Spatial statistics in star-forming regions

Brendan Retter

Submitted by Brendan Retter to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Physics, November, 2020.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signed: .....

Brendan Retter

Date: .....

Abstract

Observational studies of star formation reveal spatial distributions of Young Stellar Objects (YSOs) that are 'snapshots' of an ongoing star formation process. Using methods from spatial statistics it is possible to test the likelihood that a given distribution process could produce the observed patterns of YSOs. I determine the sensitivity of the spatial statistical tests Diggle's G function (G), the 'free-space' function (F), Ripley's K and O-ring for application to astrophysical data. To do this I applied each test to simulated data containing 2D Gaussian clusters projected on a random distribution of background stars. By varying the number of stars within the Gaussian cluster and the number of background stars I determined the ability of the tests to reject complete spatial randomness (CSR) with changing signal-to-noise. Ripley's K and O-ring were shown to be much more sensitive to Gaussian clusters than G and F.

I then apply the O-ring test to determine if column density alone is sufficient to explain the locations of Class 0/I YSOs within Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348. Star formation is known to occur more readily where more raw materials are available, a relationship that is often expressed in the form of a 'Kennicutt–Schmidt' relation where the surface density of Young Stellar Objects (YSOs) is proportional to column density to some power, μ. Using the O-ring test as a summary statistic, confidence envelopes were produced for different values of μ from probability models made using the *Herschel* column density maps. four distribution models: the best-estimate of μ for the region, The YSOs were tested against four distribution models: the best-estimate of μ for the region, $\mu = 0$ (i.e. random) above a column density threshold and zero probability elsewhere, $\mu = 1$, and the power-law that best represents the five regions as a collective, $\mu = 2.05 \pm 0.20$. Serpens South and NGC1333 rejected the $\mu = 2.05$ model on small scales of ~ 0.15 pc which implies that small-scale interactions may be influencing their distribution. On scales above 0.15 pc, the positions of YSOs in all five regions can be well described using column density alone.

Contents

List of Figures

List of Tables

Declaration

This thesis contains work published in the Monthly Notices of the Royal Astronomical Society (MNRAS) paper 'Spatial statistics in star-forming regions: testing the limits of randomness' (Retter, B., J. Hatchell, and Tim Naylor. 2019. MNRAS 487, no. 1 (July): 887–899) of which I am the first author. This work is presented in Chapters 2 and 3. Chapter 4 contains work pending submission to MNRAS in a paper in which I am first author. The majority of analysis has been performed using code which was developed by myself, but where this is not the case I have provided attribution in the relevant places.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Jenny Hatchell for all of her supervision and guidance over these past four years. I would also like to thank Tim Naylor for his secondary supervision, and all-around statistics know-how. They have both taught me so much and I'm incredibly grateful for their time, patience, and talks about science.

I also thank the Exeter Astrophysics group for being such a happy, helpful and welcoming group of people. Thank you for all of the coffee, quizzes, pub trips, game nights and film nights that made climbing up the hill to the physics building all the more worthwhile. I especially acknowledge and thank my fellow members of G27: Aarynn, Adam, Mark and Tom. I am grateful to have spent so much time with a fantastic group of people with a shared love of shenanigans.

I thank my parents, Rebecca, Derek, Robin and Sam for all of the support they have given me over these years. You have always been there for me when I needed help. I thank Celia for her enormous generosity and kindness. Finally, I would also like to thank the friends, old and new, that have formed such a large part of my life. Thanks for everything.

Brendan Retter

Exeter, U.K.

*18th November 2020

Chapter 1

Introduction

1.1 Aims and Motivations

Star formation is known to occur within molecular clouds and the distributions of young stellar objects (YSOs) within these clouds are imprints of the underlying processes behind star-formation. It is possible, then, from the locations of these objects in the early stages of star formation to infer information about the processes that may affect their distributions.

Observations of young, embedded stars and their dense cloud core precursors suggest that young stars predominantly form within the high column density regions of molecular clouds. It has also been observed that very little star formation occurs at low column densities, $\approx 6 \times 10^{21} \mathrm{N_{H_2}}$ cm^{-2} (e.g. Onishi et al. 1998; Johnstone et al. 2004; André et al. 2010; Lada et al. 2010). This lower-limit is sometimes interpreted as a threshold column density requirement for star formation, and is related to the critical line density above which filaments, structures of dense material universally observed within molecular clouds, collapses and fragments.

However, YSOs are also found at column densities lower than this threshold. In addition, while YSOs tend to be correlated with column density, other factors can also influence star formation within a molecular cloud, such as turbulence and magnetic fields. The aim of this work is to apply methods from spatial statistics in order to investigate the spatial distributions of YSOs. By doing so it is possible to determine if their distributions

are consistent with a model of star formation that depends on column density alone, or if some of these aforementioned factors need to be taken into account.

1.2 Molecular Clouds

Molecular clouds are clouds of gas and dust with masses that vary between $\sim 10^3$ and $\sim 10^6\ M_\odot$ (Williams and McKee 1997; Shu et al. 1987; Heyer and Dame 2015). Most star formation occurs within molecular clouds and the presence of dense, gravitationally collapsing cores indicates that some clouds are still undergoing active star formation. Tracking the distribution of mass within a molecular cloud is therefore important to understand how and where stars form within these clouds.

1.2.1 Column Density

Column density is a measure of the amount of mass along the line of sight and because we measure along the line of sight the result is a density per unit area rather than volume. Since most of the mass in molecular clouds is in the form of H_2, the number of H_2 molecules is a useful measure for the mass. H_2 is, however. not directly visible at cold temperatures (Lombardi et al. 2014) and so column density tracers are applied in order to map how the mass in a molecular cloud is distributed. Examples of column density tracers include molecular-line emission, dust-extinction and thermal dust-emission.

Molecular line emission functions by tracing molecules found within the gas, such as H_2CO, CS, HC_3N, NH_3 and CO (Shu et al. 1987). These tracers are well suited for identifying specific features, such as dense gas, because their emission depends on their excitation and abundances. This specificity, however, does mean that they are less well suited for tracing the general column density. Dust-extinction traces the cloud material by tracking how increasing cloud material affects the visibility of objects that are behind the cloud. An early technique of tracing dust-extinction was to segment the area into a grid and the compare the counts of stars in the grid squares on the cloud to counts of stars in areas that are not obscured by the cloud (Lynds 1967). The issue with this star-counting technique is that it suffers from uncertainties of $\sqrt{N}$, where N is the number of stars, due to Poisson counting statistics. More recent techniques such as NICE (Lada et

al. 1994), NICER (Lombardi, M. and Alves, J. 2001), and NICEST (Lombardi 2009) measure dust-extinction by observing how stars behind the cloud are reddened as a function of position. Near-infrared red extinction and star-count methods as tracers are limited in their functionality by the number of detectable background stars behind the cloud (Lombardi et al. 2014), though this can be improved with the inclusion of mid-infrared observations such as *Spitzer* (Heiderman et al. 2010).

Column density can be traced from the spectral energy distribution (SED) of the emission of the dust. For the *Herschel* data, this is done by fitting a modified blackbody function of the form (Könyves et al. 2010)

$$I_\upsilon = B_\upsilon(T_d)(1 - e^{-\tau_\upsilon}), \tag{1.1}$$

where I_υ is the observed surface brightness at frequency υ, $B_\upsilon(T_d)$ is the blackbody function at dust temperature T_d and τ_υ is the optical dust depth. The column density, Σ_{gas}, is related to the optical dust depth by $\tau_\upsilon = \kappa_\upsilon \Sigma_{\text{gas}}$, where κ_υ is the dust opacity per unit mass. For the *Herschel* column density maps (Könyves et al. 2010) κ_υ is approximated by $\kappa_\upsilon = 0.1\ (\upsilon/1000\ \text{GHz})^\beta\ \text{cm}^2/\text{g}$ and the dust emissivity index, β, is fixed to 2 (Hildebrand 1983). The column density and dust temperatures for the SED are then estimated to be the values which produce the best-fitting modified blackbody function. With *Herschel*, each position is measured at five wavelengths (70, 160, 250, 350 and 500 μm), and so SEDs can be fit for every position within a map (e.g. André et al. 2010; Könyves et al. 2010; Men'shchikov et al. 2010; Arzoumanian et al. 2011; Harvey et al. 2013) resulting in a column density map such as that in Fig. 1.1.

1.2.2 Filaments

Observations of the column density in molecular clouds have revealed the near ubiquitous presence of elongated, dense structures within the cloud commonly referred to as filaments. Examples of these include the B211/B213 filament in the Taurus molecular cloud shown in Fig. 1.2, or the filaments in IC 5146 in Fig. 1.3. Filamentary structures are seen in most molecular clouds, and while their origin is not certain, theoretical models provide insight into their underlying physics. Ostriker (1964) found that filaments modelled as an

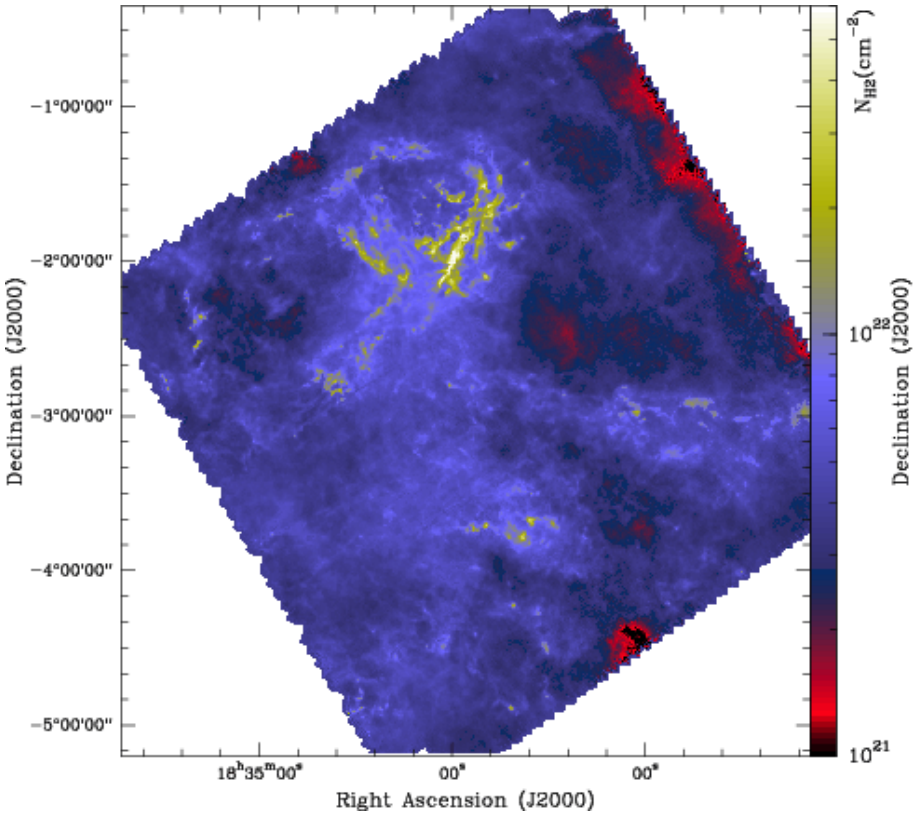

Figure 1.1: Column density map of Aquila region derived from *Herschel* data. Figure reproduced from (Könyves et al. 2010)

isothermal cylinder of gas have a density profile in the outer regions of $\rho \propto r^{-4}$, where ρ and r are the density and radius of the cylinder respectively. This is significantly steeper than the density profiles observed in filaments which follow a relation of $\rho \propto r^{-2}$ (e.g. Alves et al. 1998; Arzoumanian et al. 2011; Palmeirim et al. 2013). It was shown by Fiege and Pudritz (2000) that the inclusion of helical magnetic fields shallows this relationship to between $r^{-1.8}$ and r^{-2}.

In addition to similar density profiles, filaments appear to have a universal typical inner width of around 0.1 pc (Arzoumanian et al. 2011). The similarity between the characteristic filament width and the scales at which supersonic turbulent motions transition to subsonic suggests two potential origins: a formation due to the turbulent compression of interstellar gas, or filaments are structures formed in quasi-equilibrium with the interstellar medium pressure (Palmeirim et al. 2013).

Filaments show evidence of being highly influential in local star formation, in Orion A 71% of the pre-stellar population of dust cores are seen to be projected on top of fila-

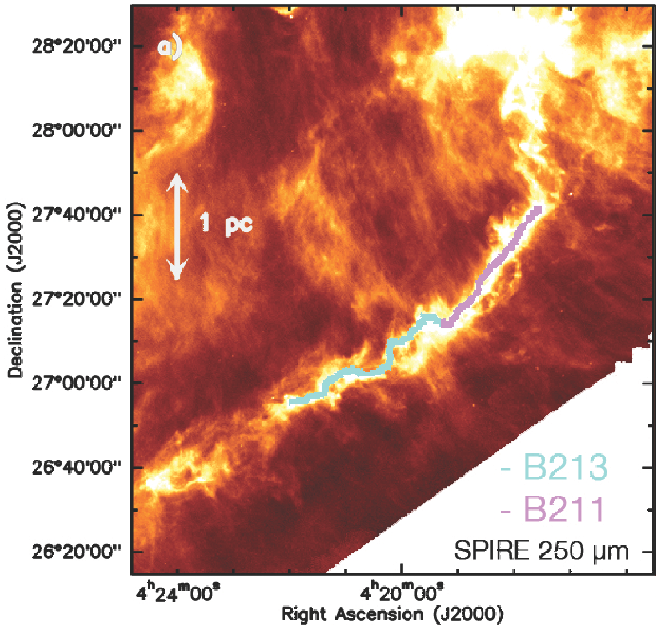

Figure 1.2: *Herschel*/SPIRE 250 μm image of the B211/B213/L1495 region in the Taurus molecular cloud. Figure reproduced from Palmeirim et al. (2013).

ments (Polychroni et al. 2013). Pre-stellar cores are starless, gravitationally collapsing, dense cores that will form a protostellar core in future, see section 1.3. If it is the case that filaments are where most stars form, it is possible that the filaments themselves are a step in the star formation process.

When modelling filaments as isothermal and unmagnetised, it was shown that there is a critical mass per unit length required for equilibrium, and when this is exceeded the filaments either fragment or collapse depending on how much larger the line mass is compared to the critical value

$$M_{\rm line} \gtrsim M_{\rm line,crit} \equiv 2c_s^2/G \sim 16\, M_\odot/{\rm pc} \tag{1.2}$$

where c_s is the isothermal sound speed for molecular gas at 10 K (Inutsuka and Miyama 1997). Numerical simulations suggest that turbulence is also a significant factor in the fragmentation of filaments and affects the form in which the filaments fragment. As the degree of turbulence increases, the length-scales of fragmentation changes to being dominated by the density perturbation scales of the turbulence (Clarke et al. 2017).

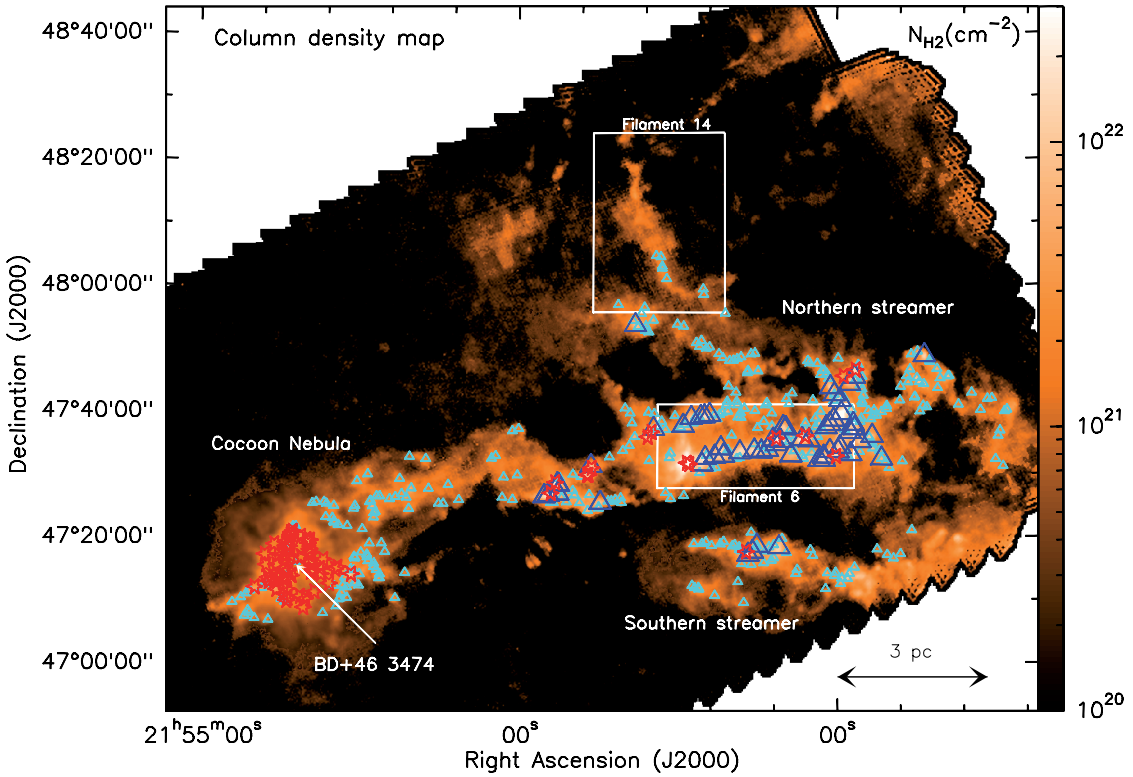

Figure 1.3: Column density map of IC 5146 overplotted with *Herschel* sources. YSOs are plotted as red stars, starless cores as blue triangles, and bound starless cores as larger blue triangles. Figure reproduced from Arzoumanian et al. (2011).

André et al. (2010) observed that almost all of the supercritical filaments in Aquila showed signs of recent or ongoing star formation, while subcritical filaments were generally free from signs of star formation. Fig. 1.3 shows an example of a filament with a mass per-unit length high enough to make it gravitationally unstable ("Filament 6"), with a greater density of bound starless cores compared to elsewhere in the region. This is not the case, however, for every molecular cloud. It was found that in Orion A North, when using a criticality parameter calculated from an isothermal, self-gravitating cylinder from Ostriker (1964), there was no correlation between the criticality of a filament and star formation (Salji et al. 2015).

1.3 YSOs

Young Stellar Objects (YSOs) are astrophysical objects in the earliest stages of star formation. These sources are often embedded within dense material which absorbs much of the emission from the central protostar and reemits the radiation in the infrared. It is this

excess of infrared radiation which, in part, identifies a source as a YSO.

YSO as a term covers a range of ages and evolutionary stages depending on how the material is distributed. Section 1.3.1 describes these evolutionary stages from a deeply embedded object with an infalling disk and envelope, up to a revealed protostar with only an optically thin disk. As the protostar evolves and the distribution of the material changes this affects the shape of the observed spectral energy distribution (SED) of the star. Section 1.3.2 describes how the shape of the SED reflects the evolutionary stage of the YSO.

1.3.1 Stages of YSO formation

YSO formation begins with the fragmentation and collapse of a filament to produce a slowly rotating dense cores within clouds of molecular hydrogen which, themselves, gently collapse under gravity. As a cloud core collapses the kinetic energy of the falling gas is released as heat which gently raises the temperature and density of the central core until it becomes opaque to the infrared radiation and the temperature rise becomes more steep. This stage, where the collapsing material is opaque to radiative cooling, is referred to as the first hydrostatic core and it persists until the temperature of the core reaches a point where collisional dissociation of the molecular hydrogen begins. In this stage the energy input required to dissociate the H_2 allows the core to continue to collapse without significantly increasing the temperature. Eventually the core, with regions of atomic and molecular hydrogen, becomes unstable and collapses once again (Stahler and Palla 2005).

This second period of collapse results in the formation of a protostar in the centre of the cloud core and triggers the beginning of the next stage of evolution. In this stage the central protostar is surrounded by an envelope of dust and gas which accretes onto the protostar in an 'inside-out' manner (Larson 1969; Shu 1977). Most of the luminosity at this stage is due to material from the envelope falling onto the accretion shock at the surface of the protostar before settling onto the central object. This accretion luminosity is given by

$$L_{\rm acc} \equiv \frac{GM_*\dot{M}}{R_*}, \tag{1.3}$$

where M_* and R_* are the mass and radius of the protostar respectively, and $\dot{M}$ is the rate at which mass accretes onto the protostar. Due to rotation and the conservation of angular

momentum some of this forms a disk inside of the outer envelope (Terebey et al. 1984) [1]. The surface of the accretion disk then forms a part of the accretion shock as increasing amounts of material fall onto the disk first before making its way to the protostar. Stage 0 and Stage I objects are believed to be in this stage of evolution, where the boundary between the two is suggested to be when the mass of the core is equal to that of the envelope (Andre et al. 1993).

During the next phase of protostellar evolution the material from the envelope begins to preferentially fall onto the disk leading to 'openings' at the poles that allow collimated jets and bipolar outflows. As the mass of the envelope decreases, it is the remaining optically thick disk which characterises objects as being in Stage II. And, finally, before the protostar becomes pre-main-sequence the disk dissipates through processes such as planetary formation and evaporation making the protostar visible with little surrounding material. These sources with optically thin disks are characterised as Stage III.

1.3.2 YSO classification

As YSOs are deeply embedded within dense gas the evolutionary stage of a YSO is estimated based on properties that can be observed. Beginning with the earliest objects in star formation, dense cores, sources are distinguished between those without an internal source of luminosity, 'starless' cores, and those with, 'protostellar' (di Francesco et al. 2007; Enoch et al. 2008; Könyves et al. 2015). Not all starless cores will become protostars, only those which are gravitationally bound will form stars in the future. Such objects are referred to as 'prestellar' and can be distinguished from prestellar cores by their detection at submillimetre or millimetre wavelengths.

Once a central protostar is formed, observed YSOs are assigned a class that approximately follows the evolutionary stages described in Section 1.3.1. YSOs are classified on the basis of their spectral energy distribution (SED).

Unlike other sources which are reddened by the molecular cloud, photons emitted by the central protostar are absorbed and reemitted by the surrounding envelope and disk

1. For slowly rotating cores the disk forms after first core, however, in cloud cores that are rotating rapidly the disk can form before the first hydrostatic core (Bate 2011)

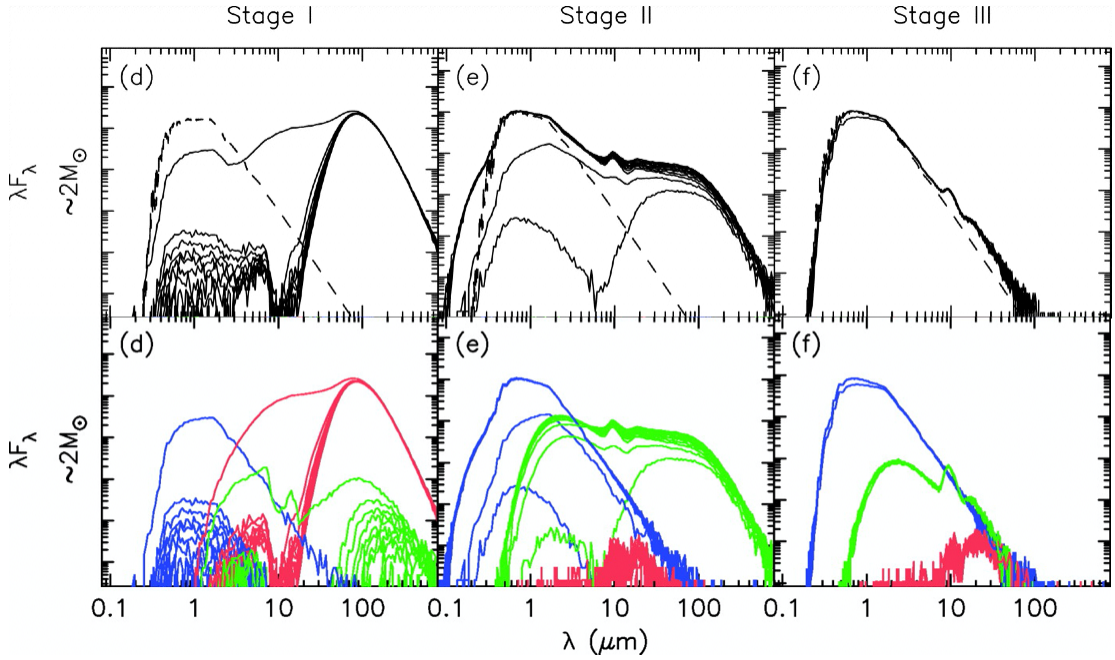

Figure 1.4: (above) example SEDs for 2 $M_{\odot}$ objects at evolutionary Stages 1, 2 and 3. (below) the same SEDs as above except the contributions to the SED from the star (blue), disk (green) and envelope (red) are shown. Figure reproduced from Robitaille et al. (2006).

until they are degraded to the longer wavelengths which are observed. The contributions of each of these components (the central protostar, disk and envelope) all contribute to the shape of the spectral energy distribution (SED) which determines the value of the spectral index,$\alpha_{\rm IR}$, given by

$$\alpha_{\rm IR} \equiv \frac{d\log(\lambda F_{\lambda})}{d\log\lambda}, \tag{1.4}$$

where F_{λ} is the flux density at wavelength λ, is the gradient of the spectral energy distribution between $\sim 2-24$ μm. Typical ranges of $\alpha_{\rm IR}$ given by Greene et al. (1994) are

Class 0/I: $0.3 \leqslant \alpha_{\rm IR}$;

Flat: $-0.3 \leqslant \alpha_{\rm IR} < 0.3$;

Class II: $-1.6 \leqslant \alpha_{\rm IR} < -0.3$;

Class III: $\alpha_{\rm IR} < -1.6$;

How the structure of the surrounding gas affects the SED, and hence the spectral index, is illustrated in Fig. 1.4 which presents simulated SEDs of YSO at Stages I, II and III as well as the individual contributions from the protostar, disk and envelope. We can see how the relative contributions of different structures in Fig. 1.4 reflect the same characteristic properties of Stage I, II and III YSOs.

An additional method of classifying YSOs from their emission spectrum is to use the bolometric temperature, $T_{\rm bol}$ (Myers and Ladd 1993). The bolometric temperature is the temperature of a blackbody which has the same mean frequency as the observed continuum spectrum. It was shown by Chen et al. (1995) that $T_{\rm bol}$ increases with the age of the source, and that each class could be associated with a range of $T_{\rm bol}$:

Class 0: $T_{\rm bol} < 70$ K;

Class I: $70 < T_{\rm bol} \leqslant 650$ K;

Class II: $650 < T_{\rm bol} \leqslant 2880$ K;

Class III: $T_{\rm bol} > 2880$ K;

Because $T_{\rm bol}$ increases with age, it is able to distinguish between Class 0 and Class I sources, whereas $\alpha_{\rm IR}$ cannot. The negative side of $T_{\rm bol}$ is that because it depends on the mean frequency of the observed spectrum it can be sensitive to the wavelengths at which data is available. If a source is lacking longer- or shorter-wavelength data then it will bias $T_{\rm bol}$ to higher or lower temperatures respectively (Visser et al. 2002; Evans et al. 2009).

Class 0 and Class I YSOs can also be distinguished by the ratio of bolometric luminosity, $L_{\rm bol}$, to submillimeter luminosity, $L_{\rm submm}$ (Andre et al. 1993), where $L_{\rm bol}$ is the luminosity measured across all available wavelengths and $L_{\rm submm}$, at wavelengths $\geqslant 350$ μm. Andre et al. (1993) propose that within a given cloud $L_{\rm submm}/L_{\rm bol}$ is approximately proportional to $M_\star/M_{\rm env}$ – the ratio of the mass of the protostellar core to the mass of the envelope – and sources with ratio $L_{\rm submm}/L_{\rm bol} \geqslant 5 \times 10^{-3}$ are classified as Class 0 protostars.

It is worth remembering that, ultimately, a YSOs stage and Class are two different properties. The former is an evolutionary stage while the latter is a label given due to observed properties, and so while classes should ideally represent the stage, they may not be accurate. For example, the observed α of a YSO is affected by column density along the line of sight which is higher when the source is obscured by a more massive envelope or behind a disk. Crapsi et al. (2008) estimate that sources with angles of inclination between 70° and 90° can be misclassified. They also estimate that this could mean as many as 34 per cent of Stage II sources could be mis-classified as class I or flat-spectrum sources due to orientation alone, though it could be possible to distinguish between Class I and II YSOs

based on the dust grains in the envelope and disk. This is because envelopes typically have cooler temperatures, 10 - 200 K, than disks (30 - 1000 K), and so envelopes are expected to contain grains coated in ice and disk grains with temperatures above 100 K are not.

1.3.3 Observational Data

As mentioned earlier YSOs spend much of their lifetime surrounded by optically thick dust and so require wavelength studies in the far-infrared and submillimetre regime to be detected. In addition, younger YSOs are typically more magnetically active than main-sequence and disk-bearing stars (Preibisch and Feigelson 2005). X-ray measurements can then be used to distinguish these YSOs from other infrared sources (Kuhn et al. 2014). X-ray data from the *Chandra X-ray Observatory* has been combined with infrared data from *Spitzer* and other sources in the Massive Young Star-forming complex in Infrared and X-ray (MYStIX) project to identify YSOs (Romine et al. 2016).

Beyond identification, spectral data is necessary to classify the YSOs into the different classes. As discussed in Section 1.3.2, YSOs are classified using a number of metrics, including $\alpha_{\rm IR}$ and $T_{\rm bol}$. YSOs classified using the spectral index are done so on the gradient of the SED between 2 - 24 μm, and so require observations in both the near- and mid-infrared to cover this range of wavelengths. The Two Micron All Sky Survey (2MASS) has measured near-infrared data for most of the celestial sphere in three near-infrared bands; J (1.25 μm), H (1.65 μm) and K_s (2.16 μm) (Skrutskie et al. 2006). Mid-infrared data has been observed in a number of molecular clouds by the *Spitzer* cores to disks (c2d) survey (Evans et al. 2003), the *Spitzer* Gould Belt survey, and the Wide-field Infrared Survey Explorer (WISE) (Wright et al. 2010). WISE observes at five bands between 3.4 and 22 μm, an observations with the *Spitzer Space Telescope* cover the wavelength ranges 3.6 to 8.0 μm with the Infrared Array Camera (IRAC), and 24 to 160 μm with the Multiband Imaging Photometer (MIPS) (Werner et al. 2004). Dunham et al. (2015) combine 2MASS, *Spitzer* and WISE photometry, in addition to SCUBA 450 and 850 μm data to construct the full SEDs for, and then classify, 2966 YSOs.

Observations in the far-infrared and submillimeter are required for colder or more embedded objects. The *Herschel* space observatory provides data in both of these regimes,

with the Photodetecting Array Camera and Spectrometer (PACS) and Spectral and Photometric Imaging REceiver (SPIRE). As discussed in Section 1.2.1, *Herschel* data can be used for tracing the gas within a molecular cloud, but it can also be used in the classification of dense cores. With data from the *Herschel* Gould Belt survey, Könyves et al. (2015) produced a census of the dense cloud cores in Aquila, classifying if they are either starless (lacking a compact internal luminosity source (di Francesco et al. 2007)) or protostellar. Similarly, Kirk et al. (2016) identified dense cores in three subregions of Orion-B (LDN 1622, NGC 2023/2024, and NGC 2068/2071) from the 850 μm emission maps from SCUBA-2.

1.4 YSO distributions within molecular clouds

1.4.1 Prestellar core distributions

As precursors to YSOs, knowledge of the distributions of prestellar cores provides both insight and comparison to the distributions of YSOs within molecular clouds. Observations support a filament-first scenario for dense core formation with prestellar cores, and dense cores in general, preferentially found in regions of high column density, for example it was found in Perseus, Serpens and Ophiuchus that 75% of protostars were found above $A_{\rm v} \geq 9.5$ mag, $A_{\rm v} \geq 10$ mag and $A_{\rm v} \geq 25.5$ mag respectively (Enoch et al. 2008).

This is potentially due to a threshold column density required for core-formation. It was found that the core formation efficiency (CFE), the ratio of core mass to cloud mass within a column density bin, of candidate prestellar cores in Orion B (Könyves et al. 2020) and Aquila (Könyves et al. 2015) both show a sharp increase at a visual extinction of $A_{\rm v} \sim 7$ mag and afterwards remains roughly constant.

It is argued that this threshold could be related to the critical line mass above which filaments are thought to become unstable and fragment into cores (André et al. 2010). As mentioned in section 1.2.2, prestellar cores and embedded protostars are preferentially observed on top of dense filaments with a mass per unit length above the critical line mass in Eqn. (Inutsuka and Miyama 1997). 71% of the pre-stellar population in Orion A were found on top of filaments (Polychroni et al. 2013) and 60% of prestellar cores in Aquila are observed on top of filaments with a critical or higher line mass (André et al. 2010).

In addition to a column density dependence of the distribution of prestellar cores, there is also evidence that prestellar cores form within clusters – a set of stars gathered in space such that there is a local overdensity. (Kirk et al. 2016) analysed the clustering properties of dense cores identified from the 850 μm emission maps from SCUBA-2 in three sub-regions of Orion B: LDN 1622, NGC 2023/2024 and NGC 2068/2071. The clusters were identified finding the minimum spanning trees of the cores (see section 1.5.2) and then pruning branches of the minimum spanning tree longer than a critical branch length (Gutermuth et al. 2009). It was found that 69%, 71% and 75% of cores in L1622, NGC 2023/2024 and NGC 2068/2071, respectively, were associated with clusters. The two-point correlation function (2PCF) and mean surface density of companions (MSDC) (explained in detail in section 1.5.3) in these regions also appear to indicate two scales in clustering, one at a characteristic separation of neighbouring clusters between 2 and 3 pc and one at a separation of more distant clusters between 4 and 5 pc. A similar application of the 2PCF in Perseus, Ophiuchus and Serpens indicates that prestellar cores in these regions are also clustered (Enoch et al. 2008).

1.4.2 Protostar distributions

1.4.2.1 Distributions as a function of age

Using the locations of Class 0/I, Class II & Class III YSOs it is possible to get an impression of how star formation within a cloud may have evolved over time. Because the different classes of YSO are expected to have different characteristic time scales, the ratios of the numbers of YSOs of different classes give a characteristic age for a region. Two ways in which a change in the distribution as a function of age could be interpreted is that star formation moved throughout the cloud over time as the gas around older stars depletes and younger stars are forming in newly dense regions of gas, or the older stars have continually migrated away from these regions over time.

Figure 1.5 shows the surface densities of the different classes of YSOs in the Rosette Molecular Cloud (RMC) superimposed on top of a column density map produced by Ybarra et al. (2013) using data from *Spitzer* and *Chandra* observations. It is clear from Figure 1.5 that star formation in the RMC has not been a continuous and uniform process.

By looking at the ratios of the different classes in the central cloud it appears that star formation has progressed from being concentrated along the north-west section towards the central dense region in the north (Ybarra et al. 2013).

As for migration it has been shown in numerical simulations that during the star formation process YSOs can be ejected from their birthplaces due to gravitational interactions with other stellar objects in an unstable multiple system (Bate et al. 2003; Bate 2009). This ejection effect should manifest as a difference in velocity between the protostar and the background molecular cloud. When looking at the radial velocities, the more prevalent these ejection interactions are, the greater the velocity dispersion of protostars should be, compared to the velocity dispersions of the molecular cloud. From a sample of 31 Class I and flat-spectrum YSOs from different clouds, Covey et al. (2006) place a limit on the effects of the stellar interactions on the radial velocity dispersion of ~ 2.5 kms^{-1}. As the velocity dispersions of the star-forming regions of the molecular cloud is ~ 1.4 kms^{-1}, the difference between the cloud and YSO velcotiy dispersions is sometimes interpreted as the approximate motion of a protostar with respect to its neighbouring cloud environment. This difference produces a value of order ~ 1 kms^{-1} (Salji et al. 2015).

In general, the distributions of earlier stage YSOs are more clustered and more associated with dense gas than their more evolved counterparts(e.g. Mairs et al. 2016) this can be seen in Fig. 1.5. This has been seen in multiple regions, including in NGC 2264, where, through application of the 'INDICATE ' statistic (See Section 1.5.1), that the cumulative distribution function of each class was significantly different (Buckner et al. 2020). This effect can also be seen in Fig. 1.6 which shows how the relative surface densities of YSOs in different column density bins changes as a function of class in Serpens South. These young objects are also the most likely to be near the locations in which they formed, given the short time for migration ($0.46 - 0.72$ Myr) (Dunham et al. 2015) and their close association with high column densities. Indeed, no statistically significant evidence was found for a diffusion of the YSOs in Orion A, Orion B, California and Taurus from their sites of formation (Lombardi et al. 2013; Lada et al. 2013).

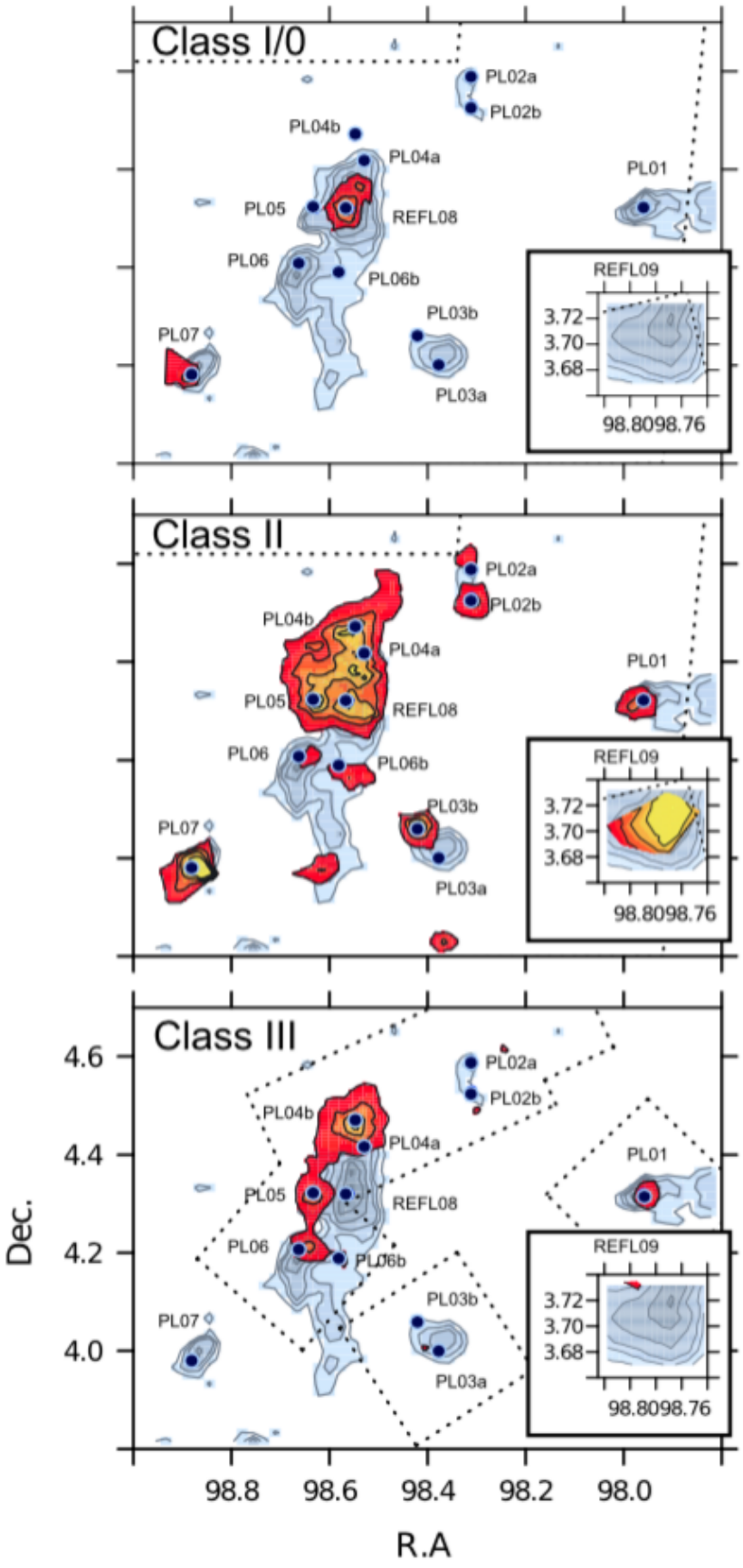

Figure 1.5: Stellar surface density maps of the Rosette molecular cloud overlaid on top of visual extinction maps. The stellar surface density contours are μ = 2.6, 4.6, 7.7 and 12.8 stars pc^{-2} and the visual extinction contours are A_v = 8, 10, 12, 14, 16, and 20. Figure reproduced from Ybarra et al. (2013)

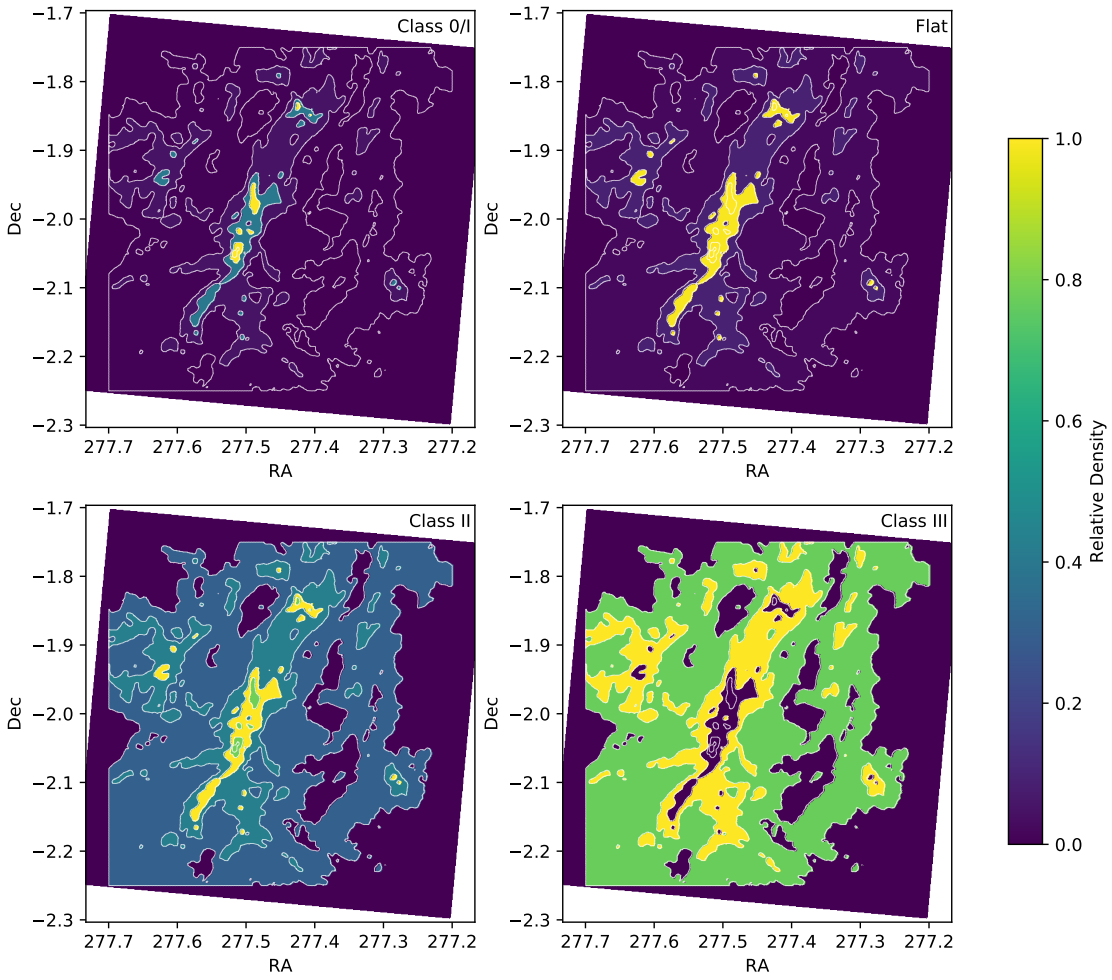

Figure 1.6: Surface densities of Class 0/I, flat, Class II and Class III YSOs in different column density bins in Serpens South. The contours (white) outline the column density bins, and the colour scale shows the YSO surface density relative to the highest density.

1.4.2.2 Kennicutt-Schmidt relation in local molecular clouds

As discussed in the previous subsection, young protostars are likely to be close to their site of formation. We can therefore use the distributions of early-stage YSOs within a molecular cloud as indicators for where star formation has recently occurred within a molecular cloud. Such observations allow us to produce and test models to better understand star formation as a process.

A relationship of particular importance is how column density relates to star formation. Observations clearly support a positive correlation between star formation and column density. On a galactic scale, the Kennicutt–Schmidt (K–S) law is frequently used to measure the relationship between star formation rate and column density. The K–S law proposes a relationship of the form $\Sigma_{\mathrm{SFR}} \propto \Sigma_{\mathrm{Gas}}^{\mu}$ where Σ_{SFR} and Σ_{Gas} are the surface densities of star formation rate and gas respectively, and μ is the power-law index. Schmidt (1959) initially proposed a power-law index of ≈ 2, and later observations suggest a lower value of $\simeq 1.4$ (Kennicutt 1998).

Despite the difference in scale, K–S relations have also been measured in local star forming regions using the surface densities of YSOs as a measure of Σ_{SFR}. Values of μ vary between regions though most values lie between $\mu = 1.8$ and $\mu = 2.3$, which can be seen in Table 1.1. Variations in μ are in part due to differences in method, for example Gutermuth et al. (2011) uses a nearest-neighbour-based method while Lombardi et al. (2013) use a parametric density model. Some variation is due to a difference in the astrophysical objects used which, as discussed in section 1.4.2.1, have different typical relationships with column density. Additionally, some variation is simply due to different power laws being present in different regions.

It is interesting that these local K–S relations produce such similar values despite the differences in methodology and region. Furthermore, in sources which look at multiple regions with the same methodology such as Pokhrel et al. (2020), Lombardi et al. (2013) and the results in chapter 4 the average power-law measured is ≈ 2.

A value of μ describes the general change in surface density of star formation as a function of column density; however, this does not mean the star formation is distributed

Table 1.1: YSO Kennicutt-Schmidt power law estimates for different regions.

Region	Power-law	Source
AFGL 490	1.8 ± 0.3	(Pokhrel et al. 2020)
Aquila North	1.8 ± 0.1	(Pokhrel et al. 2020)
Aquila South	2.3 ± 0.2	(Pokhrel et al. 2020)
Auriga-California Molecular Cloud	4	(Harvey et al. 2013)
California	3.31 ± 0.23	(Lada et al. 2017)
Cep 0B3	2.2 ± 0.1	(Pokhrel et al. 2020)
Cygnus X	1.9 ± 0.1	(Pokhrel et al. 2020)
G305	2.50 ± 0.04	(Willis et al. 2015)
G326.4	1.91 ± 0.05	(Willis et al. 2015)
G326.6	1.77 ± 0.04	(Willis et al. 2015)
G333	2.86 ± 0.03	(Willis et al. 2015)
G351	2.30 ± 0.03	(Willis et al. 2015)
Mon R2	2.1 ± 0.1	(Pokhrel et al. 2020)
MonR2	2.67 ± 0.02	(Gutermuth et al. 2011)
NGC 2264	1.8 ± 0.1	(Pokhrel et al. 2020)
NGC 6634	2.08 ± 0.08	(Willis et al. 2015)
Ophiuchus	1.87 ± 0.03	(Gutermuth et al. 2011)
Ophiuchus	1.9 ± 0.1	(Pokhrel et al. 2020)
Orion A	1.99 ± 0.05	(Lombardi et al. 2013)
Orion A	2.2 ± 0.1	(Pokhrel et al. 2020)
Orion B	2.16 ± 0.10	(Lombardi et al. 2014)
Orion B	2.3 ± 0.2	(Pokhrel et al. 2020)
Perseus	2.4 ± 0.6	(Zari et al. 2016)
Perseus	2.1 ± 0.1	(Pokhrel et al. 2020)
S140	1.8 ± 0.2	(Pokhrel et al. 2020)

Note: Different sources use different methods and different astrophysical objects to produce power-law estimates.

evenly throughout the entire cloud; for example, in Fig. 1.5 we can see that the class 0/I YSOs are isolated to two small regions. Column density appears to be a strong predictor of where stars are forming in molecular clouds; however, if column-density-independent effects are significant enough they may influence the observed distribution of YSOs in clouds. Using spatial analysis it is possible to analyse the distributions of YSOs within molecular clouds to determine if they are consistent with being distributed evenly according to a K–S law.

1.5 Methods of Spatial Analysis in Astrophysics

Statistics which describe the spatial distributions of objects in astrophysics are important for informing models of star formation. The methods described in this section have been used to examine the distribution of objects relative to one-another and to their environ-

ment in order to identify types of spatial distributions.

1.5.1 Nearest-neighbour methods

Nearest-neighbour methods utilise the distance between a star and its nth closest neighbouring star. Nearest-neighbour distances are useful as characteristic distances as they scale naturally depending on the number and separations of objects in the study window. Since changing n changes the scales of interaction to which the tests are sensitive nearest-neighbour methods are sensitive to both large inter-cluster-scale distances, as well as smaller distances between cluster members. Though the downside of small-n values is that they are more sensitive to small-scale irregularities within the distribution and so n is typically a compromise between resolution and sensitivity. The downside of nearest-neighbour methods is that some stars will be counted more times than others as a given star can be the nearest-neighbour to multiple stars.

Some nearest-neighbour statistics are used to test local, small-scale patterns by looking at the environment within the nth nearest-neighbour distance for each individual star. Examples of local statistics include the stellar surface density maps of Gutermuth et al. (2011) discussed in section 1.4.2.2, where the stellar surface density at grid position $[i, j]$ is given by $\Sigma(i.j) = \frac{n}{\pi r_n^2(i,j)}$, where r_n is the distance to the nth nearest star, and 'INDICATE' (Buckner et al. 2019), a statistic which uses the number of stars within a 5th-nearest-neighbour distance to determine if individual stars are locally clustered compared to a random field.

Other nearest-neighbour techniques look at the distribution of nearest-neighbour distances to describe more global features of the data-set. These often include looking for characteristic scales such as the median distance (Masiunas et al. 2012) or peaks in histograms of nearest neighbour distributions (Gutermuth et al. 2009). Some studies compare the distribution of nearest-neighbour distances to distributions measured from random realisations to compare to randomness (e.g. Teixeira et al. 2006).

1.5.2 Minimum Spanning Tree

The minimum spanning tree (MST) is the unique set of edges connecting a collection of stars[2] that minimises the total edge length with no cycles, producing a network. The no cycles condition means that there are no connected loops within the tree, and so each star is connected to each other star via a single path. Similarly to the nearest-neighbour methods which look at the distributions of nearest-neighbour distances, global statistics can be applied to the distribution of branch lengths with a MST. In fact, the MST contains every first nearest-neighbour distance, with some additional edges to link isolated sub-clusters, and so should produce similar results.

The advantage of the MST is that it connects every star into a network, effectively linking members into a single cluster. Clusters can then be separated from the tree by 'pruning' specific edges, and, because there are no cycles, each sub-cluster is a MST to which global statistics can be applied. This means that the results of the global and local statistics applied to sub-clusters are directly correlated to the methods used to identify clusters as they both stem from the MST.

Because the MST can identify clusters, methods have been developed that describe how stars are distributed within a cluster. One example is the Cartwright and Whitworth $Q-$parameter (Cartwright and Whitworth 2004), which is the ratio of the normalised mean edge length from the MST to the normalised mean point separation. The measured values of Q then correspond to a different type of structure within the cluster, either fractal substructure ($0.45 \lesssim Q < 0.8$), a uniform distribution ($Q \approx 0.8$), or a smooth radial distribution ($0.8 \lesssim Q < 1.5$).

Another example of a global test statistic is the total edge length of the MST. Cantat-Gaudin et al. (2018) calculate the total edge length, $l_{\rm obs}$, for a cluster and compare it to the mean total edge length from 2000 random circular distributions with the same number of stars. If $l_{\rm obs}$ is smaller than the random mean by more than one standard deviation the distribution is classified as non-random. I recommend, instead, testing if $l_{\rm obs}$ is among the $2k(s+1)$th most extreme total edge lengths measured from the random distributions,

2. In general the objects that are connected with a MST are referred to as nodes but I will use 'star' to keep language consistent between methods.

where s is the number of random simulations and $k = \alpha/2$, where α is the desired significance level of the test (see section 2.3). This method has two main advantages: the first is that it does not assume a Gaussian distribution of total edge lengths in the simulation, and the second is that it is a two-sided test which can identify both clustering and inhibition.

1.5.3 Pair correlation functions

Pair correlation functions look at the complete set of inter-point distances to describe how points are distributed relative to one another. The two most common forms of pair correlation function in astrophysics are the two-point correlation function (2PCF) (Davis and Peebles 1983) and the closely related Mean Surface Density of Companions (MSDC) (Larson 1995).

The 2PCF, $w(r)$, is defined as the excess probability of finding a star at a distance r from another star, and can be estimated with the natural estimator (using the notation of Enoch et al. (2008))

$$w(r) = \frac{H_s(r)}{H_r(r)} - 1, \tag{1.5}$$

where $H_s(r)$ is the number of pairs with separation between $\log(r)$ and $\log(r + dr)$, and H_r is the number of pairs within the same separation bin for a random distribution. The MSDC is the surface density of stars as a function of distance from another star; it is related to the 2PCF by

$$\Sigma_{\mathrm{MDSC}}(r) = \lambda[1 + w(r)], \tag{1.6}$$

where λ is the average stellar density across the entire study region. Spatial features of the distribution can then be interpreted directly from the shape of the pair-correlation function, or by comparison of the results to expected distributions under a null hypothesis.

Results from the 2PCF can tell us some direct information about the distribution. We can see from Eqns. 1.5 and 1.6 the expected values of the 2PCF and MSDC for a random distribution of stars are 1 and λ respectively. If measured values are higher than randomness it indicates scales that show clustering, and lower values indicate scales showing regularity or inhibition. Because pair-correlation functions use all of the inter-star distances, they are able to cover a large range of spatial scales, and so are particularly good

at indicating scales with interesting features, like clustering.

What requires care, however, is interpreting the underlying distribution measured by the pair-correlation function because multiple distributions can produce the same result. As an example, it was found that the MSDC of the pre-main sequence stars in the Taurus–Auriga molecular cloud and the Orion Trapezium cluster showed two distinct power-law slopes with a break at ~ 0.4 pc (Larson 1995; Simon 1997) and a power-law of ≈ -2 on large-scales. This break in power-law was interpreted as a change from a binary regime on small scales to a fractal distribution at large scales. It was shown, however, by Bate et al. (1998) that the MSDC of Taurus–Auriga can also be fit by a model with randomly distributed clusters of stars, and, Orion Trapezium can be fit by a sphere of stars with a uniform volume density in the core and a volume density proportional to r^{-2} outside of the core.

1.5.4 Summary

I have introduced three of the types of methods most commonly used within astrophysics to analyse the spatial distributions of YSOs, and provided examples as to how they have been applied. As I have shown, each of these methods measure a different property of the spatial distribution and can be applied in different ways depending on the objective of the study.

Nearest-neighbour methods are useful for finding characteristic scales associated with different levels of interaction because, unlike pair-correlation functions which look at all of the inter-point distances, nearest-neighbour methods use the subset of inter-point distances defined by n. The MST is a useful tool in studies of clusters – especially where only a subset of the points within a study region are of interest – and, finally, pair-correlation functions are good at identifying clustering and regularity at different scales.

While statistics can be constructed for the MST, such as the total branch length, these are comparable in sensitivity to tests using the first-nearest-neighbour distances except they require the additional step of constructing the MST. The nearest-neighbour methods and pair-correlations functions have the additional advantage of analytical solutions for

uniform random distributions; these allow for easier comparisons of measured distributions to randomness.

In the next chapter I introduce four methods from spatial statistics, two nearest-neighbour methods, one pair-correlation function and the integral of the pair-correlation function. I have chosen these methods as they are statistics designed to analyse complete spatial location data (i.e. they look at every star within a study region rather than a subset). I will also introduce *spatial point processes*, which are models I can use to distribute points within a study region, and how, by considering star formation as a spatial point process, the statistics I have chosen will help in analysing the distributions of YSOs.

Chapter 2

Introduction to Spatial Statistics

2.1 Introduction

Spatial statistics provides a framework to understand the processes which underlie the spatial distributions of objects. Methods from spatial statistics have been applied in a wide range of fields including ecology (Barot et al. 1999; Wiegand et al., 2009), geophysics (Johnson 2017) and astrophysics (Davis and Peebles 1983; Larson 1995; Kauffmann et al. 1999). This chapter will introduce some core concepts of spatial statistics which will be applied in chapters 3 and 4.

There are two important concepts in spatial statistics: the spatial point process and the spatial point pattern. A spatial point process is a mechanism which is used to generate a number of points, called events, in a study region and a spatial point pattern is a realisation of this mechanism. The objective of applying spatial statistics is to consider properties of an observed distribution of events and infer details of the spatial point process that may have produced the spatial point pattern. In astrophysics, for example, we can consider star formation to be a stochastic process and populations of YSOs to be events and ask what form of spatial point process could have produced observed YSO distributions.

The first spatial point process that is often considered is that of complete spatial randomness (CSR) (see section 2.4.1). With CSR, the probability of placing an event is uniform across the entire space, and the number of events within an arbitrary region fol-

lows a Poisson distribution with mean $\lambda|A|$ where λ is the first-order intensity and $|A|$ is the area contained within A. While distributions of objects of interest are typically not random, a preliminary comparison to CSR is useful as a comparison to CSR reveals whether objects are more or less aggregated than random. This can be important as in any random distribution of points one would expect to find small areas of aggregation which by eye could appear significant. A rejection of CSR using spatial statistics then informs us that processes of interest are present, and whether these processes result in statistically significant aggregation or regularity.

What follows analysis of CSR is the application of models which are informed by knowledge of the system (see Section 2.4). These models contain first-order effects, second-order effects, or some combination of both. First-order effects affect the number of events as a function of global position, while second-order effects affect the number of events as a function of position relative to other events.

In section 2.2 four of the most commonly applied summary statistics, Diggle's G function (G), the 'free-space function' (F), Ripley's K function (K) and the O-ring statistic (O), are discussed. The generation of global confidence envelopes which are used to test null hypotheses are introduced in section 2.3 and, finally, the two models which the distributions of YSOs will be tested against, CSR and the inhomogeneous Poisson point process in section 2.4.

2.2 Summary Statistics

When examining spatial point patterns, one method of analysis is to calculate a summary statistic for the pattern and determine whether the measured value is consistent with some null hypothesis. As discussed in Section 4.1, often the first null hypothesis is that of CSR, a pattern which can be produced with a homogeneous Poisson point process. For CSR the distribution of the number of events in regions $\boldsymbol{s}$ within the study space follows a Poisson distribution with mean $\lambda|\boldsymbol{s}|$, where λ is the first-order intensity of the process and $|\boldsymbol{s}|$ is the area contained within $\boldsymbol{s}$.

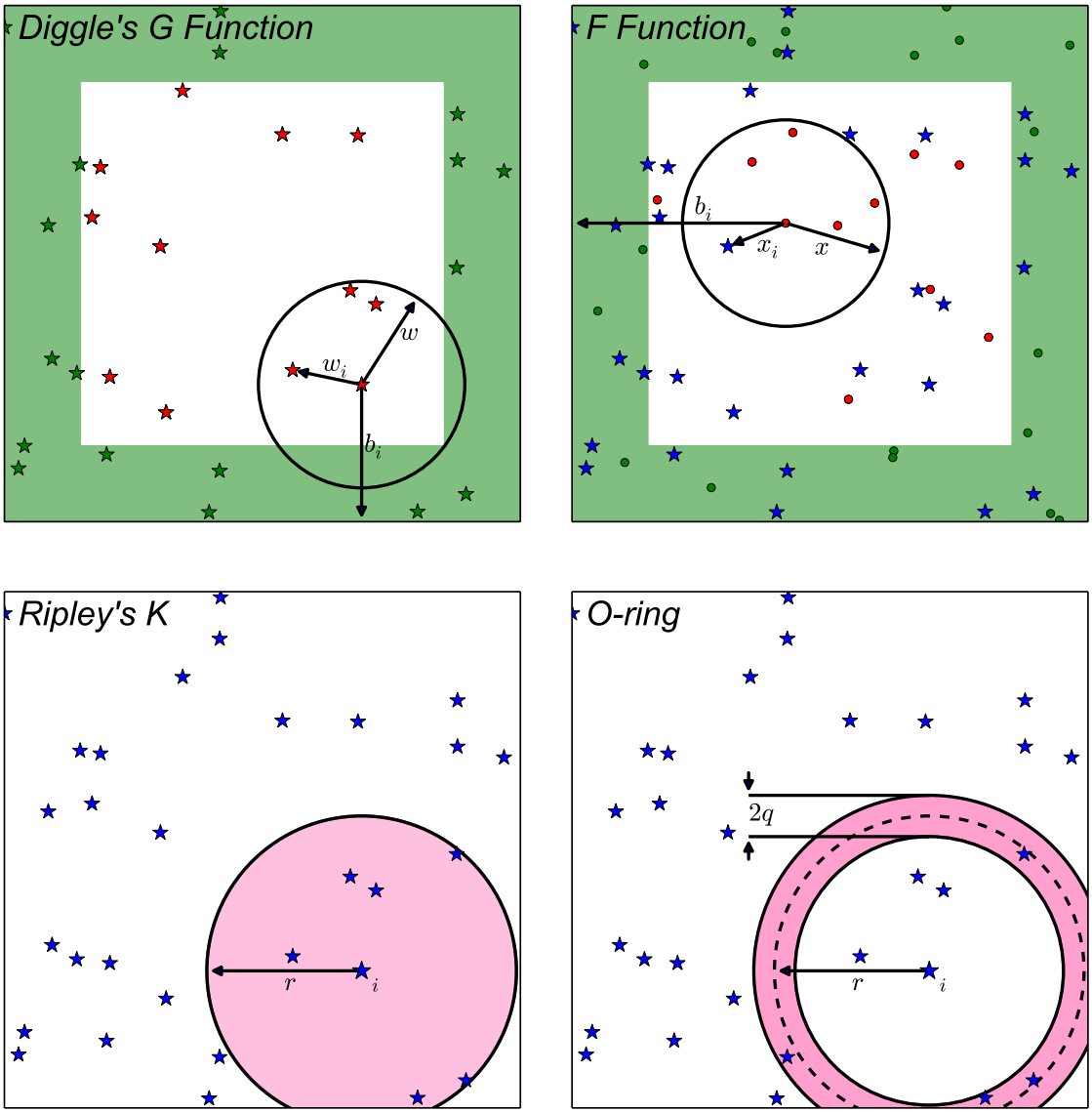

Figure 2.1: Visual schematic of summary statistics G, F, Ripley's K and O-ring. The stars show locations of events for one realisation of CSR, the circular markers are the arbitrary positions used by F. The highlighted border in G and F shows the regions where stars/positions will be excluded by the edge-correction method. The shaded areas in Ripley's K and O-ring are those which intersect between the study region and the circle/annulus where points are counted for a given r.

2.2.1 First-Order Statistics

First-order effects produce large-scale variation in the positioning of events, where the first-order intensity, and therefore the mean number of events per unit area, is a function of position, $\lambda(x)|\boldsymbol{s}|$. The first-order intensity is given by the equation (Diggle 2013)

$$\lambda(x) = \lim_{|dx|\to 0} \left\{ \frac{\mathrm{E}[N(dx)]}{|dx|} \right\}, \tag{2.1}$$

where $\lambda(x)$ is the first-order intensity at position x, dx is the small region containing x, $\mathrm{E}[N(dx)]$ is the expected number of events contained within dx and $|dx|$ denotes the area within dx. Processes where $\lambda(x)$ is a constant, λ, are called said to be stationary. These effects are typically environmental as the probability of an event being in region $\boldsymbol{s}$ has no dependence on neighbouring stars. For CSR the probability of an event having a nearest neighbour distance less than or equal to w is $1 - e^{-\pi\lambda w^2}$, which can be tested by looking at the distribution of event-event and position-event nearest neighbour distances using Diggle's G function and the free-space function.

The first test, Diggle's G function, is an estimate of the cumulative probability distribution of nearest neighbour distances between events. For a given distance, w, the uncorrected $\mathrm{G}(w)$ is the number of events with a nearest neighbour closer than w divided by the total number of events (Diggle 2013); see Fig. 2.1. However, given that events closer to the boundaries of the test region have fewer neighbours than more centrally located events an edge correction is needed. For G and F use the border method of edge-correction, wherein only events with a distance to the closest boundary, b_i, greater than w are used to estimate $\mathrm{G}(w)$ (Dale and Fortin 2014). For rectangular-shaped windows this method is relatively easy to implement. The region where $b_i \leq w$ is shown in Fig. 2.1 by a green border. It is worth noting that the border method is one of many edge-correction methods including the Kaplan-Meier estimator (Baddeley and Gill 1997).

The free-space function (F) is similar to G except it is an estimate of the cumulative probability distribution of nearest neighbour distances between randomised positions in the study window and their nearest event, which will be referred to as x_i (see Fig. 2.1). This makes F more sensitive to patterns with empty space and aggregation, hence the name

'free-space' function. Similar to G, F was also calculated using the border edge-correction method. In equation form G and F are written as (Gignoux et al. 1999)

$$\mathrm{G}(w) = \frac{\#\{w_i \leq w, b_i > w\}}{\#\{b_i > w\}} \tag{2.2}$$

$$\mathrm{F}(x) = \frac{\#\{x_i \leq x, b_i > x\}}{\#\{b_i > x\}}, \tag{2.3}$$

where #{...} is shorthand for the number of positions or events that satisfy the condition. For the case of CSR the expected value for G(w), E[G(w)], has the value, E[G(w)] = $1 - e^{-\pi\lambda w^2}$ (Dale and Fortin 2014; Feigelson and Babu 2012). E[F(x)] is identical. For CSR G and F are equal, however departures from CSR cause these values to differ due to their sensitivities to clustering and empty space respectively. G and F may also be combined to form other measures, such as the Lieshout-Baddeley J function, $J(r) = \frac{1-\mathrm{G}(w)}{1-\mathrm{F}(x)}$ which is less sensitive to edge-effects compared to G and F (Feigelson and Babu 2012).

2.2.2 Second-Order Statistics

Second-order tests look at the distributions of pairs of event, observing the change in probability compared to a random distribution of an event being within, or at, a distance r from another star.

Ripley's K is one of the most commonly used spatial statistics. Multiplying K by the first-order intensity of the point pattern gives the expected number of events within the distance r of an arbitrary event excluding the central event (Wiegand and Moloney 2004), $\lambda\mathrm{K}(r) = \mathrm{E}[\#\{\text{points in area } \boldsymbol{s}\}]$, where $\boldsymbol{s}$ is a circle of radius r centred on an arbitrary event. Numerically K can be estimated with (Dale and Fortin 2014)

$$\mathrm{K}(r) = \frac{|\boldsymbol{A}|}{N^2} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} h_i(r) I_r(i,j), \tag{2.4}$$

where N is the number of events, $|\boldsymbol{A}|$ is the area of the test region $\boldsymbol{A}$, $h_i(r)$ is a weighting allocated to each event for edge-correction purposes and $I_r(i,j)$ is a selection function taking the value of 1 if $d_{ij} \leq r$ and 0 otherwise. d_{ij} is the distance between points i and j. In this case $h_i = |s|/|\boldsymbol{A} \cap s_i|$, the inverse of the proportion of the circle s_i around event i

that intersects with $\boldsymbol{A}$. This area can be calculated algebraically for rectangular windows and computationally for arbitrarily shaped regions (see Section 2.2.3). Fig. 2.1 presents $(s_i \cap \boldsymbol{A})$ as a highlighted area around point i, and points inside the highlighted area satisfy $I_r(i,j)$. $\hat{\mathrm{K}}(r)$ is not visually intuitive, however the difference between $\hat{\mathrm{K}}(r)$ and the expected value if the pattern was CSR, $\lambda\pi r^2$, gives a statistic that, while it does not contain more information, reduces the variance and is more visually intuitive (Dale and Fortin 2014),

$$\mathrm{L}(r) = \sqrt{\mathrm{K}(r)/\pi} - r. \tag{2.5}$$

For this convention, positive values of $\mathrm{L}(r)$ imply clustering/aggregation and negative values imply overdispersion of the points. For this and the next chapter where Ripley's K is applied the L function will be used for plotting results, however the results will be discussed under the name Ripley's K.

The O-ring statistic, sometimes referred to as the neighbourhood density function (Perry et al. 2006) or the mean surface density of companions (MSDC) (Larson 1995), is similar to K except it is calculated using annuli instead of circles. The statistic is a measure of the average density that an event would observe at a given distance r,

$$\mathrm{O}(r) = \frac{|\boldsymbol{A}|}{N^2} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} h_i(r) I_r(i,j) \tag{2.6}$$

where $I_r(i,j)$ is a selection function, taking the value of 1 if $r - q \leq d_{ij} \leq r + q$, with q being the half-width of the annulus (see Fig. 2.1), and h_i is a weighting for edge correction. h_i is the inverse of the proportion of the area of the annulus that lies within the boundaries. With K, each circle of a given radius contains all of the points that were previously counted at smaller radii and as such each K measurement is not independent of those measured before it; however, annuli do not need to overlap, and, as such, the values for $\mathrm{O}(r)$ are able to be uncorrelated across the set of radial values r_i. This is true for separations greater than $2 \times q$. Using non-correlated separations allows an analytical approach when creating confidence envelopes (see Section 2.3). For CSR $\mathrm{E}[\mathrm{O}(t)] = \lambda$, so values larger than λ indicate a greater-than-average density at that distance, and vice-versa. O-ring requires

a choice of width which introduces the problem of binning. A rule of thumb is to begin with

$$q = \rho/\sqrt{\lambda} \tag{2.7}$$

with ρ taking values between 0.1-0.2 (Law et al. 2009; Yongtao 2006) and change as appropriate to maximise smoothing of the data while minimising loss of information.

2.2.3 Calculating Weights

When analysing spatial statistics it is often not the case that every possible spatial point has been mapped or labelled within the study window. This causes an artificial discontinuity at the boundaries of where the data was collected. To counter these edge effects a number of methods are used. In this section I will explain the methods of computing one method in particular, the area of the intersect between a circle and a study window when the circle is constrained to be centred within the study window.

A general brute-force method suggested in Dale and Fortin (2014) is to place regularly spaced sectors around the point and then count the number that are within the boundary. This method works for even irregular boundaries and the accuracy of the solution can be improved by adding more sectors. The negative side of this method is that a larger number of sectors comes at a greater computational cost, which is made more severe by patterns with more events. If the study region is broken up into cells then calculating the amount of area within the study region is similarly calculated by counting the area within the cells which both lie within the circle and the study window.

If, however, the study region is a rectangle then I recommend the following analytical method which provides quicker code times and more accurate solutions. There are four different ways in which a circle with a radius less-than-or-equal-to half the length of the shortest side of the rectangle can interact with the boundaries, as shown in Fig. 2.2, and even more for larger radii. An effective edge-correction algorithm needs to be able to account for all of these. Fortunately, most of the area calculation can be performed by

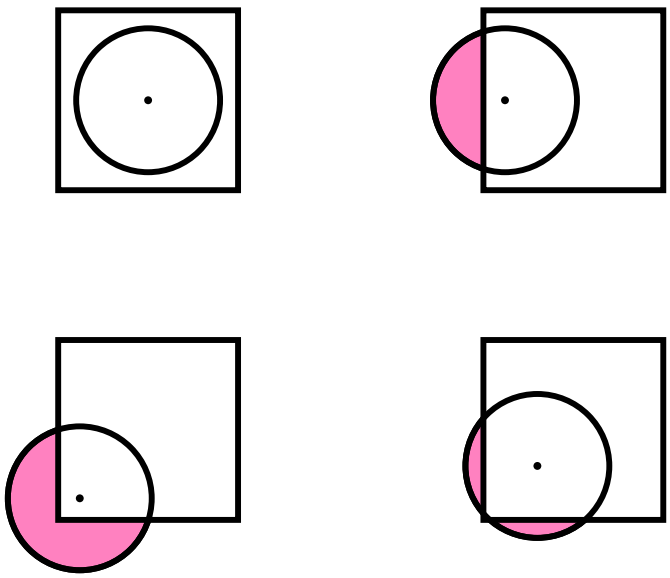

Figure 2.2: The different ways in which a circle can overlap with a rectangular study window when the radius of the circle is less than half of the side length.

simply finding the area above a chord,

$$A = \frac{\pi r^2}{2} - r^2 \arcsin\left(\frac{h}{r}\right) - h\sqrt{r^2 - h^2} \tag{2.8}$$

where r is the radius of the circle and h is the height of the chord. The problem is then simply finding the relevant values of h.

Because the study region is a rectangle finding chords is simply a matter of locating which boundaries are at a distance less than r with $h = b - r$, where b_i is the distance to the ith boundary. When none of the rectangle's corners have been enveloped by the circle each segment is isolated and so the area of the circle outside of the study window is simply the sum of the areas of the individual segments.

In the case where a corner is within the circle, the segments formed by the boundaries overlap and the overlapped area is then counted twice. This can be accounted for by calculating and subtracting the overlapped area, so that it is counted only once. The overlapped area can be broken down into a right-angled triangle plus a segment, as demon-

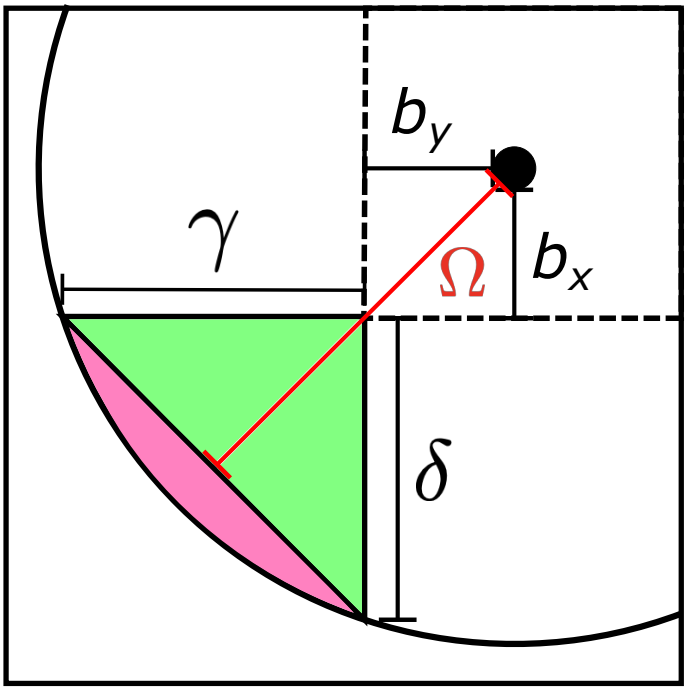

Figure 2.3: Diagram presenting the breakdown of the area contained within an overlapped corner.

strated in Fig. 2.3, where the sides of the triangle are given by

$$\gamma = \sqrt{r^2 - b_x^2} - b_y \tag{2.9}$$

$$\delta = \sqrt{r^2 - b_y^2} - b_x, \tag{2.10}$$

where b_x and b_y are the distances between the centre of the circle and two adjacent boundaries, and the height of the remaining chord is given by

$$\Omega = \sqrt{\frac{1}{2}\left(r^2 + b_x\sqrt{r^2 - b_y^2} + b_y\sqrt{r^2 - b_x^2}\right)}. \tag{2.11}$$

The final algorithm to sum the area outside of the rectangle is to add together the segments formed by crossing the boundaries of the study region and then subtract the area outside of each corner. The area inside the rectangle is then πr^2 minus the area outside the rectangle. This works for any size of circle and any number and combination of corners and edges crossed by circle and can be extended to annuli by simply applying the algorithm to the inner and outer radii of the annuli.

2.2.4 Applying to Grids

The equations discussed in Sections 2.2.1 and 2.2.2 require some adaptation to function when applied to spaces that are not continuous. Such is the case in Chapter 4 where the O-ring statistic is applied to data stored within arrays. The following methodology by Wiegand and Moloney (2004) is therefore applied to gridded data to calculate O-ring.

$$\mathrm{O}(r) = \frac{\sum_{i=1}^{N} \mathrm{Points}[\mathrm{R}_i^w(r)]}{\sum_{i=1}^{N} \mathrm{Area}[\mathrm{R}_i^w(r)]}, \tag{2.12}$$

where $\mathrm{R}_i^w(r)$ is an annulus with radius r and width $w = 2q$ centred on the ith star, and the operators $\mathrm{Points}[\mathrm{R}_i^w(r)]$ and $\mathrm{Area}[\mathrm{R}_i^w(r)]$ count the number of events and area contained within $\mathrm{R}_i^w(r)$ respectively. If border boundary conditions are applied then only the points and area within the intersection of $\mathrm{R}_i^w(r)$ and the study region are counted.

The number of events within $\mathrm{R}_i^w(r)$ in a grid is defined by

$$\mathrm{Points}[\mathrm{R}_i^w(r)] = \sum_u \sum_v \mathrm{S}(u,v)\mathrm{P}(u,v)\mathrm{I}_r(x_{u,v}, y_{u,v}, x_i, y_i). \tag{2.13}$$

Here, u and v are the row and column indices of the grid respectively. $\mathrm{S}(u,v)$ is an indicator function equal to 1 if cell (u,v) is contained within the study window and zero otherwise and $\mathrm{P}(u,v)$ is the number of events contained within (u,v). Finally, I_r is another selection function to determine if a cell is within the annulus. The Area operator within Eqn. 2.6 is

$$\mathrm{Area}[\mathrm{R}_i^w(r)] = \sum_u \sum_v \mathrm{S}(u,v)\mathrm{A}(u,v)\mathrm{I}_r(x_{u,v}, y_{u,v}, x_i, y_i), \tag{2.14}$$

where $\mathrm{A}(u,v)$ is the amount of area contained within cell (u,v).

2.3 Significance Testing

A summary statistic produces a single value (in this case at a given r) to represent a chosen facet of the data being measured. This value can be compared to the distribution of values the summary statistic takes under some null hypothesis, H_0 to determine if the null hypothesis can be rejected with some significance level, α. The distribution of a summary

statistic under a simple null hypotheses, such as that of CSR, can sometimes be found analytically (Wiegand et al. 2016); otherwise it may be sampled computationally through repeated realisations of H_0.

To test the significance of a single measurement it is sufficient to find the distribution of the statistic under the null hypothesis and determine if the measured value is among the kth-most-extreme values where, for a two-sided distribution,

$$\alpha = 2k/(n+1) \tag{2.15}$$

and n is the number of simulated patterns of H_0. However, when distributions are tested against many spatial scales their significance may be determined by the application of global confidence envelopes.

2.3.1 Confidence Envelopes

A confidence envelope represents the uncertainty in the summary statistic under H_0 as a function of radial distance. If the confidence envelope is exceeded then H_0 is considered to be rejected with the significance level attributed to the confidence envelopes.

A normal pointwise envelope tests if an observed statistic, $T_1(r)$, is among the k^{th} most extreme values among the set of $T_i(r)$ values for i = 1, ..., n+1, where n is the number of simulated patterns for a null hypothesis H_0. A pointwise envelope can then reject a null hypothesis with the probability in Eqn. 2.15 at a single distance scale r if the envelope is exceeded. This probability is, as discussed earlier, valid for testing only a single distance scale as every individual spatial scale constitutes its own independent test. The impact of testing multiple scales is that the probability of rejection increases (above $2k/(n+1)$) with every additional measurement. Functionally this means that the spatial scale being tested must be chosen in advance of plotting the envelope. Visually inspecting the pointwise envelope is equivalent to simultaneously testing every spatial scale looking for examples of statistically significant results and is an example of data dredging.

Rather than pointwise envelopes global confidence envelopes cover the range of acceptable values of the summary statistic and, if it is exceeded, rejects H_0 with the pre-

determined significance α. Unlike a pointwise envelope, a global confidence envelope is one which allows simultaneous testing of all probed radial distances while keeping the probability of rejection controlled, meaning that the graphical plot of $T_1(r)$ among the envelopes is a valid tool for testing hypotheses, as well as looking for scales at which H_0 is rejected. The global envelopes described here are *directional quantile maximum absolute difference* (MAD) envelope tests, following the description from Myllymäki et al. (2017), where the lower and upper bounds are given by

$$T^u_{\text{low}}(r) = T_0(r) - u_\alpha |\underline{T}(r) - T_0(r)| \tag{2.16}$$

$$T^u_{\text{upp}}(r) = T_0(r) - u_\alpha |\overline{T}(r) - T_0(r)| \tag{2.17}$$

respectively, where $T_0(r)$ is the expected value under H_0, $\overline{T}(r)$ and $\underline{T}(r)$ are the 2.5 per cent upper and lower quantiles of the distribution of $T(r)$ under H_0 and u_α is the parameter used to determine the confidence level of the envelopes. If the distribution of $T(r)$ is not known then it can be estimated from measurements of simulated patterns $T_i(r)$, for $i = 2, 3, \ldots, n+1$.

The u_i values are calculated using,

$$u_i = \max \left[f(r)^{-1} (T_i(r) - T_0(r)) \right] \tag{2.18}$$

where $f(r)$ is

$$f(r) = \begin{cases} \overline{T}(r) - T_0(r) & \text{if } \{T(r) \geq T_0(r)\}, \text{ or} \\ \underline{T}(r) - T_0(r) & \text{if } \{T(r) < T_0(r)\}, \end{cases} \tag{2.19}$$

and u_α is the $\alpha(n+1)$th largest value from the collection of u_is.

It is possible to produce global confidence envelopes analytically without requiring simulation when the distribution of a statistic is known (Wiegand et al. 2016). These analytical envelopes function best when the tests for each r are independent, which for O-ring

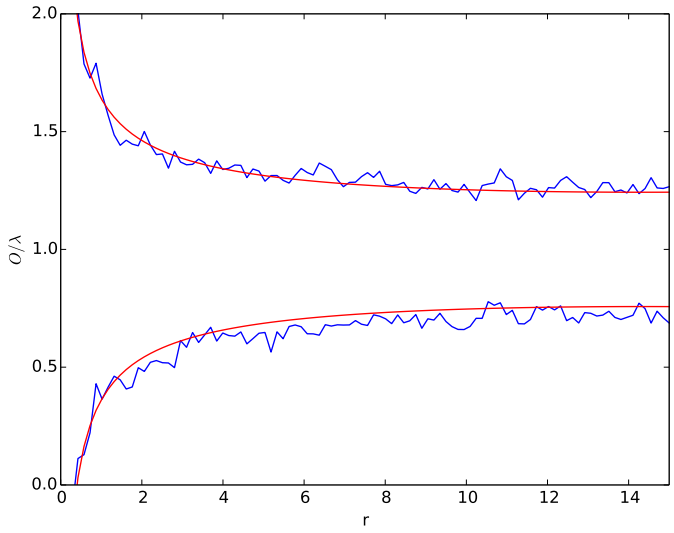

Figure 2.4: Comparison of analytical (red) vs empirical Monte Carlo (blue) confidence envelopes for O-ring for CSR with $\lambda = 0.133$ in a 30x30 arb. units area.

requires that each r position be offset from the previous by at least q and is not possible for Ripley's K due to the nature of the test. While these analytical envelopes were not used, they validate the empirical envelopes, as shown in Fig. 2.4.

2.3.2 Envelope Examples

Envelopes require three components, a null hypothesis, knowledge of how the chosen statistic is distributed under H_0 and a significance level, α. Section 2.4 describes two models which can be used as a basis of a null hypothesis as well as how to produce realisations of the processes, and Section 2.3.1 describes how to take a set of measured statistics ($T_i(r)$, for $i = 1, 3, \ldots, n + 1$) and convert them into a confidence envelope.

Fig. 2.5 demonstrates two examples of spatial point processes, along with the results measured with the methods described above. The left-hand side shows a realisation of CSR with $\lambda = 1$ keeping the number of events constant. In this example all of the statistics agree that CSR cannot be rejected as a null hypothesis for the spatial point process as they remain within the confidence envelope.

The right-hand side demonstrates a single centralised cluster produced from a uniform circular probability density function with radius $R = 3$ – an inhomogeneous Poisson process with $\lambda = 0$ outside the circle. Though there is some disagreement about which

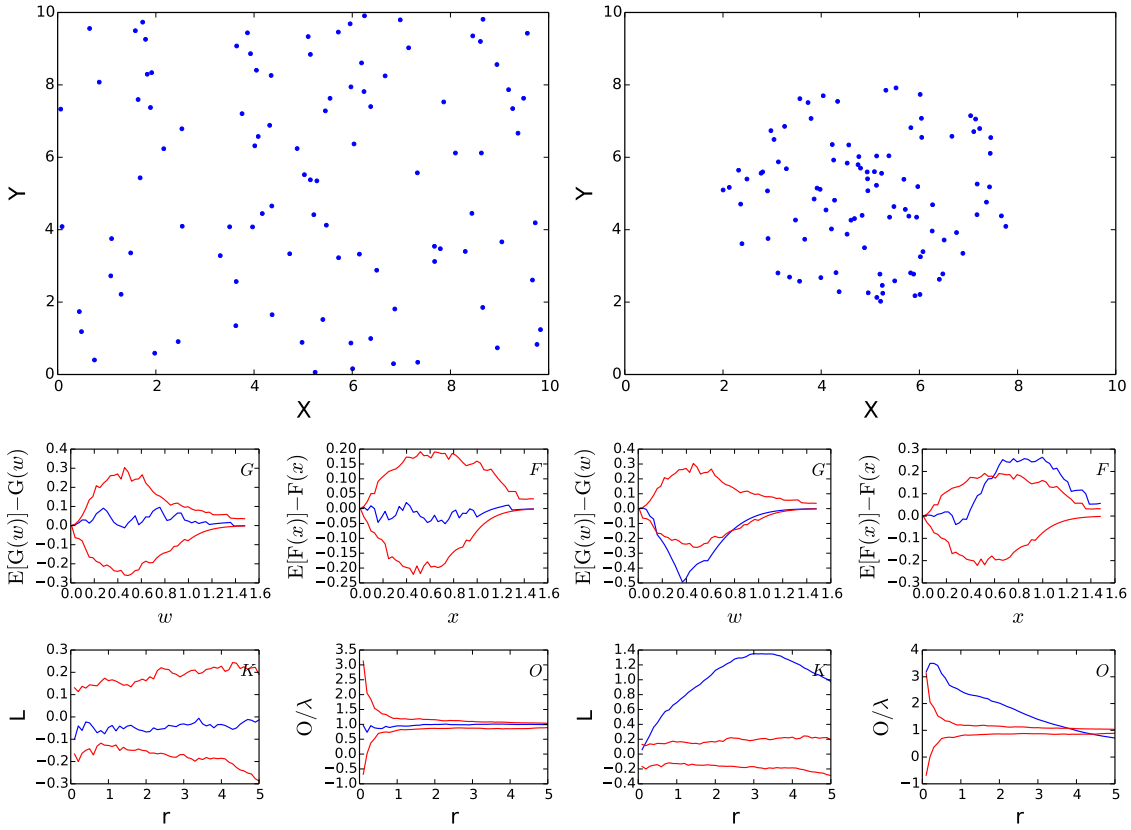

Figure 2.5: Left: (above) Realisation of CSR with $\lambda = 1.0$, (below) results of G, F, K and O-ring in blue with 95 per cent global confidence envelopes for CSR in red. Right: (above) Realisation of centralised cluster with $\lambda = 1.0$ and $R = 3$, (below) G, F, K and O-ring results in blue with 95 per cent global confidence envelope for CSR in red.

spatial scales show this deviation, all four statistics reject CSR for this pattern. K and O-ring are used to study the second-order effects of the spatial point process; however, the cluster example in Fig. 2.5 is not a product of second-order effects. This is because the star positions were drawn from a non-homogeneous uniform probability distribution, and as such the probability of placing a star was independent of the presence of any existing stars. The deviations, therefore, are due to large first-order effects which have given the impression of clustering due to second-order effects: a phenomenon known as virtual aggregation (Wiegand and Moloney 2004). Because of this, when the rejection of CSR using O-ring and K are discussed, the nature of the effect (either first or second-order) will not be commented on – only that the measured pattern is consistent or inconsistent with CSR.

This cluster example also demonstrates how virtual aggregation can manifest due to the choice of study window. As discussed, the points within the cluster are uniformly distributed and so a study window contained entirely within the cluster would detect no

significant deviation from CSR – assuming λ was recalculated for the new space. However, because the window is larger than the cluster the average density assigned to the space is lower which makes the point spacing within the cluster more statistically significant. It is therefore important to consider the boundaries when testing for a given null hypothesis.

Excursions from the envelope represent spatial scales at which the null hypothesis used to produce the envelope can be rejected. It is important, however, to understand what the tests are measuring to be able to interpret these regions. O-ring is a measure of the average density that would be observed at a distance r from an event, therefore an excursion at a given scale is indicating a significantly over or under-populated region. L, being a cumulative statistic, is measured up to a given radial distance and describes up to which scales a pattern rejects the null hypothesis of the envelope. For this reason O-ring is typically easier to interpret though both statistics contain the same information, as O-ring is related to the differential of K. Therefore, any scales indicated by excursions are both a function of the statistic and the null hypothesis represented by the envelope.

2.4 Models

Measured summary statistics are compared to their distributions under an assumed null hypothesis spatial point process. Numerous types of models exist within the literature that encompass different mechanisms by which events may be distributed, these include clustering, inhibition and environmental mechanisms as well as their combinations.

2.4.1 Complete Spatial Randomness

Complete spatial randomness (CSR) is a stationary Poisson point process where the probability of placing an event is unaffected by either location or the presence of other events. One measure of a stationary Poisson point process is that the first-order intensity (given by Eqn. 2.1) of the process is a constant value across the entire space.

Another aspect of a homogeneous Poisson process is that the counts of events, N, within areas, $\boldsymbol{s}$, follow a Poisson distribution with a mean $\lambda\boldsymbol{s}$, i.e.

$$P_{\lambda\boldsymbol{s}}(N) = \frac{(\lambda\boldsymbol{s})^N e^{-(\lambda\boldsymbol{s})}}{N!}. \tag{2.20}$$

Once again, this is because the first-order intensity, λ, is a constant across the entire study region.

Because CSR represents a state of complete non-interaction it is often the first model which distributions of events are tested against as the rejection of CSR could imply that the locations of events (or stars) is influenced in some way. For example, a pattern with an O-ring statistic which exceeds the upper bound of a CSR confidence envelope tells us that the events are more clustered than would be expected from a random distribution, as well as the scales at which the events are clustered.

What remains for the homogeneous Poisson processes is an estimate of λ. The first-order intensity can be estimated from the data; however, unless the estimated function, $\hat{\lambda}$, is equal to the λ of the underlying point process, the simulated patterns will be made by a different random process to the original. This is possible because the number of events in a realisation of a Poisson process is expected to follow a Poisson distribution. If $\hat{\lambda}$ is not equal to λ the Monte Carlo tests are invalid which manifests as a more conservative test. The alternative is to keep the number of points used in the realisations constant to produce an exact Monte Carlo test(Baddeley et al. 2014).

2.4.2 Inhomogeneous Poisson Point Processes

Unlike CSR, the first-order intensity of an inhomogeneous Poisson point process is a function of position, $\lambda(x)$, but is similar to CSR in that the probability of placing an event is independent of neighbouring events. Because of these two points first-order models are sometimes attributed to environmental factors, as one might expect to find a gradient in the concentration of a population following the availability of materials, nutrients or habitability.

Given a first-order intensity function, the number of events in $\boldsymbol{s}$ follows a Poisson distribution with a mean number of events given by,

$$\langle N \rangle = \int_{|s|} \lambda(x) dx. \tag{2.21}$$

Simulating an inhomogeneous Poisson process is then equivalent to sampling a pdf given

by $\lambda(x)$.

Lewis and Shedler (1979) suggests the following algorithm for efficiently producing realisations of inhomogeneous Poisson point processes by random thinning. First, produce a realisation of a homogeneous Poisson process with a first-order intensity equal to the maximum of $\lambda(x)$. Then loop over each of the events with coordinates (X, Y), generating a random number, U, from the interval $(0, 1)$, and retaining the event if $U \leq \lambda(X, Y)/\lambda^\star$, where $\lambda^\star$ is the maximum of λ.

For a study region represented by a grid, however, I propose the following algorithm which does not require thinning and allows for cells which cover different areas. First, produce a normalised pdf by calculating for each cell the probability of finding an event in u, given that the event has survived placement,

$$\text{prob}(u|\text{S}) = \frac{\lambda(u)|u|}{\sum_v \lambda(v)|v|}, \tag{2.22}$$

where u and v are cell indices, $|u|$ and $|v|$ are cell areas and S indicates the event has survived placement and was not lost while thinning. Then produce an array by cumulatively summing over $\text{prob}(u|\text{S})$, $B_u = \Sigma_{v=1}^{u}\text{prob}(v|\text{S})$. This maps each cell to an interval between 0 and 1: $(0, B_1)$ for $u = 1$ and $[B_{(u-1)}, B_u)$ otherwise. To position an event a random value, U, is generated and an event is positioned in the cell with the relevant interval.

We can see in Eqn. 2.22 that $\text{prob}(u|\text{S})$ is equal to the expected fraction of events within u and so makes intuitive sense, and I now present the derivation of Eqn. 2.22 to show that it produces the desired behaviour. The joint-distribution to be sampled is composed of two parts: a probability proportional to area the cell and a probability proportional to the first-order intensity in the cell.

Using the product rule I present the probability of placing an event in u and surviving,

$$\text{prob}(\text{S}, u) = \text{prob}(\text{S}|u) \times \text{prob}(u), \tag{2.23}$$

and by applying Bayes' theorem we can see that the probability of finding an event in u

given that it survived placement is

$$\text{prob}(u|\text{S}) = \frac{\text{prob}(\text{S}|u) \times \text{prob}(u)}{\text{prob}(\text{S})}. \tag{2.24}$$

The likelihood and prior of Eqn. 2.24 are the probability of an event surviving placement, given that it has has been placed in u and the probability of placing an event in u due to its area respectively. These two terms are given by

$$\text{prob}(\text{S}|u) = \frac{\lambda_u}{\sum_v \lambda_v}, \tag{2.25}$$

and

$$\text{prob}(u) = \frac{A_u}{\sum_v A_v}. \tag{2.26}$$

Note that the probability given to survival in Eqn. 2.25 is not that of Lewis and Shedler (1979) and is instead the normalised first-order intensity.

We can calculate prob(S) by marginalising prob(S, u) over all cells to find

$$\text{prob}(\text{S}) = \frac{\Sigma_u \lambda_u A_u}{\Sigma_v \lambda_v \Sigma_v A_v}, \tag{2.27}$$

Substituting Eqns. 2.25, 2.26 and 2.27 into Eqn. 2.24 gives us the final result of Eqn. 2.22.

The left panel of Fig. 2.6 shows the positions of the Class 0/I YSOs in Serpens South with an example realisation of Eqn. 2.22 with $\lambda(u)$ equal to

$$\lambda(u) \propto \Sigma_{\text{Gas},u}^{2.05}, \tag{2.28}$$

where $\Sigma_{\text{Gas},u}$ is the column density in cell u.

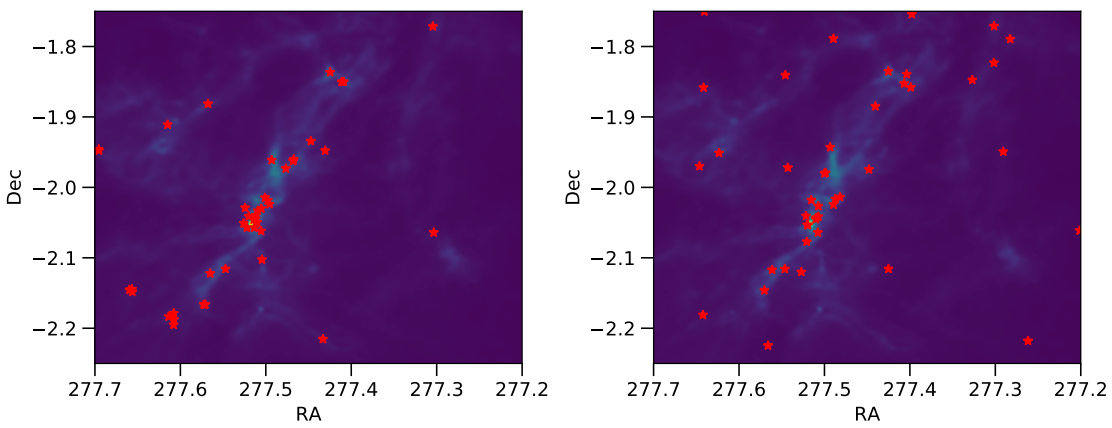

Figure 2.6: (left) Class 0/I YSOs in Serpens South and (right) random realisation of YSOs with $\mu = 2.05$ (see Section 4.4.4.3) plotted on *Herschel* 18.2″ column density maps.

Chapter 3

Applied Spatial Statistics

3.1 Introduction

In astrophysics, studies of star-formation and how it evolves over time consider the spatial distributions of young stars with respect to each other and the environment. In general, earlier stage YSOs are more densely clustered and situated closer to the densest regions of a cloud, while the more evolved YSOs tend to be more dispersed and further removed from dense gas (see section 1.4.2.1). However, the process producing this difference is uncertain.

YSOs which are spatially related are sometimes grouped together into clusters and properties of these clusters are investigated to provide insights into the physics underlying the distribution. These clusters often have members being chosen via some algorithm for example the minimum spanning tree (MST) (Gutermuth et al. 2009) (see section 1.5.2). Using the methods from spatial statistics introduced in chapter 2 it is possible to identify the presence of clustering within a dataset, and whether that clustering is statistically significant.

This can be done by testing if distributions of YSOs are consistent with the spatial point process of CSR. If we assume distributions of YSOs are realisations of a star-formation spatial point process then a rejection of CSR indicates the presence of clustering or aggregation greater than one would expect from a random distribution. This is because

CSR represents a state of complete non-interaction between stars, and between stars and their environment.

Within astrophysics, statistics such as the two point correlation function (2PCF) and MST tend to be used to classify the degree of clustering within a system rather than testing if the distribution has been produced by a random mechanism. This is not surprising, as systems within astrophysics are not usually assumed to be entirely random. On the other hand methods from spatial statistics are applied frequently in ecology and epidemiology, where it is less clear that features of interest such as trees and disease occurrences could be correlated and a statistically significant deviation from what is expected from CSR is cause for further study (Barot et al. 1999; Wiegand et al., 2009; Velázquez et al. 2016).

In this chapter I will apply the four methods introduced in chapter 2 – Diggle's G function (G), the 'free-space' function (F), Ripley's K function (K) and the O-ring statistic (O) – to datasets containing both cluster-members and randomly distributed background objects. The aim is to determine the ability of each of these tests to detect clustering when background objects contaminate the data. As a point of comparison I also apply a method of testing for CSR using the MST, which was introduced briefly in section 1.5.2 and I describe in more detail in section 3.3.4.

3.2 Trials

I tested the four statistics for their ability to reject CSR for a single cluster in the presence of background noise. To set up this experiment, a cluster with population N_c was projected on to a field of background stars of number N_{bg}. The positions of the cluster population were generated by sampling a two-dimensional Gaussian probability density function with Gaussian width $\sigma = R/2$ centred on the midpoint of the study window, where R is a characteristic radius chosen for the cluster. The background population were N_{bg} sets of (x,y) coordinates randomly distributed across the study window.

Multiple realisations of each combination of N_c and N_{bg} were tested and their summary statistics compared to 95 per cent confidence envelopes for a null hypothesis of CSR. The envelopes were generated using $N_c + N_{bg}$ points as described in Section 2.3.2. The

fraction of realisations that reject CSR are then a measure of the likelihood that a given pattern with $N_{\rm c}$ cluster members and $N_{\rm bg}$ background points will be correctly identified as non-random.

For an order of magnitude indication to determine if the tests were functioning correctly a signal-to-noise calculation was performed for the study region as whole. By assuming the number of clustered stars can be estimated by subtracting an estimation of the background population, a possible measure of the SNR for the study window is

$$\mathrm{SNR} = \frac{N_c}{\sqrt{N_c + N_{bg}}}. \tag{3.1}$$

3.3 Results

The following results display a measure of the efficacy of these statistical tests in rejecting CSR when the number of cluster points, $N_{\rm c}$, and the total number of points in the window, $N_{\rm total}$ are controlled separately. To produce the confidence envelopes 199 realisations of CSR were generated for each value of $N_{\rm total}$. To test the different numbers of cluster points and background points 30 simulations of each ($N_{\rm c}$, $N_{\rm total}$) combination were generated. The fraction of simulated observations which reject CSR, from here on referred to as the rejection fraction, is an estimate of the empirical probability that a given cluster of $N_{\rm c}$ members projected in-situ with $N_{\rm bg}$ background members would be identified as non-random. As a metric to compare the different statistics, the total number of detections was divided by the total number of trials run within a range of $N_{\rm c}$ and $N_{\rm total}$ values and presented as the detectability score. The detectability score can take values from 0 to 1, representing the extremes of no trials rejected and all trials rejected respectively. For G, F and Ripley's K, the detectability scores were calculated from 10 unique computations of the entire $N_{\rm c}$ and $N_{\rm total}$ parameter space; for each variation in the annulus width for O-ring only one computation was produced and so uncertainties are not available for those values.

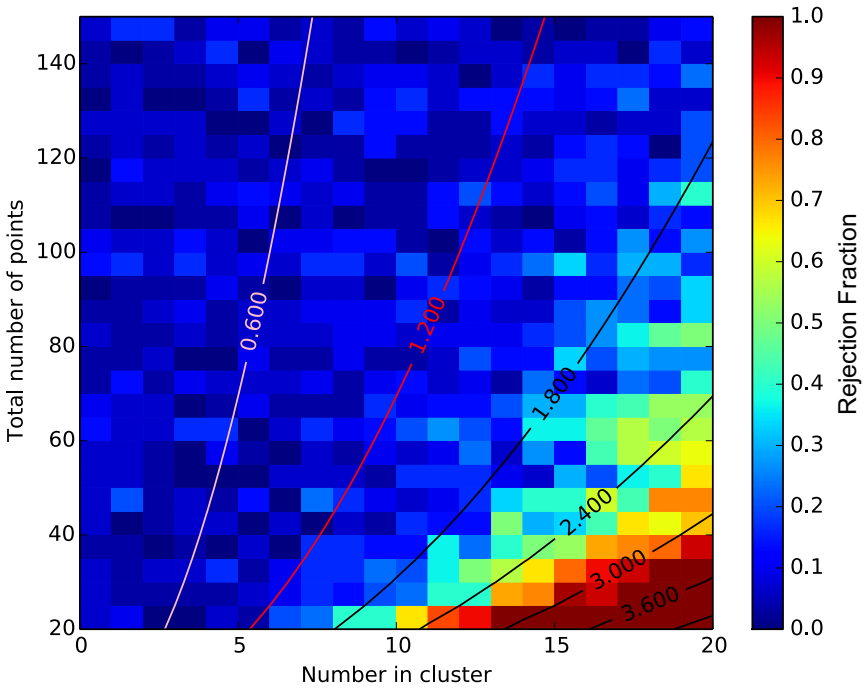

Figure 3.1: The rejection fraction with $P(H_0) < 5\%$ for Diggle's G function. The contours show the theoretical SNR (Eqn. 3.1)

3.3.1 Diggle's G and the 'empty space' function

The results using E[G(w)]-G(w) and E[F(x)]-F(x) are shown in Figures 3.1 and 3.2 respectively. For G, only the largest clusters with the least contamination consistently rejected CSR, while the results for F appear only weakly correlated with position in the parameter space. This is reflected in their detectability scores of 0.174 ± 0.004 and 0.089 ± 0.001 respectively.

3.3.2 Ripley's K

Ripley's K rejects far more clusters than F and G as shown in Fig. 3.3. The improved rejection fraction makes has a signal-to-noise effect whereby a given cluster can be masked by the background population. The row with zero cluster members are simply runs of CSR; with a confidence envelope of 95 per cent the chance of at least one false positive is 80 per cent for each value of N_{total}. The detectability score for Ripley's K was 0.550 ± 0.003.

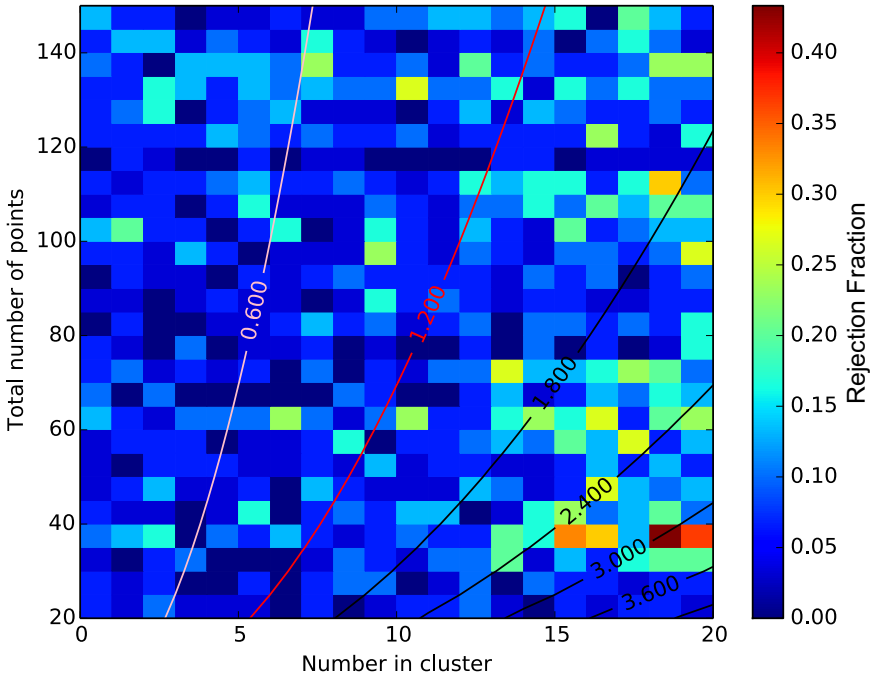

Figure 3.2: The rejection fraction with $P(H_0) < 5\%$) for F. The contours show the theoretical SNR (Eqn. 3.1). Note that the colour scale is over a much reduced range.

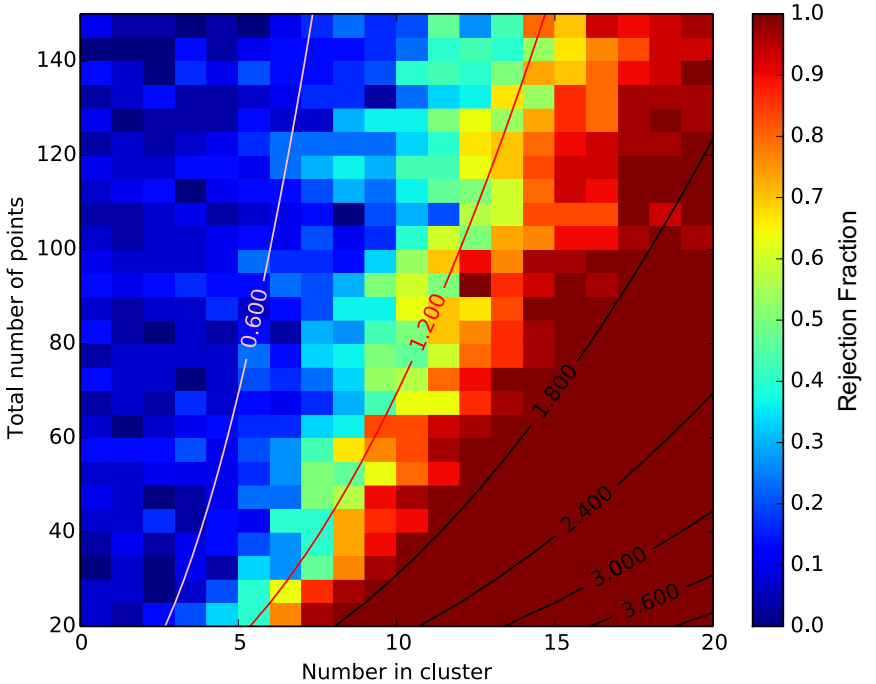

Figure 3.3: The rejection fraction with $P(H_0) < 5\%$ for Ripley's K. The contours show the theoretical SNR (Eqn. 3.1)

3.3.3 O-ring

O-ring has an additional parameter compared to the other tests — the width of the annuli. This width determines the amount of area contained within each annulus and therefore the number of cluster and non-cluster points contained within. The benefit of Equation 2.7 is that the bin widths are decided without prior knowledge of the existence or scale(s) of the cluster(s) in the study window. The results of using $\rho = 0.1$ and 0.2 in Equation 2.7 are shown in Fig. 3.8. The change in detectability scores between the two values can be seen in Table 3.1.

Having a different annulus width for each total number of points obfuscates the direct effect of the width on the detectability fraction and so Fig. 3.5 shows the same region of parameter space except the annulus width has been kept constant across all positions in the parameter space. Here the annulus width is some multiple of the radius of the cluster. The radius of the cluster is not a value often known *a priori*, however this demonstrates that an annulus with a width larger than the radius of the cluster begins to degrade the ability of the statistic to reject CSR. As is to be expected, the rejection fraction is both a function of the degree of clustering and the annulus width. Fig. 3.6 shows the effect of keeping the annulus width constant and adjusting the characteristic radius of the cluster. Overall the likelihood of rejection is decreased, shown in Table 3.1.

An alternative method for the annuli width is to use logarithmic widths, where the ratio of the outer to inner radius of each annulus is kept constant, making the width a function of the radial distance. Fig. 3.7 demonstrates four of these functions with half-widths given by

$$q = \rho r \tag{3.2}$$

for $\rho = 0.1, 0.3, 0.5,$ and 0.9. While there is an effect on the rejection fraction due to the bin width, as seen in Fig. 3.7, the results are consistent when $\rho \geq 0.3$. This is a promising result as this method requires no prior knowledge of the cluster width and no dependence on the number of points in the study region.

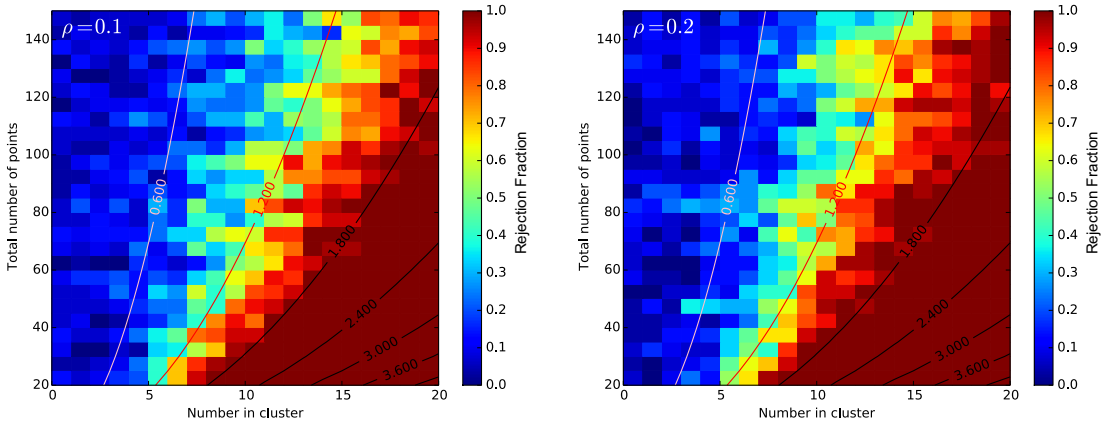

Figure 3.4: Rejection fraction as a function of annulus width for the O-ring statistic. Annulus widths decrease with the total number of points according to Equation 2.7. The contours show the theoretical SNR (Eqn. 3.1).

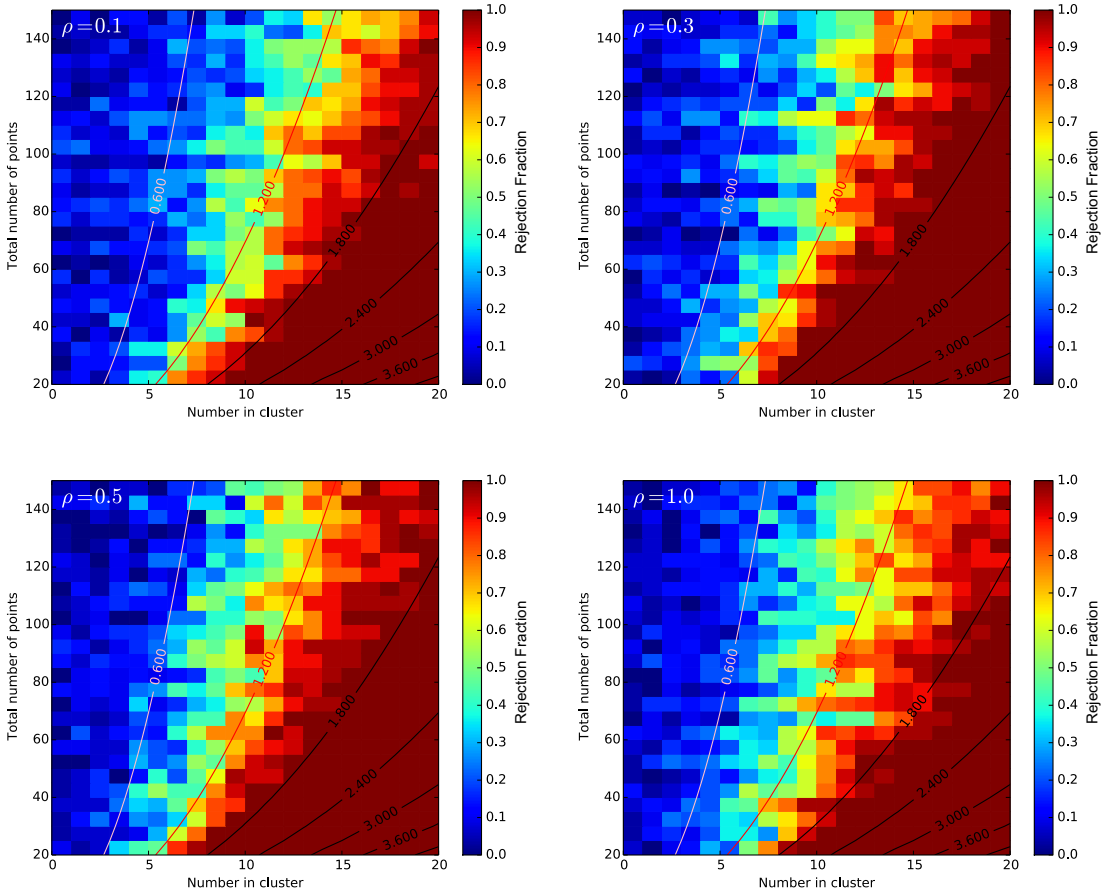

Figure 3.5: The rejection fraction with $P(H_0) < 5\%$ for O-ring statistic with fixed annuli widths. The widths used in each parameter space were given by $\rho \times R$ with $\rho = 0.1, 0.3, 0.5$ and 1.0 respectively and R is the radius of the cluster. The contours show the theoretical SNR (Eqn. 3.1).

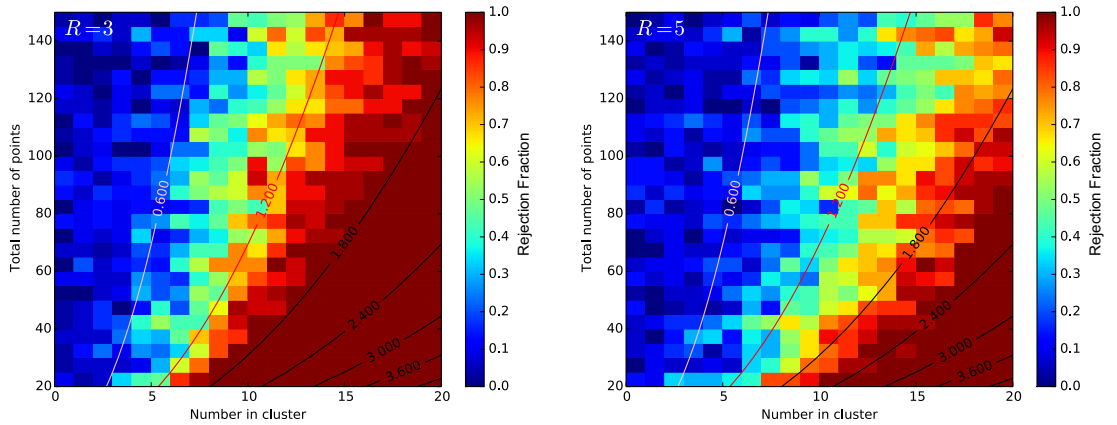

Figure 3.6: The rejection fraction with $P(H_0) < 5\%$ for O-ring statistic with fixed annuli widths for cluster radii 3 and 5. The annulus width has been kept constant for both sets of trials, $q = 1.5$ arb. units. The contours show the theoretical SNR (Eqn. 3.1).

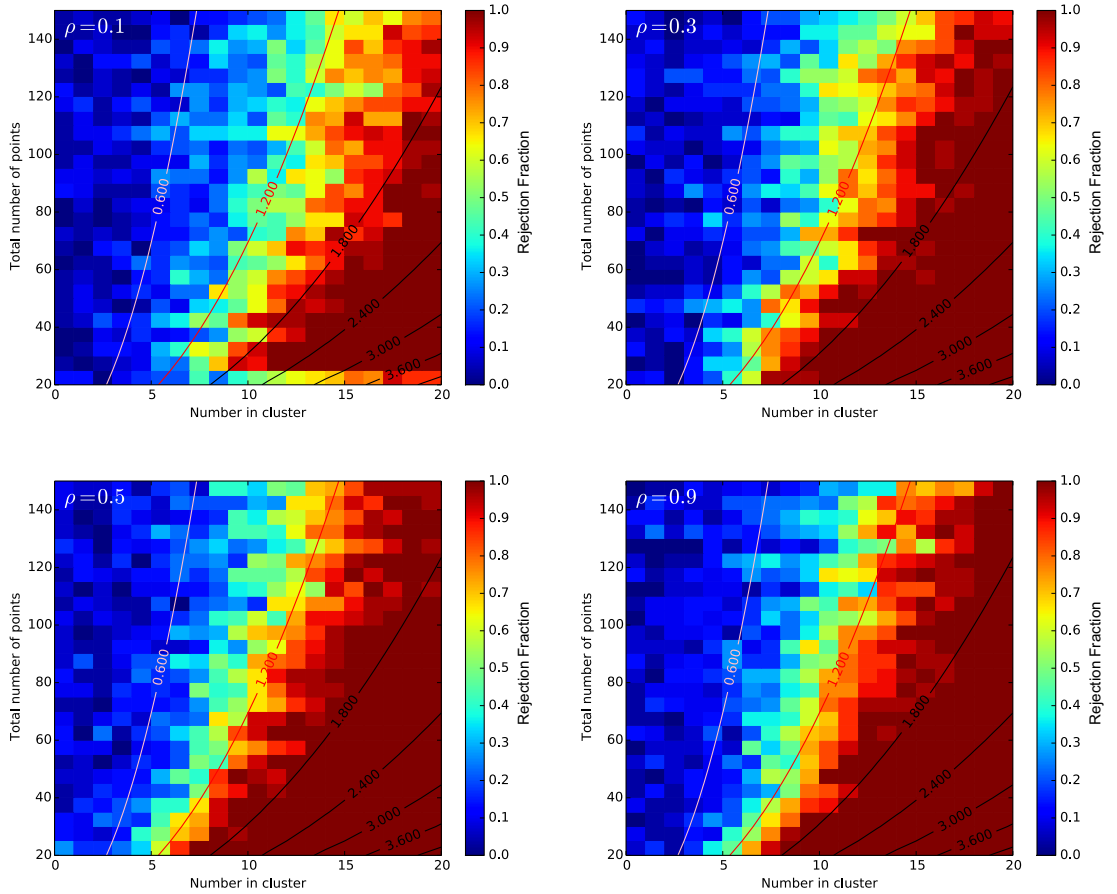

Figure 3.7: The rejection fraction with $P(H_0) < 5\%$ for O-ring statistic with logarithmic annuli widths. Within each panel, the annulus width is given by Equation 3.2 and the contours show the theoretical SNR (Eqn. 3.1). Between panels, ρ increases from top left to bottom right with (a) $\rho = 0.1$; (b) $\rho = 0.3$, (c) $\rho = 0.5$; and (d) $\rho = 0.9$.

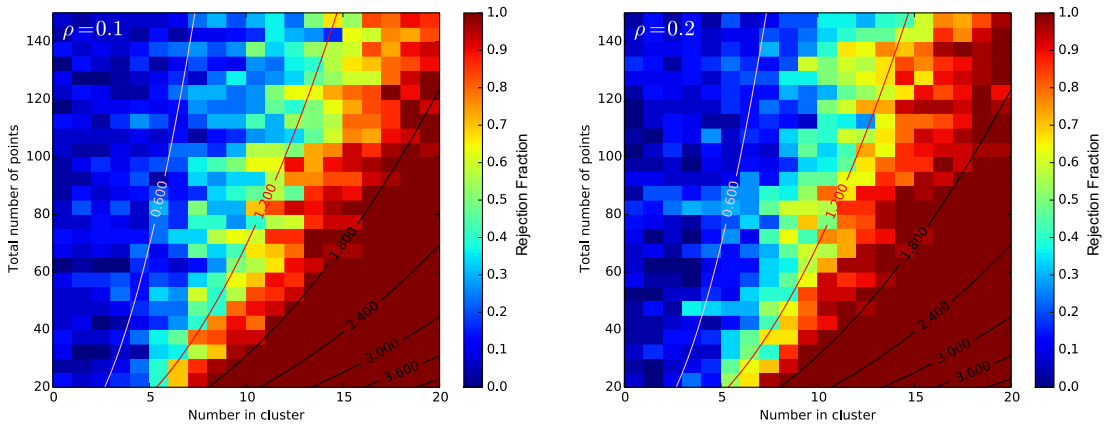

Figure 3.8: Rejection fraction as a function of annulus width for the O-ring statistic. Annulus widths decrease with the total number of points according to Equation 2.7. The contours show the theoretical SNR (Eqn. 3.1).

3.3.4 Minimum Spanning Tree

One method of using the MST to test for mass segregation by Allison et al. (2009) has been adapted to test if a distribution of stars is random. The total edge length of the MST of an observed sample is compared to the mean total length of 2000 random circular distributions with the same number of members using (Cantat-Gaudin et al. 2018)

$$\Gamma = \frac{l - l_{\text{obs}}}{\sigma_l} > 1, \tag{3.3}$$

where l is the mean total branch length calculated from the random distributions, σ_l is the standard deviation of the random distributions and l_{obs} is the branch length of the pattern being tested. To adapt this method the stars have been randomly distributed around the same rectangular region as were used for the other tests. The results of using Equation 3.3 are shown in Fig. 3.9 a). Using the kth most extreme branch lengths can produce a two-sided significance test as described in Section 2.3, using a 95 per cent confidence level produces the results shown in Fig. 3.9 b). The detectability scores for the two tests are shown in Table 3.1.

3.4 Discussion

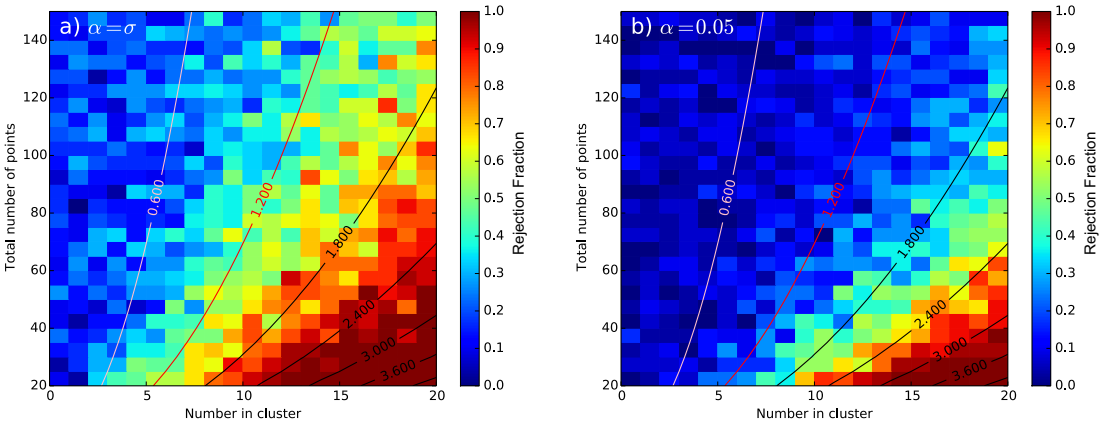

Figure 3.9: The rejection fraction for MST tests: a) Equation 3.3 and b) $P(H_0) < 5\%$ for MST total branch length test. The contours show the theoretical SNR (Eqn. 3.1).

Table 3.1: Detectability scores for tests from Section 3.3.

Cluster Radius	Test	Variation	Score
3	G		0.174 ± 0.004
3	F		0.089 ± 0.001
3	K		0.550 ± 0.003
3	O-ring	variable, $\rho = 0.1$	0.53
3	O-ring	variable, $\rho = 0.2$	0.56
3	O-ring	fixed, $\rho = 0.1$	0.54
3	O-ring	fixed, $\rho = 0.3$	0.59
3	O-ring	fixed, $\rho = 0.5$	0.59
3	O-ring	fixed, $\rho = 1.0$	0.54
3	O-ring	logarithmic, $\rho = 0.1$	0.49
3	O-ring	logarithmic, $\rho = 0.3$	0.56
3	O-ring	logarithmic, $\rho = 0.5$	0.58
3	O-ring	logarithmic, $\rho = 0.9$	0.56
3	MST	Eqn. 3.3	0.478 ± 0.001
3	MST	$P(H_0) < 5\%$	0.239 ± 0.001
5	G		0.106 ± 0.006
5	F		0.069 ± 0.003
5	K		0.438 ± 0.003
5	O-ring	fixed, $\rho = 0.3$	0.48
5	O-ring	fixed, $\rho = 0.5$	0.50
5	MST	Eqn. 3.3	0.398 ± 0.001
5	MST	$P(H_0) < 5\%$	0.179 ± 0.001

3.4.1 Comparison between methods

All four tests are capable of determining if an underlying process is random through significance testing. While additional tests are required to determine what type of clustering is present, a rejection of CSR due to a higher-than-expected average density is sufficient to determine the presence of overdensities within the data set, as well as an indication of the spatial scales. From the results in Section 3.3, the tests which reject randomness for this scenario the most sensitively are the second-order tests Ripley's K and O-ring, followed by G and then F.

G is able to reliably reject CSR when the majority of the points belong to the cluster, but the likelihood of rejection drops off rapidly with increasing number of background stars, while F is only able to reject CSR in very few cases. The most likely reason for this is the number of cluster members is relatively low compared to the number of background stars. G and F test the distribution of observed first nearest neighbour distances to distributions from realisations of CSR and to reject randomness there must be a significant shift in the distribution. Clustering produces a shift towards shorter nearest neighbour lengths, therefore the less-clustered and smaller central clusters have less effect and become more difficult to detect. G has an advantage in this test as the test positions are the stellar positions, while F utilises random positions in the window which are not necessarily located close to, or inside, the cluster.

The results of Ripley's K and O-ring are comparable, with the main difference being O-ring's dependence on bin width. As demonstrated by the detectability scores Ripley's K is the most consistent with an average score of 0.550 ± 0.003 while O-ring has a greater potential of CSR rejection when the bin width is optimised. With regards to choosing the bin width the results show that the logarithmic bins outperform Ripley's K and the variable bin widths for values of $\rho \geq 0.3$, and neither are as effective as a bin width matched to the cluster radius. For situations where a characteristic scale for the cluster is known a width approximately equal to this scale is preferred.

From the four tests discussed, the best test for CSR rejection is Ripley's K due to its lack of dependence on any other parameters. However, the results from Ripley's K

are less easy to interpret, as previously described in Section 2.3.2. Therefore, because O-ring is able to match or exceed this performance with most logarithmic bin widths, and because the results describe the density at and around a given radial distance, O-ring is recommended as the preferred test for this situation.

Both of the second-order tests show a characteristic region of consistent rejection of CSR followed by a gradient towards non-rejection. The rejection fraction trends in a similar fashion to the contours of constant signal-to-noise given by Equation 3.1 even when the overall detectability decreases due to having a more dispersed cluster, as seen in Fig. 3.6. This is indicative of the way the tests function. By using all of the interpoint distances the effect on the density up to or at a given distance due to a cluster will be lessened by having a larger cluster which is more dispersed. Similarly, by increasing the size of the study region to include more background stars the average measurement for K and O-ring will be reduced closer to the background value making it more difficult to exceed the envelope.

3.4.2 Comparison to other tests

Part of determining the ability of these tests is comparing them to established tests within the literature. Two of the most commonly used methods to determine clustering or randomness in astrophysics are the two-point correlation function (2PCF) and the minimum spanning tree (MST).

3.4.2.1 Two-Point Correlation Function

The O-ring test (also known as Mean Surface Density of Companions (MSDC)) is related to the 2PCF $\xi(r)$ by (Cressie 1993)

$$O(r) = (1 + \xi(r))\lambda. \tag{3.4}$$

The tests are therefore equivalent and it is the usage of the test that is the main difference. The results of the 2PCF are generally used to ascertain a radial dependence on density, and by inspecting this radial dependence describe the underlying mechanism that could have produced the results. In contrast the methods used in this paper simulate an underlying mechanism and determine how well the results of the mechanism fit the data.

For example, the MSDC for the Taurus, Ophiucus and Orion Trapezium star forming regions shows two distinct clustering regimes (Simon 1997), one for small-scale clustering of binaries and another for larger scale clustering, between which is a break point that varies for each star forming region. This is consistent with an underlying fractal structure over a subset of stellar densities, however, as Bate et al. (1998) demonstrates, different distributing processes can produce MSDC lines that fit the data equally well.

3.4.2.2 Minimum Spanning Tree

The total branch length of a MST can be used to test for CSR. Using a significance level of $\alpha = 0.05$ produces a detectability comparable to Diggle's G function. This is to be expected as the minimum spanning tree contains all of the star-star nearest neighbour edges – as well as a number of edges to ensure that it is a spanning tree – and G uses just the star-star nearest neighbour distances. Compared to O-ring and Ripley's K the test is less sensitive even with the reduced significance level of 0.16. In addition to a change in α another difference between the results of Eqn. 3.3 and the Monte Carlo method from Section 2.3 is that the tests are one-tailed and two-tailed respectively. This means that Eqn. 3.3 is unable to reject CSR when the pattern shows inhibition between stars and so care must be taken in interpreting its results when CSR is not rejected.

While it is less sensitive than the second-order tests, the MST does have some advantages which make it worth investigating further. As mentioned earlier there exist hierarchical clustering techniques to generate clusters (Yu et al. 2015) which can identify cluster members which is something G, F, K and O-ring cannot do in the form described in this paper. Further, there are measures of the MST that exist such as the graph distance matrix, a symmetric matrix containing the number of edges on the path between any pair of points in the tree, which to my knowledge have not yet been applied to significance testing but could be useful in cluster identification and significance testing.

3.4.3 Observing Cluster Scales

As discussed, with the O-ring test it is possible to identify a characteristic scale of the cluster by measuring at which point the function re-enters the CSR confidence envelope.

However, this observed scale is dependent on many factors including how centrally condensed the cluster is, the noise level in the background, and the width of the O-ring annuli. As a test, I determined the characteristic scale for the runs used in $\rho = 0.3$ subplot of Fig. 3.7, which has an input cluster radius of 3, and present these in the top plot in Fig. 3.10. The scales in this figure represent the shortest length scale at which the measured O-ring statistic re-entered the confidence envelope after exceeding it. To improve readability and confidence I have ignored any radius measurements where fewer than 10 successful radius measurements were found. We can see from this figure that the observed characteristic radii are all representative of the input cluster, being within a factor of ≈ 2. This is especially promising when considering that these clusters were both Gaussian in nature and projected on a noise background.

To investigate how the observed scale is a function of SNR, the middle and bottom plots in Fig. 3.10 show estimates of the SNR at the scale observed in the top plot. The centre plot shows the estimated SNR within a circle centred on the cluster with radius R, and the bottom plot shows the estimated SNR for an annulus centred on the cluster with radius R with width $\rho \times R$. For both of these plots, R is the scale in the detected radius (top) plot. The circle estimate SNR approximates the overall cluster SNR while the annulus SNR approximates the SNR that may be perceived by the O-ring test. We can see from this figure that, while the absolute SNR values are not equal, the apparent distributions of SNR values are similar and even share a similar relative range of approximately a factor of 2 between the highest and lowest observed SNR. It is interesting to note that an annulus SNR of at least 1 is required to produce consistent and representative radius measurements. However, we can also see that the SNR for both methods increases with observed radius, and so it appears, therefore, that while there is potentially a minimum SNR, there is no SNR threshold which causes the O-ring statistic to re-enter the confidence envelope.

3.5 Application to Astronomical Data

In addition to the simulated data, three sets of astrophysical data were tested. The first was the locations of Young Stellar Objects around the Serpens South star forming region (Gutermuth et al. 2008) from the Dunham et al. (2015) catalogue, using the limits $277.2 \leq$

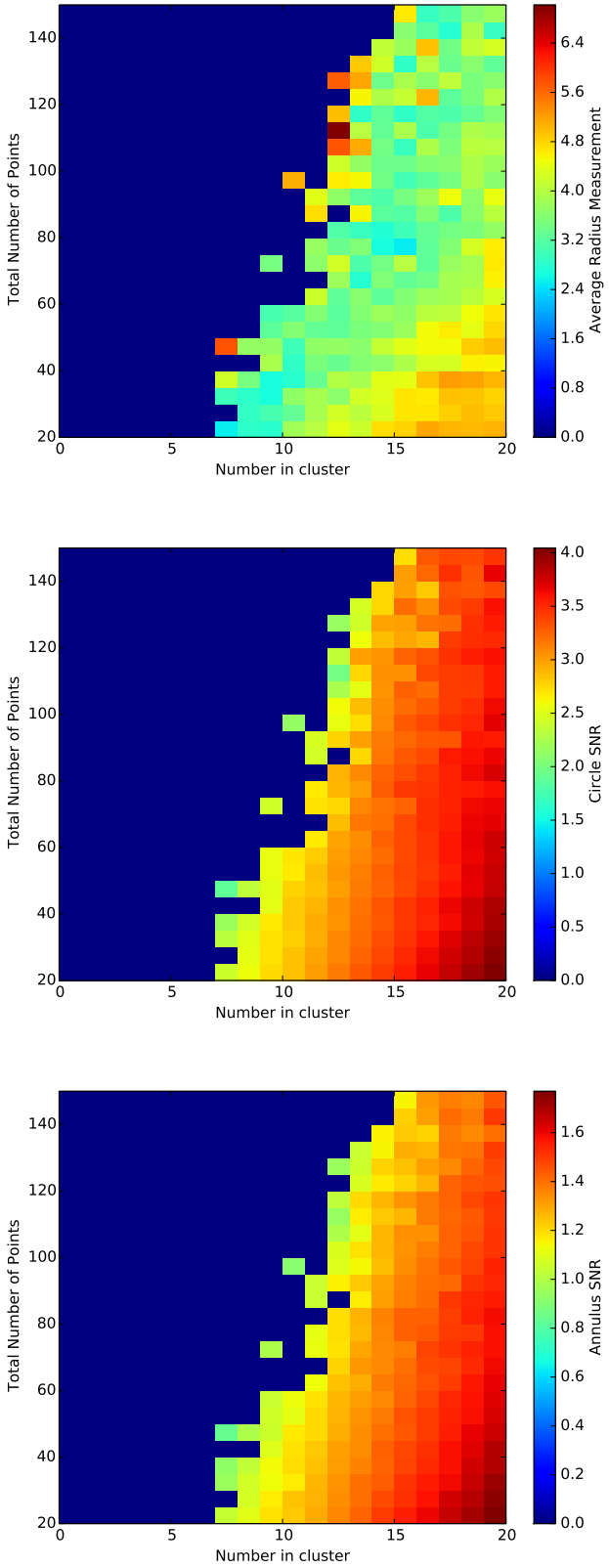

Figure 3.10: Top: Average output cluster radius measurement for O-ring test with logarithmic widths with $\rho = 0.3$ measured on cluster with radius $R = 3$. Middle: Signal to noise ratio in circle at the scale indicated in the top figure. Bottom: Signal to noise ratio at the scale indicated in the top figure. For this figure sections with fewer than 10 successful radius measurements have been ignored.

RA ≤ 277.7 and $-2.25 \leq$ Dec ≤ -1.75. Fig. 3.11 shows the locations of the YSOs and the results of applying G, F, K and O-ring. Each summary statistic exceeds the 95 per cent confidence envelope, therefore rejecting CSR as an appropriate model for the distribution of the YSOs with 95 per cent confidence.

The second test on real data was performed on members of the *Spitzer* catalogue [1] that were randomly chosen (without replacement) from the same field as the YSOs in the previous test.

Within this region there are 2601 *Spitzer* Sources identified with the object type 'star_F0I', and this number represented a good compromise between computational time and number of sources tested. Ripley's K is invariant to random thinning and so this allows for a less-computationally intensive method of testing the region as a whole. Fig. 3.12 (left) shows the positions of the randomly chosen members as well as the results. K and O-ring both reject CSR with 95 per cent confidence, and both statistics also display clustering followed by inhibition. It is likely that this is due to the extinction of the cloud; the outline of which can be faintly seen in Fig. 3.12 (left) when compared to Fig. 3.11.

The third test was on 2601 random *Spitzer* Sources from a region of equivalent area that was translated in declination by 0.5° with respect to the previous two. Unlike the previous two tests, these sources, shown in Fig. 3.12 (right), were chosen from an off-cloud region. The results show that CSR cannot be rejected as a null hypothesis for the distribution of these randomly chosen members.

The rejection of CSR runs for the YSO distribution was to be expected and demonstrates that the tests are able to reject genuine patterns that were not produced by randomly distributing stars in a window. It is interesting to note that O-ring shows the presence of two subsets of radial distances which exhibit overdensities though the effect causing this, either first or second-order, is not known as the distribution is not homogeneous and isotropic, so the appearance of clustering could be due to virtual aggregation as discussed in Section 2.2. The rejection of CSR for the random on-cloud members due to extinc-

1. The full *Spitzer* Gould Belt Survey catalogue of infrared sources produced using the *Spitzer* Cores to Disks (c2d) methodology: see Harvey et al. 2007; Evans et al. 2007 available from https://irsa.ipac.caltech.edu/data/SPITZER/C2D/doc/c2d_del_document.ps

tion shows that these tests function not only for clustering processes but for inhibiting processes, while the off-cloud random members demonstrates that there are examples of astrophysical data which are consistent with CSR.

3.6 Previous tests on spatial statistics

The work in this chapter sits among a number of other works that have explored the efficacy of spatial statistical tests. In a similar way to how I have tested the sensitivity of G, F, K and O to detecting clusters in the presence of noise, other works typically choose a facet of spatial statistics and explore how this would impact their results when applied to real-world data.

The results in this chapter fit among works which similarly look to quantify the sensitivity of these tests when presented with data that contains clustering or regularity (e.g. Diggle 1979; Gignoux et al. 1999; Thönnes, Elke and van Lieshout, Marie-Colette 1999; Ho and Chiu 2006; Ho and Chiu 2009). Diggle (1979) performed one of the earliest tests on these spatial statistics to determine their sensitivities to clustering and regularity. This early test was performed by generating 100 realisations of two spatial point processes. The first was a Poisson clustering process where cluster centres are randomly generated and offspring randomly generated from these centres in a normal distribution. The second was a sequential inhibition process where points may only be generated further than some minimum distance, δ, from all previously positioned points. Gignoux et al. (1999) nicely summarise the findings of Diggle (1979) in the form of a hierarchy: "against aggregation power(F) > power(K) > power(G) and power(K) > power(G) > power(F) against regularity." While regularity was not tested in this chapter, my results disagree for aggregation and would place F as the least powerful of the tests; this difference is likely due to the difference in model tested. Unlike most studies I have tested a parameter space of two mixed spatial point processes: a centralised cluster with a CSR background. This is because I am interested in testing which of the statistical tests is the most sensitive to clustering with the presence of background noise.

Other works focus on different aspects of applying spatial point processes, such as

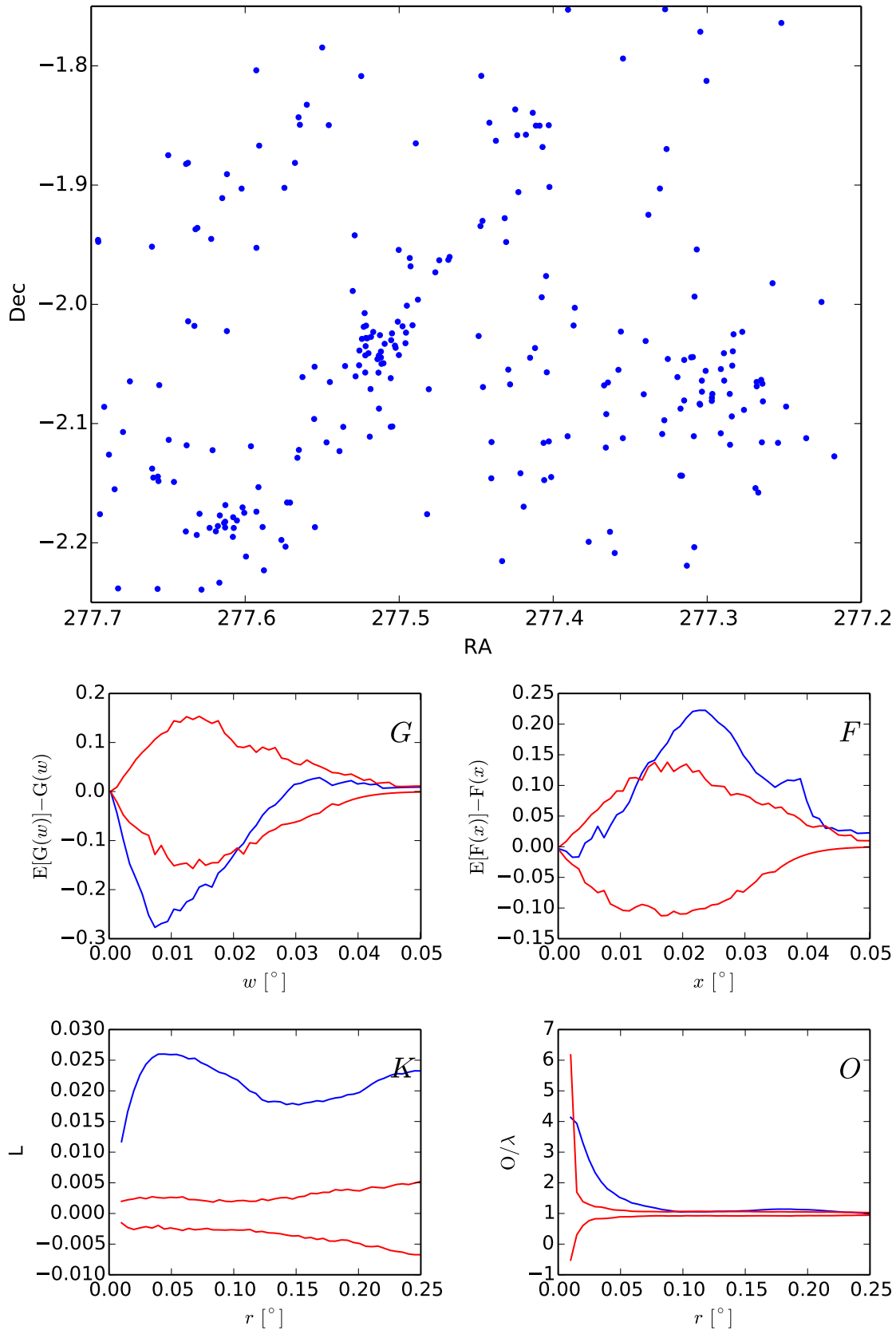

Figure 3.11: (above) Positions of YSOs within Serpens South, (below) results of G, F, K and O-ring with 95 per cent global confidence envelopes for CSR.

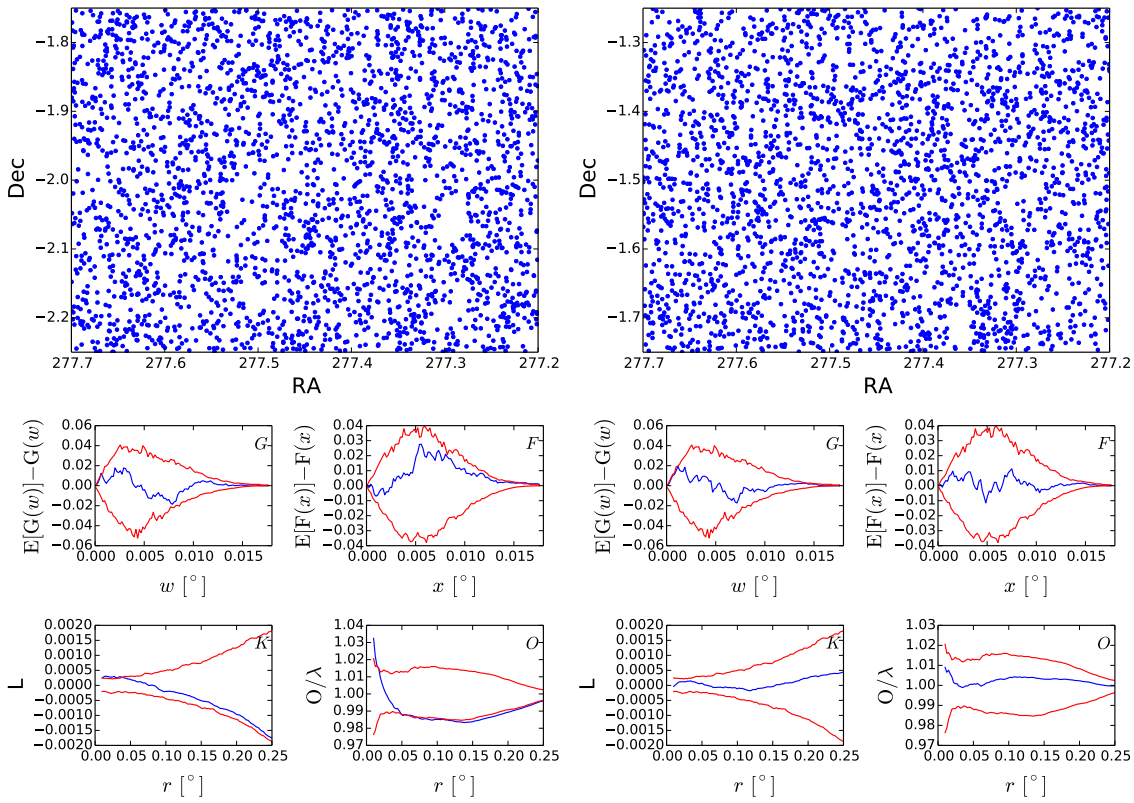

Figure 3.12: Left: (above) Positions of 2601 *Spitzer* catalogue members within Serpens South; (below) results of G, F, K and O-ring with 95 per cent global confidence envelopes for CSR. Right: (above) Positions of 2601 *Spitzer* catalogue members offset from Serpens South by 0.5° in Dec ($277.2 \leq \text{RA} \leq 277.7$ and $-1.75 \leq \text{Dec} \leq -1.25$); (below) G, F, K and O-ring results with 95 per cent global confidence envelope for CSR.

the important problem of edge-correction. In this work I use the border method for G and F and the weighted area method for Ripley's K and O-ring; however, there are arguments for the use of different methods, or even not applying edge-correction at all. It is clear that different methods of edge-correction change, and in some cases improve, the estimate of the statistic(e.g. Baddeley and Gill 1997; Bate et al. 1998; Baddeley et al. 2000; Yamada and Rogerson 2003). However, when it comes to the detection of clustering or regularity in the dataset it has been shown that edge-correction may provide no advantage in the case of Ripley's K (Yamada and Rogerson 2003). Or, for G and F, the loss of information by excluding the points may outweigh this improvement, particularly at low number of points (between 10 and 30) (Gignoux et al. 1999).

3.7 Conclusion

I have adapted spatial statistics methods for testing distributions of stars from ecology and them tested for their ability to reject randomness. Centralised clusters were generated and projected on top of a population of randomly distributed background members and, by varying the number of stars in the cluster and background, a parameter space of the empirical probability of CSR rejection was produced for each statistical test.

(i) The rejection of randomness for a given cluster radius approximately follows contours of signal-to-noise calculated over the parameter space of the number of stars in the cluster and background. A larger cluster radius over the same region of parameter space shows a decrease in the likelihood of rejection.

(ii) The O-ring test is equivalent to the two-point correlation function and Ripley's K and O-ring are better tests for CSR than the total branch length of the minimum spanning tree.

(iii) Three example star fields were tested against complete spatial randomness. The null hypothesis of randomness was rejected for the distribution of young stellar objects within Serpens South, as well as randomly selected *Spitzer* sources from the same region. The statistics from randomly selected *Spitzer* sources sampled from an off-cloud region were not consistent with complete spatial randomness.

For future applications the recommended test is the O-ring test using logarithmically spaced overlapping bins. Ripley's K and the O-ring test with a constant bin width show a slightly reduced detectability compared to O-ring with logarithmic bins, making them both good tests but not optimal. I take this result forward into the next chapter where I use the O-ring statistic with logarithmically spaced bins as a summary statistic when analysing the distributions of Class 0/I YSOs.

Chapter 4

Protostellar Distributions

4.1 Introduction

Star formation is known to occur within molecular clouds and it is has been found that star formation occurs more readily in regions of greater gas column density where more material is available for star formation. This effect can be observed on larger scales such as giant molecular clouds and galaxies where Kennicutt–Schmidt (K–S) relations show a higher surface density of star formation rate at higher column densities (Kennicutt 1989). It is observable on molecular cloud scales with similar star formation surface density relations (Heiderman et al. 2010), and finally, smaller, filamentary scales where prestellar cores are more frequently observed coincident on the sky with high-density filaments (André et al. 2010). Since star formation surface density is correlated with column density, is it possible to describe the distribution of where star formation has occurred using only column density information?

The answer to this question requires a quantification of the amount that star formation is enhanced by increasing column density as well as a model for how this leads to star formation being distributed throughout a cloud. As described in chapter 1, the K–S law is a method of relating the surface density of star formation rate to column density and is

often presented in the form of a power-law,

$$\Sigma_{\rm SFR} = C_{\rm r}\Sigma_{\rm Gas}^{\mu}, \tag{4.1}$$

where $\Sigma_{\rm SFR}$ is the star formation rate surface density and $\Sigma_{\rm Gas}$ is the gas surface density. The parameters $C_{\rm r}$ and μ quantify the relation between the surface densities of star formation and gas. Table 1.1 presents some previous measurements of μ in local star forming regions. Most μ values typically range between 1.5 and 2.5 (Heiderman et al. 2010; Gutermuth et al. 2011; Lombardi et al. 2014; Pokhrel et al. 2020) with some more extreme values with $\mu > 3$ (Lada et al. 2017).

A value of μ describes the change in the surface density of star formation as a function of column density, but it is missing information as to how the star formation is distributed throughout the cloud. To illustrate this point Fig. 4.1 presents two distributions of early-stage Young Stellar Objects (YSOs) which share the same value of μ. The YSOs in the left-hand cloud are evenly distributed throughout the cloud according to column density, while those on the right are biased towards the lower-right of the region. A distribution such as that in the right-hand cloud could be due to a column-density-independent effect which has strongly influenced the distribution of star formation, such as magnetic fields or stellar feedback. If such effects are significant column density may not be sufficient to describe the distribution of star formation on local scales.

In Chapter 3 the efficacy of the summary statistics G, F, Ripley's K and O-ring were tested and then applied to the locations of the YSOs in Serpens South to determine if they were consistent with complete spatial randomness (CSR). It was shown that the YSO locations were found to be inconsistent with CSR, and this chapter expands on this result by using the O-ring summary statistic to test different inhomogeneous Poisson Point Processes where the probability of placing a star at a given location is affected by the local column density. In doing so it will be possible to determine if Class 0/I YSO positions, and therefore locations of star formation, are equivalent to randomly sampling a two-dimensional probability distribution based on the observed gas density.

While spatial statistics are useful for testing models, I employ Bayesian statistics

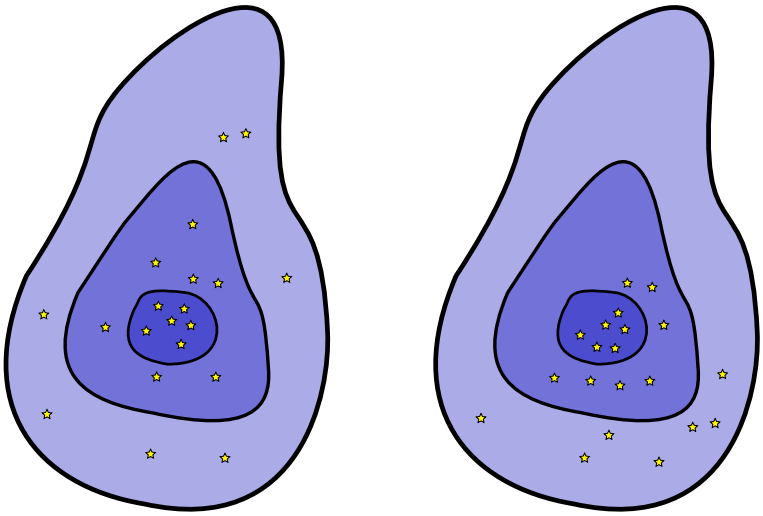

Figure 4.1: Illustration of two populations of YSOs with the same power-law relationship with column density. The YSOs in the cloud on the left are evenly distributed throughout the cloud while those on the right are clustered towards the lower-right portion of the cloud.

for parameter fitting. For this reason, I introduce a Bayesian method of estimating μ by measuring the surface density of Class 0 and Class 1 Young Stellar Objects (YSOs) within column density bins in Section 4.2.

Class 0/I YSOs are used in this chapter as a tracer of star formation as they are the in the earliest stage of YSOs evolution. This can be observed in the relative distributions of YSOs where the Class 0 and Class I YSOs whose positions tend to be more correlated with dense cloud than the more evolved Flat and Class II and Class III sources (Mairs et al. 2016; Buckner et al. 2020) (see section 1.4.2.1). Prestellar cores would likely be good tracers for star formation, but unfortunately it is not easy to determine which starless cores are likely to evolve to become stars and so Class 0/I YSOs are the youngest objects that are identifiable as definite precursors to stars. In this chapter I will look at the YSOs in five local star forming regions which, due to their proximity and number of YSOs, are good laboratories for testing distribution functions: Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348.

By applying the Bayesian method described in Section 4.2, the parameter μ was measured for each region individually as well as for the set of regions as a whole to estimate a global value of μ, the results of which are presented in Section 4.4.1. In addition to measuring μ for each region, I apply the same Bayesian method to find the best-estimates of C_r for each region in Section 4.4.2.

Using these results, in Section 4.4.4 I test the distributions of YSOs in each of these regions against Eqn. 4.1 with the estimated values of μ for each region. I will also present the results of testing the distributions of YSOs for general models with $\mu = 0$, $\mu = 1$ and the global value of $\mu = 2.05$. Each of these models test for a different potential physical description for how the distribution of stars is correlated with column density. The first model uses $\mu = 0$ above a threshold visual extinction $A_v = 6$ to test if there is a relationship between the amount of column density and surface density of YSOs, or if, once some threshold is reached, the YSOs are simply dispersed randomly. The second model, $\mu = 1$, is motivated by the distribution of prestellar cores which appear to be linearly correlated with column density (Könyves et al. 2020). With this model I look to determine if the observed distribution of YSOs is consistent with the distribution they may have had at an earlier stage in their evolution. The third model is to explore how well, or even if, a single power-law, can simultaneously represent multiple star-forming regions.

It was found that, when considering the number of regions that reject the model, the region-specific μ values were the most successful at describing the distributions of Class 0/I YSOs. As for the general models, the best-performing model was the global model of $\mu = 2.05$ which was consistent with 3 out of the 5 regions it was applied to, and the worst performing model was $\mu = 1$ – which was rejected by every region. As a further test I apply the $\mu = 2.05$ model to the Class II YSOs in all five regions in Section 4.4.5. This is to show that these statistics have enough discriminatory power to distinguish between two, potentially similar, distributions within the same study region.

Finally, the discussion of these results is presented in Section 4.5. There I discuss, based on my results, how effective a column-density-only model is at describing the distributions of YSOs. I also explore how the rejection of the $\mu - 1$ model by Class 0/I YSOs could imply an environmental dependence on the evolutionary time-scales for prestellar cores and/or Class 0/I YSOs.

4.2 Bayesian Statistics

From inspection of Eqn. 4.1, μ, for a single region, can be estimated from the straight-line gradient in a plot of $\log(\Sigma_{\rm SFR})$ versus $\log(\Sigma_{\rm Gas})$ (Heiderman et al. 2010; Gutermuth et al. 2011). A plot of the YSO surface density versus column density for the five regions studied in this chapter is presented in Fig. 4.2 and from this it is clear that there is a correlation between YSO surface density and column density in each of these regions. The gradients, μ values, are not dissimilar and the effect of different $C_{\rm r}$ values can be seen where the YSO surface density functions have different y-intercepts in different regions. One problem with measuring gradients is that the number of YSOs in each $A_{\rm v}$ bin are small enough that care needs to be taken when dealing with the uncertainties. Care must also be taken when estimating μ values that represent multiple regions as combining area and YSO counts to find an average density assumes every region has the same value of $C_{\rm r}$, which is not always true. For these reasons a Bayesian method of estimating μ is described in this section which can be extended to calculate joint values of μ for sets of regions.

To model the distribution of Class 0/I YSOs with respect to the observed column density in a star forming region I introduce the following equation,

$$\hat{\lambda}(\mathrm{N_{H_2}}) = C_{\rm r} \times \mathrm{N_{H_2}}^{\mu}, \tag{4.2}$$

where $\hat{\lambda}(\mathrm{N_{H_2}})$ is the estimate of the number density of Class 0/I YSOs at column density $\mathrm{N_{H_2}}$, $C_{\rm r}$ is a region-specific constant that normalises the number of Class 0/I YSOs such that Eqn. 2.21 is satisfied for a given region, and μ is the global power law affecting the distribution of YSOs with respect to the column density. We then look to find the most likely values of $C_{\rm r}$ and μ.

First, consider a sub-region, m, of the star-forming region that has an approximately constant column column density, $\mathrm{N_{H_2},m}$. If we assume the probability that this sub-region contains a number of YSOs, N_m, follows a Poisson distribution, it can then be shown that the probability of counting N_m YSOs is given by

$$\mathrm{prob}(N_m|A_m, \mathrm{N_{H_2},m}, C_{\rm r}, \mu) = \frac{(C_{\rm r}\mathrm{N_{H_2,m}}^{\mu} A_m)^{N_m} e^{-(C_{\rm r}\mathrm{N_{H_2,m}}^{\mu} A_m)}}{N_m!}, \tag{4.3}$$

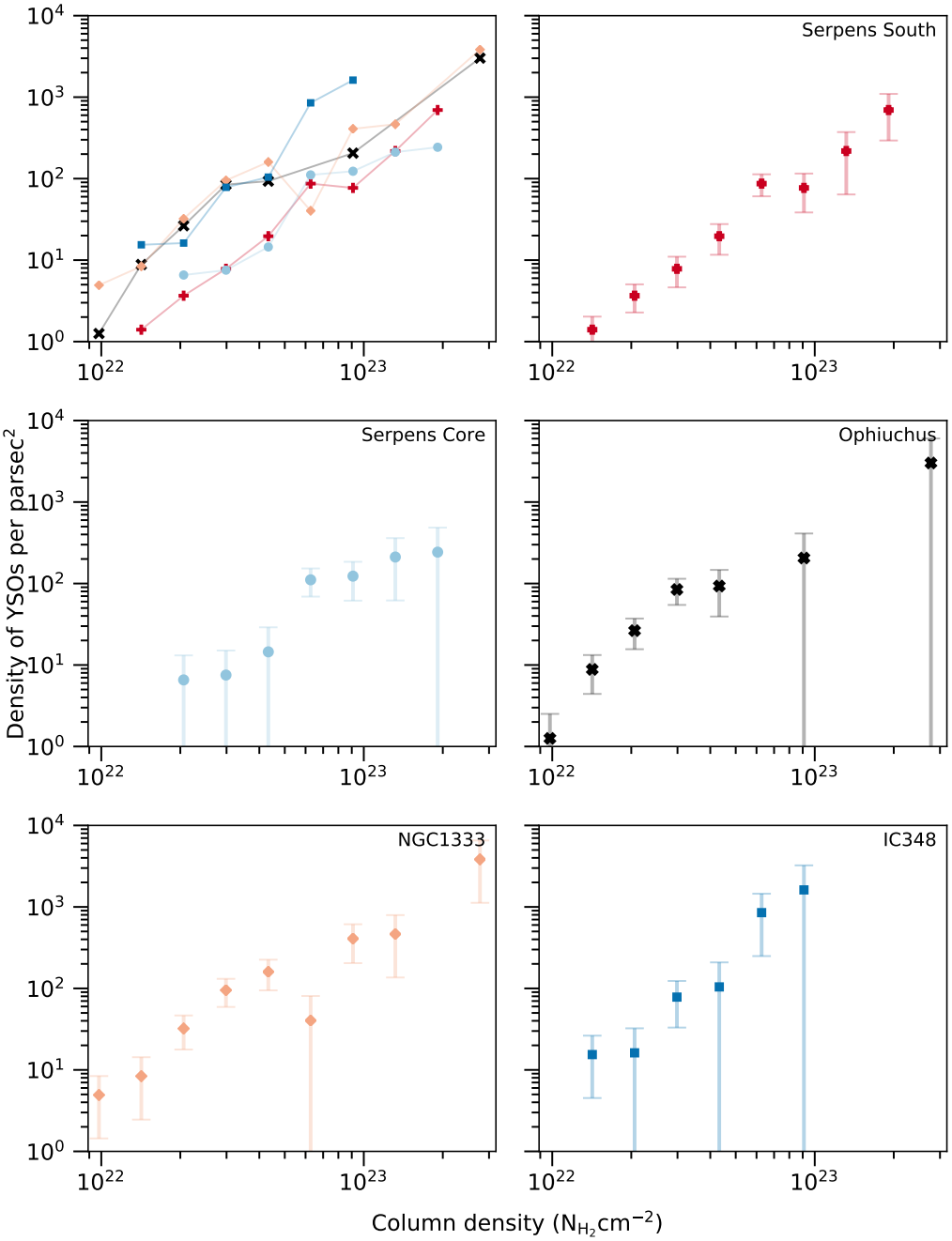

Figure 4.2: YSO surface density measurements within column density bins in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348. Uncertainties on Ophiuchus are the Poisson uncertainty on YSO counts to give an idea of YSO surface density uncertainties.

where A is the area of m. Repeating the experiment with M different sub-regions of the star-forming region results in the probability

$$\text{prob}(\{N\}|\{A\}, \{\text{N}_{\text{H}_2}\}, C_\text{r}, \mu) = \prod_{m=0}^{M} \frac{(C_\text{r}(\text{N}_{\text{H}_2,m})^\mu A_m)^{N_m} e^{-(C_\text{r}(\text{N}_{\text{H}_2,m})^\mu A_m)}}{N_m!}, \tag{4.4}$$

where $\{N\}$ is a vector of N, etc. Given the condition of similar column densities, this equation functions with any form of subdivision of the star-forming region. For this chapter I use contours of column density for each sub-region.

If we assume a uniform prior for μ and a Jeffrey's prior for C_r, we find that the naïve prior of μ and C_r is inversely proportional to C_r, i.e.

$$\text{prob}(C_\text{r}, \mu) = \begin{cases} \frac{1}{C_\text{r}} & \text{for } C_\text{r} \geq 0 \text{ and } \mu \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

With Bayes' theorem we may then construct an equation that can be used to find the probability associated with a combination of C_r and μ:

$$\text{prob}(C_\text{r}, \mu|\{N\}, \{A\}, \{\text{N}_{\text{H}_2}\}) \propto \frac{1}{C_\text{r}} \prod_{m=0}^{M} \frac{(C_\text{r}(\text{N}_{\text{H}_2,m})^\mu A_m)^{N_m} e^{-(C_\text{r}(\text{N}_{\text{H}_2,m})^\mu A_m)}}{N_m!}. \tag{4.6}$$

This joint pdf can be marginalised to find the marginal probabilities of μ and C_r for the star-forming region separately.

As an extension, we may then question if the power in Eqn. 4.2 is not specific to a single star-forming region, but is instead a universal property shared across different star-forming regions, each with unique values of C_r. It would be desirable, then, to combine the measurements from multiple star-forming regions, each with their own sub-regions, to produce a single best-estimate for μ. These measurements can be included by modifying the prior in Eqn. 4.5 to be proportional to the inverse of the product of C_r values,

$$\text{prob}(\{C_\text{r}\}, \mu) \propto \prod_{i=1}^{\#\ s.f.\ regions} \frac{1}{C_{\text{r},i}}, \tag{4.7}$$

and adding the measurements to the product in Eqn. 4.6 to produce the equation

$$\mathrm{prob}(\{C_\mathrm{r}\}, \mu | \{N\}, \{A\}, \{\mathrm{N_{H_2}}\}) \propto \prod_{i=1}^{\#\ s.f.\ regions} \frac{1}{C_{\mathrm{r},i}} \prod_{m=0}^{M} \frac{(C_{\mathrm{r},i}(\mathrm{N_{H_2},}m)^{\mu} A_m)^{N_m} e^{-(C_{\mathrm{r},i}(\mathrm{N_{H_2},}m)^{\mu} A_m)}}{N_m!}. \tag{4.8}$$

The addition of more regions increases the number of dimensions of the pdf which increases the computational difficulty of sampling the pdf without the use of techniques such as Markov chain Monte Carlo (MCMC) (Hastings 1970; Goodman and Weare 2010; Foreman-Mackey et al. 2013).

4.3 Spherical Projection

The column density data used in this chapter are sections of the celestial sphere which have been stored in a tangent plane projection. This type of projection, also known as the gnomonic projection, affects the areas contained within pixels and the distances between YSOs. Both of these quantities must be taken into account in order to measure the O-ring statistic, estimate the quantity μ using the methodology in Section 4.2 and accurately reproduce first-order spatial point processes.

These projected maps are stored in the FITS file format (Wells et al. 1981) which contains the keywords necessary to convert pixel coordinates to coordinates in the projection projection plane or celestial sphere. The set of keywords that allow for coordinate conversion are known as the FITS "world coordinate system" (WCS) and with these the pixel coordinates of most maps can be converted to projection and celestial coordinates (Calabretta and Greisen 2002; Greisen and Calabretta 2002). Software is available to perform transform between these coordinates systems; the results in this work use the library WCSLIB [1].

While it is possible to use projection coordinates entirely, they are representations of sections of a spherical surface and so it may be more simple and intuitive to use the native spherical coordinates. For example, with spherical coordinates the angular distance, $\Delta\sigma$,

1. https://www.atnf.csiro.au/people/mcalabre/WCS/wcslib/

between two points can be calculated using the haversine formula,

$$\Delta\sigma = 2\arcsin\sqrt{\sin^2\left(\frac{\Delta\delta}{2}\right) + \cos\delta_1\cos\delta_2\sin^2\left(\frac{\Delta\alpha}{2}\right)} \tag{4.9}$$

where α and δ refer to the right ascension (RA) and declination (Dec) in radians.

Some projections distort the shape or size of areas on the sphere and so the amount of angular area represented by a given pixel on the map is not necessarily consistent across the projection plane. The *Herschel* maps used in Section 4.4, use the gnomonic projection which is designated the AIPS code 'TAN'. A gnomonic projection is produced by projecting points on the surface of a sphere onto a tangential plane from the perspective of an observer at the sphere's centre. With this type of projection the amount of distortion increases as a function of latitude measured from the tangent point, θ. This distortion increases the projected area covered by objects on the sphere by $1/\cos^3(\theta)$ and, therefore, decreases the amount of area on the sphere covered by a pixel in the projection by $\cos^3(\theta)$. Finally, the amount of area covered by a pixel is given by,

$$|u| = |O|\cos^3(\theta_u), \tag{4.10}$$

where $|u|$ is the area covered by pixel u, $|O|$ is the area of a pixel at the tangent point and θ_u is the latitude of u. The angle θ_u is equal to $\Delta\sigma$ between the tangent point and u. Unlike angular distance, which requires only the angular coordinates for two points, the area covered by a pixel depends on the type of projection. As a general solution, however, it may be possible to calculate the world coordinates of the corners of each pixel and use those to approximate the areas, though this assumes the pixels are small enough to be approximately flat.

4.4 Application to Star-Forming Regions

In this chapter μ values are determined for the star-forming regions Serpens South, Serpens Main, Ophiuchus, NGC1333 and IC348, as well as a joint μ value for all regions. The joint μ value and $\mu = 0$ and $\mu = 1$ were then tested using 95 per cent confidence envelopes as described in Chapter 2.

Table 4.1: Summary of cloud properties used in this chapter.

Region	No. of Class 0/I YSOs	RA limits	Dec limits	Distance (pc)
Serpens South	44	$277.2° \leq \mathrm{RA} \leq 277.7°$	$-2.25° \leq \mathrm{Dec} \leq -1.75°$	484 [a]
Serpens Core	16	$277.4° \leq \mathrm{RA} \leq 277.6°$	$1.18° \leq \mathrm{Dec} \leq 1.38°$	484 [a]
Ophiuchus	24	$246.0° \leq \mathrm{RA} \leq 248.5°$	$-25.2° \leq \mathrm{Dec} \leq -23.8°$	144 [a]
NGC1333	32	$52.0° \leq \mathrm{RA} \leq 52.8°$	$31.0° \leq \mathrm{Dec} \leq 31.8°$	293 [b]
IC348	12	$55.8° \leq \mathrm{RA} \leq 56.4°$	$31.9° \leq \mathrm{Dec} \leq 32.5°$	321 [b]

[a] (Zucker et al. 2019)
[b] (Ortiz-León et al. 2018)

The Dunham et al. (2015) YSO catalogue was used for YSO position and classification for all regions. YSOs classified as Class 0/I are those with a corrected spectral index value greater than or equal to 0.3 and $T_{\mathrm{bol}} < 650$K. Using a single catalogue and classifying with corrected spectral index and T_{bol} may lead to missing and misclassified sources; however, it provides a simple and consistent method of identifying YSO populations which can be compared between clouds. The column density data used for the different regions were the *Herschel* 18.2″ resolution maps (André et al. 2010; Palmeirim et al. 2013). Table 4.1 lists the number of Class 0/I YSOs and distances to each cloud, as well as the RA and Dec boundaries used to extract the regions.

4.4.1 Estimations of μ

In this section I present the results of estimating the power law using the Bayesian marginalisation described in Section 4.2. The joint-probability distributions for C_{r} and μ were calculated using the number of YSOs and the area contained within column density bins in each region the results of which are presented in Fig. 4.4. These joint-probability distributions were then marginalised over C_{r} to find the pdfs for μ for each region, the results of which are shown in Fig. 4.5 and Table 4.2.

To determine if these joint-probability distributions were reasonable Eqn. 2.21 was used to find the solutions where $\langle N \rangle$ was equal to the observed number of YSOs in each region. Presented in Fig. 4.4 as a dashed line are the solutions where $\langle N \rangle$ is equal to the observed number of YSOs in all regions. The intersection between Eqn. 2.21 and the high probability density regions of the joint probability distribution demonstrates that Eqn. 4.6

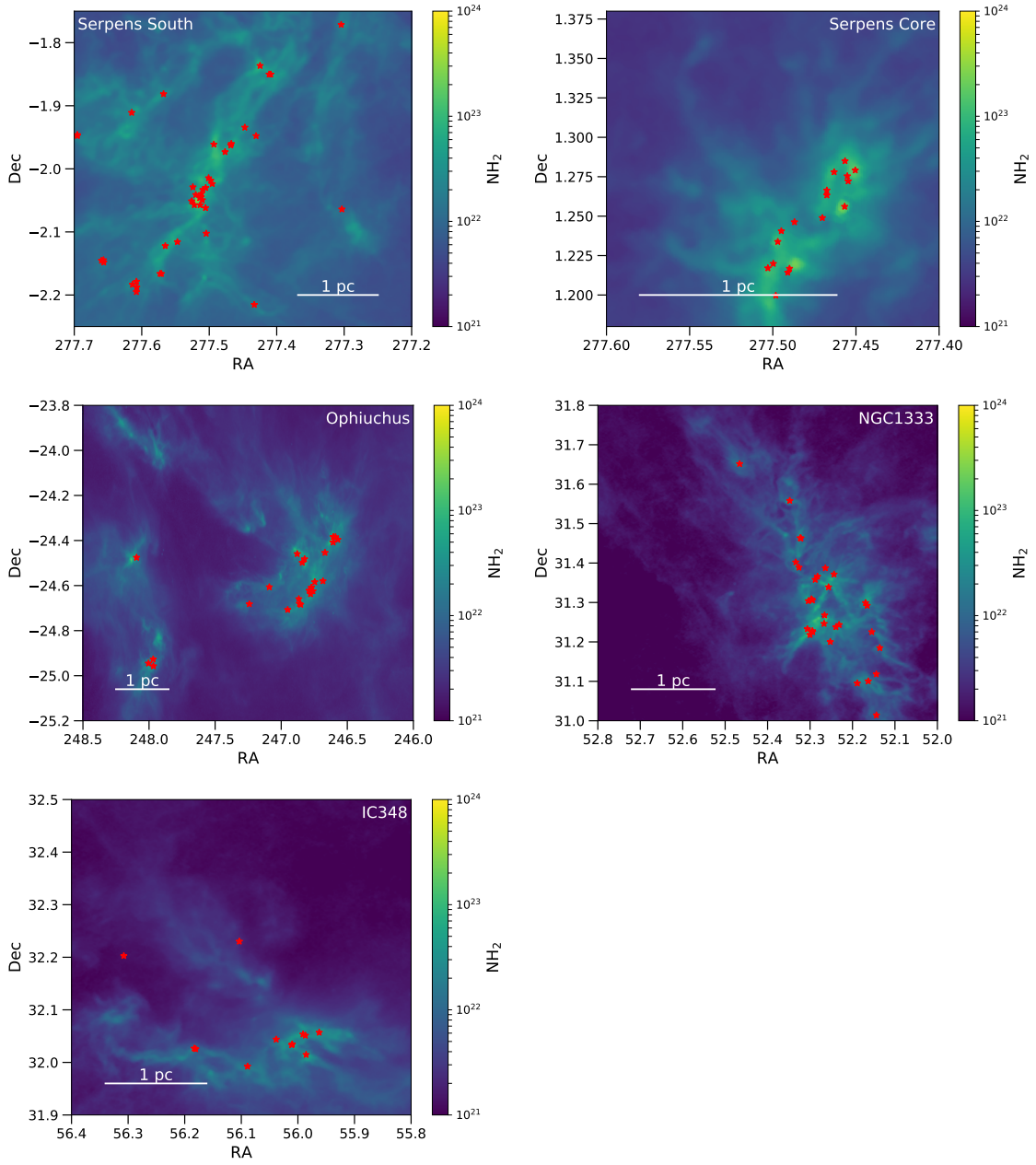

Figure 4.3: Column density images of Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 plotted with Class 0/I YSO positions. Column density data is plotted using a logarithmic scale to improve visibility of low and high column density data.

Table 4.2: μ estimates for all regions.

Region	Best estimate of μ	95 per cent confidence interval
Serpens South	2.39	$2.06 \geq \mu \geq 2.70$
Serpens Core	2.06	$1.45 \geq \mu \geq 2.76$
Ophiuchus	1.78	$1.28 \geq \mu \geq 2.20$
NGC1333	1.66	$1.28 \geq \mu \geq 2.04$
IC348	3.00	$2.03 \geq \mu \geq 4.04$
All regions	2.05	$1.85 \geq \mu \geq 2.25$

is consistent with producing C_r and μ values that approximate the number of YSOs used to calculate the joint-distribution. As can be seen in the figure, this overlapping of the two functions is consistent across all regions.

To find the most likely value of μ over all regions the YSO and area measurements were combined into a single 6-dimensional joint-probability distribution: one dimension for each of the five regions' C_r values and one for μ. This distribution was sampled and marginalised using the MCMC functionality of the Python package EMCEE (Foreman-Mackey et al. 2013) to find the pdf for μ. The best-estimate for the global power law was found to be $\mu = 2.05 \pm 0.20$, using a 95 per cent confidence interval as the uncertainty.

A value of $\mu = 2.05$ sits within the 95 per cent confidence intervals for Serpens Core, Ophiuchus and IC348 and only marginally outside those of Serpens South and NGC1333. While further testing is required to determine if the distributions of the YSOs within these regions are consistent with a global power law, the global power law appears to adequately describe the distribution of these YSOs as a set.

4.4.2 Estimations of C_r

In this section I present the results of estimating the region-specific constants for each of the five regions. The best-estimates for C_r were produced by marginalising the joint-probability distributions over μ, as demonstrated in Fig. 4.4. These results are presented in Table 4.3 as C'_r values, which are related to C_r by

$$C'_r = C_r \times 10^{22\mu} \tag{4.11}$$

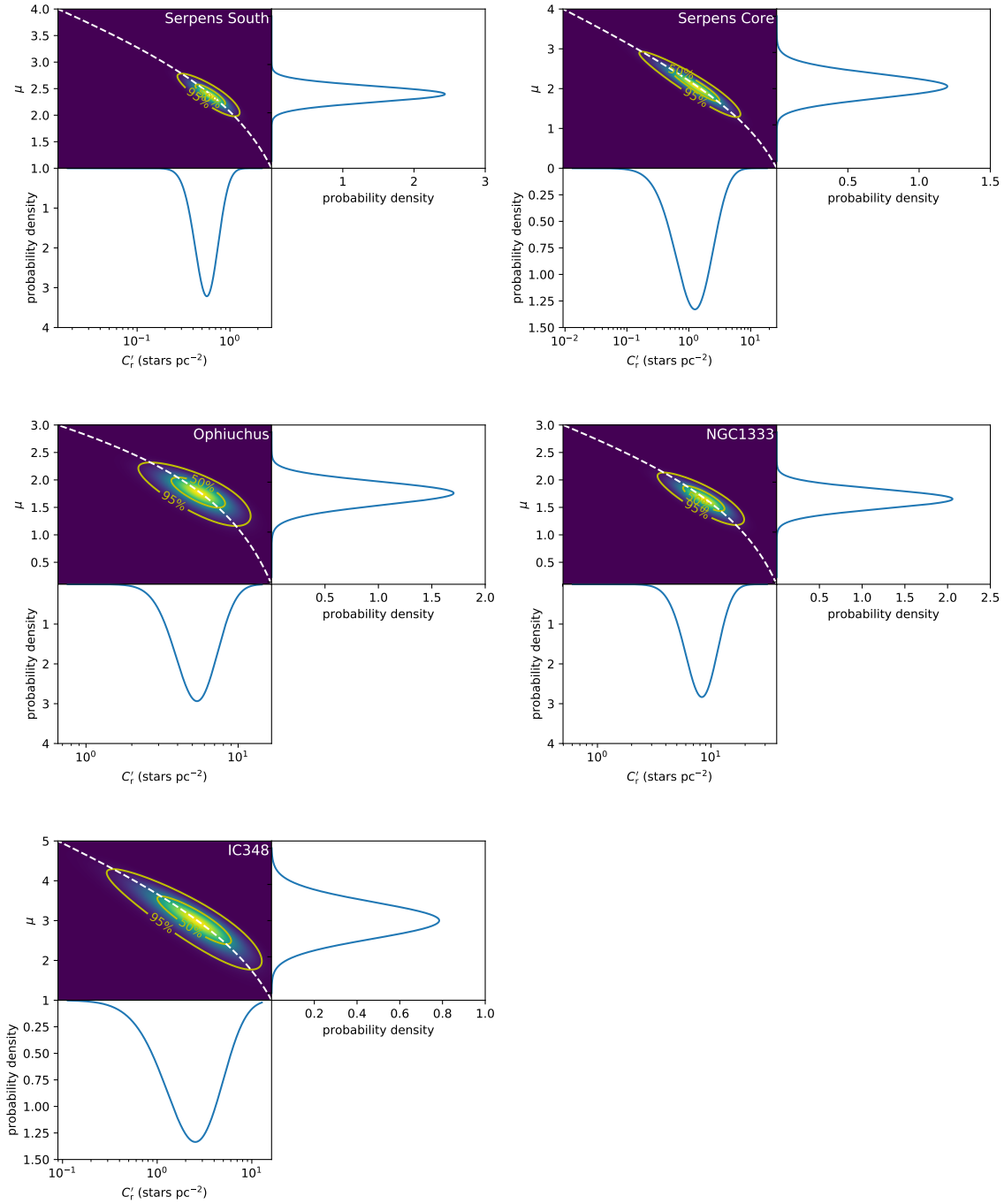

Figure 4.4: Joint-probability distribution of Eqn. 4.6 for Class 0/I YSOs in the labelled star-forming regions with the marginalised probability density functions for μ and C_r'. The contours outline the 50 and 95 per cent cumulative probabilities and the dashed line follows the solutions to Eqn. 2.21. The x-axis is C_r' (Eqn. 4.11), because the joint-probability distributions have been calculated using column density values scaled by a factor of 10^{-22}. This reduces the span of C_r values and allows the structure in the joint-probability distribution to be distinguishable. See Section 4.4.2

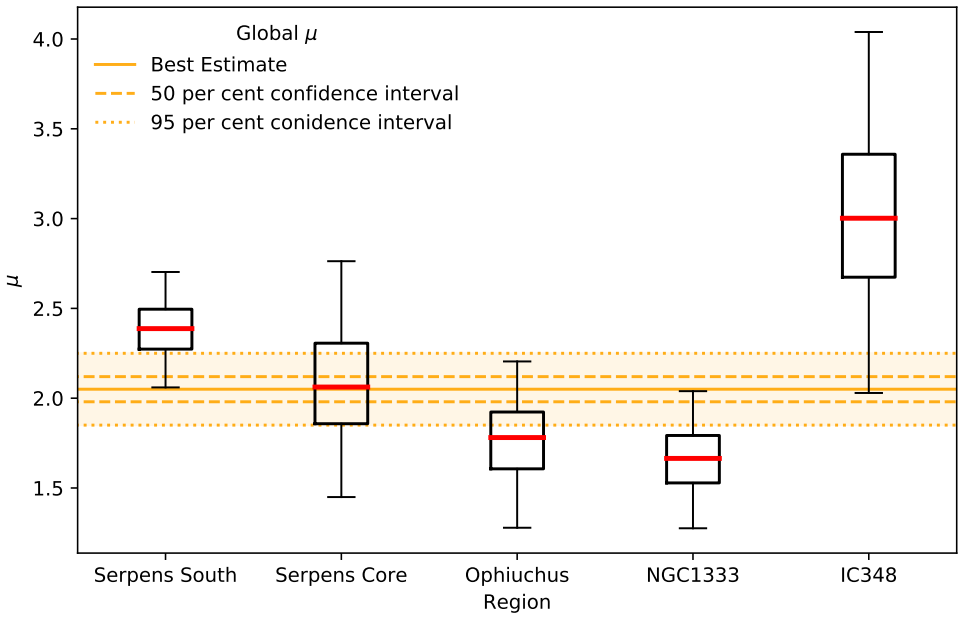

Figure 4.5: Marginalised distributions for μ for Serpens South, Serpens Core, Ophiuchus, NGC1333, IC348 and global estimate. The box and whiskers present the best estimate and the 50 and 95 per cent intervals. The orange lines present the best estimate and the 50 and 95 percent intervals for the global μ value.

or, equivalently,

$$\hat{\lambda}(\mathrm{N_{H_2}}) = C'_{\mathrm{r}} \times \left(\frac{\mathrm{N_{H_2}}}{10^{22}\ \mathrm{cm^{-2}}} \right)^{\mu}. \tag{4.12}$$

The results are presented in this form because C_{r} values scale with $\mathrm{N_{H_2}}^{-\mu}$, and with column density values of order $\sim 10^{22}$ cm^{-2} the uncertainties in μ lead to a large range in the magnitudes of potential C_{r} values. It is because of such large ranges of potential C_{r} values that the joint-probability densities of Fig. 4.4 were presented using C'_{r}.

While C_{r} (and C'_{r}) values are important for estimating the amount of star formation within a region with Eqn. 4.2, C_{r} does not affect where in a cloud the stars will be positioned in the model – as can be seen by substituting Eqn. 4.2 into Eqn. 2.22 – and the observed number of YSOs in a region can be used when simulating a spatial point pattern. In addition, care must be taken in the interpretation of the meaning of the values. If we take the logarithm of Eqn. 4.2,

$$\log(\Sigma_{\mathrm{SFR}}) = \log(C_{\mathrm{r}}) + \mu \log(\Sigma_{\mathrm{Gas}}), \tag{4.13}$$

Table 4.3: Estimates of C'_r from marginalisation over μ for all regions.

Region	Best estimate of C'_r (stars pc^{-2})	95 per cent confidence interval
Serpens South	0.56	$0.31 \geq C'_r \geq 0.94$
Serpens Core	1.26	$0.25 \geq C'_r \geq 3.93$
Ophiuchus	5.36	$2.71 \geq C'_r \geq 9.31$
NGC1333	8.31	$4.10 \geq C'_r \geq 14.73$
IC348	2.53	$0.48 \geq C'_r \geq 7.5$

we can see that for a given value of μ, C_r is equal to the expected YSO surface density when the gas density measure is equal to one, in this case $\Sigma_{Gas} = 1$ cm^{-2}. From this we can see that to compare C_r values is to compare *expected* YSO surface densities at unit Σ_{Gas} and since Σ_{Gas} can be any density measure, C_r can be measured at any column density.

Fig. 4.6 shows the expected YSO surface densities using the best-estimates from Tables 4.2 and 4.3 – this is reasonable as, in these regions, the best individual estimates of μ and C_r are approximately equal to the best joint estimates for μ and C_r. From this figure it can be seen that the YSO surface density in these regions is well represented by a power-law with column density using the results of Eqn. 4.8. As discussed, since C_r is the expected YSO surface density at a chosen reference column density, we can also see in Fig. 4.6 how the region with the highest C_r depends on this choice of reference.

While it is not possible to remove the μ dependence from C_r, it is possible to find the best-estimates for C_r in each region assuming the same value of μ across all regions. Table 4.4 presents the best estimates of C'_r assuming $\mu = 2.05$, and Fig. 4.7 shows the new expected YSO surface density functions. We can see from these results the effect of different star-formation efficiencies on regions which are assumed to have the same column density dependence.

I

4.4.3 Application to simulated protostar spatial distributions

In this section we apply the O-ring statistic with 95 per cent global confidence envelopes to two sets of simulated YSO distributions in Serpens South, presented in Fig. 4.8. Both sets of simulated data contain the same number of YSO positions as Class 0/I YSOs observed in

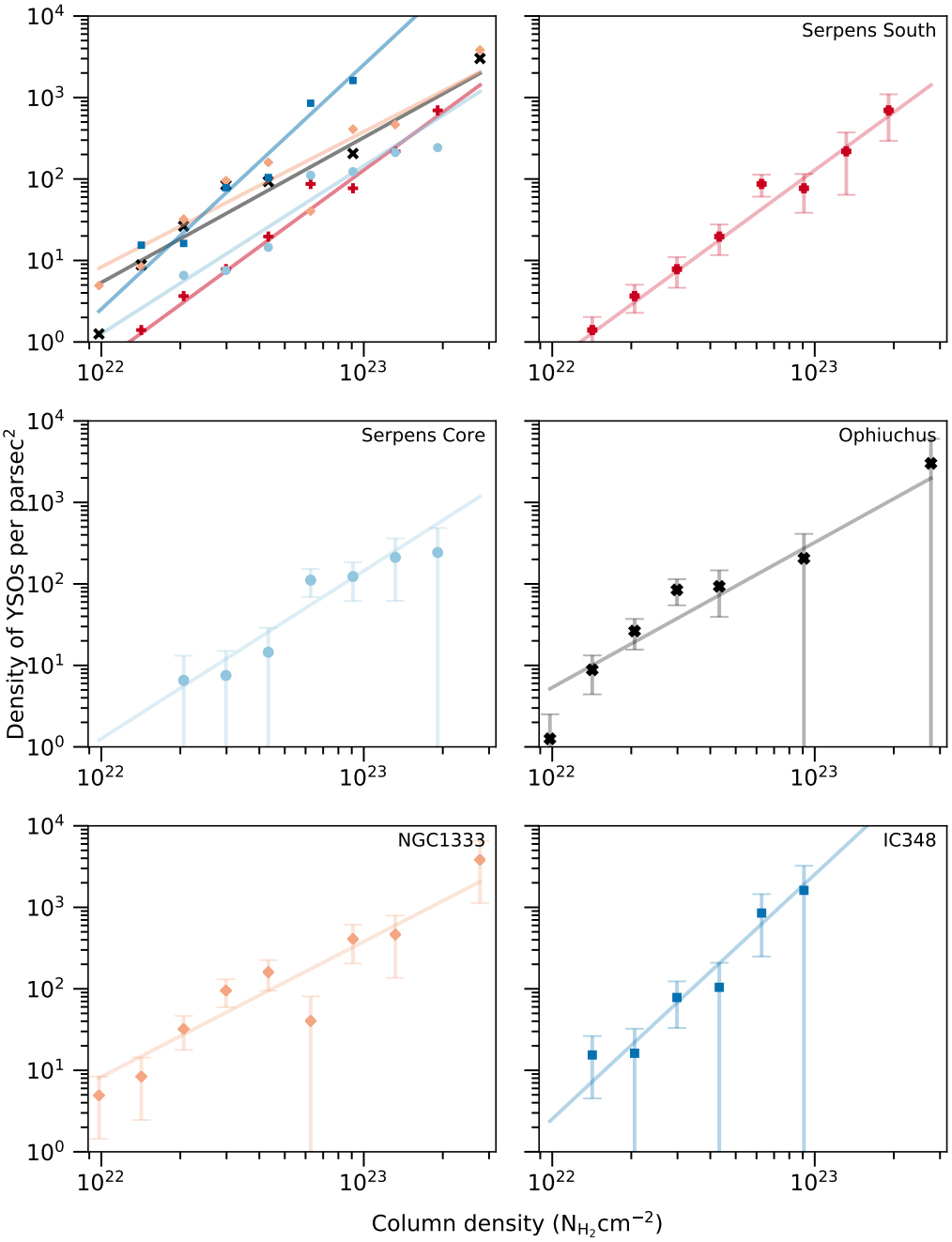

Figure 4.6: YSO surface density measurements within column density bins in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 with straight lines showing best estimates of μ and C_r in each region. Uncertainties on Ophiuchus are the Poisson uncertainty on YSO counts to give an idea of YSO surface density uncertainties

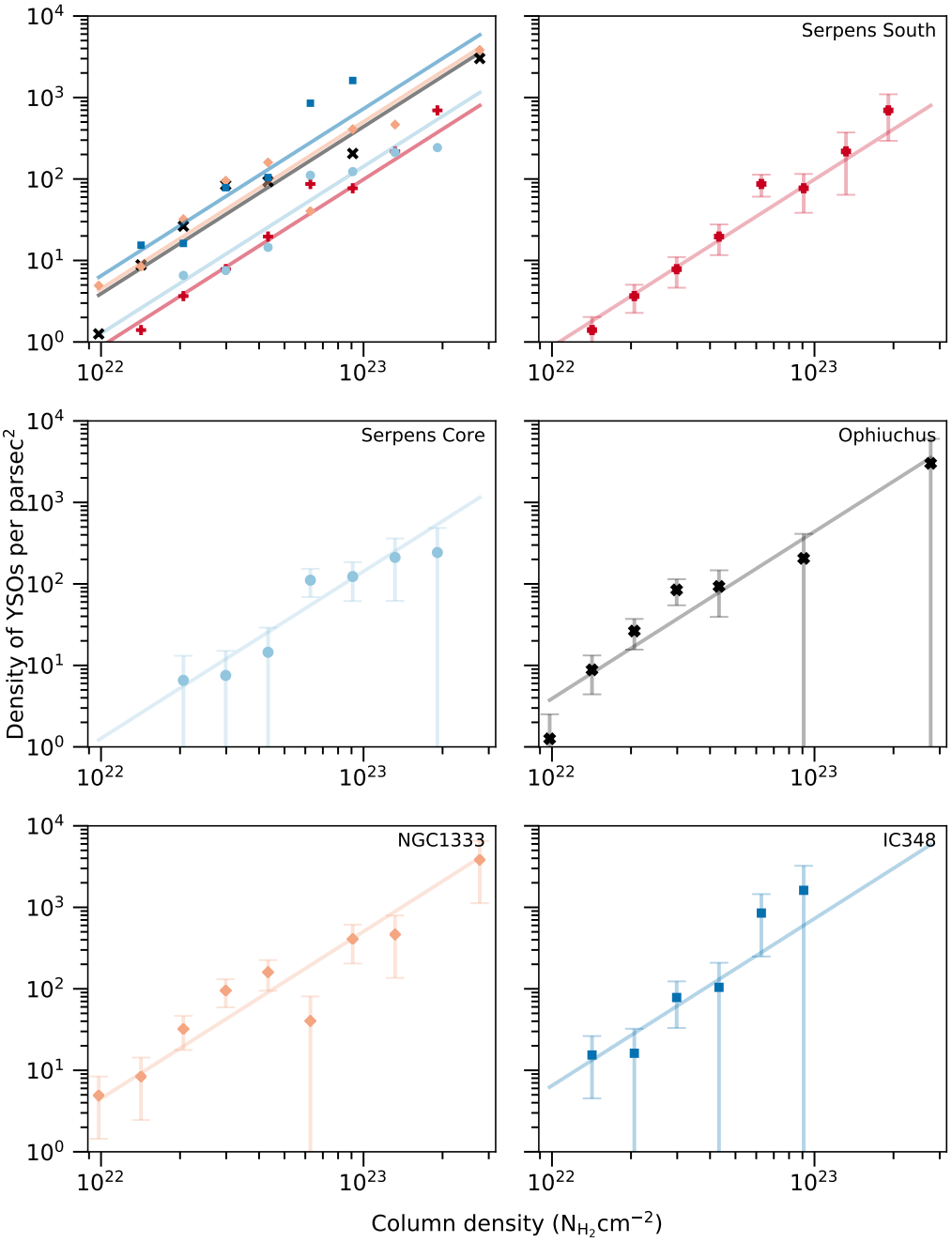

Figure 4.7: YSO surface density measurements within column density bins in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 with straight lines showing best estimates of C_r assuming $\mu = 2.05$ in each region. Error bars are the Poisson uncertainty on YSO counts to give an idea of YSO surface density uncertainties

Table 4.4: Estimates of C'_r for all regions for $\mu = 2.05$.

Region	Best estimate of C'_r (stars pc^{-2})	95 per cent confidence interval
Serpens South	0.88	$0.64 \geq C'_r \geq 1.16$
Serpens Core	1.27	$0.74 \geq C'_r \geq 1.93$
Ophiuchus	3.91	$2.50 \geq C'_r \geq 5.62$
NGC1333	4.50	$3.06 \geq C'_r \geq 6.22$
IC348	6.50	$3.11 \geq C'_r \geq 10.85$

Serpens South and were generated using Eqn. 4.2 with $\mu = 2.05$ using the *Herschel* column density data for Serpens South. The left-hand, or unbiased, distribution was generated using a probability distribution which spanned the entire study region, while the right, or biased, distribution was generated using a probability map covering only the south-west portion of the map.

Using the same methods applied to the star forming regions in Section 4.4.1, μ was measured for the unbiased and biased distributions to find $\mu = 2.06$ and $\mu = 2.01$ respectively. Similar to the illustration in Fig. 4.1, Fig. 4.8 presents two distributions of YSOs with similar values of μ that have been produced from different models.

The lower portion of Fig. 4.8 presents the results of using the O-ring statistic to test for Eqn. 4.2 with $\mu = 2.05$. We can see from these results that the unbiased distribution did not reject, and so is consistent with, this model – this is to be expected as the unbiased distribution is a realisation of said model. We can also see that the biased distribution rejects this model as the O-ring statistic exceeds the envelope. These results show how the O-ring test is able to reject the spatially biased distribution of YSOs, whereas a power-law measurement like μ does not have this type of discriminatory power.

4.4.4 Application to protostar spatial distributions

The distribution of protostars in Perseus, Ophiuchus and Serpens were tested against four distribution models: a minimum threshold of 6×10^{21} N_{H_2}cm^{-2} for YSOs to be placed but with no other dependence on column density, a power law dependence of $\mu = 1$, a power law dependence of $\mu = 2.05$, and a power-law equal to the Bayesian-estimated power-law for the region. The distributions of protostars were tested for their consistency with a distribution model using the O-ring statistic as a summary statistic and 95 per cent global

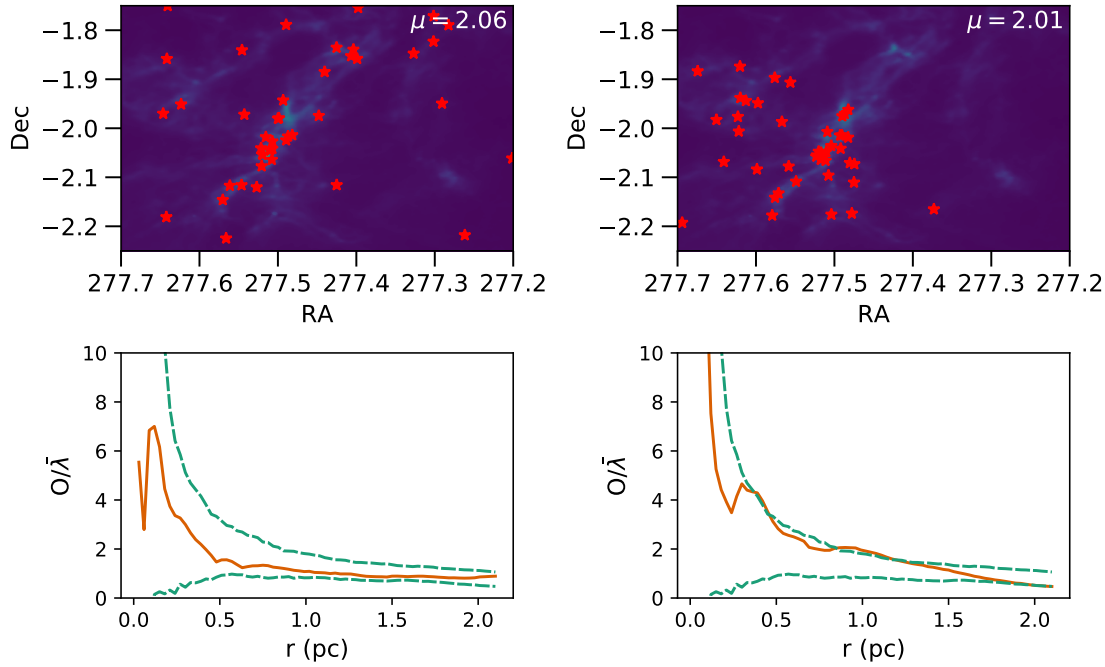

Figure 4.8: Two distributions of YSOs generated using the *Herschel* column density data for Serpens South and Eqn. 4.2 with $\mu = 2.05$ with the 95 per cent confidence envelope tests for Eqn. 4.2 with $\mu = 2.05$. The left-hand distribution was generated using the entire study region, while the right, was generated using a probability map covering only the south-west portion of the map.

confidence envelopes.

The O-ring statistic for the distribution of Class 0/I YSOs were measured in each region at a set of spatial scales,

$$r = \{x | x = n\Delta r, x \leq R\ , n \geq 1\}, \tag{4.14}$$

where $\Delta r = 0.03$ pc and R is equal to the half the length of the shortest axis (either RA or Dec). In other words, r values are linearly spaced in intervals of Δr from $r = \Delta r$ up to the largest scale that is less than, or equal to, half of the length of the shortest axis. Following the results from Chapter 3, the widths of the annuli used in the O-ring test are logarithmic and are equal to

$$w = 0.6 \times \rho, \tag{4.15}$$

where ρ is the set of spatial scales for each region, r, converted into degrees.

Confidence envelopes were generated for each of the four models described ear-

lier using 99 realisations of the first-order processes. Each realisation was produced by sampling the first-order intensity map a number of times equal to the number of YSOs observed in the region. The O-ring statistic was measured for each realisation at the spatial scales r with annuli widths w.

All of the confidence envelopes, along with the measured O-ring statistic for the Class0/I YSOs, are presented in Fig. 4.9 and discussed in the following subsections. The y-axis of the subplots in Fig. 4.9 are the measured O-ring statistic divided by the YSO density of the study window encompassing the star forming region i.e.

$$\bar{\lambda} = \frac{\sum_u \sum_v \mathrm{P}(u,v)}{\sum_u \sum_v \mathrm{A}(u,v)}. \tag{4.16}$$

As such the y-axis represents how many more times clustered, or less clustered, the YSOs are compared to CSR in the same window.

The envelopes applied in this section are unlike those that were applied in Chapter 3 in that they are produced by inhomogeneous Poisson Point processes, rather than CSR. This subtle difference makes the interpretation of the scales at which the confidence envelope is exceeded more difficult to interpret as they simply only indicate the scales on which the models are an inadequate description of the data. And so, while the scales at which the envelopes are exceeded indicate scales at which interesting physics may be occuring, because we are not comparing against CSR we don't expect the scales of rejection to necessarily correspond to a cluster size.

4.4.4.1 CSR in gas above cutoff value

The simplest null hypothesis is that there is no correlation between molecular cloud material and YSOs. CSR is unlikely to be a successful model as protostars and prestellar cores are known to be associated with dense material within molecular clouds (André et al. 2010), the measured power laws in Section 4.4.1 are greater than zero and it was shown that Serpens South is inconsistent with this model in Fig. 3.11. A more sensible, simple relationship between column density and YSOs is one in which some material is required for stars to form but the amount of material has no impact on the number of YSOs.

The impetus for this model is the potential existence of a threshold density required for star formation, which has been measured in Taurus, Ophiuchus and Perseus (Onishi et al. 1998; Johnstone et al. 2004; André et al. 2010). As such the spatial point process used for the confidence envelopes uses a uniform probability for forming stars in any pixel with a visual extinction above $A_v = 6$ (assumed to be equal to $6 \times 10^{21} N_{H_2} cm^{-2}$) and zero otherwise.

The study windows covered in this work are too limited in size to come to any conclusions on the existence, or value of, a star-formation column density threshold. A more complete study of star-formation thresholds would require a greater array of thresholds, and study windows covering more low-column density space; however, this is outside of the scope of the work in this chapter which is primarily interested in determining the suitability of powers of μ.

The envelopes for this model, presented in the first column of Fig. 4.9, are exceeded by every region except NGC1333. For most regions this model produced too few pairs of YSOs at small scales, as evidenced by the measured O-ring statistics exceeding the upper bound of the envelope. The O-ring statistic for IC348, however, exceeds the lower bound of the confidence envelope at a separation of 0.7 pc, due to too few YSOs at that separation.

4.4.4.2 Envelopes with $\mu = 1$

A power law of $\mu = 1$ means that the surface density of Class 0/I YSOs is directly proportional to the column density. This is a worthwhile test to perform as it is the simplest relationship in which the surface density of YSOs increases with column density. It is also of interest as within Orion B the distribution of prestellar cores have been observed to follow a linear relationship with column density above a visual extinction threshold of $A_v \sim 7$ (Könyves et al. 2020).

The results of applying this model are presented in the second column of Fig. 4.9. Serpens South, Serpens core, Ophiuchus and IC348 all exceed the envelope at small scales due to YSOs being more clustered at that scale than typically measured with a μ equal to 1. NGC1333 also exceeds the envelope though at a more intermediate scale of 1.3 pc.

4.4.4.3 Envelopes with $\mu = 2.05$

Following the results discussed previously in Section 4.4.1 the third model tested was that of the global value of $\mu = 2.05$. This power is the best estimate of a model where the distribution of Class 0/I protostars is proportional to column density with a consistent power across the five star-forming regions examined in this chapter.

The 95 per cent confidence envelopes presented in the third column of Fig. 4.9 show that Serpens South and NGC1333 both exceed the envelopes at spatial scales around 0.15 pc and therefore reject the model. IC348, Serpens Core and Ophiuchus remain entirely within the envelopes and are therefore consistent with the model. While still rejected by two regions, this was the most successful value of μ tested.

4.4.4.4 Envelopes with best estimate for μ

The final test performed on each region was using the best-estimate for μ calculated in Section 4.4.1. Unlike the previous models where one value of μ was applied to all of the regions, with this test each region was tested against a different value for μ.

Confidence envelopes were produced for each region using the best-estimates of μ presented in Table 4.2 and the results are presented in Fig. 4.10. Serpens South, Serpens Core, NGC1333 and IC348 all remain within their respective envelopes, however Ophiuchus rejects the model on small scales of 0.06 pc.

4.4.5 Application to Class II YSOs

Class II YSOs are more evolved than Class 0/I sources and tend to be less associated with the dense gas material (Mairs et al. 2016); it is likely, then, that the surface-density of Class II YSOs should follow a different power law with column density to Class 0/I YSOs, if any at all. To show that the O-ring statistic with 95 per cent confidence envelopes has enough discriminatory power to distinguish between YSO surface-density models the $\mu = 2.05$ model was applied to the Class II YSOs in each region.

The Class II YSOs were selected from the Dunham et al. (2015) catalogue with $-1.6 \leq \alpha < -0.3$ and $T_{\rm bol} > 100$K. Due to there being different numbers of Class II YSOs compared

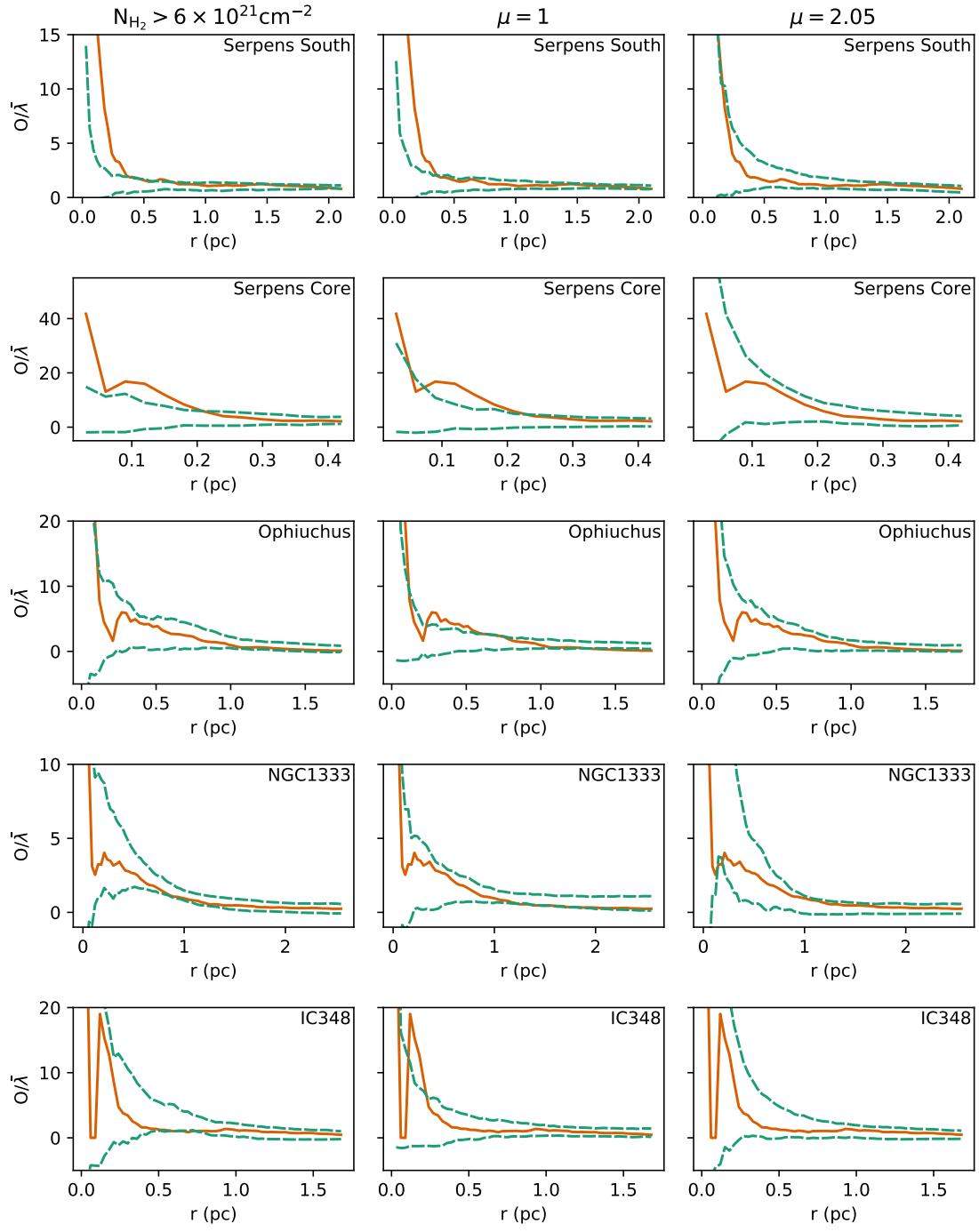

Figure 4.9: Measured $O/\hat{\lambda}$ vs r for Class 0/I YSOs in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 with 95 per cent confidence envelopes for different YSO surface-density models.

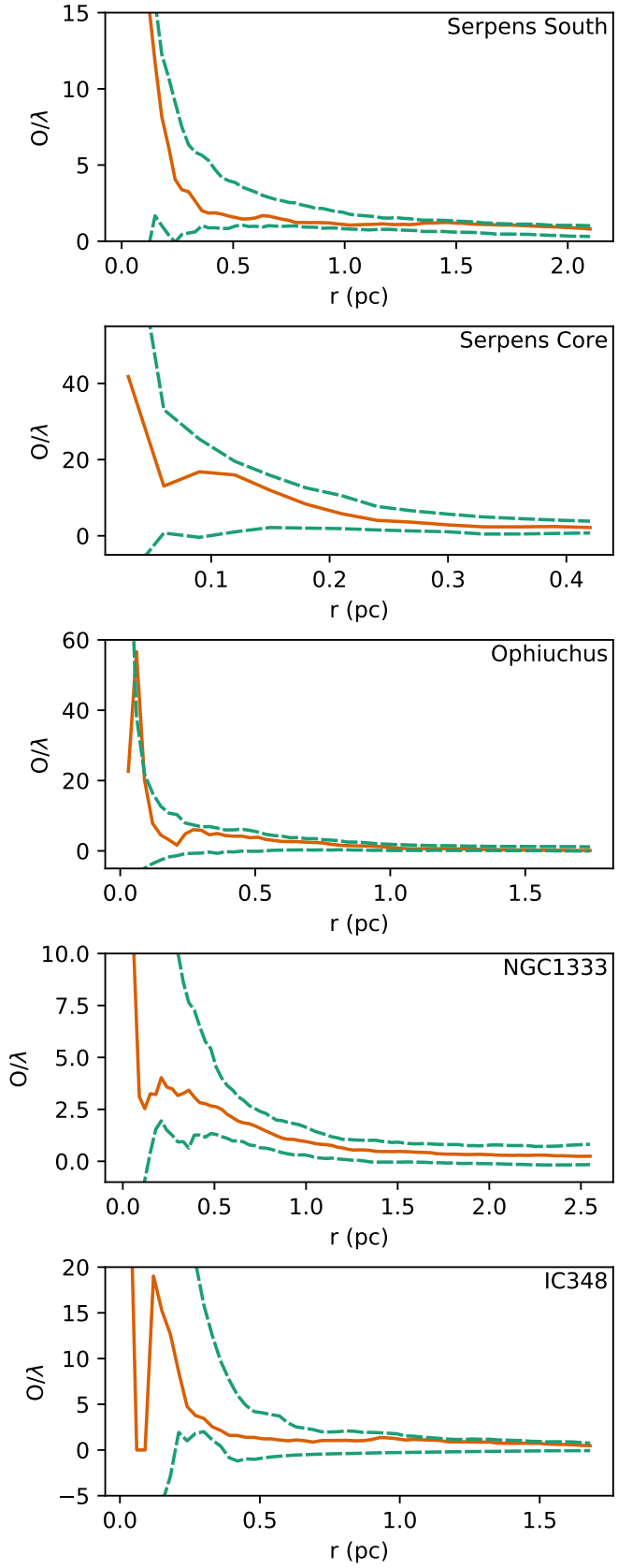

Figure 4.10: Measured $O/\hat{\lambda}$ vs r for Class 0/I YSOs in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 with 95 per cent confidence envelopes using the best-estimate for μ in each region from Table 4.2.

to Class 0/I the confidence envelopes were recalculated using 99 realisations.

Presented in Fig. 4.11 are the measured O-ring statistics and $\mu = 2.05$ model confidence envelopes for the Class II YSOs in each region. The measured $O/\bar{\lambda}$ values show that, except for NGC1333, Class II YSOs are less clustered compared to CSR at small scales than Class 0/I YSOs within the same region. Serpens South, Serpens Core, Ophiuchus and IC348 exceed the 95 per cent confidence envelopes and therefore reject the $\mu = 2.05$ model, while NGC1333 stays within the envelopes.

4.5 Discussion

In Sections 4.4.1 and 4.4.2 terms relating to possible power law relationships between the column density and the surface density of Class 0/I YSOs were estimated. In Section 4.4.4 methods from spatial statistics were used to determine if, and how many, star forming regions from the set were consistent with the stellar distribution models tested. These are complementary and independent methods as one does not necessarily imply the other; the first test assumed a model and found the parameters that best fit the model, while the spatial statistics tests determined the suitability of the proposed models.

4.5.1 Measured YSO surface density relations

The power-law relationship between the surface density of Class 0/I YSOs and column density was measured in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348, the results of which are presented in Tables 4.2 and 4.3. The power, μ, was estimated by marginalising the joint-probability distributions of Eqn. 4.6 for each region over the region-specific constant, C_r and vice-versa. As discussed, μ defines the relative likelihoods of forming YSOs at different column densities within a region, and C_r is a region-specific constant which normalises the number of YSOs formed within the region.

The region-specific constants were measured for the individual regions, as discussed in Section 4.4.2; however, it was also discussed that because the region-specific constants depend on μ and the units of Σ_{Gas}, comparison of C_r between regions of different μ is difficult to interpret. The dependence on μ can be mitigated by considering C_r

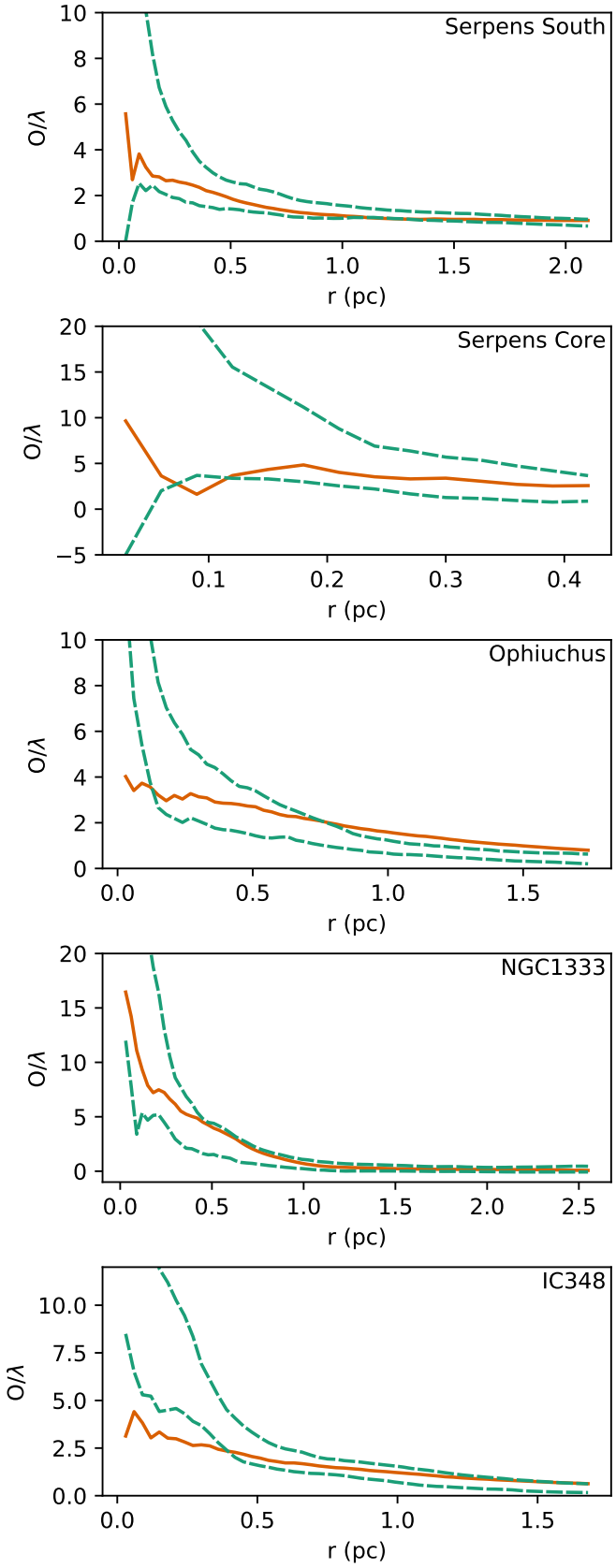

Figure 4.11: Measured $O/\hat{\lambda}$ vs r for Class II YSOs in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 with 95 per cent confidence envelopes for a $\mu = 2.05$ YSO surface-density model.

values when regions are assumed to have the same μ, and it was shown in Fig. 4.7 that different regions which have the same value of μ can have different values of C_r. Such differences in C_r are due to factors external to column density which cause different amounts of star-formation, for example age, magnetic fields, turbulence and different cloud orientations or geometries. And so, while C_r is important for estimating the YSO densities using Eqn. 4.2, the column density dependence is what is being tested using the O-ring statistic and so discussion will be focussed on μ.

Values of μ measured for the star forming regions in this work are consistent with studies looking for YSO surface density relationships in other star-forming regions (Gutermuth et al. 2011; Rapson et al. 2014; Willis et al. 2015; Lada et al. 2013; Lombardi et al. 2013; Lombardi et al. 2014; Pokhrel et al. 2020). Even high values of μ, such as that of IC348, have been measured such as Perseus with $\mu = 3$ (Hatchell et al. 2005) and $\mu = 3.8$ (Gutermuth et al. 2011) and the California Nebula with $\mu = 3.31$ (Lada et al. 2017) – though it is shown in Fig. 4.9 and Table 4.2 that IC348 is consistent with a much lower value of μ. There is some overlap between the regions tested in this chapter and those tested in other works: Ophiuchus with a μ of 1.78 is exceptionally close to the value of 1.87 and 1.9 measured by Gutermuth et al. (2011) and Pokhrel et al. (2020) respectively, and IC348, within the Perseus molecular cloud shows a similar power law to Gutermuth et al. (2011), though NGC1333 does not. It is interesting that these power laws show such similarity given the differences in the methods of measuring the power law, the column density measures, identifying the YSOs and the star-forming regions used. There is even potential that some of the higher values of μ, such as that of IC348, may be reduced in future with increasing resolution as happened with Orion B (Lombardi et al. 2014).

In addition to measuring μ for each region, Eqn. 4.8 was used to estimate the power-law value which best represents the YSO distributions in all five regions simultaneously by marginalising over the region-specific constants: $\mu = 2.05$. Unlike taking an average value of μ, which requires measured values of μ and an assumption as to how they should be weighted, this method directly uses the available data to estimate the parameter. Given this difference, it is interesting how similar this value is to the weighted mean $\mu = 2.06 \pm 0.14$ (with 95 per cent confidence intervals) for these regions and the mean value of $\mu \sim 2$

for the 12 regions studied in Pokhrel et al. (2020).

Compared to the individual estimates of μ, presented in Table 4.2 and Fig. 4.5, $\mu = 2.05$ appears to represent the ensemble of Bayesian fitted μ values well. While this is to expected, given that this value of μ was estimated using the YSO and area counts that were used to produce the values of μ for each region, it did not use the values of μ themselves and so demonstrates that the combination of measurements produces a value that is reasonable and representative. We can also see in Fig. 4.7 that for most regions $\mu = 2.05$, combined with the appropriate estimate of C_r, provides a good, visual representation of the YSO surface density measurements despite not being the best-estimate for the most regions.

4.5.2 Testing YSO distributions against spatial distribution models

A measured μ value describes how the surface density of YSOs changes in general across the entire study region. Measuring a power-law, however, does not mean that the YSOs are evenly distributed according to column density throughout the cloud. This was demonstrated in Section 4.4.3, where a value of μ measured from an evenly distributed population of YSOs could be reproduced in a population of stars distributed over only half of the cloud. From this it is reasonable to say that μ is a useful metric to describe YSO distributions but is not enough on its own to say whether YSOs have a relationship with cloud material of the form Eqn. 4.2. By utilising the spatial information, spatial statistics can test if observed distributions of YSOs are consistent with a power-law relationship with column density.

4.5.2.1 Class 0/I YSOs

One should expect the surface density of YSOs to be affected by column density. From a physics standpoint this makes sense as a greater reservoir of material has the potential to form more stars, and from an observational standpoint the values of μ measured star-forming regions are all greater than zero. The O-ring tests confirms this as Serpens South, Serpens Core, Ophiuchus and IC348 all have Class 0/I YSO populations inconsistent with YSOs positioned independently of column density above $A_v = 6$. It was also confirmed by

the O-ring test that this relationship is likely superlinear as all five regions rejected $\mu = 1$ and subsequent tests with higher values of μ all had fewer rejections.

Each region was also tested against the global estimate of $\mu = 2.05$, the 95 per cent confidence envelopes for which are presented in Fig. 4.9. These results show that of the five regions tested, Serpens Core, Ophiuchus and IC348 have Class 0/I YSO populations that are consistent with the $\mu = 2.05$ model. While it is unsurprising that Serpens Core is consistent with $\mu = 2.05$, given its power law was estimated to be $\mu = 2.06$, this is a more interesting result for Ophiuchus and IC348 as their estimates for μ were 1.78 and 3.00 respectively.

Serpens South and NGC1333 rejected the $\mu = 2.05$ model. This was due to over-clustering and regularity for Serpens South and NGC1333 respectively. Interestingly, the outcome of the envelope tests – with Serpens South and NGC1333 rejecting the $\mu = 2.05$ model while the other regions do not – is mirrored in the μ values measured in Table 4.2. The power $\mu = 2.05$ is within the 95 per cent confidence intervals for Serpens Core, Ophiuchus and IC348 individually while it is marginally outside the interval for Serpens South and NGC1333. It is perhaps due to the proximity of 2.05 to the 95 percent confidence intervals of Serpens South and NGC1333 that the O-ring values exceed the confidence envelopes over such a small set of spatial scales at ~ 0.15 pc.

Finally, each region was tested against its best-estimate for μ. Unlike the other models which assume a single value of μ, this model contains μ as an adjustable parameter for each region. By having five additional adjustable parameters in total, one should expect the number of YSO distributions that are consistent with the model to increase. This was observed in Fig. 4.10 where it was found that Serpens South, Serpens Core, NGC1333 and IC348 all have YSO populations consistent with their best-estimates of μ. From these results then we can see that Eqn. 4.2, using the Bayesian estimates of μ, is generally supported by spatial statistics. Though Ophiuchus rejected $\mu = 1.78$ at 0.06 pc, on similar scales to the regions which rejected $\mu = 2.05$.

4.5.2.2 Class II YSOs

While μ values were not measured for the Class II YSOs in these regions, by looking at the measured O-ring statistic and $\mu = 2.05$ envelopes in Fig. 4.11 it is clear that the two populations are not equally dependent on column density. The Class II YSOs in Serpens South, Serpens Core, Ophiuchus and IC348 are all inconsistent with the a $\mu = 2.05$ model, while those in NGC1333 remain within the envelope. This increase in rejection by older YSOs, in combination with lower $O/\bar{\lambda}$ values and generally flatter O-ring results as a function of radial separation, shows that there is a change in the separation of YSOs as a function of their age. These results also demonstrate that these tests have enough discriminatory power to distinguish between two distinct but related populations within the same region - Class 0/I and Class II YSOs.

4.5.3 Potential for a universal column density model

A question proposed at the beginning of this chapter was if it is possible to describe the locations of YSOs within a molecular cloud with a model that only uses column density. After applying four different models to the Class 0/I YSOs in five star forming regions every region was found to be consistent with at least one model. The answer to this question, therefore, appears to be 'yes' as the parameters for a given model can be tweaked in order to be consistent with a given set of YSOs. Given that an individual region can be described using a column density model, the next question is whether it is possible to describe the distributions of YSOs within multiple molecular clouds using the same column density model.

The most successful of the four models tested was that in which the best-estimate of μ calculated for each region using the Bayesian methodology from Section 4.2 was applied. Using this model, four out of the five regions were found to have YSO distributions consistent with being distributed throughout the cloud according to column density alone. It is possible, therefore, that if YSOs are distributed following column density alone that μ simply varies between star-forming regions and that there is no universal power-law distribution. However, not all of the regions were consistent with their best-estimate of μ and it is difficult to say whether this increase in consistency with the data is significant

enough to justify the addition of an adjustable parameter to the model.

As discussed in Sections 4.4.4.3 and 4.5.1, multiple regions can be consistent with the same power-law despite the best estimate of their μ values not being equivalent. Fig. 4.7 shows how a YSO surface density proportional to column density to the power of $\mu = 2.05$ represents the data quite well, and using the O-ring statistic the $\mu = 2.05$ model is able to describe the YSO distributions of Serpens Core, Ophiuchus and IC348 across all of the tested spatial scales. Out of the three models tested using a single value of μ, $\mu = 2.05$ performed the best with three regions out of five being consistent with the distribution. The first test with $\mu = 0$ above a column density threshold was only consistent with NGC1333 and $\mu = 1$ was not consistent with any of the regions.

While the Class 0/I YSOs in Serpens South and NGC1333 rejected the $\mu = 2.05$ model, this rejection was only over a small set of spatial scales between 0.12 pc and 0.18 pc, and on other scales the distribution was consistent with the model. This can be seen in Fig. 4.12 which shows the measured O-ring data from Fig. 4.9 for $r > 0.18$ pc; the O-ring statistics in Fig. 4.12 remain within the envelope across all scales and so appear consistent with the $\mu = 2.05$ model. It is worth emphasising that, while the envelopes in Fig. 4.12 have been adjusted to retain a 95 per cent significance level, the envelopes have been calculated using the same null hypothesis data from Fig. 4.9 and so is not an independent test. Fig. 4.12 does show, however, that remaining within the envelopes at larger r values is a feature of the data and not due to the envelopes being widened by the O-ring values which exceed the envelopes. From this we can say that the large-scale behaviour of the Class 0/I YSOs in all of these regions is well described by the same power-law relationship with column density.

4.5.4 Alternative universal models

Using both spatial statistics and Bayesian statistics it was shown that a power-law model with $\mu = 2.05$ provides a good approximation to the data. It is interesting to note that while Eqn. 4.2 appears to fit the measured YSO surface density data in Serpens South and NGC1333, as shown in Fig. 4.7, these regions both exceed the confidence envelopes. This could imply a situation like that discussed in Section 4.4.3 where column-density-

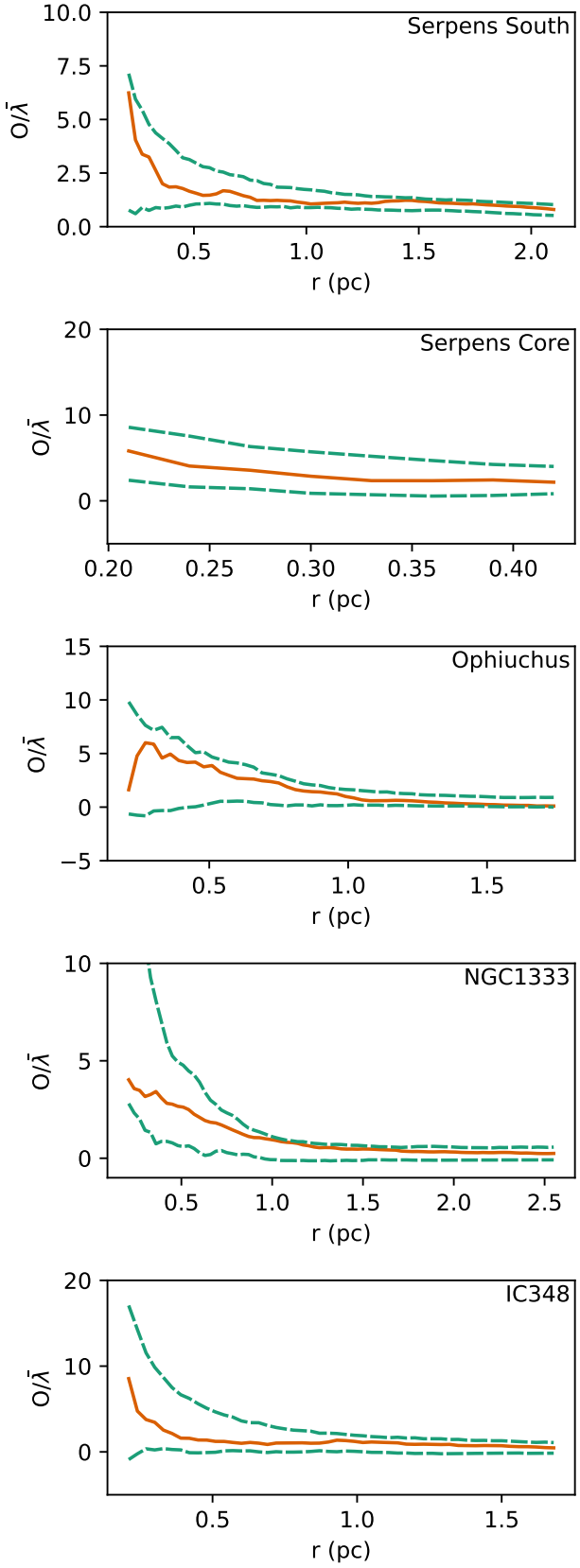

Figure 4.12: The measured $O/\hat{\lambda}$ vs r for Class 0/I YSOs in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 from Fig. 4.9 with 95 per cent confidence envelopes for a $\mu = 2.05$ YSO surface density model, using $r > 0.18$ pc.

independent effects resulted in star formation being unevenly distributed throughout the cloud.

The surface density of YSOs is not necessarily proportional to the column density to some power and so some modification of the first-order model Eqn. 4.2 may produce better results. The number of different models which could be simulated are potentially unlimited; however, the excursions from the $\mu = 2.05$ envelope were brief and otherwise are consistent with the power law and so any additional changes to the model should not have a large effect on the power-law relationship. Furthermore the scales on which these additional parameters influence the star formation need only be limited to small scales.

The results in Fig. 4.9 show that Serpens South and NGC1333 reject the model at scales close to 0.15 pc due to the over-clustering and under-clustering respectively and are otherwise consistent with the model. At most scales, therefore, the distribution of YSOs in Serpens South and NGC1333 behave similarly to this simple power-law relationship – except at a spatial scale of 0.15 pc. From Fig. 4.9 is not possible to determine exactly what this rejection means without further testing, however some possibilities will be discussed here.

The first option is that a global first-order model for star formation between clouds requires a different power law. This appears unlikely as the O-ring statistic shows over-clustering in Serpens South and under-clustering in NGC1333. Any increase in μ would lead to increased clustering at smaller scales while a decrease in μ would have the opposite effect, neither of which would necessarily represent Serpens South and NGC1333 simultaneously. Fig. 4.9 shows the results for a power law of $\mu = 1$ which consistently under-represent the density in Serpens South while NGC1333's O-ring statistic is consistent at small scales. Extrapolating the envelopes between the first and second columns of Fig. 4.9 provides some evidence against power laws less than 2, though more simulations would be required to conclusively determine this to be true.

A second option would be to add more parameters to the surface density model. One parameter of particular interest is a column density threshold for star formation. Indeed, from Fig. 4.9 it was shown that the YSOs within NGC1333 are consistent with

being positioned randomly in pixels with a column density greater than $6 \times 10^{21} \mathrm{cm}^{-2}$. Lombardi et al. (2013), hereby LLA, introduced a Bayesian method related to that in this chapter which uses the positions of protostars and the visual extinctions at those positions to estimate parameters in their model for protostellar surface densities. The surface density model in LLA is similar to Eqn. 4.2, except with two parameters in addition to C_r and μ (in their notation κ and β respectively): σ and A_0. σ is a diffusion coefficient term which allows for some amount of travel between the protostars' site of formation and observation and A_0 is a star formation threshold density. Lada et al. (2013) applied the method of LLA to Orion A, Orion B, California and Taurus, and found that there was no significant measurement of a diffusion coefficient and that a star-formation threshold may be more due to the distribution of material in the cloud – suggesting that the model is scale free. From the results in this work it is not possible to come to a conclusion on a model which uses both a power law and a threshold; no such model was tested and the number of YSOs from low column density regions in this work is insufficient to provide much insight on YSO distributions at low column densities. However, given the results of LLA the effect of including a column density threshold would be likely be limited.

A third option is that the power-law model cannot be applied to small spatial scales. This could be due to data-related problems, for example resolution. The spatial separations used start at, and are separated by, an interval of 0.03 pc, at these small spatial scales resolution effects become more prominent which in turn increases the likelihood that close, separate sources will be counted as a single source or a vice-versa. This is a particular problem at small radii where a small change in the number of YSOs has a large impact on the density. It could also be that the distributions of YSOs are affected by different physics at small-scales than large scales. The scales at which Serpens South and NGC1333 reject the global power law are at scales close to the filament scale of 0.1 pc (Arzoumanian et al. 2011), and average core separations in filaments of 0.14 pc (Könyves et al. 2020). Given that YSOs form within collapsing filaments it is possible that the structure of the filaments in which these Class 0/I YSOs form affects their distributions.

Finally, it may also be the case that a model with second-order components is needed to capture the nature of the distribution of star formation in star-forming regions. In a first-

order model the clustering of YSOs is a product of a general increase in YSO density due to a change in environment, in this case column density. It is also possible that clusters of protostars are not simply a function of increased density but are instead a product of a cluster-formation process which preferentially generates clusters in higher-density regions. Such behaviour can be represented through application of second-order effects which raise or lower the probability of forming a star as a function of distance from another star. This could be a YSO disrupting the column density of its immediate surroundings for example in NGC1333 (Knee and Sandell 2000), or it could be clusters of protostars forming within a dense core or filament for example in Perseus (Tobin et al. 2016).

4.5.5 Changing evolutionary timescales with column density

As discussed in Section 4.4.4.2 the surface density of prestellar cores in Orion B has been observed to follow a linear relationship with column density above a visual extinction threshold of $A_v \sim 7$ (Könyves et al. 2020) and observations have found that the YSOs in Orion B follow a power-law with $\mu \approx 2$ (Lombardi et al. 2014; Pokhrel et al. 2020). Fig. 4.9 shows that the YSOs in Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348 are not consistent with a linear model and so, while Orion B was not explored in this work, it is unlikely the YSOs in Orion B will be consistent with the linear power-law followed by the prestellar cores.

This leads naturally to the question of why these should be different if prestellar cores are expected to evolve into Class 0 YSOs. One reason for this could be that the prestellar cores and Class 0/I YSOs have time-scales that are affected by environment in different ways, or, the converse argument, for these distributions to be the same it would require the evolutionary time-scales of prestellar cores and Class 0/I YSOs to share the same dependence on the environment.

To see why this is the case a simple model of the rates of change of surface density over time, similar to that in nuclear decay, is introduced using a subset of eqns. (2)–(7) from Kristensen and Dunham (2018). Assuming prestellar cores are produced at a constant rate, γ (which may be a function of local column density), and evolve into Class 0/I YSOs with

a lifetime τ_{PC} the change in prestellar core surface density is

$$\frac{d\Sigma_{PC}}{dt} = \gamma - \frac{\Sigma_{PC}(t)}{\tau_{PC}}, \tag{4.17}$$

where Σ_{PC} is the surface density of prestellar cores. Similarly, assuming Class 0/I YSOs evolve into Flat or Class II YSOs with a lifetime $\tau_{0/I}$ the surface density of Class 0/I YSOs, $\Sigma_{0/I}$, is

$$\frac{d\Sigma_{0/I}}{dt} = \frac{\Sigma_{PC}(t)}{\tau_{PC}} - \frac{\Sigma_{0/I}(t)}{\tau_{0/I}}. \tag{4.18}$$

The solutions to Eqns. 4.17 and 4.18 are

$$\Sigma_{PC} = \gamma\tau_{PC}\left(1 - e^{-t/\tau_{PC}}\right), \tag{4.19}$$

and

$$\Sigma_{0/I} = \gamma\tau_{0/I}\left(1 - \frac{\tau_{PC}}{\tau_{PC} - \tau_{0/I}}e^{-t/\tau_{PC}} - \frac{\tau_{0/I}}{\tau_{0/I} - \tau_{PC}}e^{-t/\tau_{0/I}}\right) \tag{4.20}$$

respectively, where it is assumed that at $t = 0$ the surface density of prestellar cores and protostars are zero.

Everything inside the brackets of Eqns. 4.19 and 4.20 is unitless and column density independent. The column density dependence of a population, therefore, is defined by the product of prestellar core formation rate and the lifetime of the population. For simplicity the solutions to the steady-state condition, where Eqns. 4.17 and 4.18 are both equal to zero, are

$$\Sigma_{PC} = \gamma\tau_{PC} \tag{4.21}$$

and

$$\Sigma_{0/I} = \gamma\tau_{0/I}. \tag{4.22}$$

From inspection it can be seen that for Σ_{PC} and $\Sigma_{0/I}$ to share the same column density dependence, their lifetimes must also be equally dependent on column density. In other words, if the prestellar cores and Class 0/I YSOs in a region have different column density dependences this could be due to different column density dependences of the evolutionary timescales of prestellar cores and protostars.

As discussed, observations in Orion B find that Σ_{PC} is linearly proportional to Σ_{Gas} while $\Sigma_{0/I}$ is proportional to Σ_{Gas} to a power of about two; additionally, measurements in Monoceros R2 show $\Sigma_{PC} \propto \Sigma_{Gas}^{2}$ (Sokol et al. 2019) and $\Sigma_{0/I} \propto \Sigma_{Gas}^{2.67}$ (Gutermuth et al. 2011). And, while the power-law relations between column density and the surface densities of prestellar cores and protostars were not measured, the two-point correlation function (2PCF) was applied to the prestellar cores and protostars in Serpens South, Ophiuchus and Perseus by Enoch et al. (2008). The 2PCF showed that the protostars in these regions were more clustered than the prestellar cores and that, while the prestellar cores were not as clustered, their relative densities fell off more slowly as a function of spatial separation. This clustering behaviour is what would be expected from protostars that have a higher μ than prestellar cores: greater clustering on small scales near high-density material compared to a more dispersed population. And so, while the exact values of μ may differ between regions these observations, combined with Eqns. 4.21 and 4.22, strongly suggests that τ_{PC} and $\tau_{0/I}$ must have different dependences on column density due to interactions with the environment such as on-going accretion.

To gain some insight into how the relative time-scale depends on column density in Orion B I substitute in the observed relations of $\Sigma_{0/I} \propto \Sigma_{Gas}^{2}$ – from this and other measurements discussed in Section 4.5.3 – and $\Sigma_{PC} \propto \Sigma_{Gas}$ from Könyves et al. (2020):

$$\gamma\tau_{PC} \propto \Sigma_{Gas}, \tag{4.23}$$

$$\gamma\tau_{0/I} \propto \Sigma_{Gas}^{2} \tag{4.24}$$

and

$$\frac{\tau_{PC}}{\tau_{0/I}} \propto \Sigma_{Gas}^{-1}, \tag{4.25}$$

where Eqn. 4.25, the ratio of Σ_{PC} and $\Sigma_{0/I}$, states that the difference in column density dependence between τ_{PC} and $\tau_{0/I}$ is a factor of Σ_{Gas}. This suggests that prestellar cores evolve more quickly at higher column densities than Class 0/I YSOs. There are different ways to interpret this: (i) prestellar cores evolve on shorter time-scales at higher column densities; (ii) Class 0/I YSOs remain embedded in their envelope longer at higher column densities; (iii) alternatively, both are column density dependent in some form with prestellar cores

ultimately evolving faster than Class 0/I YSOs at higher column densities.

It is very likely that both τ_{PC} and $\tau_{0/I}$ are column density dependent. For prestellar cores their lifetime is often compared to the free-fall time of a spherically-symmetric mass:

$$t_{ff} \propto \rho^{-1/2}, \tag{4.26}$$

where ρ is the density of the sphere. Eqn. 4.26 shows that, since free-fall time is proportional to volume density to a power $-1/2$, higher density objects collapse more quickly. Numerical simulations have shown that Bonner–Ebert spheres have higher central densities and are quicker at collapsing within higher density environments (Kaminski et al. 2014). It appears possible that τ_{PC} is lower at higher column densities.

For Class 0/I protostars to take longer to evolve at higher column densities it would require that they remain embedded within their envelopes for longer compared to their lower-column-density counterparts. It may be the case that Class 0/I protostars are able to remain embedded while material is available for accretion which would result in longer lifetimes in regions that are more dense. This is in part supported by numerical simulations where it was found that the accretion rate onto protostars was equivalent between two simulated clouds of different densities (Bate and Bonnell 2005). Assuming this to be true, a change in $\tau_{0/I}$ with respect to column density could be observable in the relative masses in protostars in regions of different column density. Indeed some evidence of this has been observed in mass segregation in YSOs and dense cores, where the most massive sources were found within regions with higher densities of sources and towards the central location of the cluster (Kirk and Myers 2011; Kirk et al. 2016). It was also noted in Bate and Bonnell (2005) that objects formed within a denser cloud showed a greater variation in the time taken for an object to accrete. As a counter argument, the same simulations also showed that dynamical interactions between objects were the dominant force in terminating accretion and objects were more likely to be ejected sooner in a higher density cloud (Bate 2012). This would imply that $\tau_{0/I}$ is smaller in higher column densities. These are, however, results from numerical simulations and observational evidence is currently insufficient to convincingly support either lengthening or shortening lifetimes.

Ultimately, from these ratios it is not possible to determine which of the terms γ, τ_{YSO} or $\tau_{0/I}$ are column density dependent, but at least two of the terms must be functions of column density, one of which must be an evolution time-scale for prestellar cores or Class 0/I YSOs. This is true for any region in which Σ_{PC} and $\Sigma_{0/I}$ are measured to have different dependencies on column density.

4.6 Conclusions

In this chapter the distribution of Class 0/I YSOs in Serpens South, Serpens core, Ophiuchus, NGC1333 and IC348 were tested against a spatial distribution model of the form

$$\hat{\lambda}(N_{H_2}) \propto {N_{H_2}}^{\mu}, \tag{4.27}$$

where $\hat{\lambda}(N_{H_2})$ is the estimate of the surface density of Class 0/I YSOs at a column density N_{H_2}, and μ is some power.

(i) It was found that four of the regions had Class 0/I populations inconsistent with $\mu = 0$ when combined with a threshold column density of $6 \times 10^{21} N_{H_2} cm^{-2}$ and zero probability elsewhere – implying that star formation is not decoupled from column density (Section 4.4.4.1).

(ii) The Class 0/I YSOs in all of the tested regions were also found to be inconsistent with $\mu = 1$ – the power law associated with the surface densities of prestellar cores (Section 4.4.4.2).

(iii) The power law index μ was measured for each region individually in Section 4.4.1, the results of which are tabulated in Table 4.2, and by combining the YSO surface density data from all regions a global μ value was measured to be 2.05 ± 0.20 where the reported uncertainty is the 95 per cent confidence interval.

(iv) The best value of μ tested was that of the global μ value 2.05 in Section 4.4.4.3, with only the YSOs in Serpens South and NGC1333 rejecting the model between 0.12 pc and 0.18 pc. It was shown that all five regions were consistent with $\mu = 2.05$ when considering radial separations greater than 0.18 pc.

Table 4.5: Table of Symbols

Symbol	Description
α	significance level
n	number of simulated patterns
$\boldsymbol{s}$	arbitrary areas in study region
w	star-star nearest neighbour distance
x	position-star nearest neighbour distance
r	radius, radial separation
$\boldsymbol{A}$	study region
q	half-width of annulus
ρ	multiplicative factor for annulus width
R	cluster radius
Σ_{SFR}	star formation rate surface density
Σ_{GAS}	gas surface density
C_{r}	region-specific constant
μ	power-law index of SFR surface density relation
λ	first-order intensity
$\mathrm{N}_{\mathrm{H_2}}$	column density cm^{-2}
N	number of YSOs
N_{c}	number of points in cluster
N_{bg}	number of background points
N_{total}	total number of points
u,v	cell indices
$\Delta\sigma$	angular distance
α	right ascension (RA), significance level
δ	declination (Dec)
H_0	null hypothesis
n	number of simulated patterns
A	area

(v) Serpens South and NGC1333 rejected the $\mu = 2.05$ model at a radial separation of $\sim$ 0.15 pc. This could be due to physical effects such as a preferential scaling for filament collapse or small-scale interactions between YSO or data-related issues, such as resolution. However, because of the generally good fit to the model any modification should be limited to small spatial scale interactions.

(vi) Class 0/I YSOs were shown to have a different relationship to column density than Class II YSOs (Section 4.4.5) showing that this relationship is not consistent over time.

(vii) In Section 4.5.5, using a toy evolution model it was determined that column density plays a role in at least two of the three terms: prestellar core formation rate, prestellar core evolutionary time-scale and Class 0/I evolutionary time-scale.

Chapter 5

Conclusion

5.1 Summary

In this thesis I tested the sensitivity of four methods from spatial statistics to clustering in the presence of background noise: Diggle's G function (G), the 'free-space' function (F), Ripley's K (K) and the O-ring statistic (O). By applying these tests to Gaussian clusters projected on a background of random stars it was found that the second-order statistics K and O were more consistently able to reject CSR than the first-order statistics G and F. While Ripley's K and the O-ring statistic performed similarly I recommended O-ring for future applications.

I then investigated the spatial distributions of Class 0/I YSOs in the local star-forming regions Serpens South, Serpens Core, Ophiuchus, NGC1333 and IC348, with respect to column density. Using a Bayesian method I estimated the power-law relationship between the surface density of YSOs and column density for each region individually and the power-law which represents all of the regions as a collective, $\Sigma_{\rm YSO} \propto \Sigma_{\rm Gas}^{2.05}$.

The Class 0/I YSOs in these five regions were tested then for their consistency against three common models, and one individual model. From the set of common models, the most successful model was $\mu = 2.05$. The YSOs in all five regions appear to be consistent with this model on large scales, however at small scales (~ 0.15 pc) the Class 0/I YSOs in Serpens South and NGC1333 were shown to be inconsistent with this model.

The most successful model overall was using the best-estimate for μ for the individual region which was only rejected by Ophiuchus. However, it unclear if this improvement is significant enough to justify the addition of five adjustable parameters – one for each region.

5.2 Proposed future work

5.2.1 Class II YSOs

The positions of Class 0/I YSOs are indicative of where they initially formed within the cloud, and so from their distribution it is possible to infer information about recent star formation within the cloud. However, as discussed in chapter 1, YSOs become less densely clustered and less associated with dense gas with increasing class. It is possible that this evolution of distributions is due to a change in the morphology of the molecular cloud over time, or the later-class YSOs have travelled further from their sites of formation. Application of the methods discussed in this work could provide some insight into what causes these distributions to change over time.

If this evolution is due to a change in the morphology of the molecular cloud then we would expect the locations of more evolved YSOs, such as the Class IIs, to have had a relationship with column density similar to that of Class 0/Is at the time of formation. The way to test this is to produce first-order intensity maps that are estimations of earlier cloud morphologies and produce confidence envelopes for these models. One potential option to explore is estimating the first-order map based on the current distribution of Class II YSOs. Wiegand and Moloney (2004) achieve this by calculating the first-order intensity within a moving kernel of a given radius. Fig. 5.1 shows the estimated first-order intensity maps for the Class 0/I and Class II YSOs in Serpens South using a square kernel. We can see from this figure how spatially different the distributions of each class are. Fig. 5.1 also shows the limitations of this method where the density of points is low and the kernel becomes more granular.

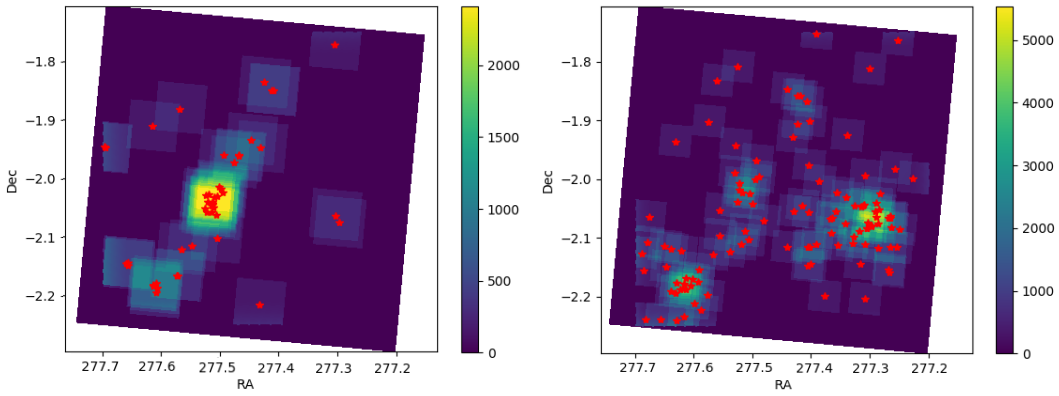

Figure 5.1: Estimated first-order intensity maps for Class 0/I YSOs (left) and Class II YSOs (right) in Serpens South using a square kernel.

5.2.2 Column density threshold

Another area in which this work can be extended is in the estimation of a column density threshold. The study regions which are looked at in chapter 4 were selectively chosen to test K–S relations, and so including low column density regions with few YSOs were not strictly necessary. However, by extending the study region to contain more low column density material it is possible to measure the YSO surface density and observe the K–S law at column densities below the typical threshold of $A_V \approx 6$. In addition, with the inclusion of more lower column density regions it becomes possible to test column density threshold models using the methods in Chapter 4.

5.2.3 Numerical simulations

Another possible extension of this project is in the use of these methods to aid, and be aided by numerical simulations of star formation such as those of (Bate 2011). Numerical simulations of star formation are able to consider the three-dimensional structure of the cloud, whereas the results of this thesis apply to the two-dimensional projection of the cloud. As such, the methods in this work could be useful as a means for testing the physical accuracy of three-dimensional simulations by looking at the spatial distributions of the simulated YSOs when projected into 2D.

The analysis of such a work would follow a similar path to that in Chapter 4. First, checking if the best-estimate of μ is similar to that of observed distributions, and then comparing against confidence envelopes for simulations with this newly-calculated estimate

of μ, and $\mu = 2.05$. A rejection of either, or both, of these models may not be a significant result; however, a significant and consistent deviation from the confidence envelopes could indicate that the simulation may be missing some key physics. This is because the excursions from the envelopes in Fig. 4.9 were brief and over only small scales.

On the other side, numerical simulations could help inform as to how projection affects can alter observed power-laws. In addition, by including the radial distances to objects, it would be possible to produce similar models that relate the star-formation rate to the volume density of star-forming material.

5.2.4 Cluster scales

Statistics such as the pair-correlation function have expected values under a null hypothesis and comparison to these values can indicate scales at which clustering or regularity is present. However, even a random pattern will contain some degree of clustering simply due to random fluctuations of points within the space. Comparison to these values is therefore not a reliable indicator that the pattern contains significant clustering. We can, however, use global confidence envelopes of CSR as tests for general clustering, and the regions at which the statistic exceeds the envelope are those that contain significant clustering.

To illustrate this, I have reproduced Fig. 5.2 showing only the results of O-ring. From this figure we can see that there is an initial region of clustering which ends at an angular separation of $\sim 0.1°$ and another region with a peak at $\sim 0.2°$. Comparing these values to the YSOs in Fig. 5.2 these separations coincide with the size of the individual subclusters and the distance between the subclusters. Features such as peaks in correlation functions are already used to identify characteristic clustering scales, however the global confidence envelope provides a new threshold, and a means to identify sets of scales that are above it. Applying this method of calculating scales to YSO distributions would be an interesting extension to the work in this thesis.

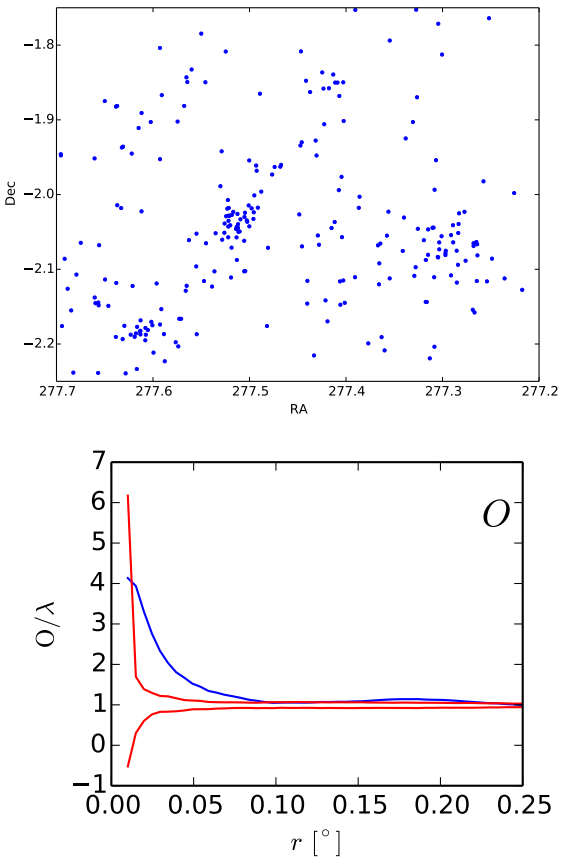

Figure 5.2: (above) Positions of YSOs within Serpens South, (below) O-ring statistic with 95 per cent global confidence envelopes for CSR.

5.2.5 Marked spatial point process analysis

The spatial analysis methods applied in chapters 3 and 4 use only the location data of the YSOs. Another direction in which this work could be expanded is to include the quantitative data, or mark, attached to the YSOs, such as mass, $T_{\rm bol}$ or α. Points which contain data are called a marked variable and the process is called a marked point processes (Feigelson and Babu 2012). The marks will not necessarily follow the same distribution as the YSOs. Analysis of the marked processes could indicate features such as a strong correlation in mass on small scales where more massive YSOs are more centrally concentrated.

5.3 Acknowledgements

This research has made use of the Starlink software (Currie et al. 2014) which is supported by the East Asian Observatory. The figures in this paper have been produced using matplotlib: a 2D graphics package used for Python for application development, interactive scripting, and publication-quality image generation across user interfaces and operating systems (Hunter 2007). This research made use of Astropy,[1] a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018).

This research has made use of NASA's Astrophysics Data System.

This work is based (in part) on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. The Spitzer data used in this thesis are available in NASA/IPAC Infrared Science Archive at https://irsa.ipac.caltech.edu/data/SPITZER/C2D/images/ and the Dunham et al. (2015) Young Stellar Object source data are available at https://doi.org/10.1088%2F0067-0049%2F220%2F1%2F11 .

This research has made use of data from the Herschel Gould Belt survey (HGBS) project (http://gouldbelt-herschel.cea.fr). The HGBS is a Herschel Key Programme jointly carried out by SPIRE Specialist Astronomy Group 3 (SAG 3), scientists of several institutes in the PACS Consortium (CEA Saclay, INAF-IFSI Rome and INAF-Arcetri, KU Leu-

1. http://www.astropy.org

ven, MPIA Heidelberg), and scientists of the Herschel Science Center (HSC). The *Herschel* Gould Belt survey (HGBS) data available in HGBS Archive, at http://www.herschel.fr/cea/gouldbelt/en/ .

Bibliography

Allison, R. J., S. P. Goodwin, R. J. Parker, et al. 2009. 'Using the minimum spanning tree to trace mass segregation.' MNRAS 395 (May): 1449–1454.

Alves, João, Charles J. Lada, Elizabeth A. Lada, Scott J. Kenyon, and Randy Phelps. 1998. 'Dust Extinction and Molecular Cloud Structure: L977.' ApJ 506, no. 1 (October): 292–305.

Andre, P., D. Ward-Thompson, and M. Barsony. 1993. 'Submillimeter continuum observations of Rho Ophiuchi A - The candidate protostar VLA 1623 and prestellar clumps.' ApJ 406 (March): 122–141.

André, P., A. Men'shchikov, S. Bontemps, et al. 2010. 'From filamentary clouds to prestellar cores to the stellar IMF: Initial highlights from the Herschel Gould Belt Survey.' A&A 518 (July): L102.

Arzoumanian, D., P. André, P. Didelon, et al. 2011. 'Characterizing interstellar filaments with Herschel in IC 5146.' A&A 529 (May): L6.

Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, et al. 2013. 'Astropy: A community Python package for astronomy.' A&A 558 (October): A33.

Baddeley, A. J., M. Kerscher, K. Schladitz, and B. T. Scott. 2000. 'Estimating the J function without edge correction.' *Statistica Neerlandica* 54 (3): 315–328.

Baddeley, Adrian, Peter J. Diggle, Andrew Hardegen, et al. 2014. 'On tests of spatial pattern based on simulation envelopes.' *Ecological Monographs* 84, no. 3 (August): 477–489.

Baddeley, Adrian, and Richard D. Gill. 1997. 'Kaplan-Meier estimators of distance distributions for spatial point processes.' *Ann. Statist.* 25, no. 1 (February): 263–292.

Barot, Sébastien, Jacques Gignoux, and Jean-Claude Menaut. 1999. 'DEMOGRAPHY OF A SAVANNA PALM TREE: PREDICTIONS FROM COMPREHENSIVE SPATIAL PATTERN ANALYSES.' *Ecology* 80 (6): 1987–2005.

Bate, M. R. 2009. 'Stellar, brown dwarf and multiple star properties from hydrodynamical simulations of star cluster formation.' MNRAS 392 (January): 590–616.

Bate, M. R., I. A. Bonnell, and V. Bromm. 2003. 'The formation of a star cluster: predicting the properties of stars and brown dwarfs.' MNRAS 339 (March): 577–599.

Bate, M. R., C. J. Clarke, and M. J. McCaughrean. 1998. 'Interpreting the mean surface density of companions in star-forming regions.' MNRAS 297 (July): 1163–1181.

Bate, Matthew R. 2011. 'Collapse of a molecular cloud core to stellar densities: the formation and evolution of pre-stellar discs.' MNRAS 417, no. 3 (November): 2036–2056.

———. 2012. 'Stellar, brown dwarf and multiple star properties from a radiation hydrodynamical simulation of star cluster formation.' MNRAS 419, no. 4 (February): 3115–3146.

Bate, Matthew R., and Ian A. Bonnell. 2005. 'The origin of the initial mass function and its dependence on the mean Jeans mass in molecular clouds.' MNRAS 356, no. 4 (February): 1201–1221.

Buckner, Anne S. M., Zeinab Khorrami, Marta González, et al. 2020. 'The spatial evolution of young massive clusters. II. Looking for imprints of star formation in NGC 2264 with Gaia DR2.' A&A 636 (April): A80.

Buckner, Anne S. M., Zeinab Khorrami, Pouria Khalaj, et al. 2019. 'The spatial evolution of young massive clusters. I. A new tool to quantitatively trace stellar clustering.' A&A 622 (February): A184.

Calabretta, M. R., and E. W. Greisen. 2002. 'Representations of celestial coordinates in FITS.' A&A 395 (December): 1077–1122.

Cantat-Gaudin, T., A. Vallenari, R. Sordo, et al. 2018. 'Characterising open clusters in the solar neighbourhood with the Tycho-Gaia Astrometric Solution.' A&A 615 (July): A49.

Cartwright, A., and A. P. Whitworth. 2004. 'The statistical analysis of star clusters.' MNRAS 348 (February): 589–598.

Chen, H., P. C. Myers, E. F. Ladd, and D. O. S. Wood. 1995. 'Bolometric Temperature and Young Stars in the Taurus and Ophiuchus Complexes.' ApJ 445 (May): 377.

Clarke, S. D., A. P. Whitworth, A. Duarte-Cabral, and D. A. Hubber. 2017. 'Filamentary fragmentation in a turbulent medium.' MNRAS 468, no. 2 (June): 2489–2505.

Covey, K. R., T. P. Greene, G. W. Doppmann, and C. J. Lada. 2006. 'The Radial Velocity Distribution of Class I and Flat-Spectrum Protostars.' AJ 131 (January): 512–519.

Crapsi, A., E. F. van Dishoeck, M. R. Hogerheijde, K. M. Pontoppidan, and C. P. Dullemond. 2008. 'Characterizing the nature of embedded young stellar objects through silicate, ice and millimeter observations.' A&A 486 (July): 245–254.

Cressie, Noel. 1993. *Statistics for Spatial Data.* Wiley-Interscience.

Currie, M. J., D. S. Berry, T. Jenness, et al. 2014. 'Starlink Software in 2013.' In *Astronomical Data Analysis Software and Systems XXIII*, edited by N. Manset and P. Forshay, 485:391. Astronomical Society of the Pacific Conference Series. May.

Dale, Mark R. T., and Marie-Josée Fortin. 2014. *Spatial Analysis: A Guide For Ecologists.* Cambridge University Press.

Davis, M., and P. J. E. Peebles. 1983. 'A survey of galaxy redshifts. V. The two-point position and velocity correlations.' ApJ 267 (April): 465–482.

di Francesco, J., II Evans N. J., P. Caselli, et al. 2007. 'An Observational Perspective of Low-Mass Dense Cores I: Internal Physical and Chemical Properties.' In *Protostars and Planets V*, edited by Bo Reipurth, David Jewitt, and Klaus Keil, 17. January.

Diggle, Peter J. 1979. 'On Parameter Estimation and Goodness-of-Fit Testing for Spatial Point Patterns.' *Biometrics* 35 (1): 87–101.

Diggle, Peter J. 2013. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition (Chapman & Hall/CRC Monographs on Statistics & Applied Probability).* Chapman / Hall/CRC.

Dunham, M. M., L. E. Allen, N. J. Evans II, et al. 2015. 'Young Stellar Objects in the Gould Belt.' ApJS 220 (September): 11.

Enoch, Melissa L., II Evans Neal J., Anneila I. Sargent, et al. 2008. 'The Mass Distribution and Lifetime of Prestellar Cores in Perseus, Serpens, and Ophiuchus.' ApJ 684, no. 2 (September): 1240–1259.

Evans, II, Neal J., Lori E. Allen, Geoffrey A. Blake, et al. 2003. 'From Molecular Cores to Planet-forming Disks: An SIRTF Legacy Program.' PASP 115, no. 810 (August): 965–980.

Evans, N. J., II, Michael M. Dunham, Jes K. Jørgensen, et al. 2009. 'The Spitzer c2d Legacy Results: Star-Formation Rates and Efficiencies; Evolution and Lifetimes.' *The Astrophysical Journal Supplement Series* 181 (2): 321.

Feigelson, Eric D., and G. Jogesh Babu. 2012. *Modern Statistical Methods for Astronomy: With R Applications.* Cambridge University Press.

Fiege, Jason D., and Ralph E. Pudritz. 2000. 'Helical fields and filamentary molecular clouds - I.' MNRAS 311, no. 1 (January): 85–104.

Foreman-Mackey, Daniel, David W. Hogg, Dustin Lang, and Jonathan Goodman. 2013. 'emcee: The MCMC Hammer.' *Publications of the Astronomical Society of the Pacific* 125, no. 925 (March): 306–312.

Gignoux, Jacques, Camille Duby, and Sébastian Barot. 1999. 'Comparing the Performances of Diggle's Tests of Spatial Randomness for Small Samples with and without Edge-Effect Correction: Application to Ecological Data.' *Biometrics* 55 (1): 156–164.

Goodman, Jonathan, and Jonathan Weare. 2010. 'Ensemble samplers with affine invariance.' *Communications in Applied Mathematics and Computational Science* 5, no. 1 (January): 65–80.

Greene, T. P., B. A. Wilking, P. Andre, E. T. Young, and C. J. Lada. 1994. 'Further mid-infrared study of the rho Ophiuchi cloud young stellar population: Luminosities and masses of pre-main-sequence stars.' ApJ 434 (October): 614–626.

Greisen, E. W., and M. R. Calabretta. 2002. 'Representations of world coordinates in FITS.' A&A 395 (December): 1061–1075.

Gutermuth, R. A., T. L. Bourke, L. E. Allen, et al. 2008. 'The Spitzer Gould Belt Survey of Large Nearby Interstellar Clouds: Discovery of a Dense Embedded Cluster in the Serpens-Aquila Rift.' *The Astrophysical Journal Letters* 673 (2): L151.

Gutermuth, R. A., S. T. Megeath, P. C. Myers, et al. 2009. 'A Spitzer Survey of Young Stellar Clusters Within One Kiloparsec of the Sun: Cluster Core Extraction and Basic Structural Analysis.' ApJS 184 (September): 18–83.

Gutermuth, R. A., J. L. Pipher, S. T. Megeath, et al. 2011. 'A Correlation between Surface Densities of Young Stellar Objects and Gas in Eight Nearby Molecular Clouds.' ApJ 739, no. 2 (October): 84.

Harvey, Paul M., Cassandra Fallscheer, Adam Ginsburg, et al. 2013. 'A First Look at the Auriga-California Giant Molecular Cloud with Herschel and the CSO: Census of the Young Stellar Objects and the Dense Gas.' ApJ 764, no. 2 (February): 133.

Hastings, W. K. 1970. 'Monte Carlo Sampling Methods Using Markov Chains and Their Applications.' *Biometrika* 57 (1): 97–109.

Hatchell, J., J. S. Richer, G. A. Fuller, et al. 2005. 'Star formation in Perseus. Clusters, filaments and the conditions for star formation.' A&A 440, no. 1 (September): 151–161.

Heiderman, Amanda, II Evans Neal J., Lori E. Allen, Tracy Huard, and Mark Heyer. 2010. 'The Star Formation Rate and Gas Surface Density Relation in the Milky Way: Implications for Extragalactic Studies.' ApJ 723, no. 2 (November): 1019–1037.

Heyer, Mark, and T. M. Dame. 2015. 'Molecular Clouds in the Milky Way.' ARA&A 53 (August): 583–629.

Hildebrand, R. H. 1983. 'The determination of cloud masses and dust characteristics from submillimetre thermal emission.' QJRAS 24 (September): 267–282.

Ho, L. P., and S. N. Chiu. 2006. 'Testing the complete spatial randomness by Diggle's test without an arbitrary upper limit.' *Journal of Statistical Computation and Simulation* 76 (7): 585–591.

Ho, Lai Ping, and Sung Nok Chiu. 2009. 'Using Weight Functions in Spatial Point Pattern Analysis with Application to Plant Ecology Data.' *Communications in Statistics - Simulation and Computation* 38 (2): 269–287.

Hunter, J. D. 2007. 'Matplotlib: A 2D graphics environment.' *Computing In Science & Engineering* 9 (3): 90–95.

Inutsuka, S.-i., and S. M. Miyama. 1997. 'A Production Mechanism for Clusters of Dense Cores.' ApJ 480 (May): 681–693.

Johnson, Joel P. L. 2017. 'Clustering statistics, roughness feedbacks, and randomness in experimental step-pool morphodynamics.' *Geophysical Research Letters* 44 (8): 3653–3662.

Johnstone, Doug, James Di Francesco, and Helen Kirk. 2004. 'An Extinction Threshold for Protostellar Cores in Ophiuchus.' ApJL 611, no. 1 (August): L45–L48.

Kaminski, Erica, Adam Frank, Jonathan Carroll, and Phil Myers. 2014. 'ON THE ROLE OF AMBIENT ENVIRONMENTS IN THE COLLAPSE OF BONNOR-EBERT SPHERES.' *The Astrophysical Journal* 790, no. 1 (July): 70.

Kauffmann, Guinevere, Jorg M. Colberg, Antonaldo Diaferio, and Simon D. M. White. 1999. 'Clustering of galaxies in a hierarchical universe - I. Methods and results at z=0.' MNRAS 303, no. 1 (February): 188–206.

Kennicutt, Jr., Robert C. 1989. 'The Star Formation Law in Galactic Disks.' ApJ 344 (September): 685.

———. 1998. 'The Global Schmidt Law in Star-forming Galaxies.' ApJ 498, no. 2 (May): 541–552.

Kirk, H., D. Johnstone, J. Di Francesco, et al. 2016. 'THE JCMT GOULD BELT SURVEY: DENSE CORE CLUSTERS IN ORION B.' *The Astrophysical Journal* 821, no. 2 (April): 98.

Kirk, Helen, and Philip C. Myers. 2011. 'Young Stellar Groups and Their Most Massive Stars.' ApJ 727, no. 2 (February): 64.

Knee, L. B. G., and G. Sandell. 2000. 'The molecular outflows in NGC 1333.' A&A 361 (September): 671–684.

Könyves, V., Ph. André, D. Arzoumanian, et al. 2020. 'Properties of the dense core population in Orion B as seen by the Herschel Gould Belt survey.' A&A 635 (March): A34.

Könyves, V., Ph. André, A. Men'shchikov, et al. 2010. 'The Aquila prestellar core population revealed by Herschel.' A&A 518 (July): L106.

Könyves, V., Ph. André, A. Men'shchikov, et al. 2015. 'A census of dense cores in the Aquila cloud complex: SPIRE/PACS observations from the Herschel Gould Belt survey.' A&A 584 (December): A91.

Kristensen, L. E., and M. M. Dunham. 2018. 'Protostellar half-life: new methodology and estimates.' A&A 618 (October): A158.

Kuhn, M. A., E. D. Feigelson, K. V. Getman, et al. 2014. 'The Spatial Structure of Young Stellar Clusters. I. Subclusters.' ApJ 787 (June): 107.

Lada, C. J., M. Lombardi, and J. F. Alves. 2010. 'On the Star Formation Rates in Molecular Clouds.' ApJ 724 (November): 687–693.

Lada, Charles J., Elizabeth A. Lada, Dan P. Clemens, and John Bally. 1994. 'Dust Extinction and Molecular Gas in the Dark Cloud IC 5146.' ApJ 429 (July): 694.

Lada, Charles J., John A. Lewis, Marco Lombardi, and João Alves. 2017. 'HP2 survey. III. The California Molecular Cloud: A sleeping giant revisited.' A&A 606 (October): A100.

Lada, Charles J., Marco Lombardi, Carlos Roman-Zuniga, Jan Forbrich, and João F. Alves. 2013. 'Schmidt's Conjecture and Star Formation in Molecular Clouds.' ApJ 778, no. 2 (December): 133.

Larson, R. B. 1995. 'Star formation in groups.' MNRAS 272 (January): 213–220.

Larson, Richard B. 1969. 'Numerical calculations of the dynamics of collapsing proto-star.' MNRAS 145 (January): 271.

Law, Richard, Janine Illian, David F. R. P. Burslem, et al. 2009. 'Ecological information from spatial patterns of plants: insights from point process theory.' *Journal of Ecology* 97 (4): 616–628.

Lewis, P. A. W, and G. S. Shedler. 1979. 'Simulation of nonhomogeneous poisson processes by thinning.' *Naval Research Logistics Quarterly* 26 (3): 403–413.

Lombardi, M. 2009. 'NICEST, a near-infrared color excess method tailored to small-scale structures.' A&A 493, no. 2 (January): 735–745.

Lombardi, Marco, Hervé Bouy, João Alves, and Charles J. Lada. 2014. 'Herschel-Planck dust optical-depth and column-density maps. I. Method description and results for Orion.' A&A 566 (June): A45.

Lombardi, Marco, Charles J. Lada, and João Alves. 2013. 'Fitting density models to observational data. The local Schmidt law in molecular clouds.' A&A 559 (November): A90.

Lombardi, M. and Alves, J. 2001. 'Mapping the interstellar dust with near-infrared observations: An optimized multi-band technique.' *A&A* 377 (3): 1023–1034.

Lynds, Beverly T. 1967. 'Space distribution of small dark nebulae.' In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Physical Sciences*, 51–60. Berkeley, Calif.: University of California Press.

Mairs, S., et al. 2016. 'The JCMT Gould Belt Survey: a first look at Southern Orion A with SCUBA-2.' MNRAS 461 (October): 4022–4048.

Masiunas, L. C., R. A. Gutermuth, J. L. Pipher, et al. 2012. 'A Structural Analysis of Star-forming Region AFGL 490.' ApJ 752, no. 2 (June): 127.

Men'shchikov, A., P. André, P. Didelon, et al. 2010. 'Filamentary structures and compact objects in the Aquila and Polaris clouds observed by Herschel.' A&A 518 (July): L103.

Myers, P. C., and E. F. Ladd. 1993. 'Bolometric Temperatures of Young Stellar Objects.' ApJL 413 (August): L47.

Myllymäki, Mari, Tomáš Mrkvička, Pavel Grabarnik, Henri Seijo, and Ute Hahn. 2017. 'Global envelope tests for spatial processes.' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79 (2): 381–404.

Onishi, T., A. Mizuno, A. Kawamura, H. Ogawa, and Y. Fukui. 1998. 'A $C^{18}O$ Survey of Dense Cloud Cores in Taurus: Star Formation.' ApJ 502 (July): 296–314.

Ortiz-León, Gisela N., Laurent Loinard, Sergio A. Dzib, et al. 2018. 'The Gould's Belt Distances Survey (GOBELINS). V. Distances and Kinematics of the Perseus Molecular Cloud.' ApJ 865, no. 1 (September): 73.

Ostriker, J. 1964. 'The Equilibrium of Polytropic and Isothermal Cylinders.' ApJ 140 (October): 1056.

Palmeirim, P., P. André, J. Kirk, et al. 2013. 'Herschel view of the Taurus B211/3 filament and striations: evidence of filamentary growth?' A&A 550 (February): A38.

Perry, George L. W., Ben P. Miller, and Neal J. Enright. 2006. 'A comparison of methods for the statistical analysis of spatial point patterns in plant ecology.' *Plant Ecology* 187 (1): 59–82.

Pokhrel, Riwaj, Robert A. Gutermuth, Sarah K. Betti, et al. 2020. 'Star-Gas Surface Density Correlations in 12 Nearby Molecular Clouds. I. Data Collection and Star-sampled Analysis.' ApJ 896, no. 1 (June): 60.

Polychroni, D., E. Schisano, D. Elia, et al. 2013. 'Two Mass Distributions in the L 1641 Molecular Clouds: The Herschel Connection of Dense Cores and Filaments in Orion A.' ApJL 777 (November): L33.

Preibisch, T., and E. D. Feigelson. 2005. 'The Evolution of X-Ray Emission in Young Stars.' ApJS 160 (October): 390–400.

Price-Whelan, A. M., B. M. Sipőcz, H. M. Günther, et al. 2018. 'The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package.' AJ 156 (September): 123.

Rapson, V. A., J. L. Pipher, R. A. Gutermuth, et al. 2014. 'A Spitzer View of the Giant Molecular Cloud MON OB1 EAST/NGC 2264.' ApJ 794, no. 2 (October): 124.

Robitaille, T. P., B. A. Whitney, R. Indebetouw, K. Wood, and P. Denzmore. 2006. 'Interpreting Spectral Energy Distributions from Young Stellar Objects. I. A Grid of 200,000 YSO Model SEDs.' ApJS 167 (December): 256–285.

Romine, G., E. D. Feigelson, K. V. Getman, M. A. Kuhn, and M. S. Povich. 2016. 'Young Stellar Populations in MYStIX Star Forming Regions: Candidate Protostars.' *ArXiv e-prints* (September).

Salji, C. J., J. S. Richer, J. V. Buckle, et al. 2015. 'The JCMT Gould Belt Survey: constraints on prestellar core properties in Orion A North.' MNRAS 449 (May): 1769–1781.

Schmidt, Maarten. 1959. 'The Rate of Star Formation.' ApJ 129 (March): 243.

Shu, F. H. 1977. 'Self-similar collapse of isothermal spheres and star formation.' ApJ 214 (June): 488–497.

Shu, F. H., F. C. Adams, and S. Lizano. 1987. 'Star formation in molecular clouds - Observation and theory.' ARA&A 25:23–81.

Simon, M. 1997. 'Clustering of Young Stars in Taurus, Ophiuchus, and the Orion Trapezium.' ApJL 482 (June): L81–L84.

Skrutskie, M. F., R. M. Cutri, R. Stiening, et al. 2006. 'The Two Micron All Sky Survey (2MASS).' AJ 131 (February): 1163–1183.

Sokol, Alyssa D., R. A. Gutermuth, R. Pokhrel, et al. 2019. 'Early science with the Large Millimetre Telescope: An LMT/AzTEC 1.1 mm Survey of dense cores in the Monoceros R2 giant molecular cloud.' MNRAS 483, no. 1 (February): 407–424.

Stahler, S. W., and F. Palla. 2005. *The Formation of Stars*, 865.

Teixeira, Paula S., Charles J. Lada, Erick T. Young, et al. 2006. 'Identifying Primordial Substructure in NGC 2264.' ApJL 636, no. 1 (January): L45–L48.

Terebey, S., F. H. Shu, and P. Cassen. 1984. 'The collapse of the cores of slowly rotating isothermal clouds.' ApJ 286 (November): 529–551.

Thönnes, Elke and van Lieshout, Marie-Colette. 1999. 'A Comparative Study on the Power of van Lieshout and Baddeley's J-Function.' *Biometrical Journal* 41, no. 6 (September): 721–734.

Tobin, John J., Kaitlin M. Kratter, Magnus V. Persson, et al. 2016. 'A triple protostar system formed via fragmentation of a gravitationally unstable disk.' Nature 538, no. 7626 (October): 483–486.

Velázquez, Eduardo, Isabel Martı nez, Stephan Getzin, Kirk A. Moloney, and Thorsten Wiegand. 2016. 'An evaluation of the state of spatial point pattern analysis in ecology.' *Ecography* 39 (11): 1042–1055.

Visser, Anja E., John S. Richer, and Claire J. Chandler. 2002. 'Completion of a SCUBA Survey of Lynds Dark Clouds and Implications for Low-mass Star Formation.' AJ 124, no. 5 (November): 2756–2789.

Wells, D. C., E. W. Greisen, and R. H. Harten. 1981. 'FITS - a Flexible Image Transport System.' A&AS 44 (June): 363.

Werner, M. W., T. L. Roellig, F. J. Low, et al. 2004. 'The Spitzer Space Telescope Mission.' ApJS 154 (September): 1–9.

Wiegand, Thorsten, Pavel Grabarnik, and Dietrich Stoyan. 2016. 'Envelope tests for spatial point patterns with and without simulation.' *Ecosphere* 7 (6): e01365.

Wiegand, Thorsten, Isabel Martinez, and Andreas Huth. 2009. 'Recruitment in Tropical Tree Species: Revealing Complex Spatial Patterns' [in English]. *American Naturalist* (1427 E 60TH ST, CHICAGO, IL 60637-2954 USA) 174, no. 4 (October): E106–E109.

Wiegand, Thorsten, and Kirk A. Moloney. 2004. 'Rings, circles, and null-models for point pattern analysis in ecology.' *Oikos* 104 (2): 209–229.

Williams, Jonathan P., and Christopher F. McKee. 1997. 'The Galactic Distribution of OB Associations in Molecular Clouds.' ApJ 476, no. 1 (February): 166–183.

Willis, S., A. Guzman, M. Marengo, et al. 2015. 'The Schmidt Law in Six Galactic Massive Star-forming Regions.' ApJ 809, no. 1 (August): 87.

Wright, E. L., P. R. M. Eisenhardt, A. K. Mainzer, et al. 2010. 'The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance.' AJ 140 (December): 1868-1881.

Yamada, Ikuho, and Peter A. Rogerson. 2003. 'An Empirical Comparison of Edge Effect Correction Methods Applied to K-function Analysis.' *Geographical Analysis* 35 (2): 97–109.

Ybarra, J. E., E. A. Lada, C. G. Román-Zúñiga, et al. 2013. 'The Progression of Star Formation in the Rosette Molecular Cloud.' ApJ 769 (June): 140.

Yongtao, Guan. 2006. 'A Composite Likelihood Cross-validation Approach in Selecting Bandwidth for the Estimation of the Pair Correlation Function.' *Scandinavian Journal of Statistics* 34 (2): 336–346.

Yu, Meichen, Arjan Hillebrand, Prejaas Tewarie, et al. 2015. 'Hierarchical clustering in minimum spanning trees.' *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25, no. 2 (February): 023107.

Zari, Eleonora, Marco Lombardi, João Alves, Charles J. Lada, and Hervé Bouy. 2016. 'Herschel-Planck dust optical depth and column density maps. II. Perseus.' A&A 587 (March): A106.

Zucker, Catherine, Joshua S. Speagle, Edward F. Schlafly, et al. 2019. 'A Large Catalog of Accurate Distances to Local Molecular Clouds: The Gaia DR2 Edition.' *The Astrophysical Journal* 879, no. 2 (July): 125.