

# Measuring anticlustering in regions of triggered star formation

A. Cartwright and A. P. Whitworth

School of Physics and Astronomy, Cardiff University, 5 The Parade, Cardiff CF24 3AA, UK  
 e-mail: Annabel.Cartwright@astro.cf.ac.uk

Received 15 April 2009 / Accepted 11 June 2009

## ABSTRACT

*Context.* Statistical methods for quantifying clustering have not yet been tested against anti-clustered data.

*Aims.* We investigate the performance of the normalised correlation length,  $\bar{s}$ , and  $Q$  when applied to anti-clustered data.

*Methods.* We tested the operation of  $\bar{s}$  on simulated data and recently published observations.

*Results.*  $\bar{s}$  successfully distinguishes anticlustered data in simulations and indicates anticlustering in an area where it is speculated that triggered star formation has occurred. The existence of radial anticlustering can cause the  $Q$  clustering analysis method to malfunction. Users of  $Q$  should therefore use  $\bar{s}$  to rule out anticlustering before applying  $Q$ .

**Key words.** stars: formation – HII regions – methods: statistical

## 1. Introduction

Cartwright & Whitworth (2004) devised a measurement,  $Q$ , which enables clustering to be both quantified and classified as either radial or multi-level subclustering. This measure is now widely used (e.g. Bastien et al. 2009; Sanchez & Alfaro 2009; Schmeja et al. 2009) and  $Q$  is calculated as  $\bar{m}/\bar{s}$ , where  $\bar{m}$  is the mean edge length of the minimal spanning tree, and  $\bar{s}$  the correlation length. However, it is possible that star clusters exist in which the arrangement of stars is “anti-clustered”; that is, a spherical distribution of stars has lower density or a void in the centre, and the density of stars increases with radius.

In this paper we investigate the behaviour of  $Q$  when applied to the analysis of anticlustered data. We develop the normalised correlation length,  $\bar{s}$ , and demonstrate its effectiveness in identifying and quantifying radial anticlustering. In Sect. 1 we describe the methods used to produce artificial test data and Sect. 2 explains the calculation of  $\bar{s}$ . Section 3 shows the results obtained using  $\bar{s}$  on artificial data, and these are discussed and compared with the results obtained from real data in Sect. 4. In Sect. 5 we draw our conclusions.

## 2. Methodology

Three different types of artificial star cluster were created, using random numbers  $\mathcal{R}$  to generate the individual star positions. The first type (2D0) is a circular cluster (i.e. two-dimensional disc) with statistically uniform surface density. This is the distribution expected from circular samples of randomly distributed data points. The second type (3D $\alpha$ ) are spherical clusters (i.e. three-dimensional clusters) with volume density  $n \propto r^\alpha$  and  $\alpha = -2, -1, 0, 0.5, 1.0,$  or  $1.5$ . Thirdly a fractal or multiscale clustering is used to create clusters of type F( $dim$ ),  $dim = 3, 2.5, 2.0$  or  $1.5$ .  $dim$  indicates the Fractal Dimension, or space filling dimension, of the cluster, a value of 3.0 indicating homogeneous, unclustered space filling, and 1.5 the most sub-clustered distribution. All of the random artificial clusters are created with 100 to 300 stars. For a full description of the methods used for creating these clusters, see Cartwright & Whitworth (2004). For each type of cluster 1000 realisations are analysed, so that means and standard deviations can be obtained for the

parameters extracted. The three-dimensional clusters are projected onto an arbitrary plane for analysis. The data for the real star clusters used are illustrated in Fig. 2 and the sources listed in Table 2.

Each cluster, real or artificial, is centred on the mean position of its stars, and distances are then normalised by setting the radius of the cluster (i.e. the distance from the mean of the stars' positions to the most distant star in the cluster) to unity. In the rest of this paper, all distances are normalised in this way.

## 3. The normalised correlation length

The normalised correlation length, also known as the “first moment”, is a single measurement characteristic of the distribution function  $p(s)$ , where  $p(s)ds$  gives the probability that the projected separation between two cluster stars chosen at random is in the interval  $(s, s + ds)$ . To obtain  $p(s)$  empirically, we define  $i_{\max}$  equal bins in the range  $0 < s < 2R_{\text{cluster}} = 2$ , so all the bins have width  $\Delta s = 2/i_{\max}$ . Thus the  $i$ th bin corresponds to the interval  $(i-1)\Delta s < s < i\Delta s$ , and to the mean value  $s_i = (i-1/2)\Delta s$ . Then we count the number of separations  $\mathcal{N}_i$  falling in each bin, and put

$$p(s_i) = \frac{2\mathcal{N}_i}{\mathcal{N}_{\text{total}}(\mathcal{N}_{\text{total}} - 1)\Delta s}, \quad (1)$$

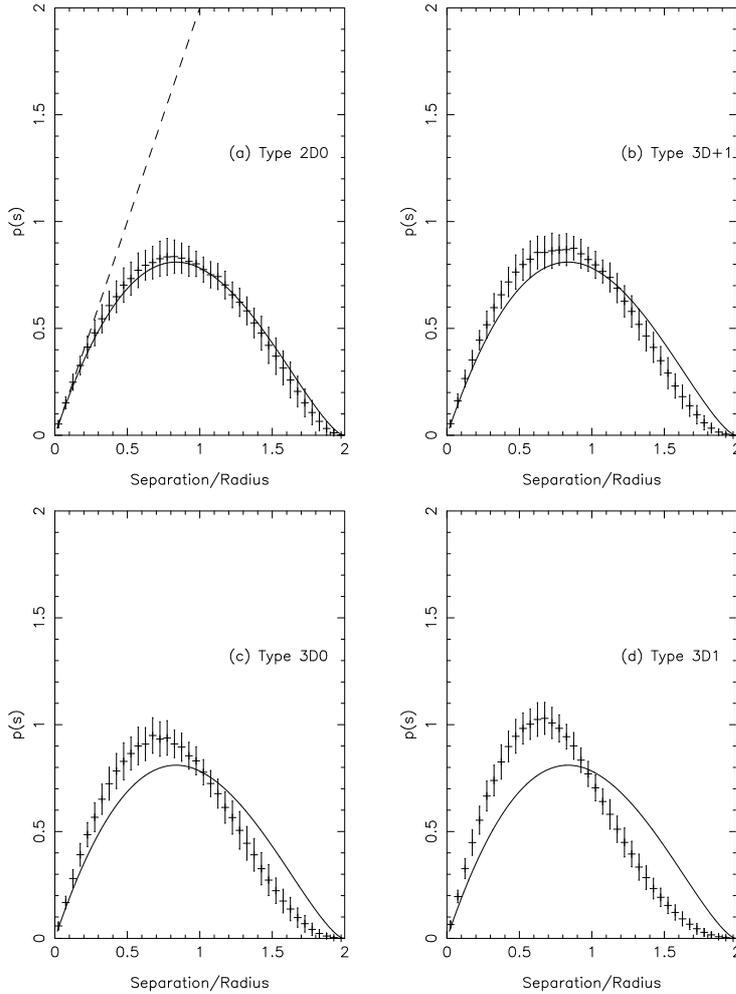
where  $\mathcal{N}_{\text{total}}$  is the total number of stars in the cluster, and hence  $\mathcal{N}_{\text{total}}(\mathcal{N}_{\text{total}} - 1)/2$  is the total number of separations.

Figure 1a presents the results obtained from 1000 clusters of type 2D0, a disc with uniform surface-density. The plotted points give the mean  $\bar{p}(s_i)$  from the 1000 realizations, and the error bars give the width of the bin and the  $1\sigma$  standard deviation. If there were no edge effects (i.e. if the uniform surface-density extended to infinity in two dimensions), we would have  $p(s) = 2s$ , and this is indeed a good fit to  $\bar{p}(s_i)$  at small  $s$  values, as indicated by the dashed line in Fig. 1a.

For a circle of unit radius populated with complete spatial randomness,  $p(s)$  is given by (Diggle 2003):

$$p(s) = 1 + \frac{1}{\pi} \left[ 2(s^2 - 1) \cos^{-1}(s/2) - s(1 + s^2/2) \sqrt{1 - s^2/4} \right]$$

for all  $0 \leq s \leq 2$ .



**Fig. 1.** Distribution function  $p(s)$  for separations between randomly chosen stars in artificial (non-fractal) cluster of type **a)** 2D0,  $N \propto r^0$ ; **b)** 3D+1,  $N \propto r^1$ ; **c)** 3D0,  $n \propto r^0$ ; **d)** 3D1,  $n \propto r^{-1}$ . The solid line is the value of  $p(s)$  for a star cluster of type 2D0 having an infinite number of stars, and is included for reference.

The solid line in Fig. 1a shows that this function fits the plotted points well, and it is included in all the other plots for reference, i.e. to emphasize the differences between the distributions.

The overall structure of the cluster is well represented by  $p(s)$ , as can be seen from Figs. 1b through 1d, which show the results obtained for three other types of artificial star cluster. Figure 1c shows how  $p(s)$  is slewed towards smaller  $s$  values for a projected sphere with constant density. This is due to the extra depth, and therefore larger number of stars, in the central region of the projected cluster, which results in more small separations between stars. This bias towards shorter separations is even more pronounced for Fig. 1d, a projected sphere with a centrally concentrated density,  $N \propto r^{-1}$ .

Figure 1b gives  $p(s)$  for a sphere with density which increases with radius,  $N \propto r^{-1}$ . It can be seen that  $p(s)$  for this configuration lies somewhere between those shown in Figs. 1a and 1c. By comparison,  $p(s)$  for a sphere with multiscale sub-clustering shows many peaks, corresponding to the distances between subclusters and the mean distances within them (Fig. 2 in Cartwright & Whitworth 2004).

The Normalized Correlation Length is the mean separation  $\bar{s}$  between stars in a cluster, and is therefore the mean value of the distributions in Fig. 1. Table 1 gives mean values of  $\bar{s}$  and their standard deviations, for the artificial cluster types tested. It can be seen that increasing values of  $\bar{s} > 0.8$  are measured for anticlustered data, while  $\bar{s} < 0.8$  is monotonically decreasing for clustered data, whether radially or fractally clustered.

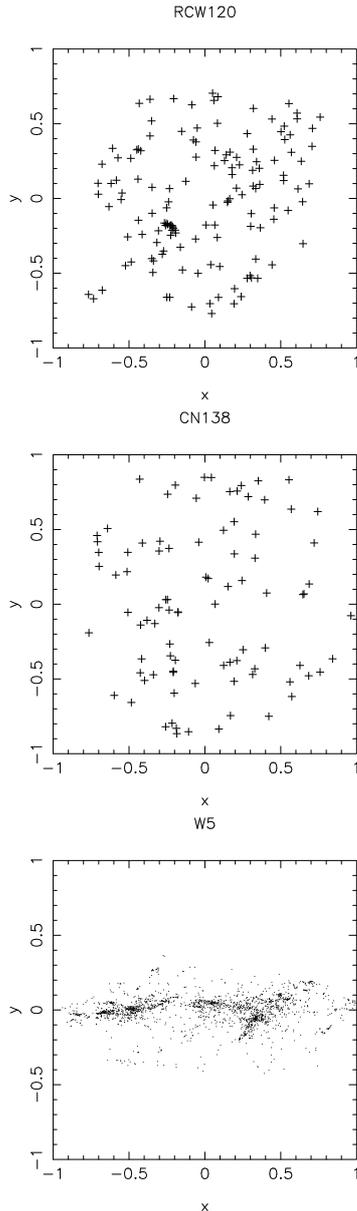
In order to test the robustness of  $\bar{s}$  with varying numbers of stars in the clusters, means and standard deviations were

**Table 1.** Clustering measures obtained for artificial star clusters.

Cluster type or name	$\bar{s}$	$\bar{m}$	$Q$
2D Uniform			
2D0 ( $N \propto r^0$ )	$.88 \pm .03$	$.65 \pm 0.02$	$.74 \pm .02$
Radially clustered			
3D-2 ( $n \propto r^{-2}$ )	$.60 \pm .03$	$.55 \pm .02$	$.93 \pm .03$
3D-1 ( $n \propto r^{-1}$ )	$.73 \pm .03$	$.61 \pm .02$	$.84 \pm .02$
3D0 ( $n \propto r^0$ )	$.80 \pm .02$	$.63 \pm .02$	$.79 \pm .02$
Radially anticlustered			
3D05 ( $n \propto r^{0.5}$ )	$.82 \pm .03$	$.63 \pm .02$	$.77 \pm .02$
3D1 ( $n \propto r^{1.0}$ )	$.83 \pm .03$	$.63 \pm .02$	$.76 \pm .02$
3D15 ( $n \propto r^{1.5}$ )	$.85 \pm .03$	$.64 \pm .02$	$.75 \pm .02$
Fractally clustered			
F2.5 ( $D = 2.5$ )	$.74 \pm .09$	$.54 \pm .05$	$.73 \pm .06$
F2.0 ( $D = 2.0$ )	$.67 \pm .13$	$.41 \pm .04$	$.61 \pm .08$
F1.5 ( $D = 1.5$ )	$.62 \pm .18$	$.27 \pm .07$	$.45 \pm .09$

calculated for 1000 examples of each cluster type with 85, 100, 300 and 500 members. The results are in Table A.2 in the Appendix.

Additionally, clusters of 1000 members were created and then 100 different subsamples of 200 members were randomly selected and analysed to produce  $\bar{s}$ . The results are in Table A.1 in the appendix. This table also shows the variation in the calculated centre (mean position) of the cluster, depending on which stars are selected from a cluster. Column 1 lists the cluster type and Col. 2 the mean and standard deviation of  $\bar{s}$ . Columns 3 and



**Fig. 2.** Position data for the protostars referred to in the text. Top, RCW120, middle CN138. Bottom the types I and II protostars from the W5 region. All the positions have been normalised by centring on the mean position of the cluster members, and then setting the distance from the centre to the furthest star to be one distance unit.

4 give the mean and standard deviation of the mean position of the cluster, in units of the radius of the original spherical distribution of points.

## 4. Results

Table 1 shows that  $\bar{s}$  decreases monotonically as distributions become more clustered, whether radially or multiscale, and increases monotonically as they become radially anticlustered. A value of  $s = 0.8$  corresponds to a statistically uniform density. (In the limit, numerical investigations revealed that  $\bar{s}$  reaches a maximum value of 1.17 for a ring of stars viewed face on.)  $\bar{s}$  can therefore be used to distinguish between clustered and anticlustered distributions. Tests also confirm the robustness of  $\bar{s}$  and its invariance with number of stars in a cluster and the selection or omission of points within the sample (Tables A.2 and A.1).

**Table 2.** Real star clusters.

Name	Members	Sources	$\bar{s}$
CN 138	85	Watson (2009)	0.84
RCW 120	65	Deharveng (2009)	0.67
W5	2053	Koenig (2008)	n/a

Table 1 also lists the mean edge length  $m$  of the minimal spanning tree and the derived  $Q = \bar{m}/\bar{s}$  for each cluster. Cartwright & Whitworth (2004) showed that  $Q$  can be used to distinguish the two types of clustering, radial and multiscale, and indeed this can be seen in the table,  $Q > 0.8$  indicating fractal-type, multiscale clustering, and  $Q < 0.8$  indicating radial clustering. However, the anticlustered data also yield a  $Q < 0.8$ , and so a simple measurement of  $Q$  does not distinguish fractal sub-clustering and anticlustering.

The confusion is avoided by first using  $s$  to distinguish between clustered and anti-clustered data, and then applying  $Q$  for data which is apparently clustered, to establish whether the clustering is radial or multiscale.

Table 2 reports the value of  $\bar{s}$  obtained for young stars in the RCW 120 and CN 138 regions, as reported by Deharveng et al. (2009) and Watson et al. (2009). A value of 0.67 for  $\bar{s}$  for RCW120 does not indicate anticlustering. However, a value of 0.84 for the CN 138 region does indicate anticlustering.

The values of  $\bar{s}$  for clustered, anti clustered and statistically random distributions are very close. A statistical test is therefore performed in order to establish the probability of obtaining a value of  $\bar{s} = 0.84$  from a random or fractally clustered distribution, and compare this with the probability of obtaining it from an anticlustered distribution. The test is performed as follows. First we wish to reject the hypothesis that the observed data are randomly arranged, either in 2 or 3 dimensions. 1000 simulations are performed using the 3D0 cluster type, with the same number of points as CN 138 (85) yielding 1000 values of  $\bar{s}_i$ . These are placed in size order, and compared with the value obtained for the CN 138 data. As each value of  $\bar{s}_i$  has equal probability, the position of the observed value of  $\bar{s}$  in the ordered set of random observations indicates the probability that a value greater than or equal to the observed value can be obtained from a 3D0 type cluster (see e.g. Diggle 2003). Out of 1000 simulations of the 3D0 clusters, only 60 yielded  $\bar{s} \geq 0.84$ , giving a confidence level of 94% that the 3D0 hypothesis can be rejected. Similarly, the 2D0 hypothesis, that the stars are randomly arranged in 2 dimensions, and the Fractal hypothesis, that CN 138 is fractally clustered with a dimension of  $D = 2.5$ , may be rejected with 99% confidence. The values of  $\bar{s}$  for radial density gradients proportional to  $r$  and  $r^{1.5}$  are closest to the measured value for CN 138. Therefore, on the basis of the measured value of  $\bar{s}$ , the protostars in the CN 138 region are found to be anticlustered with a density proportional to  $r^\alpha$  where  $\alpha = 1.25 \pm 0.25$ .

## 5. Discussion

Tests on artificial star clusters show that the correlation length  $\bar{s}$  is reliably found to be  $\bar{s} = 0.8 \pm 0.02$  for statistically uniform-density spherical distributions of points viewed in a 2-dimensional projection.  $\bar{s}$  increases monotonically above 0.8 as distributions become more radially anticlustered, and  $\bar{s}$  decreases below 0.8 as the distributions become more clustered, either radially or with multiscale, fractal characteristics.  $\bar{s}$  can therefore be used to distinguish between clustered and anticlustered distributions of points. Tests show that  $\bar{s}$  is robust and remains constant with changing numbers of stars in a cluster and is

also largely unaffected if different members are selected or omitted from a cluster before analysis (Tables A.2 and A.1; see also Bastien et al. 2009).

If the mean edge length of the minimal spanning tree  $m$  is also calculated, and used to obtain  $Q = \bar{m}/\bar{s}$ , then  $Q$  permits the two types of clustering to be distinguished. However, Table 2 shows that confusion may arise if  $Q$  is calculated for a distribution which is anticlustered, as the  $Q$  value is likely to be less than 0.8, falsely indicating a fractally subclustered distribution. It is important therefore, that users of  $Q$  check the value of  $\bar{s}$  before proceeding to the calculation and interpretation of  $Q$ .

The fact that the correlation length  $s$  of a cluster can indicate a region of anticlustering is in itself a useful tool. The formation of stars produces stellar winds and ionizing radiation which are predicted to disrupt the surrounding molecular clouds, triggering further star formation. Elmegreen & Lada (1977) proposed the collect and collapse mechanism, whereby expanding HII regions would sweep up spherical shells of matter from the interstellar medium, which would then collapse to form stars. Recent work has identified such bubbles, together with accurate census data of their associated star forming regions (Koenig et al. 2008; Deharveng et al. 2009; Watson et al. 2009). If the collect and collapse mechanism occurs in nature, one would expect the spatial distribution of the observed stars to bear the imprint of this process. The star formation rate should increase with the distance away from the centre of the expanding bubble, and should therefore be radially anticlustered.

Tests on observations of protostars which are believed to be forming in an expanding HII region, and therefore may be anticlustered, produce mixed results. Data from the CN 138 region (Watson et al. 2009) do show an increased  $\bar{s}$ , consistent with anticlustering, and a density gradient proportional to  $r^\alpha$  where  $\alpha = 1.25 \pm 0.25$ . The possibility of a random arrangement of stars producing this value of  $\bar{s}$  can be ruled out with 94% confidence. Data from the RCW120 region (Deharveng et al. 2009), however, give a lower value of  $\bar{s}$ , consistent with clustering. In this case, it can be seen from Fig. 2 that there are several small clusters in the data, and the existence of these small clusters would tend to reduce  $\bar{s}$  cancelling out the signature of the radial anticlustering.

Koenig et al. (2008) have also published observations of protostars forming in association with expanding HII bubbles, and these form arcs and ‘elephants’ trunks’ which are very suggestive of triggered star formation.  $\bar{s}$  would not be able to diagnose anticlustering in this situation, as the arcs do not form complete circles.  $\bar{s}$  is only able to detect anticlustering when the stars are forming in a fairly homogeneous arrangement, and when there are no tight knots of stars.

It should also be noted that a random arrangement of stars in two dimensions, or cluster type 2D0, will also give increased values of  $\bar{s}$ . It is therefore necessary to ensure that observational criteria are used to include only the stars in the three dimensional volume of interest, before applying the correlation length analysis.

## 6. Conclusion

Users of the  $Q = \bar{m}/\bar{s}$  parameter should ensure that they calculate the correlation length  $\bar{s}$  first, and exclude the possibility of anticlustered data, before proceeding to the calculation of  $Q$  and its interpretation. The correlation length  $\bar{s}$  increases monotonically as clusters are more radially anticlustered.  $\bar{s}$  can therefore be used to estimate the radial density gradient of star clusters whether the gradient is positive or negative. It can also be used

to rule out spatial randomness. The technique fails if the region being studied has subclusters of stars, or if only a part of a spherical bubble of stars exists.

*Acknowledgements.* The author is a Royal Society Dorothy Hodgkin Research Fellow. Thanks to C. Watson and L. Deharveng for providing early access to their data, and to the anonymous referee for helpful comments and questions.

## Appendix A: Stability tests on $\bar{s}$

**Table A.1.** Variance of  $\bar{s}$  and calculated centre (mean position) of cluster, when calculated from subsets of 200 points randomly selected from 1000 cluster members.

Cluster type	$\bar{s}$	$x$	$y$
2D Uniform			
2D0 ( $N \propto r^0$ )	.87 ± .03	.01 ± 0.03	.03 ± .03
Radially clustered			
3D-2 ( $n \propto r^{-2}$ )	.59 ± .02	.006 ± .02	.010 ± .02
3D-1 ( $n \propto r^{-1}$ )	.73 ± .02	.009 ± .03	.007 ± .02
3D0 ( $n \propto r^0$ )	.79 ± .02	.004 ± .03	.012 ± .03
Radially anticlustered			
3D05 ( $n \propto r^{0.5}$ )	.82 ± .03	.03 ± .03	.006 ± .03
3D1 ( $n \propto r^{1.0}$ )	.84 ± .03	.02 ± .03	.008 ± .03
3D15 ( $n \propto r^{1.5}$ )	.85 ± .03	.01 ± .03	.02 ± .03
Fractally clustered			
F3.0 ( $D = 3.0$ )	.79 ± .02	.01 ± .02	.003 ± .03
F2.5 ( $D = 2.5$ )	.75 ± .03	.18 ± .02	.15 ± .03
F2.0 ( $D = 2.0$ )	.63 ± .03	.08 ± .03	.08 ± .03
F1.5 ( $D = 1.5$ )	.67 ± .03	.09 ± .08	.14 ± .02

**Table A.2.** Variance of  $\bar{s}$  with numbers of objects in cluster.

Cluster type or name	$\bar{s}$	$\bar{s}$	$\bar{s}$	$\bar{s}$
	$N = 85$	$N = 100$	$N = 300$	$N = 500$
2D Uniform				
2D0 ( $N \propto r^0$ )	.85 ± .04	.86 ± 0.03	.88 ± .02	.88 ± .02
Radially clustered				
3D-2 ( $n \propto r^{-2}$ )	.60 ± .03	.59 ± .03	.59 ± .02	.59 ± .01
3D-1 ( $n \propto r^{-1}$ )	.73 ± .03	.73 ± .03	.73 ± .02	.73 ± .02
3D0 ( $n \propto r^0$ )	.79 ± .03	.79 ± .03	.80 ± .02	.80 ± .02
Radially anticlustered				
3D05 ( $n \propto r^{0.5}$ )	.81 ± .03	.81 ± .03	.82 ± .02	.82 ± .02
3D1 ( $n \propto r^{1.0}$ )	.83 ± .03	.83 ± .03	.84 ± .02	.84 ± .02
3D15 ( $n \propto r^{1.5}$ )	.84 ± .04	.84 ± .03	.85 ± .02	.86 ± .02
Fractally clustered				
F3.0 ( $D = 3.0$ )	.79 ± .03	.79 ± .04	.80 ± .02	.80 ± .01
F2.5 ( $D = 2.5$ )	.74 ± .06	.75 ± .05	.73 ± .05	.74 ± .05
F2.0 ( $D = 2.0$ )	.70 ± .09	.70 ± .08	.70 ± .08	.69 ± .07
F1.5 ( $D = 1.5$ )	.68 ± .10	.69 ± .09	.68 ± .09	.70 ± .10

## References

- Bastien, N., Gieles, M., Ercolano, B., & Gutermuth, R. 2009, MNRAS, 392, 868  
 Cartwright, A., & Whitworth, A. P. 2004, MNRAS, 348, 589  
 Cartwright, A., Whitworth, A. P., & Nutter, D. 2006, MNRAS, 369, 1411  
 Deharveng, L., Zavagno, A., Schuller, F., et al. 2009, A&A, 496, 177  
 Diggle, P. J. 2003, Statistical Analysis of Spatial Point Patterns, Hodder Arnold  
 Elmegreen, B. G., & Lada, C. J. 1977, ApJ, 214, 725  
 Koenig, X. P., Allen, L. E., Gutermuth, R. A., et al. 2008, ApJ, 688, 1142  
 Sanchez, N., & Alfaro, E. J. 2009, ApJ, 696, 2086  
 Schmeja, S., Gouliermos, D. A., & Klessen, R. S. 2009, ApJ, 694, 367  
 Watson, C., Corn, T., Churchill, E. B., et al. 2009, ApJ, 694, 546  
 Whitworth, A. P., Bhattal, A. S., Chapman, S. J., Disney, M. J., & Turner, J. A. 1994, MNRAS, 268, 291