A systematic survey for infrared star clusters with $|b| < 20°$ using 2MASS

D. Froebrich,[1,2][*] A. Scholz[3] and C. L. Raftery[4]

[1]*Dublin Institute for Advanced Studies, 5 Merrion Square, Dublin 2, Ireland*
[2]*Centre for Astrophysics and Planetary Science, The University of Kent, Canterbury, CT2 7NH*
[3]*University of Toronto, 60 St George Street, Toronto, Canada*
[4]*Trinity College Dublin, College Green Dublin 2, Ireland*

ABSTRACT

We used star density maps obtained from the Two-Micron All-Sky Survey (2MASS) to obtain a sample of star clusters in the entire Galactic Plane with $|b| < 20°$. A total of 1788 star cluster candidates are identified in this survey. Among those are 681 previously known open clusters and 86 globular clusters. A statistical analysis indicates that our sample of 1021 new cluster candidates has a contamination of about 50 per cent. Star cluster parameters are obtained by fitting a King profile to the star density. These parameters are used to statistically identify probable new globular cluster candidates in our sample. A detailed investigation of the projected distribution of star clusters in the Galaxy demonstrates that they show a clear tendency to cluster on spatial scales in the order of 12–25 pc, a typical size for molecular clouds.

Key words: methods: statistical – globular clusters: general – open clusters and associations: general.

1 INTRODUCTION

Star clusters are the building blocks of galaxies. By investigating the basic properties of clusters, e.g. morphology, mass function and spatial distribution, we can trace the long-term evolution of stars in the Milky Way. The pre-requisites for such studies are large-scale surveys providing the means to detect large homogeneous samples of open (OpCl) and globular clusters (GlCls).

So far, about 1100 OpCls are known in the Galaxy, and their properties have been compiled in large OpCl data bases (see e.g. Lynga 1995; Dias et al. 2002; Mermilliod & Paunzen 2003). Most of these clusters, however, have been detected in the framework of optical surveys, mainly based on photographic plates. Since interstellar and cloud extinction are a strong factor at wavelengths shorter than 1 μm, optical cluster catalogues may be highly incomplete. Near-infrared (NIR) surveys are probably more appropriate to provide a more complete census of clusters, since they are able to probe more distant regions in the Galaxy due to much lesser sensitivity to interstellar extinction, and to detect the youngest clusters, which are still embedded in molecular clouds. Recent studies have shown that these embedded infrared OpCls might outnumber optical visible clusters by an order of magnitude [see Lada & Lada (2003) for a recent review]. Infrared OpCls are thus an important complement to the classical cluster catalogues.

Similarly, the known sample of GlCls is very likely incomplete because it is mostly based on visual inspection of photographic plates at optical wavelengths. Of the ∼150 GlCls known in our Galaxy

(e.g. Harris 1996), only very few have been discovered using infrared data (e.g. Hurt et al. 2000; Kobulnicky et al. 2005). Similar to OpCls, however, many GlCls remain undiscovered due to obscuration from dust close to the Galactic Plane. Based on the spatial distribution of the known GlCls, Ivanov, Kurtev & Borissova (2005) estimate a lower limit of 10 ± 3 for the number of unknown GlCls near the Galactic Plane ($|Z| \leqslant 0.5$ kpc) and within 3 kpc from the Galactic Centre. Moreover, the sample of off-plane GlCls is likely to be incomplete as well, as indicated by the recent serendipitous discoveries of the two off-plane clusters GC Whiting1 (Carraro 2005) and ESO 280-SC06 (Ortolani, Bica & Barbuy 2000). This is relevant, because a complete GlCl sample is of fundamental importance as a probe for the variety of processes that may have contributed to the formation of the Galaxy: rapid protogalactic collapse, accretion, cannibalism, galaxy collisions, star bursts (e.g. van den Bergh 1993; West et al. 2004). The diversity of these processes is obviously reflected by the diversity of GlCls, implying that a complete census of the GlCls is crucial for the understanding of the formation of the Galaxy.

The Two-Micron All-Sky Survey (2MASS), (Skrutskie et al. 2006) provides an excellent data base to identify and analyse large samples of infrared clusters. Recent work by Dutra et al. (2003) and Bica et al. (2003) proved that 2MASS indeed contains hundreds of previously unidentified infrared clusters. A systematic search for 2MASS clusters in the entire Galactic Plane, as well as a comprehensive analysis of their properties inferred from NIR photometry alone, has not yet been carried out. In a previous paper, we presented a relative extinction map of the Galactic Plane ($|b| < 20°$) derived from star counts in the 2MASS catalogue

[*]E-mail: df@cp.dias.ie

(Froebrich et al. 2005). The original aim of this project was to search for dark globules. It required obtaining a star density map of the entire region. Such a map is additionally well-suited to detect OpCl and GlCls. In this paper, we report about a systematic search and analysis of Galactic star clusters based on 2MASS star counts.

This paper is structured as follows. In Section 2.1, we introduce our cluster-selection method and discuss the reliability of our cluster candidates. A description and discussion of the spatial distribution of the clusters is put forward in Section 3. Finally, we describe our method of statistically classifying the newly detected cluster candidates to search for the best new Galactic GlCl candidates (Section 4).

2 CLUSTER SELECTION

2.1 Method and results

We selected all stars from the 2MASS point source catalogue with Galactic Latitude $|b| < 20°$. To ensure high photometric accuracy, only objects with a quality flag better than 'C' (i.e. only stars with a photometric accuracy better than five times the signal-to-noise ratio) are used. The area with $|b| < 20°$ was then divided in 288 regions of 10° length in Galactic Longitude and 5° height in Galactic Latitude. In each of these fields, the completeness limit was determined as the peak in the histogram of the brightness distribution of all stars in each of the three filter bands JHK. The star density at this brightness was then measured in each of the bands. Independent of the completeness limit, the star density was determined by counting the number of stars in 3.5×3.5 arcmin sized boxes every 20 arcsec.

The star density maps were searched automatically for local density enhancements using the SExtractor software (Bertin & Arnouts 1996). As cluster candidates, we selected all objects with a local star density above the 4σ level of the background and an extent of more than 11 arcmin^2. Only objects detected either in J and H or in H and K were selected. Objects detected only in one of the three bands, or in J and K only, were discarded. The redundant information in different filters allows us a reliable object identification and safely excludes spurious detections. This automated procedure resulted in the detection of 961 cluster candidates.

A random inspection of the star density maps, however, revealed that the automatic detection missed quite a number of cluster-like objects slightly fainter than the 4σ detection limit and detected several objects that are obviously not star clusters (spikes from bright stars, dark cloud edges). This is caused by density gradients and variable signal-to-noise ratio in the $10° \times 5°$ fields due to Galactic structure and dust clouds. Hence, an additional manual selection was performed by visually inspecting the JHK star density maps.

In a first step, all obvious non-cluster objects were rejected from the automatically obtained catalogue. These rejected objects are mostly edges of dark clouds in regions with steep star density gradients close to the Galactic Plane. We rejected 179 objects from the automatically detected sample. Then, to obtain our final source sample, we manually selected all remaining objects possessing the same visual appearance in the star density maps as known star clusters. As for the automated detection procedure, the cluster had to be visible in the H band and at least in one of the other two bands. This resulted in a total of 1788 cluster candidates. 782 (44 per cent) of those were detected by the automated procedure. The ratio of the number of clusters detected manually and automatically is more or less constant along the Galactic Plane. Hence, the manual selection procedure introduces no additional selection effects that change systematically along the Galactic Plane.

We searched the SIMBAD data base[1] for known clusters within three times the core radius (see Section 4.1) of each candidate. All entries with classification as GlCl?, Cl*, GlCl, OpCl, *inCl, Gl?, Cl*, GlC, OpC, As* or *iC were selected. We identified 681 known OpCls and 86 known GlCls among the cluster candidates in our survey field. This includes, e.g., the GlCl Glimpse-C01 which was recently discovered with *Spitzer* data by Kobulnicky et al. (2005).

Please note that the SIMBAD classification is not 100 per cent reliable for all individual objects. However, single misclassifications will not influence the statistical analysis performed in this paper. One possible reason for such misclassifications is erroneous coordinates of known clusters in the SIMBAD data base. We found, for example, that the SIMBAD coordinates of the known OpCl Berkeley 51, which is 3.9 arcmin away from one of our new cluster candidates (Cl 0197, core radius 0.6 arcmin), are probably inaccurate. An inspection of the 2MASS images reveals no star cluster like object at the SIMBAD coordinates of Berkeley 51, but clearly a star cluster at the coordinates of our candidate (see also Fig. 1). Some other examples for potentially erroneous coordinates in the SIMBAD data base are the objects Cl 0191–Berkeley 49, Cl 0361– Berkeley 98, Cl 1476–VDBH 63. We have therefore searched the vicinity ($r < 10$ arcmin) of all our new cluster candidates to find such objects. In Tables A1 and B1, we provide the names of all known star clusters closer than 10 arcmin and outside three times the candidate core radius. There are 68 (7 per cent) cases of such a close-by known cluster outside three times the core radius of our candidate.

In total, 1021 of our cluster candidates have no known entry in the SIMBAD data base. In the Appendix Tables A1 and B1, we summarize the properties of our candidates. Note that these tables will be available online only.

2.2 Reliability of our cluster candidates

How reliable are the new cluster candidates, i.e. how significant is the contamination of our sample with random star density fluctuations of field stars? Here, we will discuss several approaches to quantify the contamination of our sample statistically.

Spatial distribution. We analyse the spatial distribution of our cluster candidates in Section 3.1. A comparison of the distributions of known OpCls and new cluster candidates shows that about 50 per cent of the new candidates are distributed homogeneously in our map (see left-hand panel of Fig. 3). Thus, they do not follow the distribution of known OpCl, and are therefore likely contaminating objects (see Section 3.1 for details). This provides a first estimate of the contamination in our sample of 50 per cent.

Detection method. We detect 73/86 (85 per cent) of the know GlCls and 435/681 (64 per cent) of the known OpCls in our sample automatically. In contrast, only 274/1021 (27 per cent) of the new candidates are detected automatically. If we assume that the ratio of automatically/manually detected known OpCls applies also for the new candidates, we should have found only 428 new clusters. Hence, we can estimate a contamination of $(1021–428)/1021 = 58$ per cent. Since new clusters are likely to possess few members (because they remained undiscovered so far), this number can be considered to be an upper limit. Thus, it is in good agreement with the above estimated contamination of 50 per cent. It also indicates that the automatically detected cluster candidates are less contaminated. We

---

[1] This cross-identification was performed in 2006 May.
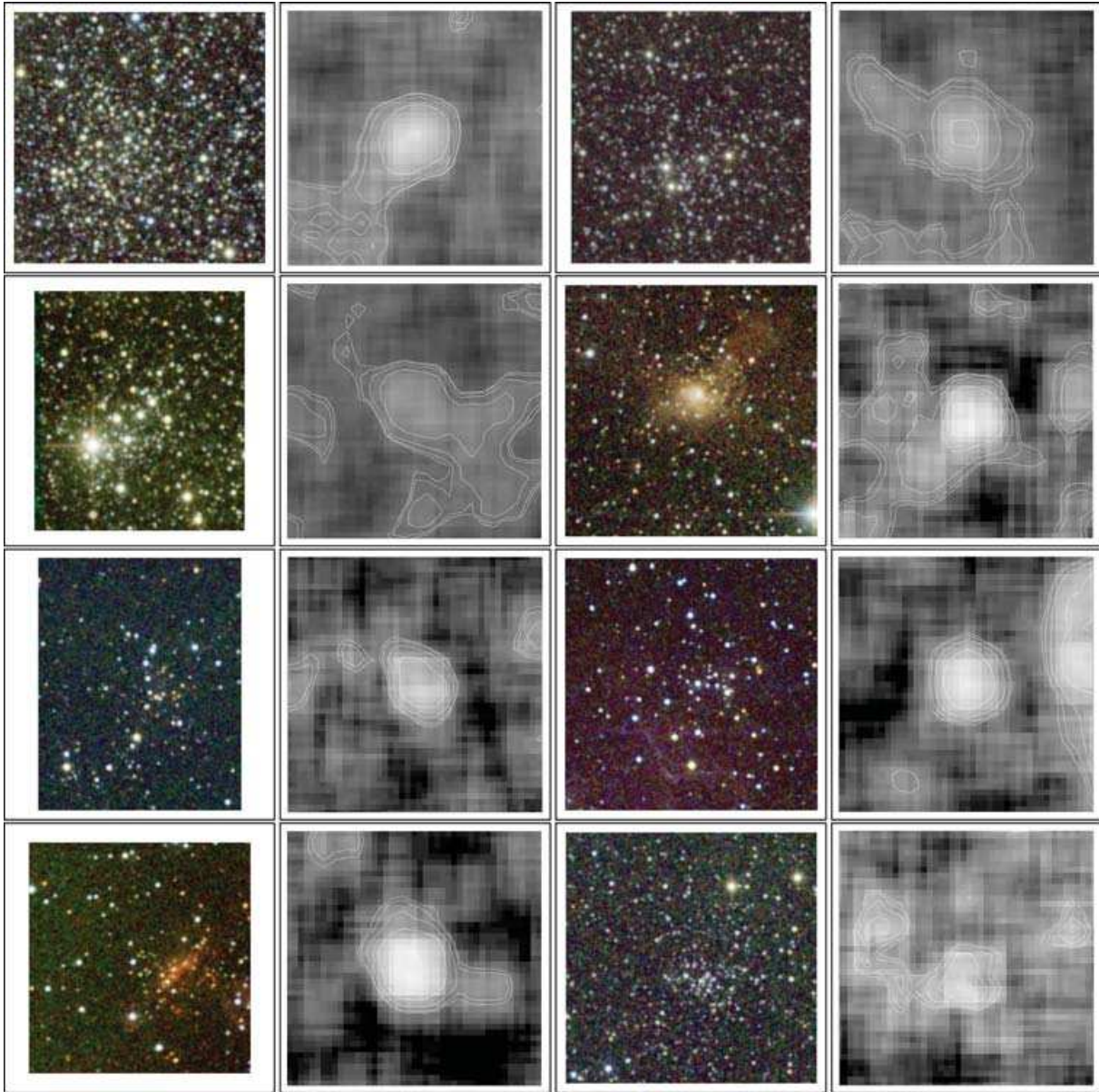
Figure 1. 2MASS *JHK* colour (left-hand side) and *K*-band star density (right-hand side) maps for the new cluster candidates 0190 (Qual. flag 1), 0191 (0), 0197 (1), 0426 (3), 0488 (3), 0784 (4), 0849 (4), 1476 (3); from top left to bottom right. The size of the boxes around the *JHK* colour images is 5 arcmin; north is up and east is towards left. In the star density maps, bright regions indicate enhanced star density. Contrary to the *JHK* images, the *K*-band star density maps are in Galactic coordinates, possess a size of 0°.25 and are centred on the cluster position. Note that in the *JHK* images, the clusters are not always centred due to the available 2MASS image data.

have marked all cluster candidates in Tables A1 and B1 that are detected automatically.

Cluster pairs. In Section 3.2, we investigate the probability $P(r)$ to find pairs of clusters with a given separation r. If one subsample of our clusters is dominated by contamination, we should clearly see a difference in the probability distribution for cluster pairs. We compared $P(r)$ for (i) the known OpCls in our sample, (ii) all new cluster candidates, (iii) the probable cluster candidates, (iv) the manually detected clusters and (v) the automatically detected clusters. We find that there is no statistically significant difference between $P(r)$ for all subsamples. This indicates that the contamination, albeit present, does not influence our analysis in Section 3.2.

Cluster morphology. A further possibility to determine the contamination is the morphological analysis of our cluster candidates

in Section 4.1. We fit radial star density (King) profiles to all identified objects, and analyse the quality of those fits. We find that the known GlCls give the best fit: 81/86 (94 per cent) have a quality flag better than 3 (see Section 4.1 for how this flag is determined). In the case of the OpCls, we have 483/681 (71 per cent) with a good quality flag. For our new cluster candidates, we obtain for 455/1021 (45 per cent) a good fit. Assuming again the percentage of the OpCls for the new candidates, we should have had a good fit for 726 objects. This indicates a contamination of $(1021–726)/1021 = 29$ per cent, somewhat less than the estimates above. In Tables A1 and B1, we list for each of the new cluster candidates the quality flag.

Visual inspection. We performed a visual inspection of 2MASS images for a number of randomly selected new cluster candidates. We inspected *JHK* colour images of 60 cluster candidates and
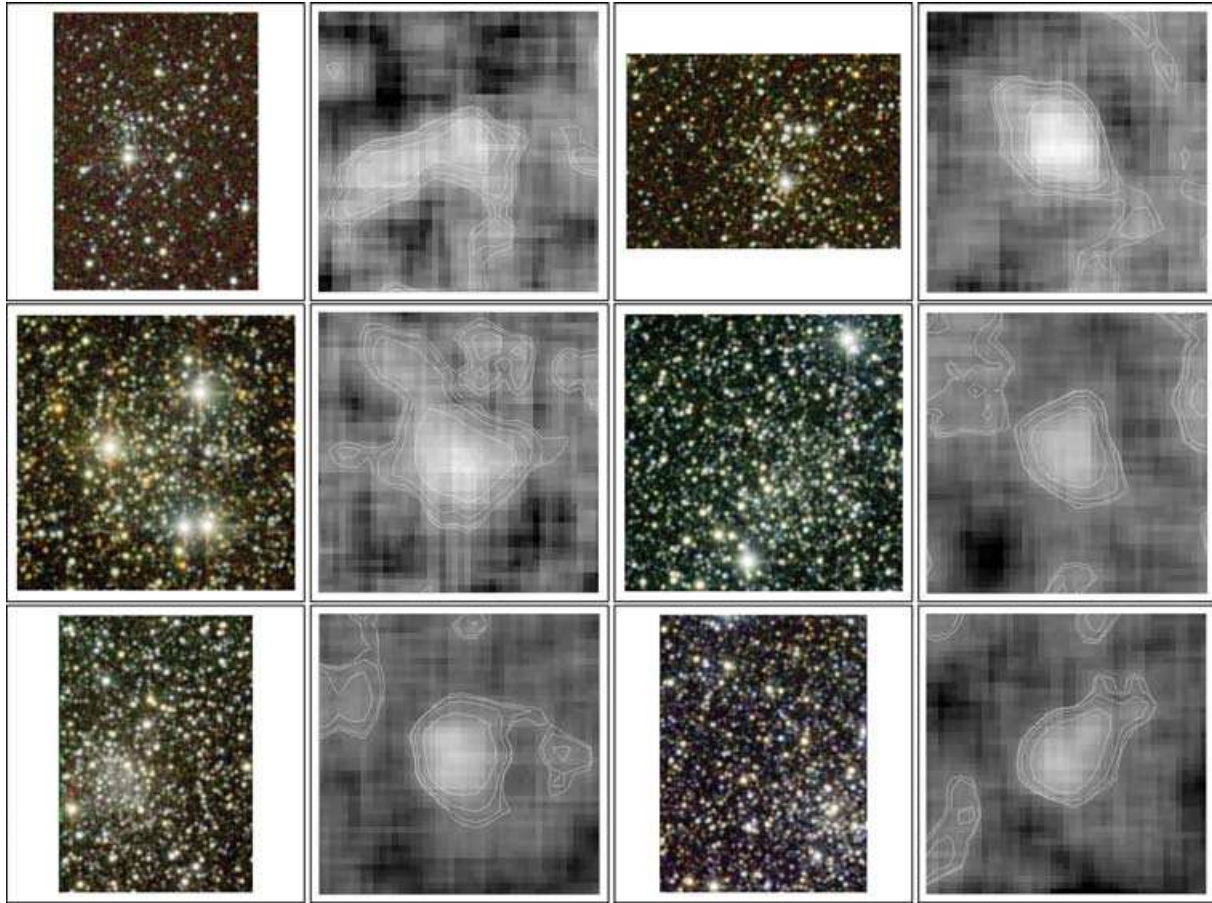
Figure 2. As Fig. 1 but for the new cluster candidates 1483 (1), 1530 (2), 1656 (3), 1716 (2), 1735 (4), 1744 (3); from top left to bottom right.

selected all objects where an obvious cluster of stars can be identified visually. We find that for objects with a quality flag better than 4, about 30 per cent of the cluster candidates show a clear cluster-like appearance in the 2MASS images. To exemplarily demonstrate the appearance of our candidates in 2MASS images, in Figs 1 and 2 we present *JHK* colour images and for comparison *K*-band star density maps for a number of candidates, where a clear clustering of stars is apparent in the *JHK* image. The 30 per cent of clear cluster candidates would indicate a contamination of 70 per cent. We note, however, that the visual impression in images can be misleading if the cluster consists of many faint stars. Hence, the star density maps are a better guide for cluster identification. Similarly, an increased number of bright stars in an image might lead to the impression of a cluster, even if the star density map shows no peak.

In summary with three independent estimates, we find contamination rates between 30 and 60 per cent, and thus probably around half of our newly identified candidates are no real clusters. Since we account for this in our statistical analysis in Section 3.2, it does not influence our results. For individual clusters, the determined quality flag, based on statistical estimates (see Section 4.1), can be used to judge the likelihood that the candidate is a real cluster. Verification of the nature for each individual clusters requires the analysis of colour–magnitude diagrams and thus the identification of cluster members. Since we concentrate here on the statistical analyses of our cluster sample, this is beyond the scope of this work and will be addressed in a future paper.

However, using our quality flag and the position of the cluster candidates, we have identified the most probable new star clusters in our sample. In particular, we selected all cluster candidates that have $|b| < 4°$ and possess a quality flag of less than five (73 per cent of the known OpCls fulfil this condition). These are 473 objects, their properties are listed in Table A1, and we will refer to them as probable clusters. The majority of the remaining objects are probably only local star density enhancements, their properties are listed in Table B1, and they are referred to as possible clusters.
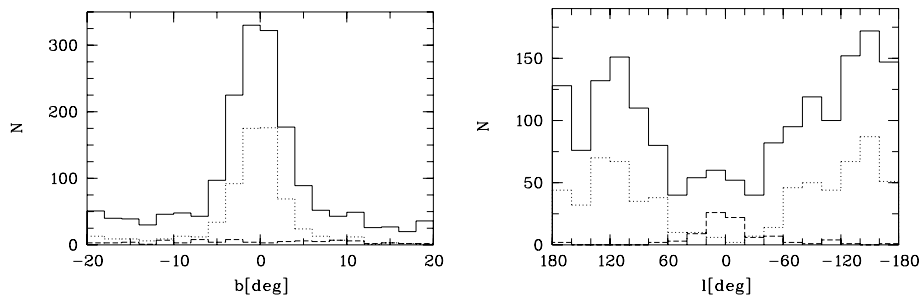
3 OVERALL DISTRIBUTION OF STAR CLUSTERS

3.1 Distribution, contamination and completeness

In Fig. 3, we plot the distribution of our clusters in the search area. The upper panel shows the known GlCls contained in our sample. The concentration towards the Galactic Centre is seen clearly. The middle panel shows the known OpCls identified in our sample. As expected, this type of cluster is concentrated towards the Galactic Plane and the main star-forming regions. However, there is a clear lack of objects towards the Galactic Centre indicative of the fewer known OpCls in this area, as well as a selection effect of our cluster detection method (evident also in the right-hand panel of Fig. 4). In the lower panel of Fig. 3, the new cluster candidates of our sample are shown. They follow the principle distribution of the OpCls,
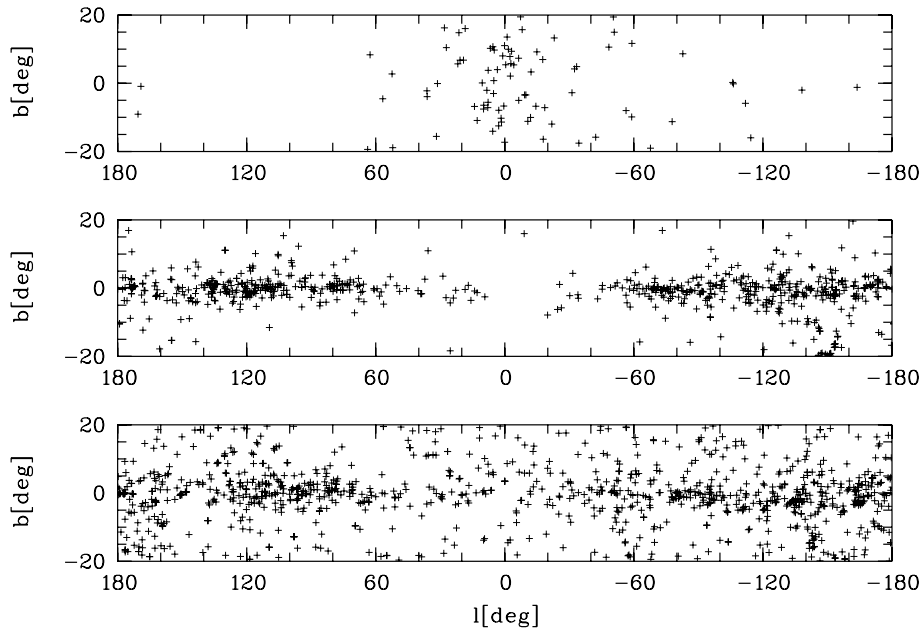
Figure 3. Distribution of the known clusters and new cluster candidates in our search area. Top: known GlCls. Middle: known OpCls. Bottom: new star cluster candidates. The different distributions of the two types of stars clusters can be seen clearly. Also, the lack of known OpCls near the Galactic Centre is evident in the middle panel.

Figure 4. Distribution of our star cluster sample perpendicular (left-hand panel) and along (right-hand panel) the Galactic Plane. The solid line shows the histograms for all our objects. The dotted line represents the known OpCls and the dashed line the known GlCls. Again, the apparent lack of clusters toward the Galactic Centre is evident.

reflecting the fact that most of the new candidates are expected to be OpCls. This panel also shows that there is a component of about 500 objects which are homogeneously distributed. This population is also evident in the histogram of the cluster distribution across the Galactic Plane (left-hand panel of Fig. 4). Assuming that the distribution of the known OpCls is representative, we can conclude that those 500 cluster candidates are most likely not star clusters, but only local star density enhancements. Thus, the contamination of our list of new clusters is likely to be about 50 per cent, implying that ~500 of our new cluster candidates are real, as already stated in Section 2.2. We note the contaminating objects, which are homogeneously distributed, do not influence the statistical analysis of the clustering properties in Section 3.2.

By counting the cluster candidates in bins of 60° length along the Galactic Plane, we can estimate the number of potentially missing objects. In the bins beside the Galactic Centre, we find 150 and 180 clusters, respectively. The region $180° < l < 240°$ possesses the highest number of clusters (470), while the remaining three bins contain about 320 objects each. The bin with 470 clusters is close to the star-forming regions of Orion, Taurus and Perseus and might represent a local exception from the overall cluster density.

Nevertheless, there are about 300 clusters missing in the area ±60° around the Galactic Centre. These clusters are not detected due to the high background star density (low-density contrast between star cluster and background) and hence provide a promising target area for future projects like the UKIDDS survey.

This shows that our detection method is subject to selection effects. Only star clusters that possess a significant density enhancement towards the centre are detected. As a result, the high star densities towards the Galactic Plane and especially towards the Galactic Centre will hamper our detection rate. Hence, any statistical analysis of OpCl properties will have to be constrained to regions outside ±60° from the Galactic Centre. Clusters with few stars and clusters that are spread out over a large area on the sky are not detected either. Those types of clusters, however, are not expected to be distributed in a way different from the detected objects, and hence will not influence our analysis in Section 3.2.

3.2 Clustering of clusters

Here, we investigate the projected distribution of the star clusters in our sample. In particular, we are interested in modelling this
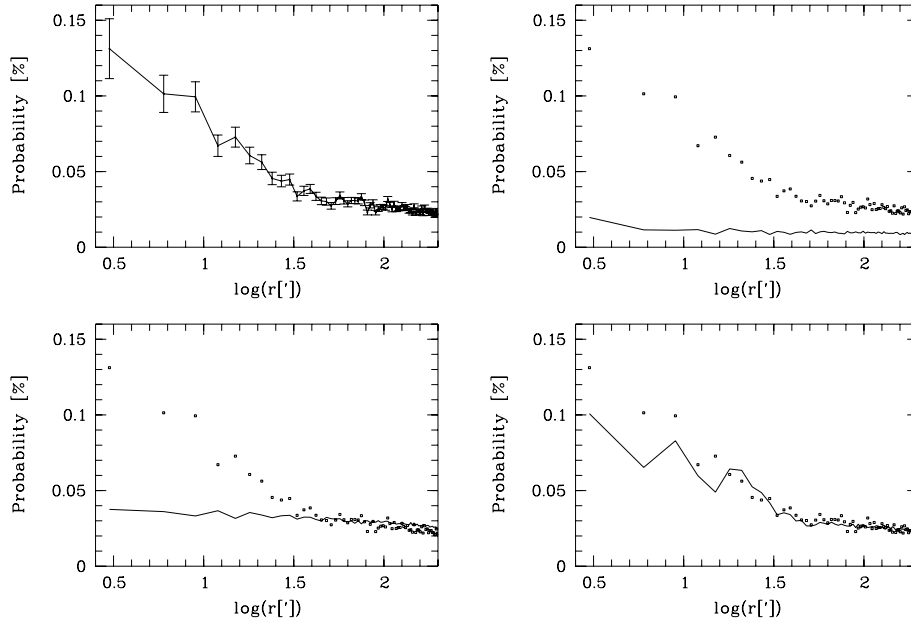
Figure 5. Probability $P(r)$ to find pairs of clusters with a separation r, normalized to the area and number of clusters. Upper left: probability for the observational sample. Only star clusters which are more than $60°$ away from the Galactic Centre are included. Error bars are statistical errors based on the number of clusters taken into account for each individual data point. Upper right: probability distribution for a homogeneous distribution of star clusters (solid line). Overplotted as dots are the observational data points (shown without error bars). Lower left: probability distribution for a model assuming a homogeneous distribution of 400 clusters. The remaining objects are homogeneously distributed along the Galactic Plane and possess a Gaussian distribution with a width of $5°$ perpendicular to it. Lower right: probability distribution assuming a homogeneous distribution of 400 clusters. 600 clusters are grouped in pairs of two. These pairs, as well as the remaining clusters, possess a homogeneous distribution along the Galactic Plane and Gaussian distribution with a width of $5°$ perpendicular to the Galactic Plane. Within the pairs, the clusters are distributed homogeneously within $0°.7$.

two-dimensional distribution using simple assumptions. To quantify the distribution of clusters, we determined the probability $P(r)$ to find pairs of clusters with a given separation r. This is done by counting the number of clusters $\mathcal{N}_i(r)$ in a ring with a radius r and a thickness δr for each cluster i. This has to be normalized by the total number of clusters ($\mathcal{N} - 1$) and the area of the ring $2\,\pi r\,\delta r$. These individual probabilities $P_i(r)$ are then averaged to obtain the probability $P(r)$ as follows:

$$P(r) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} P_i(r) = \frac{1}{\mathcal{N}(\mathcal{N}-1)} \sum_{i=1}^{\mathcal{N}} \frac{\mathcal{N}_i(r)}{2\pi r\,\delta r}. \tag{1}$$

Note that we used $\delta r = 3$ arcmin as thickness of the ring. Since we encountered a significant drop in the number of detected clusters towards the Galactic Centre ($\pm 60°$), we exclude this area from the analysis. This leaves $\mathcal{N} = 1461$ cluster candidates. There are only 13 known GlCls in this subsample, as well as 632 known OpCls and 816 new candidates.

In the upper left panel of Fig. 5, we plot $P(r)$ for our sample of clusters. There are two distinct regions in the diagram. For separations larger than about $0°.7$, we find a linear decreasing trend of P with $\log r$. At smaller r, the probability rises much steeper with decreasing radius. If we use only the 632 known OpCls to determine $P(r)$, in order to investigate if the contamination of ~ 50 per cent influences the distribution, we obtain the same qualitative behaviour but with a slightly larger scatter due to the smaller number of clusters. Indeed, all subsamples of clusters (manually detected objects, automatically detected objects, known OpCls, new cluster candidates) show the same qualitative and quantitative behaviour. This might be due to the fact that contaminating objects, i.e. random density enhancements, are not expected to show any clustering, and

thus contribute only a constant to the plot shown in Fig. 5. In any case, the results of the following analysis are unaffected by contamination.

How does this observed distribution compare with models for the distribution of the same number of clusters? The most simple model would be a homogeneous distribution of the clusters in the investigated area. The resulting probability plot for this model is shown in the upper right panel of Fig. 5. Note that we determined $P(r)$ as the average of 10 different homogeneous model distributions to minimize the scatter. The same number of repeats is also used for all other models investigated in this section. Obviously, the homogeneous model is not appropriate, which is also apparent in Fig. 3. In contrast to the observational sample, this model shows a probability that is almost independent of the distance r, and also much smaller. This is due to the fact that a homogeneous distribution leads on average to larger distances between the star clusters. In fact, an ideal homogeneous distribution would result in $\mathcal{N}_i(r)$ being proportional to the separation r, and therefore a constant value for $P(r)$ (see equation 1). The good agreement of our homogeneous model distribution with a constant value reflects the high quality of the random number generator used. The larger deviation for very small r is due to a very small number of cluster pairs at such a small separation.

Can we obtain better agreement with a more realistic approach? From the histograms in Fig. 4, we can deduce the general properties of the cluster distribution. There is a homogeneous component of about 400 clusters in the investigated area. The remaining objects show a more or less homogeneous distribution along the Galactic Plane (in the region more than $60°$ away from the Galactic Centre), and a Gaussian distribution with a width of $5°$ perpendicular to the Galactic Plane. The lower left panel of Fig. 5 shows the probability

distribution for this model. For r larger than $0°.7$, the model matches the probability of the observational data quite well. It reproduces the linear decline of P with $\log r$, and also the absolute values for P are in agreement with the observations. At small r, the model continues to show the same linear trend, in contrast to the increase in the slope in the observational sample.

What causes the difference between the model and the observations for smaller separations? Apparently at these shorter distances, it is much more probable to find another cluster than predicted by the simple model, obtained from the histograms of the cluster distribution. This is clear evidence for strong, local clustering of star clusters. We incorporate this in our models by again using 400 clusters as homogeneous component. Additionally, we sort 600 of the remaining clusters in pairs of two. Those pairs have a size of $0°.7$, in agreement with the change of slope in the probability distributions. Within the $0°.7$ box, the coordinates of the two clusters are distributed homogeneously. All pairs as well as all clusters that are not part of the homogeneous component are distributed as in the above model (homogeneous along and Gaussian, with a width of $5°$, perpendicular to the Galactic Plane). We experimented with the number of clusters in each group, the group size and the number of clusters that are part of groups. The size of the group determines the radius where we see a change in the slope of $P(r)$. If we put more than two clusters in each group, the slope at small r is much larger than in the observations. The number of groups influences the absolute value of $P(r)$.

This model is now in very good agreement with our observational sample. It very well predicts the correct absolute values for the probability at all separations. The change in the slope at $0°.7$ is also captured. At very small r (below 3 arcmin), there is still some discrepancy between model and observations, which might be due to substructuring on very small scales. In total, however, our simple model approach yields surprisingly good agreement between simulated and observed cluster frequency. All simulations without pairing (i.e. clustering) of objects are clearly not sufficient. This provides strong evidence for clustering of star clusters on scales of the order of $0°.7$. Assuming a mean distance of 1–2 kpc of the clusters, we find that the clustering of star clusters appears on scales of approximately 12–25 pc. Since this is in the same order as the size of typical star-forming molecular clouds, from which clusters are believed to form, this might indicate that the observed distribution of clusters still contains information about the birthplaces of stars in the Galaxy.

4 CLASSIFICATION OF NEW CLUSTER CANDIDATES

4.1 Morphological analysis

For every cluster candidate in our sample, we fit the radial star density profile in the three filters *JHK* at the completeness limit of 2MASS (determined locally for every cluster). Note that we include all objects with a quality flag better than 'E'. We used the King profile (King 1962)

$$\rho(r) = \rho_b + \rho_c r_{cor}^2 \left[\left(r_{cor}^2 + r^2 \right)^{-\frac{1}{2}} - \left(r_{cor}^2 + r_{tid}^2 \right)^{-\frac{1}{2}} \right]^2 \qquad (2)$$

for the star density ρ to determine the core and tidal radius (r_{cor}, r_{tid}) and the central and background star density (ρ_c, ρ_b).

Integrating the cluster star density $\rho(r) - \rho_b$ from zero to the tidal radius and substituting $x \equiv r_{tid}/r_{cor}$, we can determine the total

number N of stars in the cluster by

$$N = \pi \rho_c r_{cor}^2 \left[\ln(1 + x^2) - 4 + \frac{4\sqrt{1 + x^2} + x^2}{1 + x^2} \right]. \qquad (3)$$

This total number of stars depends on the magnitude down to which stars are included in the star density map, i.e., the local completeness limit near the cluster m_{cl}. To compare the number of cluster members between different star clusters, we have to convert N to a common magnitude limit m_{all}. This number N_c will be used in Section 4.3 to distinguish between OpCls and GlCls, and can in principle be estimated for each cluster using

$$N_c = N \times 10^{-\frac{C}{2.5}(m_{cl} - m_{all})}. \qquad (4)$$

The variable C in equation (4) denotes an unknown parameter. It describes how the number of stars in a cluster changes with stellar luminosity, i.e. a combination of the mass and luminosity function. It depends, for example, on the distance, type, age and extinction of the cluster, as well as on m_{cl}. The variable C only scales equation (4); thus the particular value of it is not important and we will set it to one for the analysis in Section 4.3. Note that different values only marginally influence the discrimination of the two cluster types.

4.2 Cluster parameters and uncertainties

We list the determined cluster parameters for the new cluster candidates in Tables A1 and B1 in the Appendix. The tables will be available online only and contain the following columns. (1) Unique ID for each new cluster candidate; (2) Galactic Longitude; (3) Galactic Latitude; (4) Right Ascension (J2000); (5) Declination (J2000); (6) Core radius in the *H* band; (7) Tidal radius in the *H* band; (8) Central star density in the *H* band; (9) Intensity contrast – central star density/sqrt(background star density) in the *H* band; (10) Number of stars in the cluster in the *H* band; (11) Corrected number of stars in the cluster in the *H* band (see Section 4.1); (12) $\log R_N$ (see Section 4.3 for details); (13) Quality flag (see below) and (14) Name of possible known cluster with erroneous coordinates in SIMBAD.

What are the uncertainties of the parameters? The cluster coordinates are determined by a Gaussian fit of the star density peak. The star density maps are created by counting stars in 3.5 arcmin sized boxes with an oversampling of 10. Considering this, the accuracy of the positions is in the order of the pixel size of our maps, i.e. 20 arcsec or $0°.005$. However, in cases of extended or not centrally condensed clusters, the uncertainty could reach $0°.01$. The error of the core and tidal radii can be estimated as follows. We determine the scatter of the fit radii from the mean value in all three filters *JHK*. As error for the radii, we assume a value that is larger than the scatter of the best 2/3 of the objects. For the core radius, we find that 2/3 of the objects possess a scatter of less than 15 per cent. In case of the tidal radius, the situation is worse. The best 2/3 of the objects possess a scatter of less than 33 per cent. Hence, for many objects the tidal radius is not well determined.

To give a measure to judge the quality of the radial star density profile fit and the reliability of the cluster parameters of each new cluster candidate, we introduce a quality flag. This flag consists of two parts. (i) An integer number ranging from zero to six, indicating how many of our quality tests are negative for this particular object and (ii) a six digit binary number, allowing to identify which of the quality tests are negative. The following bits are used: (i) the cluster is detected automatically (0) or manually (1); (ii) the scatter of the three fitted core radii is larger than 30 per cent (1) or smaller (0); (iii) the scatter of the three fitted tidal radii is larger than 66 per cent

(1) or smaller (0); (iv) the contrast of the central to background star density [Icentral/sqr(Iback)] in all filters is larger than five (0) or not (1); (v) the ratio of tidal to core radius in all three filters is between 3 and 45 (0) or not (1); (vi) the core radius in all three filters is larger than twice the error of the coordinates, i.e. $0\overset{\prime\prime}{.}01$ (0) or not (1).

4.3 Classification: OpCl versus GlCl

Our large and homogeneous sample of known star clusters gives us the opportunity to distinguish between OpCls and GlCls among the new cluster candidates based on statistical differences in their morphological properties, as determined in Section 4.1. In particular, we are interested in identifying the most probable new GlCl candidates within the vast number of unknown objects. For this purpose, we investigated all possible combinations of two morphological properties (determined as described in Section 4.1) for the known clusters in our sample. We then selected the combinations that showed the best discrimination of OpCls and GlCls. The three best combinations for this purpose are plotted in Fig. 6 (presented are data from *H*-band images), which shows how the corrected number of stars N_c, the core radius and the central star density relate to each other for the known OpCls (dots) and GlCls (plus signs) in our sample. It is evident in Fig. 6 that the two types of clusters occupy different parts of the parameter space, even if some GlCls are well within the OpCl regime.

The three parameter combinations shown in Fig. 6 are available in three different filters *JHK*, resulting in nine different parameter combinations that we can use to distinguish OpCls and GlCls. The high redundancy in all these different relationships as well as the large samples of known clusters in our data base ensures a reliable analysis of the statistical differences between OpCls and GlCls.

In each plot, we count how many known GlCls and OpCls are in a box around each star cluster candidate in our sample. The box has a size of 0.3 in logarithmic units. This number is then normalized by the total number of known clusters in our sample. These ratios are a measure for the probability that this cluster candidate is a GlCl ($P_{\mathrm{plot}_i}^{\mathrm{Gl}}$) or OpCl ($P_{\mathrm{plot}_i}^{\mathrm{Op}}$). Combining these individual probabilities from all plots using

$$P^{\mathrm{Op/Gl}} = \left(\prod_{i=1}^{9} P_{\mathrm{plot}_i}^{\mathrm{Op}/\mathrm{Gl}}\right)^{1/9}, \qquad (5)$$

we determined a measure for the probability of each new cluster candidate to be a GlCl (P^{Gl}) or OpCl (P^{Op}).

Almost all new cluster candidates are most likely OpCls. We note, however, that the sample of known OpCls is very incomplete, especially towards distant clusters and objects with few stars. Compared to this, the sample of GlCls is almost complete, only about 10 per cent are believed to be missing towards the Galactic Centre (Ivanov et al. 2005). Therefore, the probability P^{Op} is a lower limit. The probability P^{Gl}, however, opens the possibility to identify the best candidates for the last missing GlCls in our Galaxy. Given the incompleteness of the OpCl sample, however, P^{Gl} has to be considered an upper limit.

4.4 Best new GlCl candidates

Here, we investigate the feasibility of our method to determine a measure for the probability that new cluster candidates in our sample are GlCls or OpCls and to select new GlCl candidates. For each cluster, we determined the ratio $P^{\mathrm{Op}}/P^{\mathrm{Gl}}$. If we had no information at all about the morphological properties, this ratio would be
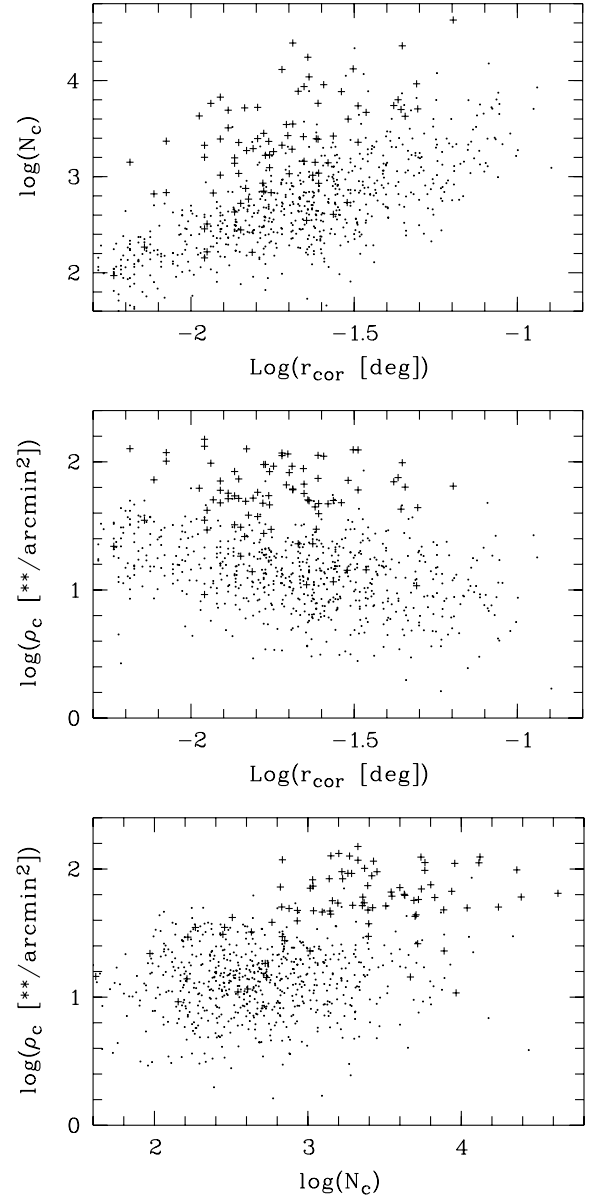

Figure 6. Example plots of cluster properties (*H*-band data) for know OpCls (dots) and known GlCls (plus signs) in our object sample, which were used to classify the new cluster candidates. Top: corrected number of stars versus the core radius. Middle: central star density versus the core radius. Bottom: central star density versus the corrected number of stars. It can clearly be seen that the OpCls occupy a different part of the parameter space than most of the GlCls.

equal to the ratio of the number of the two types of known clusters in our sample. Hence, by defining $R_N \equiv (P^{\mathrm{Op}}/P^{\mathrm{Gl}})/(681/86)$ we obtain an easy measure for each cluster to be GlCl or OpCl. If $\log R_N$ is negative, the object is probably a GlCl, otherwise it is an OpCl.

In Fig. 7, we plot the histogram for our cluster candidates versus $\log R_N$, our measure for the classification of clusters. The solid histogram represents the new cluster candidates in our sample and the dashed histogram the known GlCls and the dotted histogram the known OpCls. The vertical solid line marks the (statistical) borderline between GlCls and OpCls. Note that in Fig. 7 objects which have a P^{Gl} value of zero (and hence undetermined R_N) are missing.
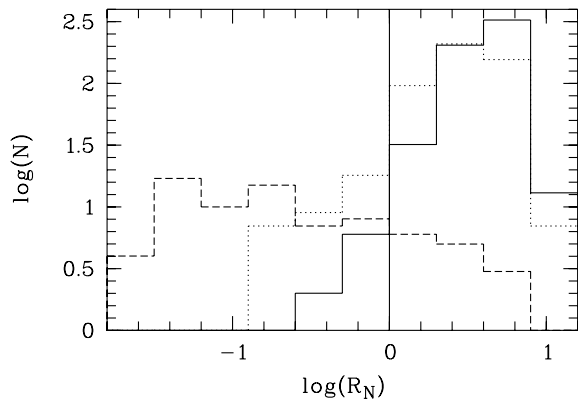
Figure 7. Distribution of the measure R_N for our sample of clusters. The value of R_N gives an indication of what type the star cluster is. We plot logarithmic histograms versus the logarithm of the measure R_N. The dividing line at $\log R_N = 0$ separates the area of GlCls (negative $\log R_N$) and OpCls (positive $\log R_N$). The solid histogram represents the new cluster candidates in our sample, the dashed histogram the known GlCls and the dotted histogram the known OpCls.

Table 1. Positions of the nine new cluster candidates in our sample that are classified as most probably being a GlCl. The $\log R_N$ values for all other new candidates and their properties can be found in the Appendix in Table A1.

Cluster	$l(°)$	$b(°)$	$\log R_N$	Qual. flag	
0001	0.029	3.474	−0.43	2	100 010
0002	0.048	3.440	−0.62	2	100 010
0055	17.992	−0.278	−0.03	4	110 110
0089	29.491	−0.981	−0.15	2	100 010
1603	298.222	−0.507	−0.11	0	000 000
1703	325.788	0.124	−0.32	3	100 110
1716	329.792	−1.589	−0.11	2	100 010
1755	348.246	0.482	−0.19	2	100 010
1767	352.602	−2.168	−0.05	2	100 001

Those objects, however, are most probably OpCls. As can be seen in the figure, most known clusters are classified correctly using this approach: out of the 86 known GlCls, only 14 (16 per cent) have $\log R_N > 0$ and thus fail our criterion. For OpCl, the agreement is even better. Only 36 out of 681 (5 per cent) OpCls are falsely classified by our approach.

There are nine new cluster candidates in our sample that possess a value of R_N less than one, and are hence promising GlCl candidates (see Table 1). Given the ratio of known GlCls and OpCls in this parameter range (72/36) and an overall contamination rate of about 50 per cent of the 1021 new cluster candidates (see Section 3), we estimate that about 25 per cent, i.e. two to three of those candidates are likely to be new GlCls. However, this estimate is subject to the uncertainties in the completeness of the known OpCls in our sample and the contamination of our new candidates. We also note that all new GlCl candidates have rather low (albeit negative) values of $\log R_N$, close to the OpCl regime, and all but one have a quality flag of two or higher. A visual inspection of 2MASS *JHK* colour images shows no clear indication of a star cluster for all but the cluster candidate 1716 (see Fig. 2). As discussed in Section 2.2, the impression of colour images can be misleading, because new clusters might be dominated by faint stars, which do not strongly affect the visual (subjective) appearance of the cluster.

Nevertheless, it will be justified to systematically investigate the new cluster candidates with the smallest $\log R_N$ values, e.g. with deep and high-resolution NIR imaging. This is particularly interesting in areas close to the Galactic Plane and the Galactic Centre, where extinction hampers the detection of GlCls at optical wavelength. The recently discovered GlCl Glimpse C01, for example, which has $\log R_N = -0.851$ according to our analysis, is situated in this area ($l = 31.31$, $b = -0.10$).

5 CONCLUSIONS

We have used star density maps determined from the 2MASS point source catalogue to obtain a complete sample of star clusters in the entire Galactic Plane ($|b| < 20°$). We used a combination of automated searches for local density enhancements and manual detection to obtain our cluster sample. In total, 1788 cluster candidates are detected by our method. There are 86 previously known GlCls and 681 OpCls among our source sample. The remainder of 1021 objects are new star cluster candidates.

An analysis of the spatial distribution of the clusters shows that the majority of the new candidates are similarly distributed as the previously known OpCls. There is, however, a component of roughly 500 objects that appear to be homogeneously distributed. These sources are therefore most likely local star density enhancements and not real star clusters. We estimate the contamination rate of our candidate sample to be about 50 per cent.

We determined how the probability of finding star cluster pairs changes with separation. Using a simple model for the two-dimensional cluster distribution, we were able to reproduce the observations. The model assumes a homogeneously distributed component of clusters (i.e. the contamination of our sample), a component that is homogeneously distributed along and Gaussian distributed perpendicular to the Galactic Plane, and a fraction of clusters in close-by pairs. These close-by pairs of star clusters are unconditionally required to reproduce the observational star cluster distribution providing strong evidence for clustering of star clusters on small scales. The size of the groups of 0.7 indicates a locally enhanced cluster density on scales of 12–25 pc, a typical size of molecular clouds.

The large and homogeneous sample of known OpCls and GlCls allowed us to use a statistical approach to classify the new cluster candidates. This was done by means of cluster properties obtained by fitting the radial star density profile of each object by a King profile. The measure for classification obtained by us classifies only 16 per cent of the known GlCls and only 5 per cent of the known OpCls wrongly. According to our criterion, there are nine promising GlCl candidates among our new cluster candidates, from which two to three are likely to be real GlCl. A systematic detailed investigation of the most promising GlCl candidates is suggested in order to identify possible new Galactic globular clusters in our sample.

REFERENCES

Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bica E., Dutra C. M., Soares J., Barbuy B., 2003, A&A, 404, 223
Carraro G., 2005, ApJ, 621, 61
Dias W. S., Alessi B. S., Moitinho A., Lépine J. R. D., 2002, A&A, 389, 871
Dutra C. M., Bica E., Soares J., Barbuy B., 2003, A&A, 400, 533
Froebrich D., Ray T. P., Murphy G. C., Scholz A., 2005, A&A, 432, 67
Harris W. E., 1996, AJ, 112, 1487
Hurt R. L., Jarrett T. H., Kirkpatrick J. D., Cutri R. M., Schneider S. E., Skrutskie M., van Driel W., 2000, AJ, 120, 1876
Ivanov V. D., Kurtev R., Borissova J., 2005, A&A, 442, 195
King I., 1962, AJ, 67, 471
Kobulnicky H. A., Monson A. J., Buckalew B. A. et al., 2005, AJ, 129, 239
Lada C. J., Lada E. A., 2003, ARA&A, 41, 57
Lynga G., 1995, VizieR On-line Data Catalog: VII/92A
Mermilliod J.-C., Paunzen E., 2003, A&A, 410, 511
Ortolani S., Bica E., Barbuy B., 2000, A&A, 361, 57
Skrutskie M. F., Cutri R. M., Stiening R. et al., 2006, AJ, 131, 1163
van den Bergh S., 1993, ApJ, 411, 178
West M. J., Côté P., Marzke R. O., Jordán A., 2004, Nat, 427, 31
Willman B., Blanton M. R., West A. A. et al., 2005, AJ, 129, 2692

APPENDIX A: PROPERTIES OF PROBABLE CLUSTER CANDIDATES

Table A1. Properties of the probable clusters in our sample of newly detected star cluster candidates. The columns contain the following. (1) Unique identifier for each cluster candidate; (2) Galactic Longitude in degrees (typical error $0°.01$); (3) Galactic Latitude in degrees (typical error $0°.01$); (4) Right Ascension (J2000) (typical error $0°.01$); (5) Declination (J2000) (typical error $0°.01$); (6) Core radius in degrees as fitted to the *H*-band radial star density profile (typical error 15 per cent); (7) Tidal radius in degrees as fitted to the *H*-band radial star density profile (typical error 33 per cent); (8) Central star density in stars per arcmin^2 as fitted to the *H*-band radial star density profile; (9) Star density contrast in the *H* band, determined as ratio of the central star density and the square root of the background star density. Values of -99 indicate objects with a fitted background star density of zero; (10) Number of stars in the cluster as obtained using equation (3) and the star density profile in the *H* band; (11) Corrected number of stars in the cluster determined using equation (4) at an apparent magnitude of 15 in the *H* band; (12) The measure $\log R_N$ as determined in Section 4.4. Values of 99 indicate cluster candidates with $P^{Gl} = 0$; (13) Quality flag determined from our analysis of the radial density profile fit in Section 4.2 and (14) Name of possible known cluster with erroneous coordinates in SIMBAD.

(1) ID	(2) l (°)	(3) b (°)	(4) RA (h: m: s)	(5) Dec. (J2000) (°: ′: ″)	(6) r^h_{core} (°)	(7) r^h_{tid} (°)	(8) I^h_{cent} (**/arcmin^2)	(9) Con.^{h}	(10) N^h	(11) N^h_c	(12) $\log R_N$	(13) Qual. flag	(14) Poss. known cluster
0001	0.029	3.474	17:32:22	−27:03:39	0.026	0.872	17	5	1654	7221	−0.379	2	100 010	
0002	0.047	3.44	17:32:32	−27:03:51	0.089	0.178	32	10	6534	28 520	−0.55	2	100 010	
0022	6.182	0.843	17:56:28	−23:11:34	0.039	0.078	16	4	642	4439	0.142	3	100 110	
0023	6.584	0.782	17:57:35	−22:52:32	0.006	0.119	25	7	119	686	0.28	3	110 001	
0031	8.906	−0.268	18:06:29	−21:22:33	0.08	3.349	15	5	15 152	41 732	99.0	2	000 110	

APPENDIX B: PROPERTIES OF POSSIBLE CLUSTER CANDIDATES

Table B1. Properties of the possible clusters in our sample of newly detected star cluster candidates. The columns contain the following. (1) Unique identifier for each cluster candidate; (2) Galactic Longitude in degrees (typical error $0°.01$); (3) Galactic Latitude in degrees (typical error $0°.01$); (4) Right Ascension (J2000) (typical error $0°.01$); (5) Declination (J2000) (typical error $0°.01$); (6) Core radius in degrees as fitted to the *H*-band radial star density profile (typical error 15 per cent); (7) Tidal radius in degrees as fitted to the *H*-band radial star density profile (typical error 33 per cent); (8) Central star density in stars per arcmin^2 as fitted to the *H*-band radial star density profile; (9) Star density contrast in the *H* band, determined as ratio of the central star density and the square root of the background star density. Values of -99 indicate objects with a fitted background star density of zero; (10) Number of stars in the cluster as obtained using equation (3) and the star density profile in the *H* band; (11) Corrected number of stars in the cluster determined using equation (4) at an apparent magnitude of 15 in the *H* band; (12) The measure $\log R_N$ as determined in Section 4.4. Values of 99 indicate cluster candidates with $P^{Gl} = 0$; (13) Quality flag determined from our analysis of the radial density profile fit in Section 4.2 and (14) Name of possible known cluster with erroneous coordinates in SIMBAD.

(1) ID	(2) l (°)	(3) b (°)	(4) RA (h: m: s)	(5) Dec. (J2000) (°: ′: ″)	(6) r^h_{core} (°)	(7) r^h_{tid} (°)	(8) I^h_{cent} (**/arcmin^2)	(9) Con.^{h}	(10) N^h	(11) N^h_c	(12) $\log R_N$	(13) Qual. flag	(14) Poss. known cluster
0009	1.864	−9.522	18:28:29	−31:53:47	0.005	0.25	16	6	69	91	0.802	4	100 111	
0010	2.149	19.617	16:40:49	−16:01:09	0.023	0.227	7	7	330	250	0.845	0	000 000	
0014	4.406	16.753	16:55:42	−15:59:40	0.009	0.25	10	8	122	92	0.779	3	110 001	
0018	5.338	5.407	17:37:41	−21:33:28	0.019	0.096	13	4	276	1586	99.0	3	110 100	
0019	5.52	6.08	17:35:39	−21:02:49	0.026	0.103	10	4	315	2180	99.0	2	000 110	

This paper has been typeset from a T$_E$X/L^{A}T$_E$X file prepared by the author.