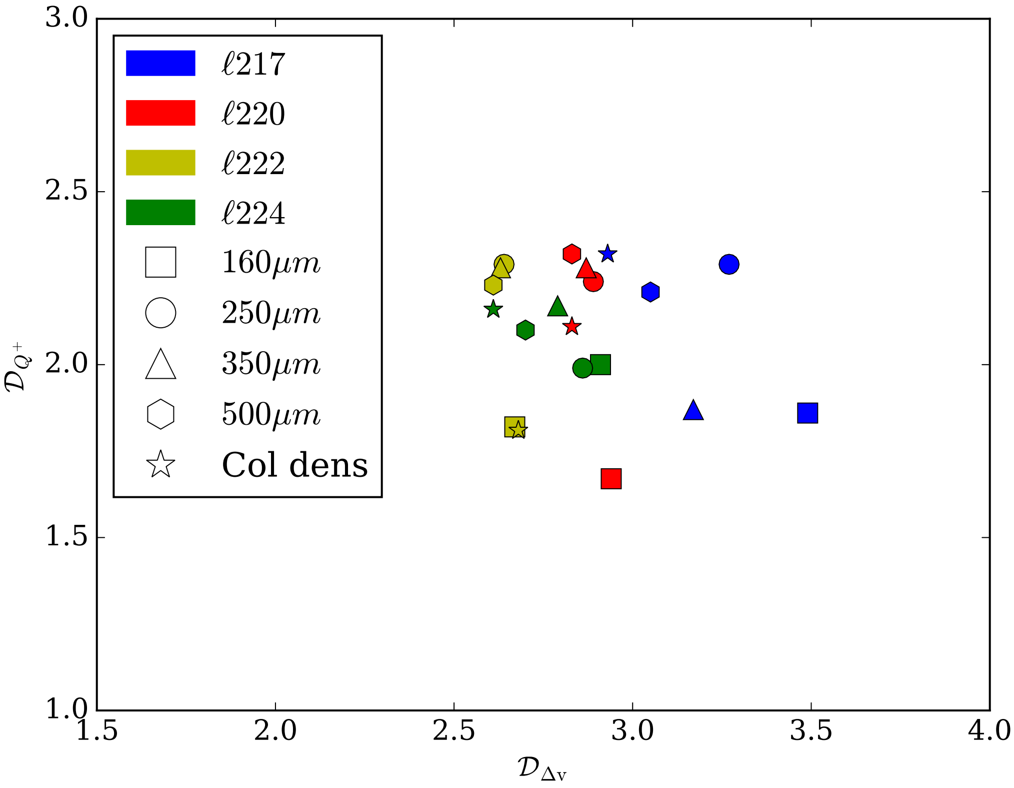

Figure 5.4. A comparison of the fractal dimension measured by the Δ-variance method (E+14) and the $\mathcal{Q}^+$ algorithm for 4 Hi-GAL tiles. The colour indicates the approximately central longitude of the tile and the symbol indicates whether the analysis used the *Herschel* image at a particular wavelength or the column density estimated from them.

5.4 CONCLUSIONS

The $\mathcal{Q}^+$ algorithm has previously been shown to robustly quantify the fractal substructure of star clusters in terms of three parameters: the fractal dimension, $\mathcal{D}_{\mathcal{Q}^+}$; the number of levels, L; and the density scaling exponent, $\mathcal{C}$. Molecular clouds also display substructure and we expect the they would impart this structure to the clusters they spawn. We therefore test methods of applying this same algorithm to continuum data to enable a direct comparison of the fractal properties of molecular clouds and young star clusters.

We describe a method for converting a continuum image into a statistically representative set of points by using it as a probability density function. We discuss the number of points that should be used as this strongly affects the range of parameters that can be recovered. We conclude that scaling the number of points by the total mass of the clouds being studied gives the best estimate of the fractal parameters.

We apply this analysis to four tiles of the *Herschel* Hi-GAL survey, testing the individual intensity maps at different wavelengths as well as the derived column densities. These tiles have previously been studied using an alternative method of calculating a fractal dimension, but we find no clear relation between the fractal dimensions found in this analysis. However, different methods of calculating fractal dimensions are known to produce different results. We hope that future analysis of a wider sample will illuminate these differences.

Chapter 6

J plots: characterising structures in the ISM [*]

An astrophysical image can be segmented into contiguous structures based on isodensity contours using dendrograms. The shape of these structures can then be analysed using the moments of inertia about their principal axes. This separates and quantifies elongated, hollow, and centrally concentrated structures, enabling us to classify the structures from the dendrogram and identify astrophysically interesting objects. This method is illustrated by examining the well known RCW 120 bubble in Hi-GAL column density maps, and quantifying the effect of turbulent accretion on the structure of filaments in simulations. We discuss the extension of this method to three dimensional data sets, both from simulations in position-position-position space and from observations or radiative transfer modelling in position-position-velocity space.

---

[*] The work presented here is published in Jaffa, Whitworth, Clarke & Howard, MNRAS, 2018, *in press*

6.1 INTRODUCTION

The multi-phase interstellar medium (ISM) is a chaotic environment in which many physical processes interact over a large range of scales. These processes sculpt the ISM into a variety of structures, and many of these structures play important roles in determining and/or revealing the locations of star formation. For example, *Herschel* observations have shown that elongated filamentary structures are a common feature in molecular clouds, in both low- and high-mass star forming regions, and also in regions that are not actively forming stars (Molinari et al., 2010; André et al., 2010; Hennemann et al., 2012; Palmeirim et al., 2013). Further studies have estimated that 60% to 75% of bound prestellar cores lie on filaments whose line mass is greater than the critical value for the onset of fragmentation (André et al., 2010; Könyves et al., 2015; Marsh et al., 2016). Feedback processes, from small scale stellar winds up to individual or clustered supernovae, can create hollow shells or bubbles which show up in observations as dense rings, sometimes with an ionising source at the centre (Churchwell et al., 2006; Simpson et al., 2012; Ingallinera et al., 2016). Finally, gravitationally bound objects, from clouds to cores and even clusters of stars, are often characterised as being centrally condensed (Cartwright & Whitworth, 2004; Li et al., 2015; Rathborne et al., 2015; Storm et al., 2016; Jaffa, Whitworth & Lomax, 2017). Each of these different types of structure can coexist in astrophysical images, and are relevant to different parts of the star formation process. We therefore need methods to identify and quantify such structures, in order to compare their properties and build a coherent picture of the interaction and interdependence of stars and the ISM.

A variety of algorithms are used in star formation to segment astronomical images into regions of interest, as discussed in Section 1.1.1. In this work we construct dendrograms to segment images into an hierarchical set of contiguous structures, using the ASTRODENDRO[†] Python package. We then analyse these structures using their principal moments of inertia. We limit the discussion to the case of 2D data; in future work we hope to extend the method to 3D data, using it (a) to explore the relationship between structures identified in PPP and PPV data-cubes from simulations, and (b) to inform the interpretation of structures identified only in PPV data-cubes from observations.

In Section 6.2 we briefly describe the use of dendrograms for segmenting astronomical images into discrete structures; we explain how the principal moments of inertia can be combined to construct J moments, which distinguish structures

---

[†] The ASTRODENDRO package is documented at http://www.dendrograms.org/

according to their degree of central concentration (or central rarefaction) and their degree of elongation; and we illustrate the use of J plots to quantify some simple synthetic images. In Section 6.3 we give examples of the application of J plots to observed and simulated data. In Section 6.4 we discuss the extension of this method to three-dimensional data, and in Section 6.5 we summarise our main conclusions.

6.2 Methods

In this section we give a brief outline of how dendrograms are used to segment a pixelated 2D grey-scale image into hierarchically nested structures. Then we describe how these structures can be analysed using J moments and J plots. In order to make the discussion less abstract, we assume that the intensity of the 2D grey-scale image being analysed represents surface-density, Σ. Hence the grey-scale contours are surface-density thresholds. For a more comprehensive discussion of dendrograms, see Rosolowsky et al. (2008).

6.2.1 Image segmentation with dendrograms

Dendrograms represent the morphology of a greyscale image in terms of hierarchically nested contiguous structures at different surface-density thresholds. At the specified minimum surface-density threshold ('min_value' parameter in ASTRODENDRO), there is usually a single contiguous structure, which is termed the 'trunk' of the tree. Any structure lower than this threshold is considered part of the background and not included in further analysis. As the surface-density threshold is increased, the trunk splits into smaller 'branch' structures.[‡] A structure that does not split into smaller structures but reaches its peak surface-density as a contiguous whole is termed a 'leaf', and corresponds to a local maximum. Leaves are only retained if (i) they comprise a minimum number of pixels ('min_pixels' in ASTRODENDRO) and (ii) there is a minimum surface-density contrast between their peak surface-density and the surface-density threshold at which they merge with another leaf or branch ('min_delta' in ASTRODENDRO). Leaves that do not satisfy both these conditions are simply absorbed into a larger structure. Thus a dendrogram analysis divides any greyscale image into a set of structures defined by their surface-density thresholds,

---

[‡] The ASTRODENDRO package imposes binary mergers, meaning that if a large structure at surface-density threshold i splits into 3 smaller structures at surface-density threshold $i + \delta i$, additional intermediate thresholds are examined to find one at which the large structure splits into two structures, one of which will split again. Other dendrogram building methods such as that described in Storm et al. (2014) do not enforce binary mergers.

and results in a two dimensional 'tree' graph which summarises the connectivity of structures at different thresholds. See, for example, Figure 6.3 which shows the dendrogram corresponding to the Hi-GAL tile shown in Figure 6.2. In the following analysis we extract the structures defined by the dendrogram, and use moments to measure their degree of central concentration and their overall shape.

We must choose carefully the parameters of the dendrogram, so as to capture the structures we are interested in. Reducing the minimum surface-density threshold ('min_value') will increase the size of the trunk, until it encompasses the entire observed area. Conversely, increasing the minimum threshold may split the dendrogram into separate trees, i.e. multiple trunks. It is recommended that this value be set to some multiple of the noise level of the image, in order to exclude small structures caused by noise, but it can be set higher if we wish to isolate denser structures. The minimum number of pixels ('min_npix') and minimum surface-density contrast ('min_delta') can change what is defined as a leaf. Reducing either of these parameters results in more leaves, and their values should therefore also be chosen to reflect the noise level of the data set. The intermediate branches of the tree are not greatly affected by any of these parameters.

6.2.2 J MOMENTS

Any point on a dendrogram (be it on a leaf, branch or trunk) corresponds to a contiguous region $\mathcal{R}$ inside which the surface-density, $\Sigma(x, y)$, exceeds or equals some threshold, $\Sigma_{\mathcal{R}}$. We can therefore compute the total area, A, and mass, M, of the corresponding object,

$$A = \int_{\mathcal{R}} dx\,dy = P\,\Delta A\,, \tag{6.1}$$

$$M = \int_{\mathcal{R}} \Sigma(x,y)\,dx\,dy = \sum_{p=1}^{p=P} \{\Sigma_p\}\ \Delta A\,. \tag{6.2}$$

Here p is the dummy ID of a pixel inside $\mathcal{R}$, P is the total number of pixels inside $\mathcal{R}$, Σ_p is the surface-density in pixel p, and ΔA is the area of a pixel; we are assuming that all pixels have the same area.

We can also compute the principal axes of this object, $\hat{\mathbf{e}}_i$ $(i = 1, 2)$, and the associated principal moments, I_i (see Appendix E). By convention, if the two principal moments are different (which is generally the case), the first principal axis, $\hat{\mathbf{e}}_1$ is the one associated with the smaller principal moment. Hence $I_1 \leq I_2$, by construction.

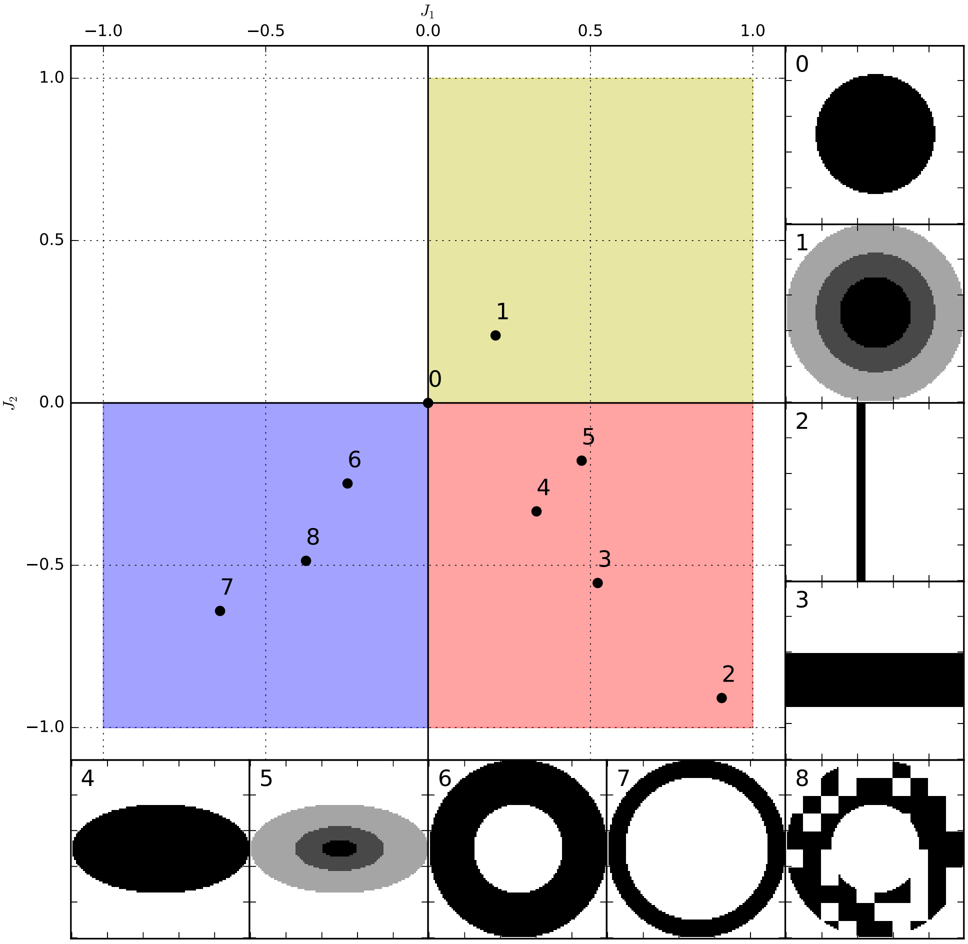

Figure 6.1. Some simple two dimensional test-structures and their positions on the J plot. A circularly symmetric disc with uniform surface-density (structure 0) occupies the centre of the plot, $J_1 = J_2 = 0$. Circularly symmetric, centrally concentrated discs (structure 1) occupy the top right quadrant, getting closer to $J_1 = J_2 = 1$ as their central concentration increases. Circularly symmetric, hollow structures (structures 6, 7 and 8) occupy the bottom left quadrant, getting closer to $J_1 = J_2 = -1$ as their thickness decreases. Elongated objects (structures 2, 3, 4 and 5) occupy the bottom right quadrant, getting closer to $(J_1, J_2) = (1, -1)$ as their aspect ratio increases. Nothing falls in the top left quadrant, as this would require $I_1 > I_2$.

Now consider the very simple case of an infinitesimally thin circular region (a disc) with radius R, and uniform surface-density $\Sigma_{\rm O}$, hence area $A = \pi R^2$ and mass $M = \pi R^2 \Sigma_{\rm O}$. Because of symmetry, the principal axes can be any pair of orthogonal axes, and without loss of generality we choose the Cartesian axes, i.e. $\hat{\mathbf{e}}_1 \to \hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_2 \to \hat{\mathbf{e}}_y$. The corresponding principal moments are

$$I_1 = I_2 \;=\; \frac{1}{2} \int\limits_{r=0}^{r=R} r^2\, \Sigma_{\rm O}\, 2\pi r dr \;=\; \frac{\pi R^4 \Sigma_{\rm O}}{4} \;=\; \frac{AM}{4\pi}, \tag{6.3}$$

where we have invoked the Perpendicular Axis Theorem.

Next, whilst maintaining circular symmetry, change the radial profile of the surface-density, $\Sigma(r)$, holding the area, A, and mass, M, constant. If the surface-density profile is made more centrally concentrated (e.g. a centrally condensed 3D core seen in projection), I_1 and I_2 are decreased, and therefore $I_1 = I_2 < AM/4\pi$. Conversely, if the surface-density profile is altered so it has a central rarefaction (e.g. a ring or 3D limb-brightened bubble), I_1 and I_2 are increased, and therefore $I_1 = I_2 > AM/4\pi$.

Finally, revert to a uniform surface-density, and again hold the area, A, and mass, M, constant, but stretch the disc into a very eccentric ellipse (i.e. a filament). This has the effect of reducing I_1 and increasing I_2, so we have $I_1 < AM/4\pi < I_2$.

Now, if we introduce

$$I_0 \;=\; \frac{AM}{4\pi}\,, \tag{6.4}$$

and define the J moments

$$J_i \;=\; \frac{I_0 - I_i}{I_0 + I_i}\,, \qquad i = 1, 2\,, \tag{6.5}$$

the uniform surface-density disc has $J_1 = J_2 = 0$; the centrally concentrated disc (3D core in projection) has $J_1 = J_2 > 0$; the centrally rarefied disc (ring or 3D limb-brightened bubble) has $J_1 = J_2 < 0$; and the eccentric ellipse (filament) has $J_1 > 0,\ J_2 < 0$. Thus, if we plot J_1 against J_2, centrally concentrated objects occupy the top right quadrant, centrally rarefied objects occupy the bottom left quadrant, and elongated objects occupy the bottom right quadrant.

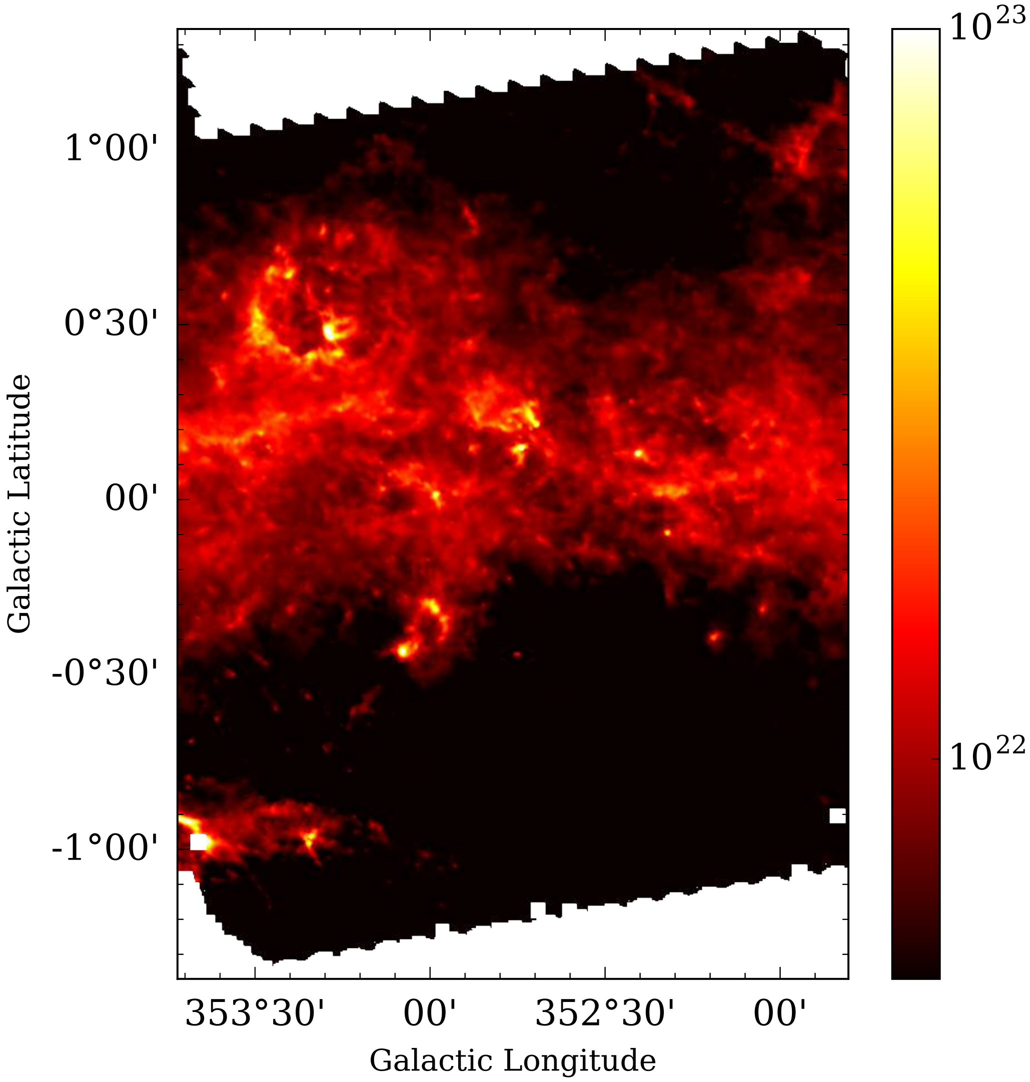

Figure 6.2. Hi-GAL surface-density tile ℓ347. The colour bar gives surface-density, on a logarithmic scale, and in units of $H_2\,cm^{-2}$.

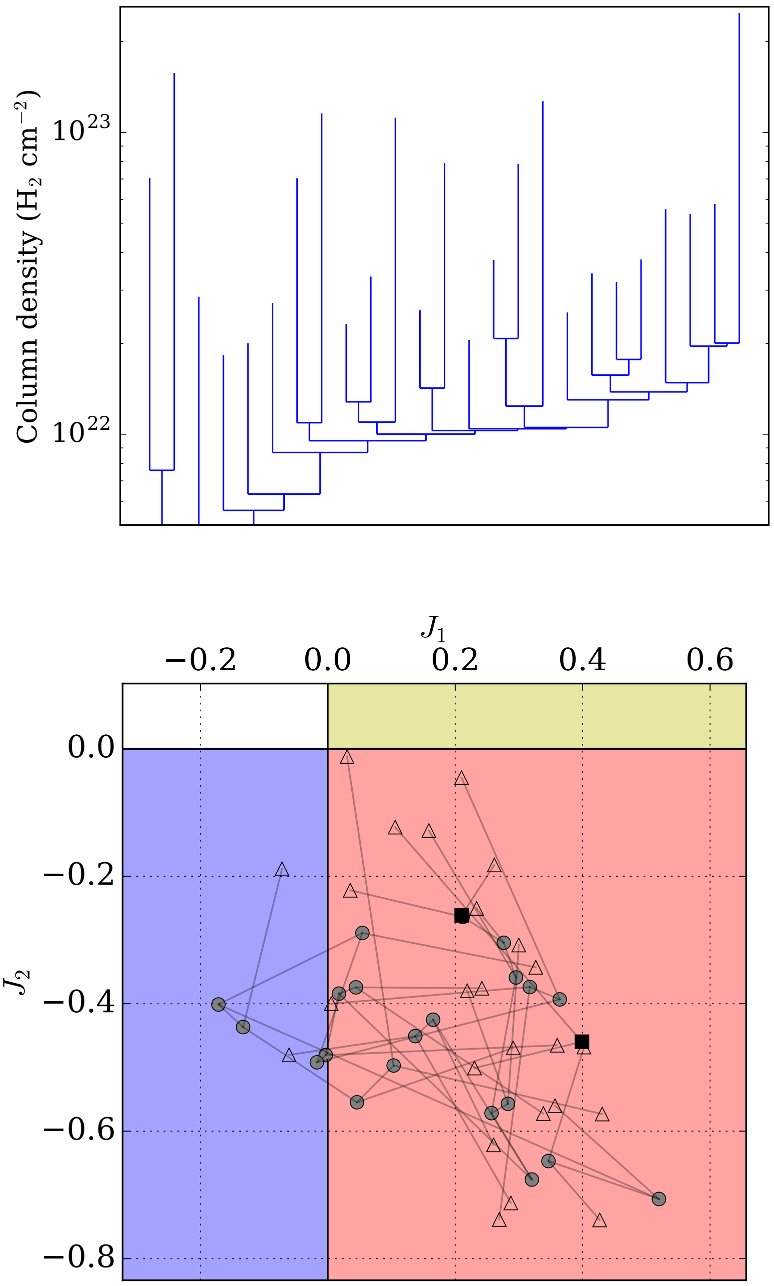

Figure 6.3. *Top:* the dendrogram of the HI-GAL tile shown in Figure 6.2 constructed using the parameters specified in the text. *Bottom:* the J plot of the identified structures. Leaves are shown as triangles, branch structures are shown as circles, and the trunks are shown as squares. Grey lines show which structures merge to form larger structures. Note that this image is focused on only the portion of the J plot occupied by the detected structures. To aid orientation, the background colouring is identical to that in Figure 6.1.

6.2.3 J PLOTS

Figure 6.1 shows a J plot of some simple test cases; filled circles mark the locations of the pixelated test-structures that are illustrated in the margin. The circle with uniform surface-density (structure 0) occupies the origin ($J_1 = J_2 = 0$). The circularly symmetric, centrally concentrated disc (structure 1) occupies the upper right quadrant ($J_1 = J_2 > 0$) as the principal moments are equal, $I_1 = I_2$, but both less than I_0; such discs move further from the origin with increasing central concentration. The circularly symmetric hollow rings (structures 6 and 7) occupy the lower left quadrant ($J_1 = J_2 < 0$) as their principal moments are equal, $I_1 = I_2$, but both greater than I_0; such rings move further from the origin with decreasing thickness (decreasing $\Delta R/R$). Elongated straight structures with uniform surface-density (structures 2, 3, and 4) occupy the lower right quadrant ($J_2 < 0 < J_1$) as $I_1 < I_0$ and $I_2 > I_0$; elongated structures move further from the origin with increasing aspect ratio. Nothing falls in the top left quadrant, or indeed above the line $J_2 = J_1$, since this would imply $I_2 < I_1$, in violation of the definition given earlier.

Structures that do not fit perfectly into one of these categories can still be represented usefully on the J plot. For example, the broken ring (structure 8) still falls in the lower left quadrant of the J plot. And the centrally concentrated ellipse (structure 5) falls between the locus for elongated structures with uniform surface-density and the locus for circularly symmetric, centrally concentrated structures, indicating that it has elements of both.

A single dendrogram structure (leaf, branch or trunk) spans a range of surface-densities, and in general its J moments will depend on the surface-density at which we analyse it. For the structures in Figure 6.1 we have treated each structure in its entirety, which is equivalent to analysing it at its base, i.e. at the lowest surface-density before it merges with another structure. However, we can also look at how the J moments of a dendrogram structure migrate on the J plot as the surface-density threshold is changed, as discussed in Section 6.3.2. For example, the circularly symmetric, centrally concentrated structure on Figure 6.1 (structure 1) would become less centrally concentrated and migrate towards the origin if we were to analyse it at higher and higher surface-density thresholds; the mapping of surface-density onto this path would depend on the details of the surface-density profile.

6.2.4 OBSERVATIONAL EFFECTS

We accept that observations of molecular cloud structures are imperfect, and there are several observational effects that should be considered when applying J plots to observed images. As we can only observe the two-dimensional projection on the sky, we may be seeing the overlapping structure of several clouds along the line of sight, especially if we are observing close to the plane of the galaxy. In molecular line observations, the radial velocity can be used as a proxy for the third spatial dimension to separate cloud structures (e.g. Colombo et al., 2016), but this assumes that each cloud is a relatively quiescent structure all moving as one. For dust continuum observations, the temperature or dust opacity index can be recovered using PPMAP (Marsh et al., 2017).

Interferometers filter out large-scale structure, so should be combined with single-dish data of the same region in order to analyse the full cloud structure. Furthermore, Alves, Lombardi & Lada (2017) highlighted the importance of the limited field of view, in that statistics should only be derived for values above the last closed contour of an image. In the dendrogram algorithm this can be set to be the min_value parameter so that the structure below this limit is not included. The min_pixels parameter also limits the smallest scale analysed, which should be related to the resolution limit of the observations to avoid spurious small structures caused by noise.

6.3 RESULTS AND DISCUSSION

6.3.1 BUBBLES IN OBSERVATIONS

The *Herschel* Hi-GAL survey has delivered an unprecedented view of the Galactic plane in five wavebands between $70\,\mu$m and $500\,\mu$m (Molinari et al., 2016), and the images in the different wavebands have been smoothed to a common resolution and fit with SEDs to obtain surface-density maps. As a simple example of the application of J plots to observational images, we take the 'tile' from the Hi-GAL survey shown in Figure 6.2, which contains the well known RCW 120 bubble centred near $\ell = 353.20^{\circ}$, $b = 0.30^{\circ}$. This HII region has been described as 'the perfect bubble' (Deharveng et al., 2009), and much work has been done studying the physical conditions in and around the bubble (Anderson et al., 2010; Pavlyuchenkov, Kirsanova & Wiebe, 2013; Rodón et al., 2015), and the likelihood of triggered star formation in the swept up shell (Motte et al., 2010; Walch et al., 2015; Figueira et al., 2017). We therefore use this well characterized bubble as a test case to see how it appears on the J plot.

To construct the dendrogram we use a minimum surface-density threshold ('min_value') of 5×10^{21} $\mathrm{H_2\ cm^{-2}}$, a minimum surface-density contrast ('min_delta') of 1×10^{22} H_2 cm^{-2}, and a minimum number of pixels (‘min_npix’) of 200. These values are deliberately chosen to be quite high, since we are looking for large, clearly defined structures. The value of the minimum surface-density threshold results in two separate trunks. In total the dendrogram contains 48 structures (trunks, branches and leaves).

For each dendrogram structure, we consider the lowest surface-density threshold (i.e. the surface-density at which it merges with another structure), and compute J_1 and J_2. The bottom panel of Figure 6.3 shows the J plot for this data set, focusing on the region of the J plot that is actually occupied. We use different symbols on the J plot to distinguish different types of dendrogram structure: the highest level structures, leaves, are shown as triangles; intermediate branch structures are shown as circles; and the trunk, the lowest level structure, is shown as a square. Lines show where smaller high-level structures merge to form a larger lower-level structure. Most of the structures fall in the lower right quadrant, indicating that they are elongated. Six structures fall in the lower left quadrant, indicating that they are ‘ring-like’, although two are very close to the boundary and therefore not clearly defined as rings. The morphologies of the 4 structures clearly defined as rings are shown in Figure 6.4. We note that structure 45 is embedded in structure 38, which in turn is embedded in structure 24.

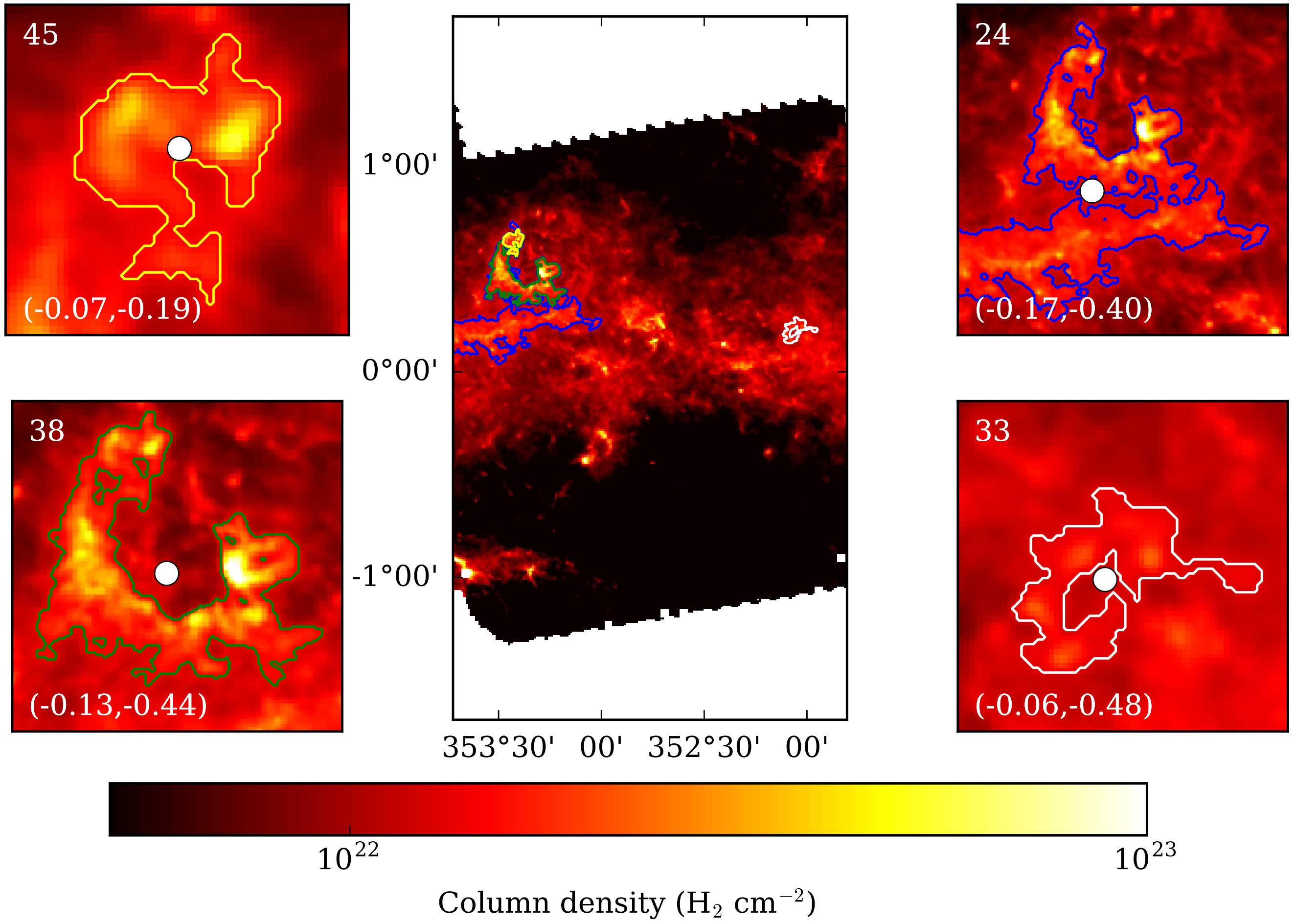

Figure 6.4. The four structures from Hi-GAL tile $\ell 347$ that are classified as rings, and their positions on the tile. In each panel, the structure ID is given in the top left corner, the J moments in the bottom left corner, and the centre of mass is marked by a white circle. We note that structure 45 is embedded in structure 38, which in turn is embedded in structure 24.

Structure 38 is the region identified as the RCW 120 bubble; in this surface-density image it is a partial ring with a lumpy boundary. Structure 45 is a small leaf structure at the top edge of the bubble wall. A visual inspection confirms a broken ring type morphology, but its astrophysical significance is unclear. Structure 24 is formed by the merging of the RCW 120 ring structure with a filamentary structure to the south. This creates an asymmetrical curved shape which the algorithm identifies as a ring, but a visual inspection would not readily identify this as a bubble candidate. Structure 33 is a collection of 4 cores that appear to be spaced regularly around a circle. These sources have been confirmed as compact sources by other surveys including the Hi-GAL point source catalogue (Elia et al., 2017) and the ATLASGAL dust condensation catalogue (Csengeri et al., 2014). However their unusually regular arrangement has not been noted previously. We can speculate that they might be a set of fragments formed out of a swept up ring or shell, but equally they might be at different distances and causally unrelated.

As demonstrated in Figure 6.1, the distance from the origin in the bottom left quadrant is associated with the thickness of the ring, although this is only an exact relation in the case of idealised test-structures. The J values, given on Figure 6.4 for each of the ring-like structures, support this interpretation: the thickest ring (structure 45) lies significantly closer to the origin than the others, and the thinnest (structure 33) lies furthest from the origin.

6.3.2 Filaments and cores in simulations

Sub-structure in filaments

Clarke et al. (2017) have performed SPH simulations of filaments forming by accreting from a turbulent medium. They perform a suite of simulations in which the initial turbulence is characterised either by purely compressive modes, or by a thermal mix of solenoidal and compressive modes (i.e. in the ratio 2:1), and they set the initial turbulent energy to be either subsonic or supersonic. For each case they perform ten simulations with different random seeds. Here we analyse the surface-density maps taken from these simulations to investigate how the fragmentation morphology appears on a J plot.

They find that when the turbulence is subsonic, filaments fragment into quasi-periodic cores and the mix of turbulent modes (purely compressive or thermal) does not significantly influence the fragmentation. When the turbulence is supersonic, the filament first fragments into numerous sub-filaments. This change in structure can be identified by eye in the three dimensional volume-density cubes, as well as in two

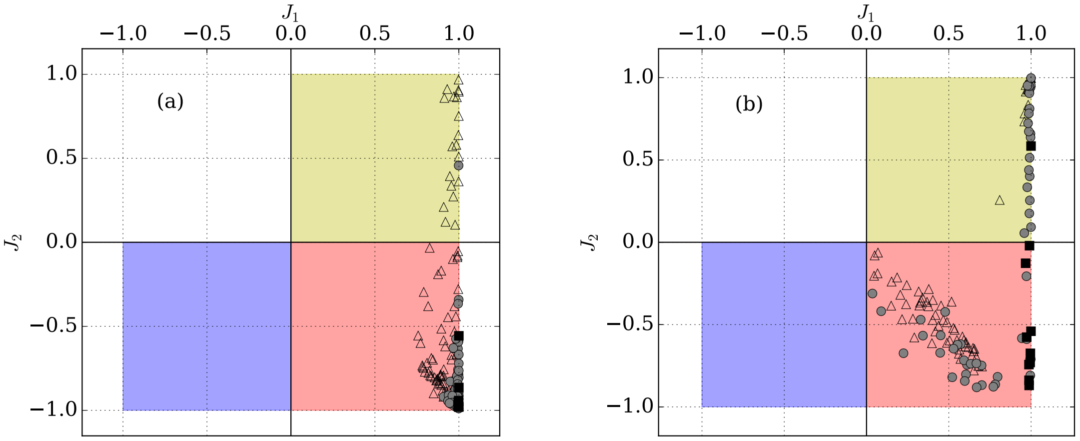

Figure 6.5. J plots of simulations of filament formation (Clarke et al., 2017) in which the turbulence in the accretion flow involves a thermal mix of modes, and is (a) subsonic or (b) supersonic. Each plot presents the results of 10 simulations, analysed at the end when 10% of the mass has been accreted into sink particles. The open triangles are leaves, the grey circles are branches, and the black squares are trunks.

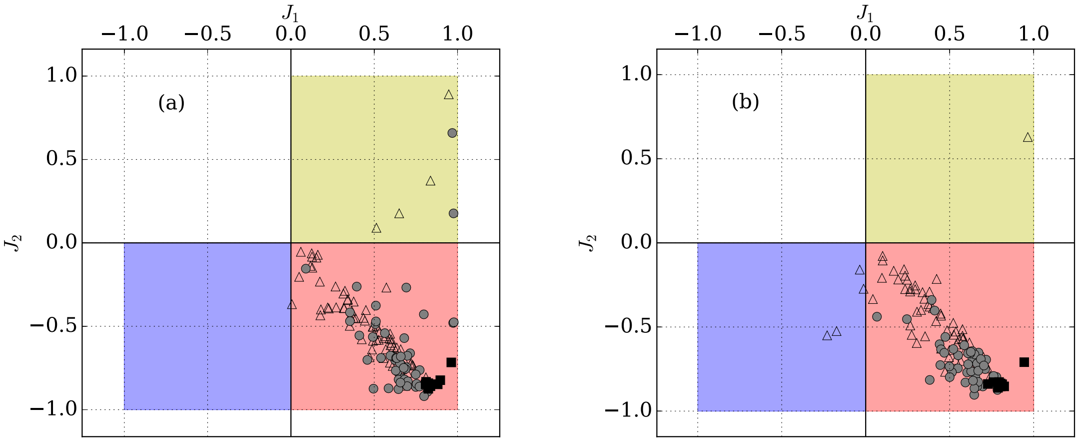

Figure 6.6. J plots for simulations of filament formation (Clarke et al., 2017) in which the turbulence in the accretion flow is supersonic, and involves (a) purely compressive modes, or (b) a thermal mix of modes. Each plot presents the results of 10 realisations, analysed after 0.5 Myr. The open triangles are leaves, the grey circles are branches, and the black squares are trunks.

dimensional projected surface-density images. Using J plots we can quantify this difference of morphology in an automated and objective way.

Figure 6.5 shows the J values of all the structures from the simulations of growing filaments in which the turbulence in the accreting material has a thermal mix of modes: (a) 10 simulations with subsonic turbulence, and (b) 10 simulations with supersonic turbulence. The dendrograms built on each surface-density image use the same parameters as Clarke et al. (2017), viz. min_value = $10^{-20}\,\mathrm{g\,cm^{-3}}$, min_pixels = 100 and min_delta = $5 \times 10^{-21}\,\mathrm{g\,cm^{-3}}$. The simulations with *subsonic* turbulence (Figure 6.5a) only create very long thin structures, similar in shape to the whole filament; the trunk of each dendrogram is shown as a black square in Figure 6.5, and the other structures (branches and leaves) are simply the spine of the filament broken up into smaller structures. In contrast, the simulations with *supersonic* turbulence (Figure 6.5b) create filamentary structures with a range of sizes and curvatures, filling more of the lower right quadrant on the J plot. These are the curved sub-filaments described in Clarke et al. (2017) and are significantly different in their morphology to the trunk structure which encompasses the whole filament.

This change in morphology is seen predominantly in the simulations that invoke a thermal mix of solenoidal and compressive turbulence; when the turbulence is purely compressive, the formation of sub-filaments is suppressed. Clarke et al. (2017) use dendrograms to quantify this. They find that in dendrograms built in three-dimensional PPP space using the volume-density distribution, the number of leaves in the dendrogram (N_{LEAVES}) is higher in the simulations using a thermal mix of turbulent modes (than those using purely compressive modes), indicating that the filaments are more sub-structured. The number of branches between the highest leaf and the trunk (N_{LEVELS}) is also higher in the simulations using a thermal mix of turbulent modes (than those using purely compressive modes), indicating that the sub-structure is more hierarchical. However, there is no significant difference in these measures if the dendrogram is built in two-dimensional PP space using surface-density images. We can use J plots to analyse the shape of structures in two dimensions. This can be applied to the structures extracted from the two-dimensional image, but also to projections of structures extracted in three dimensions. We can therefore quantify the shape of structures in the three dimensional data to test if the change in substructure also represents a change in morphology.

Figure 6.6 shows the J values of all structures from the simulations of growing filaments in which the turbulence in the accreting material is supersonic: (a) 10 simulations with purely compressive modes, and (b) 10 simulations with a thermal mix of modes. In both cases, the main population of structures is filamentary, but in

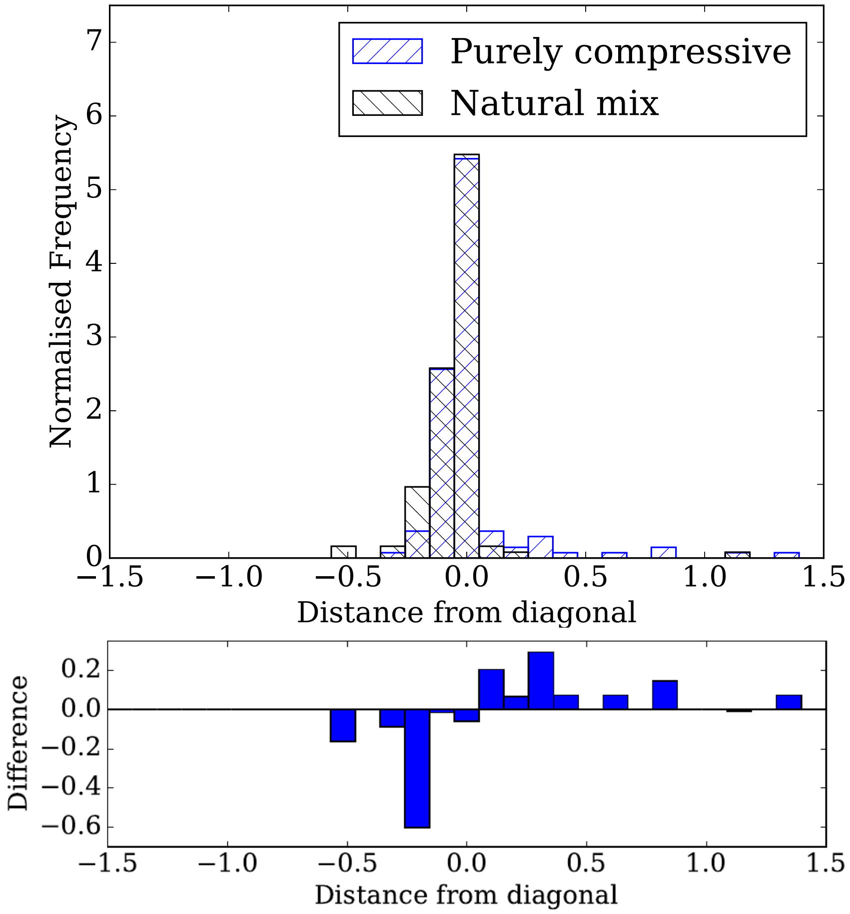

Figure 6.7. Histograms showing the perpendicular distance from each point in the J plot to the diagonal line $J_1 = -J_2$. *Top.* The black striped histogram shows the results obtained with thermal mix of turbulent modes, and the blue striped histogram shows the results obtained with purely compressive turbulent modes, both analysed at 0.5 Myr. *Bottom.* The difference between the two histograms (purely compressive minus thermal mix in each bin). This highlights the excess of structures above and to the right of the diagonal in the purely compressive case (more centrally concentrated structures), and the excess below and to the left of the diagonal in the thermal-mix case (more curved filamentary structures).

the case of purely compressive turbulent modes there are more centrally concentrated structures, while in the case of a thermal mix of turbulent modes there are more structures that are curved or even ring-like ($J_1 \leq 0$). We can illustrate this difference more clearly by looking at the distance of each point from the diagonal line $J_1 = -J_2$. This line represents the location in J space of a theoretical ‘perfect’ filament: an exactly symmetrical straight filament with uniform surface-density. Structures below and to the left of this line are more curved, and structures above and to the right of this line are more centrally concentrated. Figure 6.7 shows that, whilst both cases peak just below zero, the simulations with purely compressive turbulent modes have a significant positive tail of centrally concentrated structures, and the simulations with a thermal-mix of turbulent modes have a small negative tail indicating more curved structures. We can therefore use J plots to identify objectively, in two dimensions, differences in structure that were previously identified only by eye or in three dimensions.

Persistence of individual structures

Another way to characterise a dendrogram structure is to consider both the structure itself, and the smaller structures it contains – for example a parent-branch and all the branches and leaves at higher surface-density thresholds that are descended from the parent-branch. We can then analyse how the J moments change as the surface-density threshold is increased, from the minimum value at which the structure is first defined, to the maximum value in the densest leaf that it contains. In general, an increased J moment means that the mass is more concentrated towards the centre of mass along the corresponding axis, and a decreased J moment means that the mass is less concentrated towards the centre of mass along the corresponding axis.

In Figure 6.8 we illustrate the results obtained for two typical filaments, one in which the turbulence in the accretion flow is subsonic, and one in which it is supersonic; in both cases there is a thermal mix of turbulent modes. In the subsonic case (top panel of Figure 6.8), the filament is very narrow, straight, and continuous; a few small gaps appear at the highest surface-density threshold. In the supersonic case (bottom panel of Figure 6.8), the filament is much broader, bendier, and more fragmented; already, at a fairly low surface-density threshold, there are just two separate fragments.

Figure 6.9a shows how the J moments of the filaments in all ten *subsonic* cases evolve as the surface-density threshold is increased from $\sim 0.024\,\mathrm{g\,cm^{-2}}$ to $\sim 1.3\,\mathrm{g\,cm^{-2}}$. With one exception, the J moments remain very close to $(J_1, J_2) = (1, -1)$

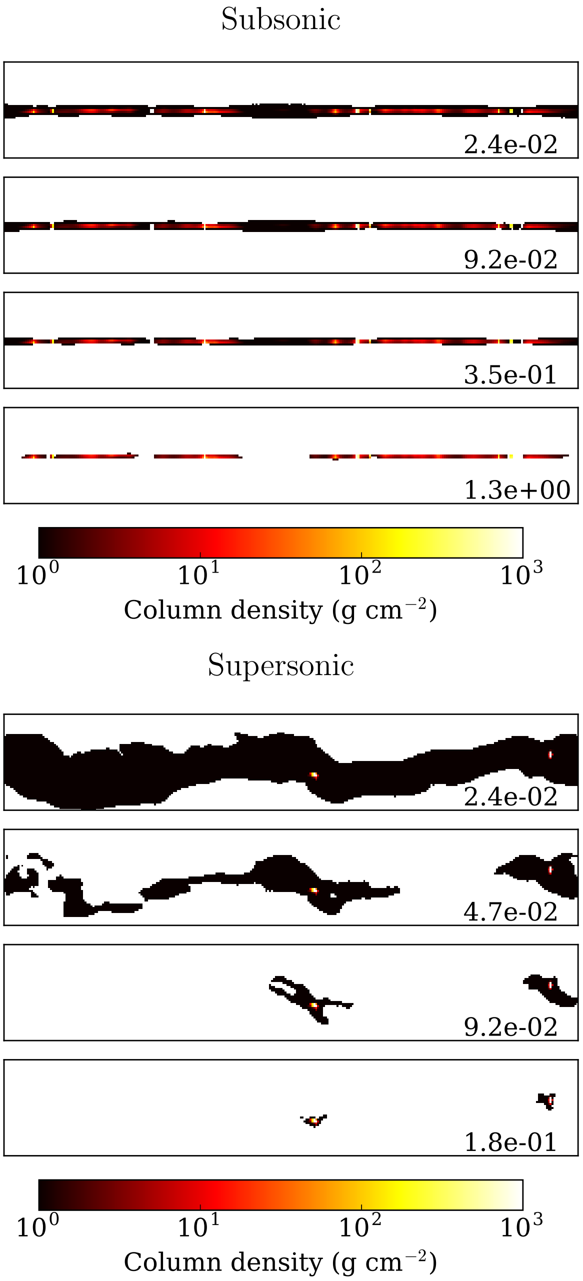

Figure 6.8. Filaments analysed at different density thresholds. *Top:* subsonic turbulence. *Bottom:* supersonic turbulence. The surface-density threshold is given in the bottom right corner of each image. Only four of the ten surface-density thresholds analysed in Figure 6.9 are illustrated. For the subsonic turbulence case we show the first, third, fifth and seventh thresholds. For the supersonic case we show the first, second, third and fourth thresholds. The filaments are 3 pc in length and the scales are the same in all images. The overall shape of the structures shown here is more important than the internal density variation.

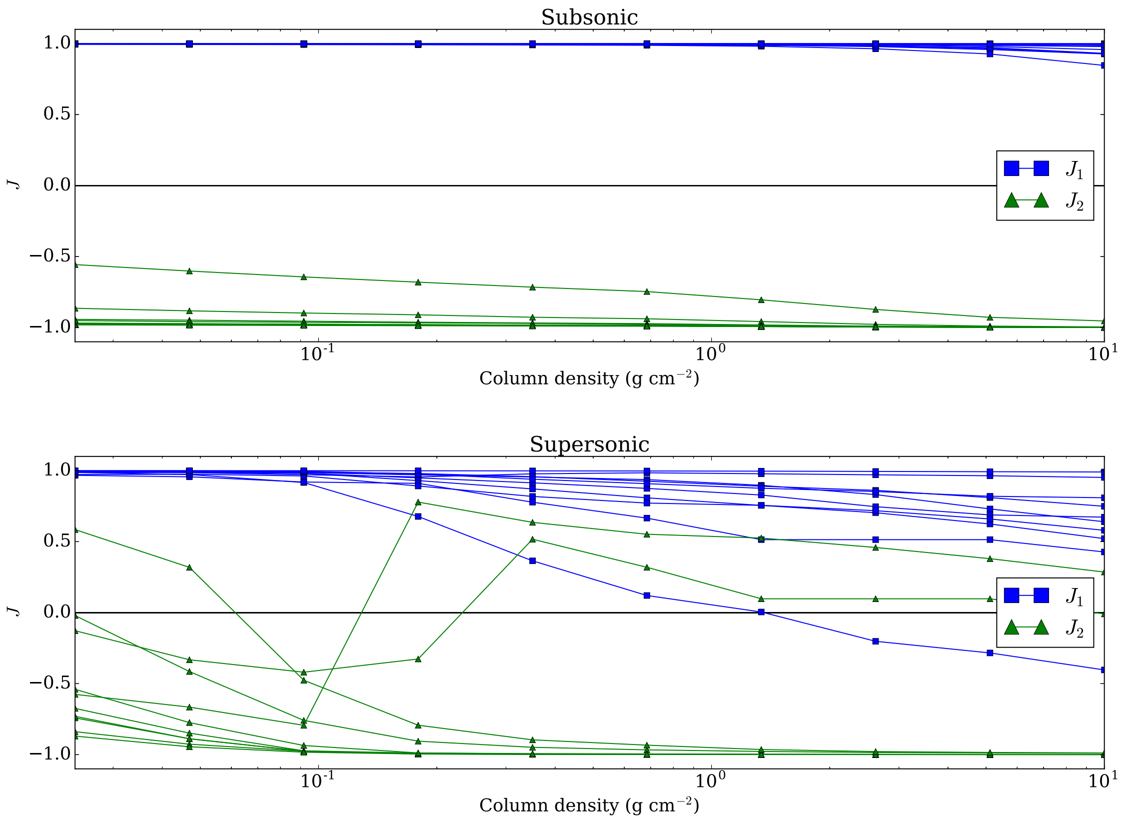

Figure 6.9. The dependence of the J moments on surface-density threshold for filaments forming in a medium with a thermal mix of turbulent modes. *Top:* 10 simulations with subsonic turbulence. *Bottom:* 10 simulations with supersonic turbulence. Filled blue squares represent values of J_1, and filled green triangles represent values of J_2.

at all thresholds, confirming objectively that the filaments are very straight and narrow, and the dense gas is rather evenly distributed along them. We describe this structure as ‘persistent’, meaning that the J moments do not change much when the surface-density threshold is increased. The one exception arises because the filament is slightly more fragmented in this particular case and so J_2 is somewhat higher. However, the pattern observed in the J plots for the nine similar filaments is very specific, and could be used to identify automatically filaments forming in a medium where the turbulence is subsonic.

Figure 6.9b shows how the J moments of the filaments in all ten *supersonic* cases evolve as the surface-density threshold is increased from $\sim 0.024\,\mathrm{g\,cm^{-2}}$ to $\sim 1.3\,\mathrm{g\,cm^{-2}}$. For eight cases there is a clear pattern: at low thresholds, J_1 is of order unity, and J_2 is negative, as we would expect for filaments; then, at higher surface-densities, as the filament breaks up into two fragments, J_1 and J_2 decrease, with J_2 quickly tending to ~ -1. The two cases that do not conform to this pattern are ones in which the filament only produces a single fragment, and therefore, as the surface-density threshold increases, J_2 abruptly increases to positive values and then decreases towards zero. This is essentially a consequence of the fact that the simulation only treats a rather short filament. Most observed filaments contain more than one fragment, so the pattern observed in the J plots for the eight similar filaments is probably more representative, and could be used to identify automatically filaments forming in a medium where the turbulence is supersonic.

6.4 THREE DIMENSIONAL ANALYSIS

We have so far discussed the analysis of only two dimensional datasets, but this method can theoretically be extended to any number of dimensions.

Observations of molecular line emission provide information about the radial velocity of the emitting gas and so lie in the three-dimensional position-position-velocity (PPV) space. A J moment analysis of PPV structures could reveal both the shape of the structure on the plane of the sky, and its distribution in velocity space. Consider two circular compact cores, one turbulent core with high line-widths, and one quiescent core with low line-widths. A two-dimensional (PP) J plot will be unable to differentiate between these structures. If a three-dimensional analysis is performed the structures can be differentiated because their shapes in PPV space would be very different. However, interpreting moments in spaces which combine different units (here position and velocity) is non-trivial and requires further study.

Another three-dimensional application is simulation output. Simulations can output density information in all three physical dimensions and so lie in position-position-position (PPP) space. J plots can be easily extended into three spatial dimensions: circular cores become spherical, filaments become cylinders and rings become bubbles. This would allow quantitative analysis of the structures formed in simulations, but could not be compared to observations without invoking some assumptions to convert the available velocity space information into a third spatial dimension.

6.5 Conclusions

We have developed a new method for analysing the structures in a segmented two dimensional image. Here we illustrate the method using dendrograms to segment the image into contiguous structures at varying surface-density thresholds. For each contiguous structure, the method first finds the area, A, mass, M, and principal moments of inertia, I_1, I_2. These parameters are then used to calculate the structure's J moments (Equation 6.5). The J moments allow one to distinguish structures that are centrally concentrated from structures that are centrally rarefied, and structures that are approximately circularly symmetric from structures that are elongated.

We apply the method to a tile from the Hi-GAL survey and show that it can identify and quantify the well-known RCW 120 bubble, as well as other ring-like structures in the same area. We also apply it to simulations of filaments growing in a turbulent medium and fragmenting, and show that J plots are able to identify and quantify, objectively, the difference in projected structure that results from changes in the nature of the turbulence. We also discuss the extension of this method to three dimensional data sets.

An analysis tool based on J plots is freely available online at
https://github.com/SJaffa/Jplots.

CHAPTER 7

SUMMARY OF THE THESIS

I don't know anything with certainty, but seeing the stars makes me dream.
- Vincent van Gogh

7.1 INTRODUCTION

Star formation is a chaotic process that involves a huge variety of complex and dynamic structures. Understanding the physical processes that sculpt these structures requires unbiased, quantitative analysis. In this thesis we aim to present and develop tools to quantify the structure in star clusters and the interstellar medium. We first review the use of fractal dimensions to describe astrophysical objects, and highlight the challenges of comparing values found by different methods. We then present two new analysis tools: the $\mathcal{Q}^+$ algorithm is designed to characterize the fractal structure of sets of points in terms of three parameters which fully constrain the fractal structure in our model; the J plots algorithm is designed to quantify the shape of pixelated structures taken from continuum images and separate centrally concentrated shapes, hollow circular shapes, and elongated filamentary shapes. We will now summarise the main conclusions and describe the future directions of this work.

7.2 FRACTALS

In Chapter 2, we review the use of fractal dimensions to describe the substructured nature of molecular clouds and star clusters. It is important to have quantitative measures of the clumpy shapes we observe, in order to compare observations or simulations in an unbiased way, and to build up a statistically robust picture of the structures involved in star formation. We describe methods for creating artificial fractal clouds or star clusters, as well as a few of the more commonly used measures of a fractal dimension.

We find that the box-counting method accurately constrains the fractal dimension of artificial star clusters created using the box-fractal method, although it is sensitive to the number of levels in the fractal hierarchy, which it cannot constrain. The Δ-variance method is able to constrain the fractal dimension (or power law index) of artificial fractal Brownian motion images, which have statistically similar structures to molecular clouds. However, when converting data sets from one type to another (greyscale images to sets of points or vice versa) we find that both methods are unable to accurately measure the fractal dimension. We conclude that the fractal dimensions measured by each method are not compatible and should not be regarded as representing the same underlying properties of the data set.

We also discuss the various relations used in the literature for inferring the real three-dimensional fractal dimension of astrophysical structures from projected two-dimensional images. We find that the different relations quoted are inconsistent and sometimes misapplied. We therefore conclude that estimates of fractal dimensions should be treated with caution and not compared unless they have been estimated using the same method.

7.3 The $\mathcal{Q}^+$ algorithm

In Chapter 3, we describe an algorithm for estimating the three-dimensional fractal structure of young star clusters from a two-dimensional image, and quantify the structure in terms of three parameters:

- the fractal dimension, $\mathcal{D}_{\mathcal{Q}+}$, reflects the clumpiness or smoothness of the distribution;
- the number of levels, $\mathcal{L}$, reflects the range of scales over which the fractal structure exists;
- the density scaling exponent, $\mathcal{C}$, reflects how the mass of the cluster is distributed over these scales.

This algorithm is calibrated against a set of analytical box-fractal clusters generated using a method adapted from Goodwin & Whitworth (2004).

The algorithm computes a set of statistics based on the Minimum Spanning Tree and complete graph and uses principal component analysis to compare the test cluster to the analytical ones and estimate the parameters that best describe its structure. We test its ability to constrain the input parameters of artificial clusters and estimate the errors on the results.

In Chapter 4, we apply this algorithm to some observed and simulated young star clusters. The analysis of observed clusters of young stars in the Gould Belt shows that real star clusters have structure across a large range of the parameter space modelled in this work, from highly substructured to smooth or centrally concentrated, and with a large difference in scales. The estimated properties of Taurus, Lupus 3, Chamaeleon I and IC 348 fit with a visual assessment of their structure, and the new algorithm reduces problems encountered using the $\mathcal{Q}$ parameter of CW04 due to their methodology not considering all 3 parameters inherent in an artificially generated fractal cluster.

Secondly, we analyse clusters produced in simulations of cloud-cloud collisions of different speeds. We find that the clusters produced in slower collisions are fractally substructured, while faster collisions produce centrally concentrated clusters.

Young clusters are expected to show similar sub-structure to the densest parts of the clouds from which they are born, but this structure will be erased by dynamical interactions. We analyse simulations of the dynamical evolution of young embedded clusters, with initial conditions taken from the MYStIX survey. The results show that the substructure is very quickly dynamically erased, usually within 1 Myr. The clusters then become increasingly centrally concentrated.

Molecular clouds also display substructure and intuitively we would expect them to pass on this structure to the clusters they spawn. We therefore test methods of applying the $\mathcal{Q}^+$ algorithm to continuum data to enable a direct comparison of the fractal properties of molecular clouds and young star clusters.

In Chapter 5, we convert a continuum images into statistically representative sets of points by using it as a probability density function. We discuss the number of points that should be used, as this strongly affects the range of parameters that can be recovered. We conclude that scaling the number of points by the total mass of the clouds being studied gives the best estimate of the fractal parameters.

We apply this analysis to four column density maps from the *Herschel* Hi-GAL survey, testing the individual wavelength intensity maps as well as the derived column densities. Comparing this to other estimates of the fractal dimension of these data sets does not show a correlation between the two measures of fractal dimension.

7.4 J PLOTS

In Chapter 6, we present a new tool for analysing structures in pixelated images. We first use dendrograms to segment an image into a set of nested structures defined by a bounding density or intensity contour. This method then finds the principal axes of each structure and calculates the moment of inertia about each principal axis. The J values are derived from these principal moments in comparison to that of a thin disk of constant density. These two J values can distinguish between roughly flat symmetric shapes, hollow ring-like shapes, elongated filamentary shapes and centrally concentrated shapes.

We apply this tool to an example tile from the Hi-GAL survey that contains the well-known RCW 120 bubble and show that it can identify and quantify the bubble, as well as other ring-like structures in the same area. We also apply it to the simulations of filament fragmentation presented in Clarke et al. (2017), which examined how turbulence affects the structures formed within an accreting filament. We show that, in projected two-dimensional images, J plots are able to identify the differences in structure which had previously been identified only by visual inspection.

7.5 FUTURE WORK

The wealth of observational data available from wide field surveys and increasingly large and complex simulations provides an excellent opportunity to utilise some of the ‘big data’ tools that have recently emerged in computer science. Machine learning tools can automate many of the tasks commonly applied in astrophysics including classification and regression. However, the usefulness of the results depends on the quality of the input data (i.e. the training data set). The statistical descriptors presented here would make an excellent feature for the application of machine learning to astronomical data.

7.5.1 FRACTALS

An empirical exploration of the relation between two- and three-dimensional measures of the fractal dimension would be important in understanding which of the relations discussed is applicable to different data sets. Secondly, fractal distributions can be nested (smaller, higher density clumps only located inside larger, lower density clumps) or not nested (the position of higher density clumps is not related to the position of larger, lower density clumps). This may explain the incompatibility of the $\mathcal{Q}^+$ and Δ-variance measures of fractal dimension, and should be explored further in relation to real astrophysical objects.

Furthermore, we have confined this discussion to artificially created clusters because they have known fractal dimensions and we assume that they are a good representation of the variation of structure we see in real clouds and clusters. Testing each of these methods on a greater range of real data would challenge this assumption and allow us to test how the results are affected by noise or whether the artificial models are too simplified.

7.5.2 $\mathcal{Q}^+$

Our analysis of the evolution of young clusters shows that the sub-structure can be erased in a very short time. Performing simulations with a range of initial conditions and taking many snapshots would allow a better statistical analysis of this process.

Analysis of more observed clusters would allow us to understand where in the parameter space real objects reside. Many surveys provide catalogues of point sources, young stellar objects, or evolved stars which can be used to statistically trace how the structure of clusters evolves with age.

When examining real observations we must also understand how observational effects such as incompleteness at lower masses or confusion of sources in very dense clusters will affect the structure we recover. Synthetic observations allow us to add these effects to simulated star clusters and test how well the $\mathcal{Q}^+$ algorithm identifies the structure under non-ideal conditions.

The statistical and unbiased comparison of fractal properties between star clusters and molecular clouds is an important step in understanding how their structures are related. Further testing of the application of $\mathcal{Q}^+$ to continuum images, such as exploring different ways to convert data sets from one type to another, would enable more robust results.

7.5.3 J PLOTS

The Hi-GAL survey has provided us with a wealth of data containing filamentary structures. The Milky Way project leveraged the power of Citizen Scientists worldwide to identify rings of a range of sizes and morphologies. Many surveys produce catalogues of point sources. Each of these could be analysed with J plots to quantify the size and shape of these structures and whether there are any trends that may reveal something about the underlying physics of the environment.

This work is currently confined to two dimensions, but simulations can give us a third spatial dimension, and molecular line observations allow us to examine the radial velocity of real astrophysical structures. We plan to extend J plots to three dimensional spaces, for example to understand how structures identified in observed position-position-velocity space may relate to their real shape in three dimensional positional space.

APPENDIX A
GRAPHS

A ‘graph’ is a collection of N points (vertices) and/or edges (lines) connecting the points in any number of dimensions. In this work the weight of an edge is simply it’s Euclidean length in two or three dimensions, but other weightings can be given for different applications.

Complete Graph

The Complete Graph (CG) is the collection of $\frac{N(N-1)}{2}$ straight edges connecting each point to every other point. The length of the edge joining points i and j on the Complete Graph is termed s_{ij}. An example is shown in the centre panel of Figure A.1.

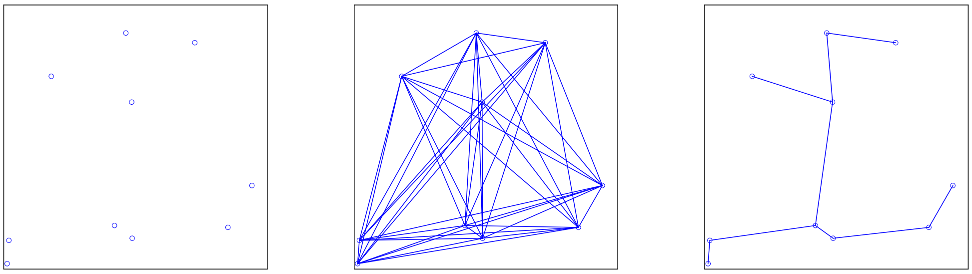

Figure A.1. Graphs demonstrated for a randomly generated set of 10 points (left): the complete graph (centre), connecting every point to every other point; and the Minimum Spanning Tree (right), the set of edges that join every point into a single set with the shortest possible total length.

Minimum Spanning Tree

A spanning tree is defined as a set of edges that connects all the points in a graph together without forming any closed loops. The minimum spanning tree (MST) is the spanning tree where the $N - 1$ edges have the smallest possible total length. An example MST can be seen in the right panel of Figure A.1.

The most popular algorithms for constructing MSTs are those of Prim (1957) and Kruskal (1958).

Prim's algorithm proceeds as follows:

1. Start from a randomly chosen point.
2. Find the shortest edge connecting that point to any other point and add the edge to the tree.
3. Find the shortest edge connecting any point in the tree to any point not in the tree and add that edge to the tree.
4. Repeat 3 until all points have been connected into the tree.

Kruskal's algorithm proceeds as follows:

1. Find the length of all edges on the complete graph and sort in ascending order.
2. Starting with the shortest edge, add edges to the tree only if they do not form a closed loop.
3. Repeat 2 until all points have been connected into the tree.

Both these algorithms will result in the same MST unless two or more edges the graph have the exact same weight. This is unlikely if the weighting used is length, but some applications may use more restricted weightings and therefore have a higher probability of two edges having the same weight. In such cases the edge to be added can be chosen at random, so while the MST for a set of points may not be unique, any statistics based on its total or average weight will be unaffected.

APPENDIX B
BAYESIAN INFERENCE

Bayesian statistics relies on conditional probabilities to estimate the probability of an event based on prior knowledge of related factors. The central equation was first proposed by Bayes & Price (1763) and developed into the modern form by Laplace (1820).

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{B.1}$$

$P(B|A)$ is the posterior probability of event B given event A. This is the term we wish to calculate. $P(A|B)/P(A)$ is the likelihood ratio, which describes the probability of event A given B, normalised by the total probability of A. $P(B)$ is the prior probability of B, and acts as a weighting function to incorporate any knowledge we already have of the probability of event B. In words, this equation updates our belief about a system (the prior) with some evidence (the likelihood ratio) to create a new belief (the posterior).

From this we can calculate the expectation value of B, E_B, and estimate the uncertainty, σ_B, on this value:

$$E_B = \frac{\sum P(B|A)B}{\sum P(B|A)} \tag{B.2}$$

$$\sigma_B^2 = \frac{\sum P(B|A)(B - E_B)^2}{\sum P(B|A)} \tag{B.3}$$

This equation is simple but its application can be very powerful and it underpins a lot of machine learning and artificial intelligence, such as spam filtering of emails.

APPENDIX C

MATRIX ALGEBRA

For an arbitrary $n \times n$ square matrix $A_{n\times n}$, we define the following:

- $A_{i,j}$ denotes the element occupying the i-th row and j-th column position, where $0 \le i, j \le n$.
- A^T is the transpose of the matrix A, which is a reflection of A in the diagonal. Formally, $[A^T]_{i,j} = [A]_{j,i}$. Therefore $[A^T]^T \equiv A$.
- $|A|$ signifies the determinant of the matrix A, defined by

  $$|A| = \sum_{i=1}^{n} (-1)^{(i+j)} |M_{i,j}| A_{i,j}, \tag{C.1}$$

  where $M_{i,j}$ is the minor of A, which is formed by removing row i and column j from A, resulting in an $n-1 \times n-1$ matrix.
- The eigenvectors, E, of a matrix are a set of vectors which satisfy the following equation:

  $$AE = \lambda E. \tag{C.2}$$

  That is when multiplied by the matrix they give a scalar multiple of themselves. (Strictly this defines right eigenvectors because they are multiplied on the right of the matrix. A similar equation can be formed for left eigenvectors.) λ is the eigenvalue associated with a given eigenvector, E.

  An $n \times n$ matrix will have n eigenvectors and n corresponding eigenvalues. These can be found by solving the equation

  $$|A - \lambda I| = 0, \tag{C.3}$$

where I is the identity matrix - an $n \times n$ matrix where all diagonal elements $(I_{i,i})$ are one and all off-diagonal elements $(I_{i,j}, j \neq i)$ are zero.

APPENDIX D
PRINCIPAL COMPONENT ANALYSIS

Principal component analysis is a mathematical technique first introduced by Pearson (1901), which allows a data set with many dimensions to be represented by a best fit in a lower dimensional space. This is used to compress data without losing too much information, and to find patterns and subsets within multi-dimensional data sets which would otherwise be difficult to identify. It is used in image compression, facial recognition, financial risk management and many other applications.

Taking a data set with n data points measuring v variables in the form of a two-dimensional matrix $X_{v\times n}$, the procedure is as follows:

1. For each variable, V_i, subtract the mean, μ_{V_i}, to create an adjusted data matrix, X', where the origin is now at the centroid of the data.

$$X'_{i,j} = X_{i,j} - \mu_{V_i} \tag{D.1}$$

2. Calculate the covariance or correlation of each variable with all other variables to form a correlation or covariance matrix, C. (The choice of correlation or covariance and its impact on the results is discussed below.)

$$C = \begin{bmatrix} co(V_0, V_0) & co(V_0, V_1) & \cdots & co(V_0, V_v) \\ co(V_1, V_0) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ co(V_v, V_0) & co(V_v, V_1) & \cdots & co(V_v, V_v) \end{bmatrix} \tag{D.2}$$

where *co* can represent either the correlation or covariance function. The diagonal elements of this matrix are the correlation or covariance of each variable with itself. If the covariance is used these terms will therefore be the variance of each variable. If the correlation is used these terms will be 1.

3. Calculate the eigenvectors and eigenvalues of the matrix, C. The eigenvectors are, by definition, an orthogonal set of linear combinations of the variables. This is equivalent to a rotation of the original data set where the eigenvector with the highest eigenvalue follows the direction of maximum possible variance of the data set.

4. Order the eigenvectors by the magnitude of their eigenvalues and choose a number, e, of eigenvectors with the highest eigenvalues (strongest variance) to form an eigenvalue matrix, $E_{e\times v}$.

5. Transform data into Principal Component space by multiplying the transpose of the eigenvalue matrix by the tranpose of the adjusted data matrix:

$$PC_e = (E_{e\times v})^T \times (X'_{v\times n})^T \tag{D.3}$$

The main limiting assumption of this method is that all parameters vary linearly, which is very often not the case. Slight deviations from linearity will only cause minor problems, but, if some or all of the variables are significantly non-linear, the principal components will not reflect the underlying variance of the data very well.

The choice of whether to use the correlation or covariance matrix relates to the ranges of the variables being used. If all variables are measured in the same units, and over similar ranges, then using the covariance matrix will give good results. However, if some variables have a much larger range then they will have a larger variance and will therefore be given more weight in the eigenvectors. This can be a problem if the difference is simply due to different units of measurement. In this case the correlation matrix should be used as this normalises each term by the variance of both variables before calculating eigenvectors.

The best choice of number of eigenvectors is subjective and depends on the data set being studied and the purpose of the analysis. The eigenvalues, W, show the contribution of each eigenvector to the variance of the data set, so a measure of the proportion of the variance represented by a subset of the eigenvectors can serve as a guide for how many principal components should be used. Since the original data set has v variables, the matrix C has size $v \times v$, and therefore v eigenvectors in total. Hence we can define the proportional variance in a subset of k eigenvectors as

$$g(k) = \frac{\sum_{e=1}^{k} W_e}{\sum_{e=1}^{v} W_e}. \tag{D.4}$$

A general guide is to choose the set of k eigenvectors with the highest eigenvalues such that they represent 90% of the variance ($g(k) \gtrsim 0.9$).

APPENDIX E

PRINCIPAL AXES AND MOMENTS

For a dendrogram structure represented by P equal pixels, each with area ΔA, we can compute the following moments,

$$\begin{aligned} M_0 &= \Delta A \sum_{p=1}^{p=P} \{\Sigma_p\}\,, & M_x &= \Delta A \sum_{p=1}^{p=P} \{\Sigma_p x_p\}\,, \\ M_y &= \Delta A \sum_{p=1}^{p=P} \{\Sigma_p y_p\}\,, & M_{xx} &= \Delta A \sum_{p=1}^{p=P} \left\{\Sigma_p x_p^2\right\}\,, \\ M_{xy} &= \Delta A \sum_{p=1}^{p=P} \{\Sigma_p x_p y_p\}\,, & M_{yy} &= \Delta A \sum_{p=1}^{p=P} \left\{\Sigma_p y_p^2\right\}\,, \end{aligned} \tag{E.1}$$

where Σ_p is the surface-density in pixel p, and (x_p, y_p) is the position of its centre.

The centre of mass is then given by

$$X = \frac{M_x}{M_0}\,, \quad Y = \frac{M_y}{M_0}\,, \tag{E.2}$$

and the moments about the Cartesian axes by

$$\begin{aligned} I_{xx} &= M_{xx} - M_x X, \quad I_{xy} = M_{xy} - M_0 XY, \\ I_{yy} &= M_{yy} - M_y Y. \end{aligned} \tag{E.3}$$

The principal moments are then

$$I_{1,2} = \left(\frac{I_{xx}+I_{yy}}{2}\right) \mp \left\{ \left(\frac{I_{xx}+I_{yy}}{2}\right)^2 - \left(I_{xx}I_{yy} - I_{xy}^2\right) \right\}^{1/2}, \tag{E.4}$$

(where the $\mp$ means that I_1 corresponds to the minus sign, and I_2 to the plus sign).

The first principal axis has the equation

$$y_1 = Y + \frac{(I_1 - I_{xx})\ x}{I_{xy}}, \tag{E.5}$$

and the second principal axis

$$y_2 = Y - \frac{I_{xy}}{(I_1 - I_{xx})\ x}. \tag{E.6}$$

BIBLIOGRAPHY

Allan D. W., 1966, IEEE Proceedings, 54

Alves J., Lombardi M., Lada C. J., 2017, A&A, 606, L2

Anderson L. D. et al., 2010, A&A, 518, L99

André P. et al., 2010, A&A, 518, L102

Balfour S. K., Whitworth A. P., Hubber D. A., Jaffa S. E., 2015, MNRAS, 453, 2471

Bashtannyk D. M., Hyndman R. J., 2001, Computational Statistics and Data Analysis, 36, 279

Bayes T., Price M., 1763, Philosophical Transactions (1683-1775), 370

Beech M., 1992, Ap&SS, 192, 103

Berry D. S., 2015, Astronomy and Computing, 10, 22

Bodenheimer P., 2011, Principles of star formation. Springer Science & Business Media

Cartwright A., Whitworth A. P., 2004, MNRAS, 348, 589

Chabrier G., 2009, Structure formation in astrophysics. Cambridge University Press

Churchwell E. et al., 2006, ApJ, 649, 759

Clarke S. D., Whitworth A. P., Duarte-Cabral A., Hubber D. A., 2017, MNRAS, 468, 2489

Colombo D., Rosolowsky E., Ginsburg A., Duarte-Cabral A., Hughes A., 2015, MNRAS, 454, 2067

Colombo D., Rosolowsky E., Ginsburg A., Duarte-Cabral A., Hughes A., 2016, SCIMES: Spectral Clustering for Interstellar Molecular Emission Segmentation. Astrophysics Source Code Library

Csengeri T. et al., 2014, A&A, 565, A75

Dale J. E., 2015, NewAR, 68, 1

Dale J. E., Haworth T. J., Bressert E., 2015, MNRAS, 450, 1199

Davis C. J., Chrysostomou A., Matthews H. E., Jenness T., Ray T. P., 2000, ApJ, 530, L115

Deharveng L., Zavagno A., Schuller F., Caplan J., Pomarès M., De Breuck C., 2009, A&A, 496, 177

Elia D. et al., 2007, ApJ, 655, 316

Elia D. et al., 2013, ApJ, 772, 45

Elia D. et al., 2017, MNRAS, 471, 100

Elia D. et al., 2014, ApJ, 788, 3

Elmegreen B. G., Falgarone E., 1996, ApJ, 471, 816

Falconer K. J., 1990, Fractal Geometry: Mathematical Foundations and Applications, first edition edn. John Wiley and Sons, London

Feigelson E. D. et al., 2013, ApJS, 209, 26

Figueira M. et al., 2017, A&A, 600, A93

Geen S., Soler J. D., Hennebelle P., 2017, MNRAS, 471, 4844

Gladwin P. P., Kitsionas S., Boffin H. M. J., Whitworth A. P., 1999, MNRAS, 302, 305

Goodwin S. P., Whitworth A. P., 2004, A&A, 413, 929

Gouliermis D. A., Hony S., Klessen R. S., 2014, MNRAS, 439, 3775

Hacar A., Tafalla M., Kauffmann J., Kovács A., 2013, A&A, 554, A55

Hennemann M. et al., 2012, A&A, 543, L3

Howard C. S., Pudritz R. E., Harris W. E., 2017, MNRAS, 470, 3346

Ingallinera A. et al., 2016, MNRAS, 463, 723

Jaffa S. E., Whitworth A. P., Lomax O., 2017, MNRAS, 466, 1082

Jeans J. H., 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Khorrami Z. et al., 2016, A&A, 588, L7

Kirk H., Myers P. C., 2011, ApJ, 727, 64

Koch E. W., Rosolowsky E. W., 2015, MNRAS, 452, 3435

Könyves V. et al., 2015, A&A, 584, A91

Kruskal J. B., 1958, Proc. Amer. Math. Soc., 7, 48

Kuhn M. A. et al., 2014, ApJ, 787, 107

Lada C. J., Lada E. A., 2003, ARA&A, 41, 57

Laplace P. S., 1820, Théorie analytique des probabilités. Courcier

Larsen S. S., 2009, A&A, 494, 539

Larson R. B., 1981, MNRAS, 194, 809

Larson R. B., 1995, MNRAS, 272, 213

Li G.-X., Wyrowski F., Menten K., Megeath T., Shi X., 2015, A&A, 578, A97

Lomax O., Whitworth A. P., Cartwright A., 2011, MNRAS, 412, 627

Mädler M., 1847, MNRAS, 7, 227

Mandelbrot B. B., 1967, Science, 156, 636

Mandelbrot B. B., 1983, The fractal geometry of nature, Revised and enlarged edition, Vol. 173. WH freeman New York

Marsh K. A. et al., 2016, MNRAS, 459, 342

Marsh K. A. et al., 2017, MNRAS, 471, 2730

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

McKee C. F., Ostriker J. P., 1977, ApJ, 218, 148

Megeath S. T. et al., 2016, AJ, 151, 5

Molinari S. et al., 2016, A&A, 591, A149

Molinari S. et al., 2010, A&A, 518, L100

Motte F. et al., 2010, A&A, 518, L77

Muench A. A., Lada C. J., Luhman K. L., Muzerolle J., Young E., 2007, AJ, 134, 411

Ossenkopf V., Krips M., Stutzki J., 2008, A&A, 485, 917

Palmeirim P. et al., 2013, A&A, 550, A38

Parker R. J., Wright N. J., Goodwin S. P., Meyer M. R., 2014, MNRAS, 438, 620

Pavlyuchenkov Y. N., Kirsanova M. S., Wiebe D. S., 2013, Astronomy Reports, 57, 573

Pearson K., 1901, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2, 559

Prim R. C., 1957, Bell System Technical Journal, 36, 1389

Rathborne J. M. et al., 2015, ApJ, 802, 125

Richardson L. F., 1961, General Systems Yearbook, 6, 139

Rodón J. A., Zavagno A., Baluteau J.-P., Habart E., Köhler M., Le Bourlot J., Le Petit F., Abergel A., 2015, A&A, 579, A10

Rosolowsky E., Leroy A., 2006, PASP, 118, 590

Rosolowsky E. W., Pineda J. E., Kauffmann J., Goodman A. A., 2008, ApJ, 679, 1338

Salji C. J. et al., 2015, MNRAS, 449, 1782

Sánchez N., Alfaro E. J., Pérez E., 2005, ApJ, 625, 849

Schisano E. et al., 2014, ApJ, 791, 27

Schmeja S., Klessen R. S., 2006, A&A, 449, 151

Scott D., 1992, Multivariate Density Estimation. 1992. Wiley, New York

Sills A., Rieder S., Scora J., McCloskey J., Jaffa S., 2018, MNRAS

Silverman B., 1986, Density Estimation for Statistics and Data Analysis. Chapman and Hall, London

Simpson R. J. et al., 2012, MNRAS, 424, 2442

Sousbie T., 2011, MNRAS, 414, 350

Storm S. et al., 2014, ApJ, 794, 165

Storm S. et al., 2016, ApJ, 830, 127

Stutzki J., Bensch F., Heithausen A., Ossenkopf V., Zielinsky M., 1998, A&A, 336, 697

Stutzki J., Güsten R., 1990, ApJ, 356, 513

Turlach B. A., 1993, in CORE and Institut de Statistique

Vázquez-Semadeni E., Kim J., Ballesteros-Paredes J., 2005, ApJ, 630, L49

Voss R. F., 1988, Fractals in nature: From characterization to simulation, Peitgen H.-O., Saupe D., eds., Springer New York, New York, NY, pp. 21–70

Walch S., Whitworth A. P., Bisbas T. G., Hubber D. A., Wünsch R., 2015, MNRAS, 452, 2794

Werner M. W. et al., 2004, ApJS, 154, 1

Williams J. P., de Geus E. J., Blitz L., 1994, ApJ, 428, 693

Winston E. et al., 2007, ApJ, 669, 493

Wright N. J., Parker R. J., Goodwin S. P., Drake J. J., 2014, MNRAS, 438, 639