Mon. Not. R. Astron. Soc. **348,** 589–598 (2004) doi:10.1111/j.1365-2966.2004.07360.x

The statistical analysis of star clusters

Annabel Cartwright* and Anthony P. Whitworth

Department of Physics and Astronomy, Cardiff University, Cardiff CF2 3YB

Accepted 2003 October 28. Received 2003 October 27; in original form 2003 September 29

ABSTRACT

We review a range of statistical methods for analysing the structures of star clusters, and derive a new measure $\mathcal{Q}$, which both quantifies and distinguishes between a (relatively smooth) large-scale radial density gradient and multiscale (fractal) subclustering.

The distribution of separations $p(s)$ is considered, and the normalized correlation length $\bar{s}$ (i.e. the mean separation between stars, divided by the overall radius of the cluster) is shown to be a robust indicator of the extent to which a smooth cluster is centrally concentrated. For spherical clusters having volume-density $n \propto r^{-\alpha}$ (with α between 0 and 2) $\bar{s}$ decreases monotonically with α, from $\sim$0.8 to $\sim$0.6. Since $\bar{s}$ reflects all star positions, it implicitly incorporates edge effects. However, for fractal star clusters (with fractal dimension D between 1.5 and 3) $\bar{s}$ decreases monotonically with D (from $\sim$0.8 to $\sim$0.6). Hence $\bar{s}$, on its own, can quantify, but cannot distinguish between, a smooth large-scale radial density gradient and multiscale (fractal) subclustering.

The minimal spanning tree (MST) is then considered, and it is shown that the normalized mean edge length $\bar{m}$ [i.e. the mean length of the branches of the tree, divided by $(\mathcal{N}_{\rm total}A)^{1/2}/(\mathcal{N}_{\rm total} - 1)$, where A is the area of the cluster and $\mathcal{N}_{\rm total}$ is the number of stars] can also quantify, but again cannot on its own distinguish between, a smooth large-scale radial density gradient and multiscale (fractal) subclustering.

However, the combination $\mathcal{Q} = \bar{m}/\bar{s}$ does both quantify and distinguish between a smooth large-scale radial density gradient and multiscale (fractal) subclustering. IC348 has $\mathcal{Q} = 0.98$ and ρ Ophiuchus has $\mathcal{Q} = 0.85$, implying that both are centrally concentrated clusters with, respectively, $\alpha \simeq 2.2 \pm 0.2$ and $\alpha \simeq 1.2 \pm 0.3$. Chamaeleon and IC2391 have $\mathcal{Q} = 0.67$ and 0.66, respectively, implying mild substructure with a notional fractal dimension $D \simeq 2.25 \pm 0.25$. Taurus has even more substructure, with $\mathcal{Q} = 0.45$ implying $D' \simeq 1.55 \pm 0.25$. If the binaries in Taurus are treated as single systems, $\mathcal{Q}$ increases to 0.58 and D' increases to 1.9 ± 0.2.

Key words: open clusters and associations: general.

1 INTRODUCTION

Since most stars are formed in clusters, it would be useful to have quantitative and objective statistical measures of their structure, with a view to comparing clusters formed in different environments, and tracking changes in structure as clusters evolve. This is particularly important for young, embedded clusters, where the structure may yield important clues to the formation process but is changing rapidly. It is also important for comparing observed clusters with numerical simulations.

At present, we do not have sufficiently robust statistical measures for this purpose. Features that are easily identified by the human eye, such as subclusters, or linear features, can be strangely elusive to objective statistical analysis. For example, it is difficult to distinguish, statistically, between a degree of fractal or random subclustering, and the existence of a density gradient (Bate, Clarke & McCaughrean 1997). This paper explores some possible measures, and evaluates their usefulness. In particular, we find a robust objective measure, which both quantifies, and distinguishes between, a smooth large-scale radial density gradient and multiscale (fractal) subclustering.

In Section 2 we describe our methodology. In Section 3 we look again at the mean surface density of companions (MSDC), a tool pioneered by Larson (1995) and subsequently used by several others (e.g. Simon et al. 1995; Bate et al. 1997; Brandner & Köhler 1998; Nakajima et al. 1998; Gladwin et al. 1999; Klessen & Kroupa 2001). We focus on measures that reflect the clustering regime (wide

*E-mail: Annabel.Cartwright@astro.cf.ac.uk separations), rather than the binary regime (close separations). In Section 4 we explore the use of the minimal spanning tree (MST) (Barrow, Bhavsar & Sonoda 1985) and its derivatives. In Section 5 we combine the MSDC and the MST to derive a single measure $\mathcal{Q}$, which is able both to quantify, and to distinguish between, a smooth radial density gradient and multiscale (fractal) subclustering.

All the measures are tested and calibrated on multiple realizations of artificial star clusters, and applied to ρ Ophiuchus, Chamaeleon, Taurus, IC348 and IC2391. Our results are discussed in Section 6 and the main conclusions are summarized in Section 7.

2 METHODOLOGY

Three different types of artificial star cluster have been created, using random numbers $\mathcal{R}$ to generate the individual star positions. The first type (2Dα) are circular clusters (i.e. two-dimensional disc-like clusters) with surface density $N \propto r^{-\alpha}$ and $\alpha = 0$ or 1. The second type (3Dα) are spherical clusters (i.e. three-dimensional clusters) having volume density $n \propto r^{-\alpha}$ with $\alpha = 0$, 1, 2 and 2.9. The third type (FD) are fractal star clusters (again three dimensional) with fractal dimension $D = 3.0$, 2.5, 2.0 or 1.5.

The different types are listed in column 1 of Table 1. All of the artificial clusters are created with 100–300 stars, as the numbers of stars within the five real clusters lie within that range. The data for the five real clusters used are illustrated in Fig. 3 and the sources are listed in Table A1.

A cluster of type 2Dα is created by positioning the stars according to

$$\begin{aligned} r &= \{(2-\alpha)\mathcal{R}_r/2\}^{1/(2-\alpha)}, \\ \phi &= 2\pi\mathcal{R}_\phi, \\ x &= r\cos(\phi), \\ y &= r\sin(\phi), \end{aligned} \qquad (1)$$

where $\mathcal{R}_r$ and $\mathcal{R}_\phi$ are random numbers in the range 0–1.

Table 1. Clustering measures obtained for artificial and real star clusters. Column 1 lists the cluster type (for artificial clusters) or name (for real clusters). Column 2 gives the normalized correlation length $\bar{s}$ (i.e. the ratio of the mean separation to the cluster radius, see Section 3). Column 3 gives the normalized mean edge length $\bar{m}$ (see Section 4). Column 4 gives the mean value of the standard deviation of the edge length, $\bar{\sigma}_m$. Column 5 gives $\mathcal{Q} = \bar{m}/\bar{s}$. For the artificial star clusters, means and standard deviations are computed from 100 realizations of each type, with $100 \leqslant \mathcal{N}_{\rm total} \leqslant 300$.

Cluster type or name	$\bar{s}$	$\bar{m}$	$\bar{\sigma}_m$	$\mathcal{Q}$
2D0($N \propto r^0$)	0.88 ± 0.03	0.65 ± 0.02	0.31 ± 0.02	0.74 ± 0.02
2D1($N \propto r^{-1}$)	0.70 ± 0.03	0.61 ± 0.02	0.38 ± 0.02	0.85 ± 0.03
3D2.9($n \propto r^{-2.9}$)	0.16 ± 0.02	0.24 ± 0.05	0.59 ± 0.07	1.50 ± 0.13
3D2($n \propto r^{-2}$)	0.60 ± 0.03	0.55 ± 0.02	0.41 ± 0.03	0.93 ± 0.03
3D1($n \propto r^{-1}$)	0.73 ± 0.03	0.61 ± 0.02	0.33 ± 0.03	0.84 ± 0.02
3D0($n \propto r^0$)	0.80 ± 0.02	0.63 ± 0.02	0.31 ± 0.02	0.79 ± 0.02
F3.0($D = 3.0$)	0.81 ± 0.03	0.64 ± 0.02	0.30 ± 0.02	0.80 ± 0.02
F2.5($D = 2.5$)	0.74 ± 0.09	0.54 ± 0.05	0.28 ± 0.03	0.73 ± 0.06
F2.0($D = 2.0$)	0.67 ± 0.13	0.41 ± 0.04	0.28 ± 0.02	0.61 ± 0.08
F1.5($D = 1.5$)	0.62 ± 0.18	0.27 ± 0.07	0.35 ± 0.07	0.45 ± 0.09
IC2391	0.74	0.49	0.30	0.66
Chamaeleon	0.63	0.42	0.45	0.67
Taurus	0.55	0.26	0.56	0.47
ρ Ophiuchus	0.53	0.45	0.39	0.85
IC348	0.49	0.48	0.46	0.98

A cluster of type 3Dα is created by positioning the stars according to

$$\begin{aligned} r &= \{(3-\alpha)\mathcal{R}_r/3\}^{1/(3-\alpha)}, \\ \theta &= \cos^{-1}(2\mathcal{R}_\theta - 1), \\ \phi &= 2\pi\mathcal{R}_\phi, \\ x &= r\sin(\theta)\cos(\phi), \\ y &= r\sin(\theta)\sin(\phi), \\ z &= r\cos(\theta), \end{aligned} \qquad (2)$$

where $\mathcal{R}_r$, $\mathcal{R}_\theta$ and $\mathcal{R}_\phi$ are random numbers in the range 0–1. Clearly, this method cannot be used for $\alpha = 3$, so to have a cluster type approximating to $\alpha = 3$ we use $\alpha = 2.9$.

A cluster of type FD is created by defining a ur-cube with side 2, and placing a ur-parent at the centre of the ur-cube. Next, the ur-cube is divided into $\mathcal{N}_{\rm div}^3$ equal subcubes, and a child is placed at the centre of each subcube (the first generation). Normally we use $\mathcal{N}_{\rm div} = 2$, in which case there are eight subcubes and eight first-generation children. The probability that a child matures to become a parent in its own right is $\mathcal{N}_{\rm div}^{(D-3)}$, where D is the fractal dimension. For lower D, the probability that a child matures to become a parent is lower, and the cluster is more ‘porous’. Children who do not mature are deleted, along with the ur-parent. A little noise is then added to the positions of the remaining children, to avoid an obviously regular structure, and they then become the parents of the next generation, each one spawning $\mathcal{N}_{\rm div}^3$ children (the second generation) at the centres of $\mathcal{N}_{\rm div}^3$ equal-volume subsubcubes, and with each second-generation child having a probability $\mathcal{N}_{\rm div}^{(D-3)}$ of maturing to become a parent. This process is repeated recursively until there is a sufficiently large generation that, even after pruning to impose a spherically symmetric envelope of radius 1 within the ur-cube, there are still more children than the required number of stars. Children are then culled randomly until the required number is left, and the surviving children are identified with the stars of the cluster. At each generation, the survival of a child is determined by generating a random number $\mathcal{R}$ in (0, 1); survival then requires that $\mathcal{R} < \mathcal{N}_{\rm div}^{(D-3)}$.

Clusters of type 2Dα are investigated for two purposes. First, we wish to clarify the effect of a sharply defined circular edge on an otherwise statistically uniform, two-dimensional distribution of stars. Clusters of type 2D0 enable us to isolate this effect. Secondly, we wish to explore how readily two-dimensional and three-dimensional distributions can be distinguished. This could be important if stars are being formed in layers, for example at a shock front.

For each type of artificial cluster, 100 realizations are analysed, so that means and standard deviations can be obtained for the parameters extracted. Three-dimensional clusters (types 3Dα and FD) are projected on to an arbitrary plane prior to analysis. Two-dimensional clusters are viewed face-on.

3 THE MEAN SURFACE DENSITY OF COMPANIONS

3.1 Log–log plots and edge effects

A widely used tool for analysing the structure of star clusters is the log–log plot of the mean surface-density of companions, $\bar{N}$ against separation, s. This tool has been pioneered by Larson (1995), building on earlier work by Gomez et al. (1993), who used the two point correlation function. Several papers have confirmed Larson’s finding that a plot of log$[\bar{N}]$ against log $[s]$, hereafter a Larson plot, can be fitted with two power-law sections, corresponding to two distinct regimes. At the smaller separations, $s < s_{\mathrm{break}}$, the companions of a star are mainly in binary and higher multiple systems, and the slope of the Larson plot is $\eta_{\mathrm{binary}} \equiv \mathrm{d}\log[\bar{N}]/\mathrm{d}\log[s] \simeq -2$. At larger separations, $s > s_{\mathrm{break}}$, companions are simply other members of the overall cluster, and may only be close due to projection. The slope here is generally larger (i.e. still negative but smaller in magnitude), $\eta_{\mathrm{cluster}} \equiv \mathrm{d}\log[\bar{N}]/\mathrm{d}\log[s] \gtrsim -1$. Larson has suggested that η_{cluster} might be related to the fractal dimension of the subclustering, $D = \eta_{\mathrm{cluster}} + 2$. In addition, he has proposed that the break point between the two straight sections, at s_{break}, might correspond to the Jeans length. However, recent analysis has cast some doubt on these interpretations. First, the break point is strongly influenced by the overall surface-density of stars (and hence by the depth of the cluster along the line of sight), as pointed out by Simon (1997) and Bate et al. (1997). Secondly, fitting η_{cluster} objectively is difficult, because at the low-s end it is distorted by the switch to the binary regime, and, more importantly, at the high-s end it is distorted by edge effects. Consequently, one is left with at best a range of the order of $2s_{\mathrm{break}}$ to $0.1R_{\mathrm{cluster}}$ and the result is sensitive to how the range is actually chosen; if the range is shortened or extended arbitrarily, the slope of the fitted line may change dramatically. Thirdly, η_{cluster} is not necessarily related to the fractal dimension of the clustering. As shown by Bate et al. (1997) and Klessen & Kroupa (2001), it may simply reflect a large-scale density gradient in the cluster.

3.2 Linear plots and edge effects

An alternative way of evaluating the data from which Larson plots are derived is to calculate the distribution function $p(s)$, where $p(s)\,\mathrm{d}s$ gives the probability that the projected separation between two cluster stars chosen at random is in the interval $(s,\ s + \mathrm{d}s)$. To do this empirically, we define i_{max} equal s-bins in the range $0 < s < 2\,R_{\mathrm{cluster}}$, so that all the bins have width $\Delta s = 2R_{\mathrm{cluster}}/i_{\mathrm{max}}$, and the ith bin corresponds to the interval $(i - 1)\ \Delta s < s < i\Delta s$, with mean value $s_i = (i - 1/2)\ \Delta s$. R_{cluster} is the overall radius of the cluster, and is defined by finding the mean position of all the stars in the cluster and then setting R_{cluster} equal to the distance to the furthest star. Then we count the number of separations $\mathcal{N}_i$ falling in each bin, and put

$$p(s_i) = \frac{2\mathcal{N}_i}{\mathcal{N}_{\mathrm{total}}(\mathcal{N}_{\mathrm{total}} - 1)\Delta s}, \tag{3}$$

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

Fig. 1(a) presents the results obtained from 100 clusters of type 2D0, i.e. a disc having statistically uniform surface-density. The plotted points give the mean $\bar{p}(s_i)$ from the 100 realizations, and the error bars give the width of the bin and the 1σ standard deviation. If there were no edge effects (i.e. if the uniform surface-density extended to infinity in two dimensions), we would have $\bar{p}(s) = 2s$, and this is indeed a good fit to $\bar{p}(s_i)$ at small s_i values, as indicated by the dashed line on Fig. 1(a). *Departures from this straight line are entirely due to edge effects.*

In fact, $\bar{p}(s)$ can be calculated semi-analytically for a disc having uniform surface-density:

$$\bar{p}(s) = \begin{cases} 2s(1-s)^2 + \dfrac{4s}{\pi}\displaystyle\int_{1-s}^{1} \theta r\ \mathrm{d}r, & 0 \leqslant s < 1; \\ \dfrac{4s}{\pi}\displaystyle\int_{s-1}^{1} \theta r\ \mathrm{d}r, & 1 \leqslant s < 2; \\ 0, & s \geqslant 2; \end{cases} \tag{4}$$

where

$$\theta = \cos^{-1}\left[\frac{r^2 + s^2 - 1}{2rs}\right] \tag{5}$$

The solid line on Fig. 1(a) shows that this function fits the plotted points well, and it is included on all the other plots for reference, i.e. to emphasize the features that are not due to edge effects.

When derived in this way, the $\bar{p}(s)$ plot contains little information concerning the distribution of binary separations, since they are all in the first bin. However, it seems to be well established that the distribution of binary separations is approximately scale free over a wide range of separations ($\eta_{\mathrm{binary}} \simeq -2$). The more critical issue, that with which we are concerned here, is the distribution of separations in the clustering regime and what it tells us about the overall structure of the cluster. This information is well represented by $\bar{p}(s)$, as can be seen from Figs 1(b)–(f), which show the results obtained for the other five types of non-fractal artificial star cluster. Fig. 1(b) shows how $\bar{p}(s)$ is slewed towards smaller s values for a disc with a centrally concentrated surface-density, $N \propto r^{-1}$. Figs 1(c)–(f) show spherical clusters having volume-density gradients $n \propto r^{-\alpha}$ with $\alpha = 0$, 1, 2 and 2.9. Again the distribution slews to smaller s values as the sphere becomes more centrally concentrated (i.e. with increasing α).

3.3 The normalized correlation length

One feature that distinguishes the plots is the location of the maximum, i.e. the separation s_{max} at which $\bar{p}(s)$ is largest. As a cluster becomes more centrally condensed, s_{max} moves to smaller values, and the amplitude of the maximum increases. However, for an individual cluster s_{max} will not be well defined, and so it is not a robust measure.

A better measure of this trend is the normalized correlation length for each cluster. The correlation length is the mean separation $\bar{s}$ between stars in the cluster, and it is normalized by dividing by R_{cluster}. The second column of Table 1 gives mean values of $\bar{s}$ and their standard deviations, for the various artificial cluster types. The $\bar{s}$ values for the five real star clusters are also given.

The shapes of the $p(s)$ plots, and hence also the $\bar{s}$ values, are independent of the number of stars in the cluster. In trials with cluster sizes of 100 to 1000 stars, $\bar{s}$ stays within one standard deviation of the mean value for 200 stars. This is at first sight surprising. A 1000-star cluster is so much more dense than a 100-star cluster, that one might expect the mean separation of stars to be smaller. However, although each star has more close neighbours, it also has more distant neighbours, and the value of $\bar{s}$ remains constant. This is an attractive feature of the normalized correlation length as a statistical descriptor for clusters. From Table 1, we see that $\bar{s}$ decreases monotonically with increasing α, and can therefore be used to estimate α for star clusters, which are presumed a priori to have radial density gradients.

Importantly, cluster types 2D1 and 3D2 are easily distinguished by their $\bar{s}$ values and their $\bar{p}(s)$ plots, despite the widespread but fallacious assumption that a three dimensional cluster with volume-density $n \propto r^{-2}$ is, when projected on the sky, similar to a two dimensional cluster with surface-density $N \propto r^{-1}$. In fact, it is clusters of types 2D1 and 3D1 (i.e. with the same exponent, $\mathrm{d}\ln[N]/\mathrm{d}\ln[r] \sim -1$, and $\mathrm{d}\ln[n]/\mathrm{d}\ln[r] \sim -1$), which are hard to distinguish.

$p(s)$ plots for the real clusters are shown on Fig. 2. IC348 and ρ Ophiuchus resemble clusters of type 3D2, both on the basis of their $\bar{s}$ values (Table 1), and the shapes of their $p(s)$ plots (Figs 2a and c). For IC2391 the $\bar{s}$ value and the $p(s)$ plot (Fig. 2b) are most like those for clusters of type 3D1.

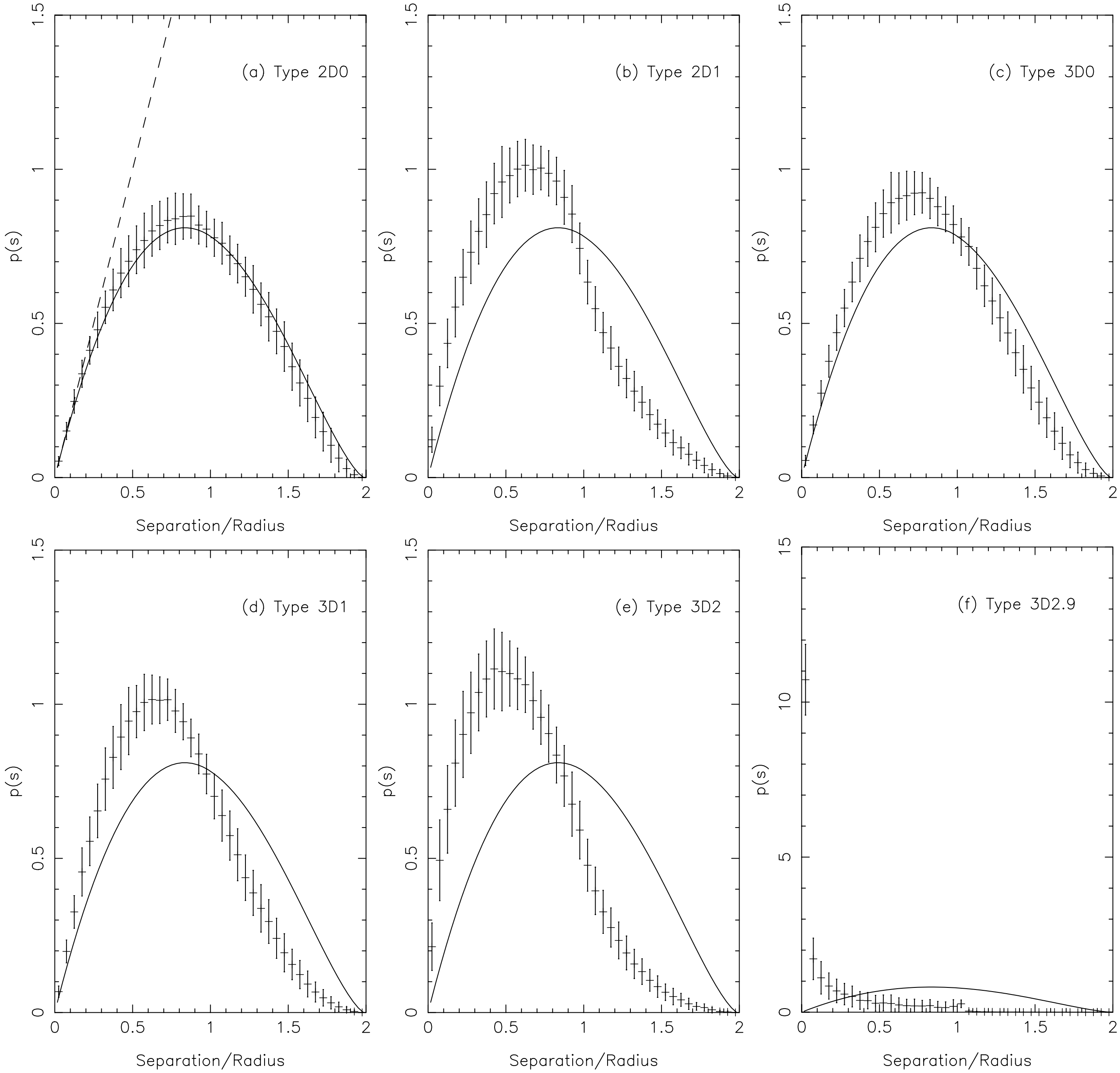

Figure 1. Distribution function $p(s)$ for separations between randomly chosen stars in artificial (non-fractal) cluster of type (a) 2D0, $N \propto r^0$; (b) 2D1, $N \propto r^{-1}$; (c) 3D0, $n \propto r^0$; (d) 3D1, $n \propto r^{-1}$; (e) 3D2, $n \propto r^{-2}$; and (f) 3D2.9, $n \propto r^{-2.9}$. The solid line is the value of $p(s)$ for a star cluster of type 2D0 having an infinite number of stars (equation 4), and is included for reference. The dashed line is $p(s) = 2s$ (see the text). s is normalized to the overall radius of the cluster, as described in the text.

3.4 The effect of subclusters on $p(s)$ and $\bar{s}$

Chamaeleon and Taurus have correlation lengths intermediate between types 3D1 and 3D2, but their $p(s)$ plots are clearly not generic. This is because they contain subclusters, as illustrated in Figs 3(d) and (e). Consequently, $p(s)$ has multiple maxima. In some cases these maxima can be identified with (i) separations between stars in the same subcluster (the maximum at the smallest separations) and (ii) separations between stars in two distinct subclusters (maxima at larger separations, corresponding to the separation between the two subclusters). If there are $\mathcal{N}_{\rm sub}$ subclusters, there can be up to $1 + \mathcal{N}_{\rm sub}(\mathcal{N}_{\rm sub} - 1)/2$ maxima, but fewer if there is degeneracy in the distances between subclusters. After smoothing, the $p(s)$ plot for Chamaeleon (Fig. 2e) shows two distinct maxima, suggesting at least two subclusters, and Fig. 3(e) does indeed show two subclusters. They are separated by $\sim$1, hence giving rise to the maximum in $p(s)$ at $s \sim 1$. However, after smoothing, the $p(s)$ plot for Taurus (Fig. 2d) shows only three well-defined maxima, suggesting at most three subclusters, whereas Fig. 3(d) shows at least eight well-defined subclusters. Evidently, the $p(s)$ plot is not a robust diagnostic of subclustering.

If we now consider artificial fractal star clusters with the same fractal dimension D, we find that there is so much variance in their individual $\bar{p}(s)$ plots that we cannot sensibly define a mean $\bar{p}(s)$ plot. However, we can still compute the mean normalized correlation length $\bar{s}$ and its variance. The results are given in Table 1. We see

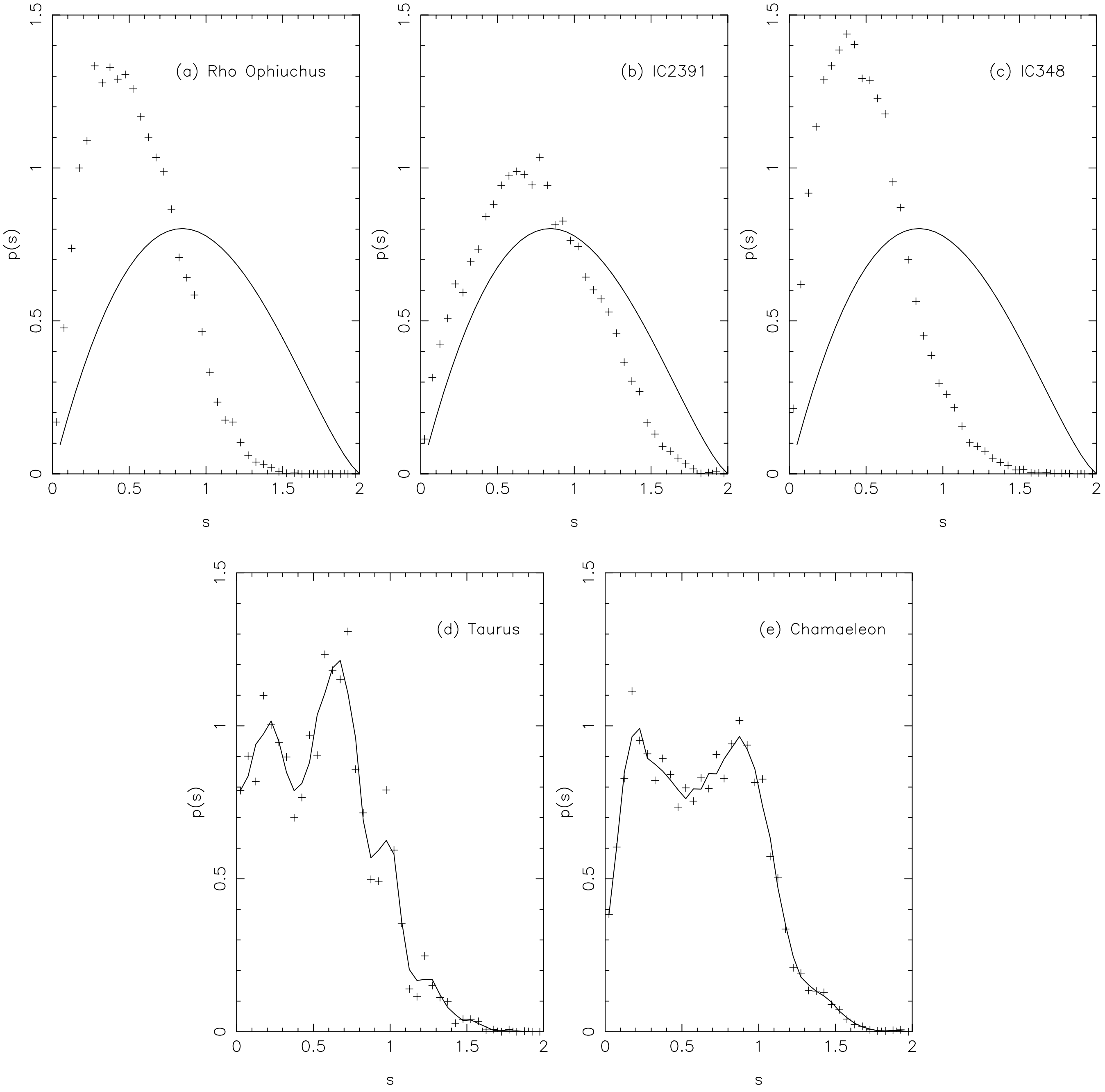

Figure 2. Distribution functions for separations between randomly chosen stars in five real star clusters. (a) ρ Ophiuchus, (b) IC2391, (c) IC348, (d) Taurus and (e) Chamaeleon. The solid line is the value of $p(s)$ for a star cluster of type 2D0 having an infinite number of stars (equation 4), and is included for reference. In (d) and (e), the solid line represents a smoothed version of the raw data, to show the existence of multiple maxima.

that $\bar{s}$ increases monotonically with increasing D and can therefore be used to estimate D for star clusters, which are presumed a priori to be fractal.

Moreover, the value of $\bar{s}$ for star clusters of type F3.0 is essentially the same as for clusters of type 3D0, as it should be. The small difference is attributable to the fact that in constructing clusters of type F3.0 the positioning of the individual stars is not completely random, whereas for type 3D0 it is.

However, the range of $\bar{s}$ for D in (1.5, 3.0) is almost identical to that for α in (0, 2). Therefore, $\bar{s}$ is degenerate and cannot on its own be used to distinguish multiscale (fractal) subclustering from a large-scale radial density gradient.

4 MINIMAL SPANNING TREES

The minimal spanning tree is the unique[1] network of straight lines joining a set of points, such that the total length of all the lines, hereafter ‘edges’, in the network is minimized and there are no

[1] Strictly speaking, if the array of points contains two or more pairs with exactly the same separation, the network may not be unique, as the points may be connected in a different order. However, even if this is the case, the total length of edges and the distribution of edge-lengths is preserved for all solutions.

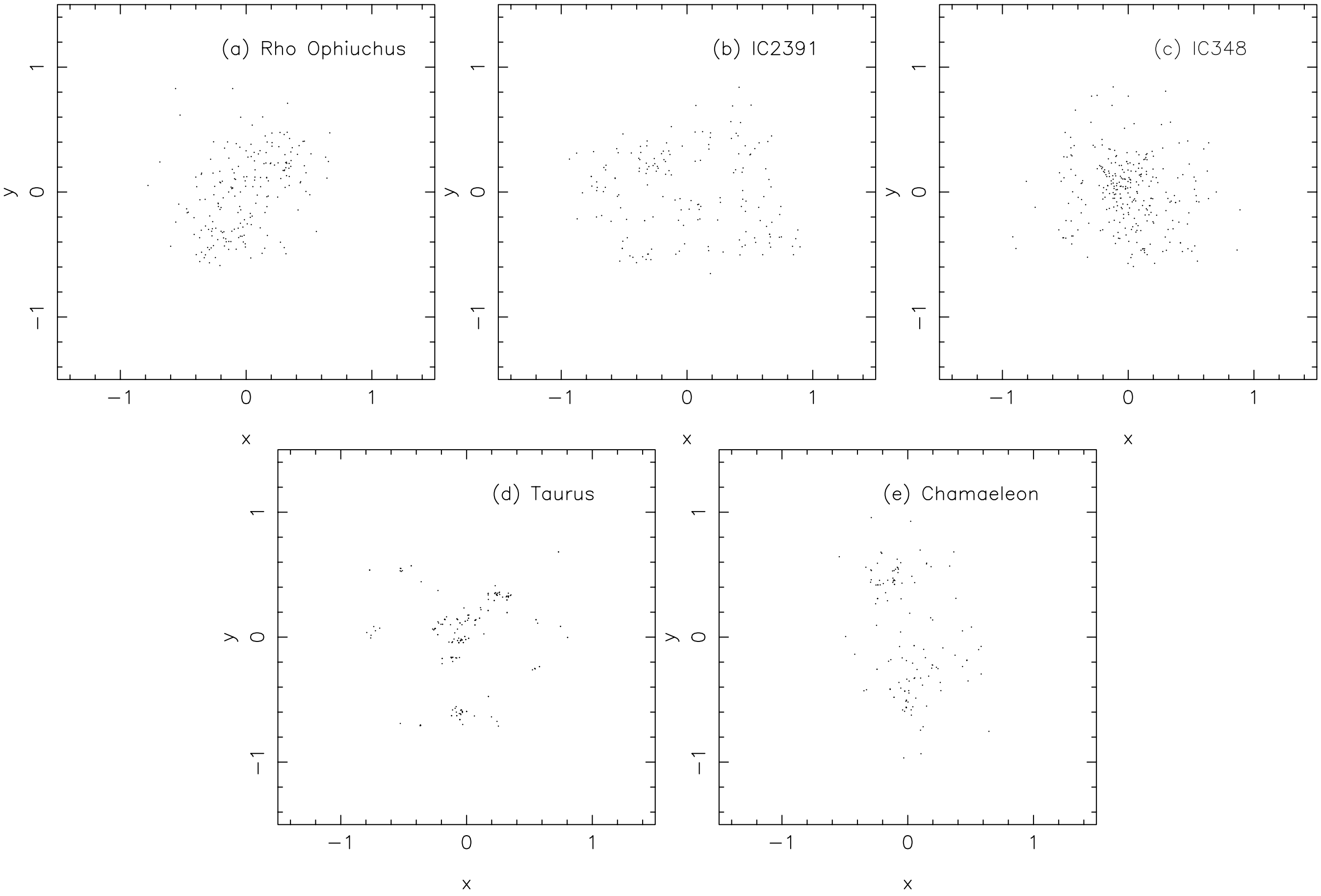

Figure 3. Raw data for all real star clusters analysed in the paper. The clusters have been centred on the mean position of all stars and scaled so that the distance from the centre to the most distant star is unity.

closed loops. The construction of such a tree is described by Gower and Ross (1969). Starting at any point, an edge is created joining that point to its nearest neighbour. The tree is then extended by always constructing the shortest link between one of its nodes and an unconnected point, until all the points have been connected. Fig. 4 shows the MSTs for the real star clusters ρ Ophiuchus, Taurus, Chamaeleon, IC348 and IC2391.

The use of MSTs as a probe of cosmological structure was explored by Barrow et al. (1985), and a further refinement, the self-avoiding random walk, was described by Baugh (1993). Although the approach seemed promising as a means of picking out clumps and filaments, the only statistical analysis of the MST for cosmological purposes of which we are aware is due to Graham, Clowes & Campusano (1995), who adopted methods developed by Hoffman & Jain (1983) and Dussert et al. (1987) and applied them to the distribution of quasars. We describe their analysis in Appendix B, and show that it does not work well for star clusters.

4.1 The normalized mean edge length

Once the MST of a cluster has been constructed it is straightforward to compute the mean edge length, $\bar{m}$. Unlike the normalized correlation length $\bar{s}$, $\bar{m}$ is not independent of the number of stars in the cluster, $\mathcal{N}_{\rm total}$. As $\mathcal{N}_{\rm total}$ increases, more short edges are created on the MST and $\bar{m}$ decreases. The expected total length of the MST of a random array of $\mathcal{N}_{\rm total}$ points, uniformly distributed over a two-dimensional area A, is asymptotically proportional to $(\mathcal{N}_{\rm total}A)^{1/2}$ (Hammersley, Halton & Beardwood 1959). As there are $\mathcal{N}_{\rm total} - 1$ edges, the mean edge length is asymptotically proportional to $(\mathcal{N}_{\rm total}A)^{1/2}/(\mathcal{N}_{\rm total} - 1)$, and so this factor should be used to normalize the mean edge length of clusters having different areas A and/or different numbers of stars $\mathcal{N}_{\rm total}$.

The resulting normalized mean edge length $\bar{m}$ has been evaluated for 100 realizations of each type of artificial star clusters, and for the real star clusters, and the results are recorded in Table 1 (column 3). Also recorded in Table 1 (column 4) is the mean of the standard deviations of the MST edge lengths, $\bar{\sigma}_m$, This quantity is used in Appendix B.

4.2 $\mathcal{Q}$

Table 1 shows that for artificial clusters of type 2Dα, 3Dα and *FD*, both $\bar{m}$ and $\bar{s}$ decrease monotonically as α increases (i.e. the degree of central concentration becomes more severe) or as D decreases (i.e. the degree of subclustering become more severe). However, $\bar{s}$ decreases more quickly than $\bar{m}$ as α is increased, while $\bar{m}$ decreases more quickly than $\bar{s}$ as D is decreased. Thus, the ratio

$$\mathcal{Q} = \frac{\bar{m}}{\bar{s}} \tag{6}$$

yields a measure, which not only quantifies, but also distinguishes between, a smooth overall radial density gradient and multiscale fractal subclustering.

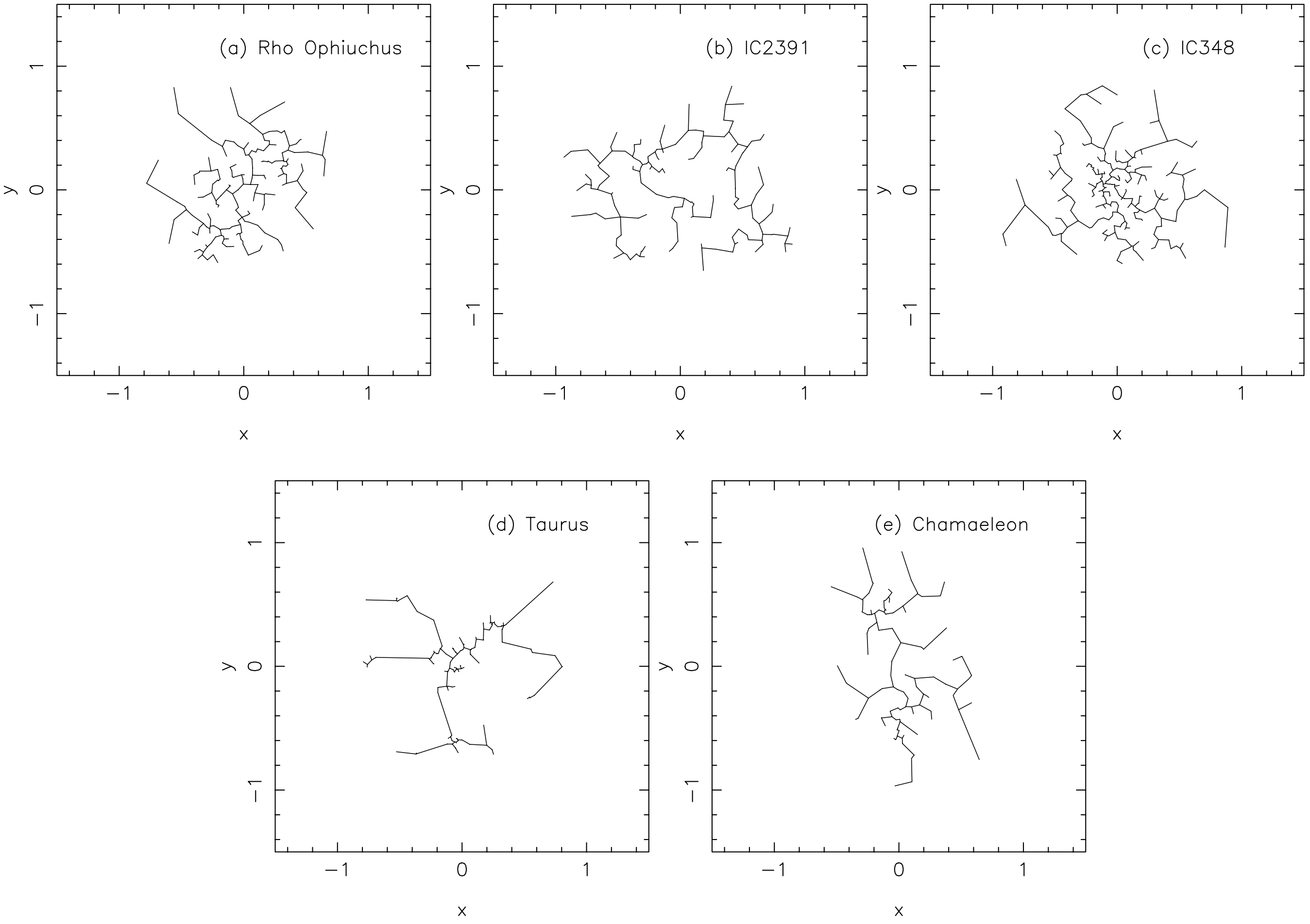

Figure 4. Minimal spanning trees for (a) ρ Ophiuchus, (b) IC2391, (c) IC348, (d) Taurus and (e) Chamaeleon.

Mean values of $\mathcal{Q}$ for the various types of artificial star cluster are recorded in Table 1 (column 5). For artificial clusters with a smooth large-scale radial density gradient (type 3Dα), $\bar{\mathcal{Q}}$ increases from $\bar{\mathcal{Q}} \simeq 0.80$ to 1.50 as the degree of central concentration increases from $\alpha = 0$ (statistically uniform number-density) to $\alpha = 2.9$ ($n \propto r^{-2.9}$). For artificial clusters with fractal substructure (type FD), $\bar{\mathcal{Q}}$ decreases from $\bar{\mathcal{Q}} \simeq 0.80$ to 0.45 as the degree of subclustering increases from $D = 3.0$ (uniform number-density, no subclustering) to $D = 1.5$ (strong subclustering).

We can therefore construct a plot (Fig. 5) of D against $\mathcal{Q}$ for $\mathcal{Q} \leqslant 0.80$, and α against $\mathcal{Q}$ for $\mathcal{Q} \geqslant 0.80$. For any real cluster we can compute its $\mathcal{Q}$ value, and then use Fig. 5 to read off its notional fractal dimension D' (if $\mathcal{Q} < 0.80$, implying subclustering), or its radial density exponent α (if $\mathcal{Q} > 0.80$, implying a large-scale radial density gradient).

The small kink at $\mathcal{Q} \simeq 0.8$ is due to the fact in constructing a cluster of type F3.0 the stars are positioned regularly (in the sense that at each generation, each subcube of space is occupied) and therefore the number-density is artificially uniform; in contrast, when we construct a cluster of type 3D0 the stars are positioned randomly, so that the density is only uniform in a statistical sense and there are Poisson fluctuations in the local density.

Fractal dimensions obtained from Fig. 5 in this way are only notional, because $\mathcal{Q}$ (or any other single measure) can reflect subclustering, but cannot capture whether the subclustering is hierarchically self-similar.

Using Fig. 5, we infer that Taurus, IC2391 and Chamaeleon have substructure with notional fractal dimensions $D = 1.5$, 2.2 and 2.25. In contrast, ρ Ophiuchus and IC348 appear to be centrally concentrated, with radial density exponents $\alpha = 1.2$ and 2.2. These

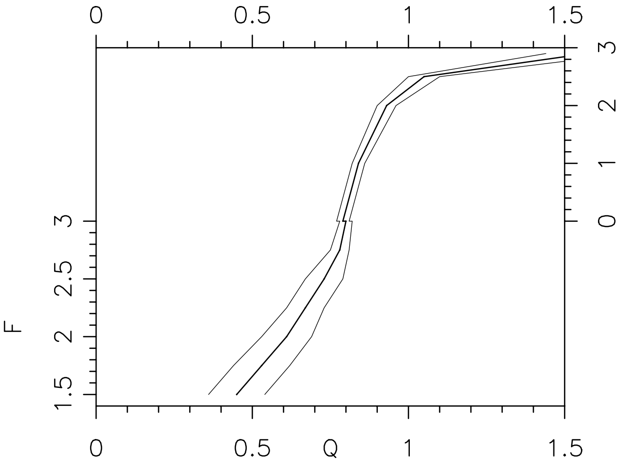

Figure 5. $\mathcal{Q}$ plot for artificial star clusters. For $\mathcal{Q} \leqslant 0.80$, the fractal dimension D should be read from the left-hand axis, and for $\mathcal{Q} \geqslant 0.80$, the radial density exponent α should be read from the right-hand axis. The small kink at $\mathcal{Q} \simeq 0.8$ is explained in the text.

inferences agree well with an intuitive reading of the raw data shown in Fig. 3.

4.3 The effect of binary companions on the MST edge length

The MST will normally link a star to its binary companion, as this will usually be the shortest way of adding one or other of the stars to the tree. Binaries create very short edges and therefore a large population of binary stars will cause a notable reduction in the mean edge length, $\bar{m}$. Of the five real clusters considered in this paper, Taurus has been subjected to particularly close scrutiny and has a larger identified population of binaries than any of the others. As the binaries are not part of the clustering regime, it is important to establish whether they are distorting the result.

Using the MST, all pairs of stars lying closer together than 10^{-4} of the cluster radius were pruned, leaving single stars. For Chamaeleon, ρ Ophiuchus, IC348 and IC2391, only 3, 0, 1 and 2 such pairs were found; Taurus, however, was pruned from 215 down to only 137 primary stars. For the pruned version of Taurus, $\bar{s}$ increased from 0.55 to 0.57, while $\bar{m}$ increased from 0.26 to 0.33 and $\mathcal{Q}$ increased from 0.47 to 0.58. Removal of the binaries resulted in the notional fractal dimension for Taurus being increased from 1.5 to 1.9. This demonstrates that in a cluster with a large binary population, it is important to prune the close companions before evaluating $\bar{s}$, $\bar{m}$ and $\mathcal{Q}$.

5 DISCUSSION

The ratio of the normalized mean edge length to the normalized correlation length, $\mathcal{Q}$, is effective in distinguishing between a smooth large-scale radial density gradient and multiscale fractal subclustering, because it is sensitive not only to the frequency of small separations between stars, but also to their spatial distribution.

The MST Edge length $\bar{m}$ is a simple average of the distances of stars to their (usually) closest neighbours. If one star is moved very close to another, the change in $\bar{m}$ will be diluted by the total number of stars, $\mathcal{N}_{\rm total}$. However, when calculating the mean distance of companions for all stars, only one star out of $\mathcal{N}_{\rm total}$ in the cluster has had one of its $\mathcal{N}_{\rm total} - 1$ companions moved very close. The change in $\bar{s}$, is therefore diluted by $\mathcal{N}^2_{\rm total}$. Thus, when small separations are scattered all over the cluster, increasing the number of small separations causes both $\bar{m}$ and $\bar{s}$ to decrease, but $\bar{s}$ decreases more slowly than $\bar{m}$. This is the case for decreasing fractal dimension.

For radially concentrated clusters, in contrast, increasing the clustering creates more small separations between stars, but these are all in the central region of the cluster. Moving another star to this area affects $\bar{m}$ in the normal way, the star having a newly short edge length between it and its nearest neighbour, and the change in the mean distance being diluted by $\mathcal{N}_{\rm total}$. However, the large number of other stars in the centre also gain another close neighbour. The decrease in $\bar{s}$ is therefore compounded and exceeds that in $\bar{m}$.

Consequently, the quotient $\mathcal{Q} = \bar{m}/\bar{s}$ successfully distinguishes between clusters that have a smooth large-scale radial density gradient and clusters that have multiscale fractal subclustering, in a way that agrees with an intuititive analysis but that cannot be accomplished using existing methods such as Larson plots or box dimension plots (BDPs). An additional advantage over these methods is that the calculation of $\mathcal{Q}$ is quantitative and objective, as no intervention is required in the normalization process, in the construction of the MST, or in choosing a range over which to calculate a slope.

We should emphasize that classical methods for evaluating the density profile of a cluster, or its fractal dimension, are not viable for clusters with ~200 members, primarily because of low-number statistics. For example, if one attempts to define the projected radial density profile for a real cluster of stars by counting stars in different annuli, the result is very noisy.

Alternatively, if one attempts to determine the mean projected radial density profile for a 200-member artificial cluster having a given radial density profile in three dimensions, using many different realizations and with a view to comparing this with a real cluster, one finds that the standard deviation is very large, and so the diagnostic power of this profile is poor.

In the same spirit one might attempt to construct the box dimension plot of a real cluster and compare it with the mean BDP of artificial star clusters having a given fractal dimension. To construct a BDP one divides the projected image of the star cluster into a grid of square cells of side l and counts the number of cells, $\mathcal{N}_{\rm occ}(l)$, which are occupied by stars. Then, by repeating this for different values of l, one obtains a plot of $\log[\mathcal{N}_{\rm occ}(l)]$ against $-\log(l)$. For a true fractal this plot is a straight line with slope equal to the fractal dimension. However, for a star cluster with only ~200 members, the plot is not linear. By treating many realizations of artificial clusters all having the same fractal dimension *and the same number of stars*, one can define a mean BDP. However, the mean BDP is not very strongly dependent on the fractal dimension and it has a large standard deviation. Therefore, the box dimension plot of a real cluster does not give a useful constraint on its fractal dimension.

It is for this reason that we have sought integral measures of cluster structure. The same philosophy informs the use of equivalent width when evaluating noisy spectral lines (for example).

We also note that a cluster cannot have a large-scale radial density gradient, and at the same time be fractally subclustered. A cluster could have a large-scale radial density gradient and *non-fractal* subclustering, but then it would require more parameters to characterize the structure, and its diagnosis would become correspondingly more difficult (if not impossible for clusters with ~200 stars).

In Table A1 we list estimates for the ages and the crossing times of the clusters we have analysed. On the basis of simple arguments, we might expect the $\mathcal{Q}$ value of a cluster to increase with time, as the substructure dissolves and the overall cluster relaxes to a radially concentrated density profile. However, this is not evident in the small sample treated here. Taurus has an age much less than its crossing time, which is consistent with its small $\mathcal{Q}$ value and low fractal dimension. On the other hand, IC2391 and Chamaeleon have ages much greater than their crossing times and yet they are still fractal with relatively low $\mathcal{Q}$ values. In contrast, ρ Ophiuchus and IC348, which have ages comparable with their crossing times, are both centrally condensed, with no discernible substructure. We should, however, caution against drawing firm conclusions from such a small sample. We also note that young clusters observed at short wavelengths (i.e. in the optical), may appear to have substructure due to patchy observation. therefore long-wavelength surveys are preferable for embedded young star clusters.

6 CONCLUSIONS

We have explored two statistical measures for analysing objectively the observed (i.e. projected) structures of star clusters. These measures are based on the MSDC and the MST. The measures are $\bar{s}$, the normalized mean separation between stars, and $\bar{m}$, the normalized mean edge-length of the MST, both of which are independent of the number of stars in the cluster. For artificial star clusters, created with a smooth large-scale radial density profile ($n \propto r^{-\alpha}$), and for artificial star clusters created with substructure having fractal dimension D, $\bar{s}$ and $\bar{m}$ both decrease with increasing α and/or decreasing D, but at different rates. Hence a cluster with a radial gradient can be distinguished from one with substructure by evaluating $\mathcal{Q} = \bar{m}/\bar{s}$. For a cluster of uniform volume-density (i.e. $\alpha = 0$ and $D = 3.0$), $\mathcal{Q} \simeq 0.80$. If the cluster is made more centrally condensed by increasing α, $\mathcal{Q}$ increases monotonically, reaching $\mathcal{Q} \simeq 1.50$ at $\alpha = 2.9$. Conversely, if the cluster is given substructure by reducing D, $\mathcal{Q}$ decreases monotonically, reaching $\mathcal{Q} \simeq 0.45$ at $D = 1.5$.

On the basis of their $\mathcal{Q}$ values, ρ Ophiuchus and IC348 have radial gradients with $\alpha \simeq 1.2 \pm 0.3$, and 2.2 ± 0.2, respectively. Chamaeleon and IC2391 have substructure with notional fractal dimension $D' \simeq 2.2 \pm 0.2$. Taurus has even more substructure, with $D' \simeq 1.55 \pm 0.25$, and if the binaries in Taurus are treated as single systems, D' increases to 1.9 ± 0.2. D' is only a notional fractal dimension, because the integral measures we have defined do not give any indication of whether the substructure is hierarchically self-similar. (Indeed, for clusters having only ~200 stars the range of separations is too small to possess hierarchical self-similarity.)

ACKNOWLEDGMENTS

We would like to acknowledge helpful discussions with Dr Norman Fry, of the Department of Earth, Ocean and Planetary Sciences, Cardiff University.

REFERENCES

Barrado y Navascues D., Stauffer J.R., Briceno C., Patten B., Hambly N.C., Adams J.D., 2001, ApJS, 134, 103

Barrow J.D., Bhavsar S.P., Sonoda D.H., 1985, MNRAS, 216, 17

Bate M.R., Clarke C.J., McCaughrean M.J., 1997, MNRAS, 297, 1163

Baugh C., 1993, MNRAS, 264, 87

Bontemps S. et al., 2001, A&A, 372, 173

Brandner W., Köhler R., 1998, ApJ, 499, L79

Briceno C., Calvet N., Gomez M., Hartmann L., Kenyon S.J., Whitney B.A., 1993, PASP, 105, 686

Dussert C., Rasigni M., Palmari J., Rasigni G., Llebaria A., Marty F., 1987, J. Theor. Biol., 125, 317

Ghez A.M., Neugebauer G., Matthews K., 1993, AJ, 106, 2005

Ghez A.M., McCarthy D.W., Patience J.L., Beck T.L., 1997, ApJ, 481, 378

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

Gomez M., Jones B.F., Hartmann L., Kenyon S.J., Stauffer J.R., Hewett R., Reid I. N., 1992, AJ, 104, 762

Gomez M., Hartmann L., Kenyon S.J., Hewett R., 1993, AJ, 106, 2005

Gower J.C., Ross G.J.S., 1969, Appl. Stat., 18, 54

Graham M.J., Clowes R.G., Campusano L.E., 1995, MNRAS, 275, 790

Hammersley J.M., Halton J.H., Beardwood J., 1959, Camb. Philos. Soc. Proc., 55, 299–327

Hartmann L., Stauffer J.R., Kenyon S.J., Jones B.F., 1991, AJ, 101, 1050

Herbig G.H., Bell K.R., 1988, Lick Obs. Bull. No. 1111

Hoffman R., Jain A.K., 1983, Patt. Rec. Lett., 1, 175

Klaoudatou I., 2002, MPhil thesis, Cardiff Univ.

Klessen R.S., Kroupa P., 2001, A&A, 372, 105

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

Lawson W.A., Fiegelson E.D., Huenemoerder D.P., 1996, MNRAS, 280, 1071

Leinert Ch., Zinnecker H., Weitzel N., Christou J., Ridgway S.T., Jameson R., Haas M., Lenzen R., 1993, A&A, 278, 129

Luhman K.L., Stauffer J.R., Muench A.A., Rieke G.H., Lada E.A., Bouvier J., Lada C.J., 2003, ApJ, 593, 1093

Nakajima Y., Tachihara K., Hanawa T., Nakano M., 1998, ApJ, 497, 721

Simon M., 1997, ApJ, 482, L81

Simon M. et al., 1995, ApJ, 443, 625

Steele J.M., 1988, Ann. Prob., 16, 1767

Waer F.M., Brown A., Mathieu R.D., Myers P.C., Vrba F.J., 1988, AJ, 96, 297

APPENDIX A: RAW DATA

Table A1 gives the sources of the positions of stars, or, in the case of ρ Ophiuchus, protostars, used in the analysis of Sections 3 and 4. These positions are plotted in Fig. 3.

APPENDIX B: DETECTING RANDOMNESS AND CLUSTERING USING THE MINIMAL SPANNING TREE

Graham et al. (1995) have applied the MST to quasar clustering on very large scales, using a method that was developed by Dussert et al. (1987) for characterizing biological structures. In Dussert's method, the mean $\bar{m}$ and standard deviation σ_m of the edge lengths of the MST are first computed and normalized by dividing by the factor $(\mathcal{N}_{\text{total}} A)^{1/2}/(\mathcal{N}_{\text{total}} - 1)$, then plotted on the (m, σ_m)-plane. Fig. B1(a), reproduced from Dussert et al. (1987), shows the theoretical locations on the (m, σ_m)-plane for different types of clustering in two-dimensions. The region of the (m, σ_m)-plane around the

Table A1. Sources of positions for cluster members and approximate ages and crossing times for clusters. (Crossing times were calculated using a typical velocity dispersion of 2 km s^{-1}.)

Name	Members	Age (Myr)	T_{cross} (Myr)	Sources
IC2391	166	53	2.5	Barrado et al. (2001)
Cham.	136	0.1–40	2.7	Lawson, Fiegelson & Huenemoerder (1996) Ghez et al. (1997)
Taurus	215	1.0	10.0	Briceno et al. (1993) Ghez, Neugebauer & Matthews (1993) Gomez et al. (1992) Hartmann et al. (1991) Herbig & Bell (1988) Leinert et al. (1993) Simon et al. (1995) Waer et al. (1988) Luhman et al. (2003)
ρ Oph.	199	0.3–2.0	1.35	Bontemps et al. (2001)
IC348	288	2.0	2.0	Luhman et al. (2003)

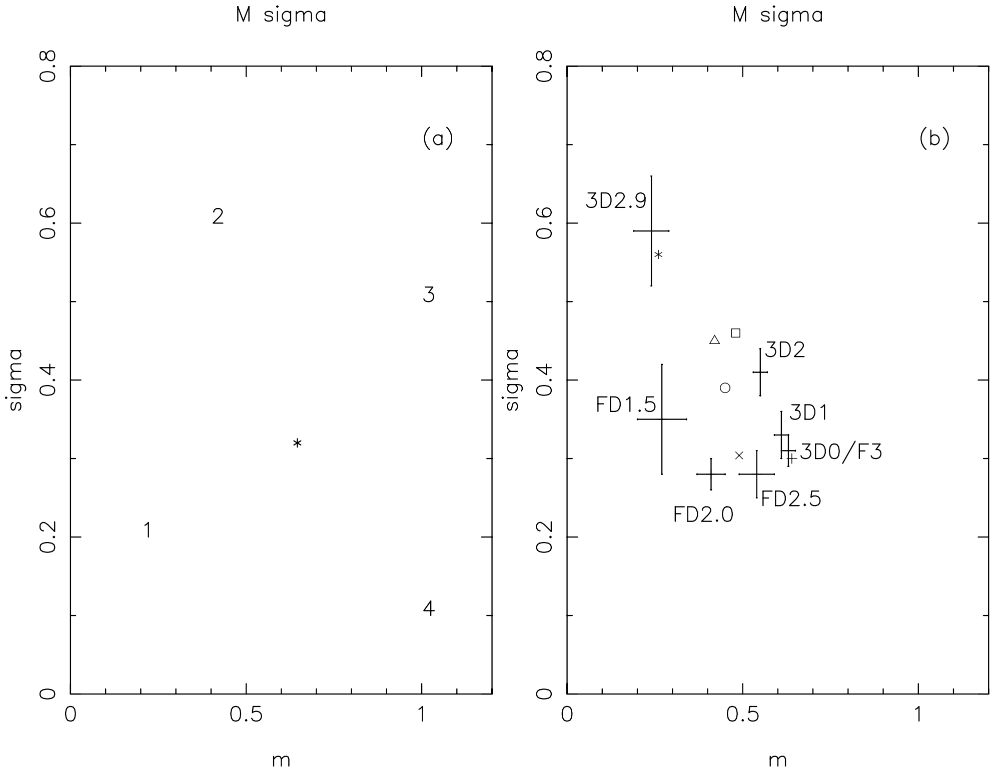

Figure B1. MST (m, σ_m)-plots. (a) (m, σ_m)-plane, showing the regions of the plane in which well characterized distributions of points converge (from Graham et al.). *, random distributions; 1, clustered structures; 2, concentration gradients; 3, quasi-periodic tilings; 4, highly organized distributions. (b) (m, σ_m)-plane, with artificial star clusters of types 3D0–3D2.9 and F1.5–F3.0 plotted. The five real clusters are indicated by symbols: $*$, Taurus; ○, ρ Ophiuchus; ×, IC2391; □, IC348; △, Chamaeleon.

central star represents the locus of a random distribution. The region of the (m, σ_m)-plane around '1' represents the locus of distributions dominated by subclustering. The region of the (m, σ_m)-plane around '2' represents the locus of distributions dominated by radial concentration gradients. The region of the (m, σ_m)-plane around '3' represents the locus of distributions dominated by quasi-periodic tilings. And the region of the (m, σ_m)-plane around '4' represents the locus of highly organized distributions (i.e. lattices).

Fig. B1(b) shows the loci on the (m, σ_m)-plane for the various artificial star cluster types and the five real clusters, and reveals some drawbacks to this plot. The locus for artificial clusters with a radial density gradient do indeed tend towards region 2 with increasing α (i.e. greater degree of central concentration), although only for $\alpha \gtrsim 2$ are they clearly distinguishable from a purely random distribution. Similarly, the locus for artificial clusters with fractal subclustering tend towards region 1 with decreasing D (i.e. greater degree of subclustering) for $D \gtrsim 2.0$. However, for $D \lesssim 2.0$, this trend is abandoned, and the locus moves towards region 2; in other words, a cluster with a low fractal dimension and hence a high degree of subclustering masquerades, on the (m, σ_m)-plane, as a cluster with a strong radial density gradient, albeit it not precisely of the form $n \propto r^{-\alpha}$. Moreover, Taurus, which to the human eye appears to have the most well-defined subclustering of all five real clusters, masquerades on the (m, σ_m)-plane as a cluster with a strong radial density gradient, $\alpha \simeq 2.7$.

We conclude that the (m, σ_m)-plane is not able to distinguish between a smooth large-scale radial density gradient and multiscale fractal subclustering.

This paper has been typeset from a TEX/LATEX file prepared by the author.