Constructing the Universe with Clusters of Galaxies, IAP 2000 meeting, Paris (France) July 2000
Florence Durret & Daniel Gerbal eds.

THE NORTHERN SKY OPTICAL CLUSTER SURVEY DETECTION OF GALAXY CLUSTERS IN DPOSS

R.R. Gal
Palomar Observatory, Caltech, MC105-24, Pasadena, CA 91125

The Northern Sky Optical Cluster Survey is a project to create an objective catalog of galaxy clusters over the entire high–galactic–latitude Northern sky. We use the object catalogs generated from the Digitized Second Palomar Sky Survey (DPOSS, Djorgovski *et al.* 1999) as the basis for this survey. We apply a color criterion to select against field galaxies, and use a simple adaptive kernel technique to create galaxy density maps, combined with the bootstrap technique to make significance maps, from which density peaks are selected. This survey eliminates many of the subjective criteria of past surveys, and utilizes more information (especially colors) than the most similar recent survey, the APM (Dalton *et al.* 1992). This paper presents the details of our cluster detection technique, as well as some initial results for two small areas totaling ~ 60 square degrees. We find a mean surface density of ~ 1.5 clusters per square degree, consistent with the detection of richness class 0 and higher clusters to $z \sim 0.3$. In addition, we demonstrate an effective photometric redshift estimator for our clusters.

1 Introduction

Clusters of galaxies are the largest bound systems in the Universe and, as such, have been used to study how matter is distributed over extremely large scales. They provide useful constraints for theories of large-scale structure formation and evolution, and are valuable (possibly coeval) samples for studying galaxy evolution in dense environments. Studies of the cluster two-point correlation function and the power spectrum are important probes of large scale structure and the scenarios of its formation. Correlations between optically and X–ray selected (e.g., ROSAT, Chandra, etc.) clusters are also of considerable scientific interest, and will help us better understand the various selection effects present in both types of cluster samples.

Most of the optical studies to date have been limited by the statistical quality of the available cluster samples. For instance, the Abell catalog (Abell 1958) suffers from a non-objective selection process, poorer plate material, a bias towards centrally concentrated clusters, especially those with cD galaxies, and a relatively low redshift cutoff ($z \sim 0.15$; Bahcall & Soneira 1983). Still, many far–reaching cosmological conclusions have been drawn from it (i.e. Bahcall & West 1992).

Other catalogs preceding ours have been generated using objective means (APM, Dalton *et al.* 1992; EDSGC, Lumsden *et al.* 1992), but they too have been hampered by the data available. The APM group, for instance, used digitized J (blue) plates from the Southern Sky Survey; the use of a single, blue band provides no color information to distinguish galaxy types, and is a poor choice for cluster detection because clusters are better delineated by redder, early-type galaxies.

This paper presents the initial results of an optical cluster survey based on DPOSS data, from two fields covering ~ 60 square degrees. We use the 3 bands (JFN) of the DPOSS survey (Djorgovski *et al.* 1999) calibrated with extensive CCD data into the Gunn gri system. From this data, we construct color-color diagrams for galaxies detected in all 3 bands. Predicated on the morphology-density relation (Dressler 1980), we select galaxies that are more likely to appear in clusters (red, early–type) based on their colors. Alternatively, this can be viewed as a means of excluding galaxies that are more common in the field. After the color cuts are applied, we use the adaptive kernel technique (Silverman 1986) to produce a surface density map of galaxies. A bootstrap method is then used to construct the corresponding significance map, from which density enhancements are selected using the FOCAS peak finding algorithm.

2 Input Data and Catalog Preparation

2.1 Input DPOSS Data

The Northern Sky Optical Cluster Survey is based on data from the Digitized Second Palomar Sky Survey (DPOSS, Djorgovski *et al.* 1999), a digitization of the POSS–II three-band photographic survey of the entire Northern sky (Reid *et al.* 1991). Each plate covers $6.6° \times 6.6°$ on the sky, with neighboring plates overlapping each other by $\sim 1.6°$. The plates are scanned at STScI (lasker *et al.* 1996), with $1''$ pixels, and the digitized data are processed into catalogs at Caltech using the SKICAT system (Weir *et al.* 1995c). The end result of plate processing is a catalog of all objects detected down to the limiting magnitude of the plate ($g_J \sim 21.5, r_F \sim 20.5, i_N \sim 20.0$), with various photometric, positional and shape parameters. This data is collected from all processed plates into the Palomar Norris Sky Catalog (PNSC), expected to contain $\sim 50x10^6$ galaxies and $> 2x10^6$ stars. Classification of each object is performed by a decision tree, which has been shown to be $> 90\%$ accurate for objects brighter than $r = 20.0^m$ (Weir *et al.* 1995a; Odewahn *et al.* 1999). The data are calibrated into the Gunn system using an extensive collection of CCD imaging data obtained at the Palomar $60''$ telescope. Typical photometric errors for extended sources at in each band are $\sigma_g \sim .21^m, \sigma_r \sim .18^m, \sigma_i \sim .37^m$ at 19th magnitude (Odewahn *et al.* 1999).

Each field in each band is processed individually. The three resulting catalogs are cross–matched to create a composite list of objects for the field. We require a detection in both the J and F bands so that we can measure at least one color. N detections are not required because the N data are not as deep as the other two bands, and suffer from large plate sensitivity variations. In practice, however, the relatively bright r magnitude limit imposed to maintain accurate classification results in most objects ($\sim 98\%$) being detected in all three bands. Finally, those areas on the plate containing saturated objects are masked. These areas often contain large numbers of falsely identified galaxies, as the plate processing software handles large, bright objects improperly. The masked areas, on average, cover 7% of the plate. These areas are not used in cluster detection.

In this paper, we present results for two fields: 447 ($14^h30^m + 30°$) and 475 ($01^h + 25°$). These fields were chosen because they are at relatively high galactic latitude ($+67°$ and $+40°$) where the effects of dust are expected to be small (but see the discussion in section 5), and because scans in all three bands were available when this project was started. Additionally, we show a preliminary density map for 2500 square degrees at the North Galactic Pole.

2.2 Selecting Cluster Galaxies Using Colors

The morphology-density relation (Dressler 1980) has been observed in galaxy clusters at low redshift for nearly two decades. Late–type galaxies are dominant in the field population, whereas early–type galaxies are preferentially seen in high density regions. Therefore, any technique that can eliminate field (i.e. late–type) galaxies on the basis of some simple observable parameter will enhance the contrast of galaxy clusters relative to the background.

Dressler & Gunn (1992) presented photometric data for seven rich clusters. They show that cluster galaxies follow a well–defined sequence in the $(g-r) vs. (r-i)$ color-color space, whereas field galaxies occupy a much larger area in this space. By defining a locus of "cluster galaxies", and using only those objects which satisfy this color criterion, it is possible to preferentially discard field galaxies from the sample. Such a procedure was utilized effectively by Odewahn & Aldering (1995) in their cluster detection program. No other large, systematic cluster survey has been able to use color information for this purpose, as the data simply did not exist.

Figure 1 shows the $(g-r) vs. (r-i)$ diagram for galaxies which are matched in the three plate catalogs (JFN) for Field 475. We use only objects classified as galaxies on the J or F plate (whichever has better seeing), with calibrated magnitudes $r < 20.0^m$, where our classification accuracy is $> 90\%$ (Weir *et al.* 1995b; Odewahn *et al.* 1999). In practice, this imposes an approximate redshift limit of $z \sim 0.3$ on our cluster detection.

Initially, we defined a strict locus of cluster galaxies in the $(g-r) vs. (r-i)$ space, similar to the prescription given by Dressler & Gunn (1992). However, inspection of our color–color diagrams shows that the i photometry is poor, resulting in large scatter in the $r-i$ color. We have therefore chosen to make color cuts in the $g-r$ space only. The chosen cuts in $g-r$ have two motivations. On the blue side, $(g-r) < 0.3$, we are excluding blue field galaxies at low z. On the red side, $(g-r) > 1.3$, we are excluding objects that have a high likelihood of being either misclassified stars, or at high redshift. These color limits are marked in Figure 1. Clearly, a significant fraction ($\sim 50\%$) of galaxies in our catalog lie outside our chosen $g-r$ range. In addition, the k–corrected colors for both E and Scd galaxies are shown, for the redshift range $0.0 < z < 0.8$, for the Gunn filters used to calibrate the data. The differences between the actual plate filter/emulsion combination and the CCD filters used for calibration do not introduce any significant effects. The sharp turn in the k–correction curves occur at $z = 0.4$, where the 4000Å break passes from the g to the r band. Not surprisingly, we can see evidence for the locus of cluster galaxies along the k–correction track for early-type galaxies, albeit with large scatter. From this figure, it is clear that we are seeing galaxies out to $z = 0.4$, although we are very incomplete above $z = 0.25 - 0.3$. This is not surprising, as an $L_\star$ early-type galaxy will have $r \sim 20^m$ at $z \sim 0.3$ for any reasonable cosmology.

3 Galaxy Density Maps and Cluster Detection

3.1 Density and Significance Maps

Once a catalog of galaxies meeting our classification, magnitude, and color criteria has been created, we proceed to detecting candidate clusters. We have chosen to use the Adaptive Kernel (AK) technique (Silverman 1986). This technique uses a two–stage process to produce a density map. First, it produces a pilot estimate $f(t)$ of the galaxy density at each point in the map. Based on this estimate, it then applies a smoothing kernel whose size changes as a function of the local density, with a smaller kernel at higher density.

This technique has many advantages over methods used by other surveys. First, the two–step process significantly smooths the low density regions, without affecting the high density peaks. The smaller kernel size at high density means that multiple clusters in high density regions (e.g., filaments) are still separated. This technique is more robust than percolation or friends–of–friends algorithms, which may link clusters in regions of high galaxy density. Second, the AK does not require an assumed luminosity function or radial profile for the cluster. With our plate data, the range in magnitude covered is small ($16 \leq m_r \leq 20$), so that fitting a luminosity function is a precarious endeavor. More importantly, perhaps, visual inspection of our candidate clusters shows that many (perhaps most) do not appear symmetric, so that any radially symmetric profile, assumed by some techniques like the matched filter (Postman *et al.* 1996) would be a poor fit. We are attempting to avoid any bias towards rich, relaxed, evolved clusters (which may occur if clusters are selected from X–ray surveys). This results in a catalog in which some clusters will have different physical properties than those discovered by other techniques and at other wavelengths.

Our density maps are created using $1'$ pixels, resulting in a final map size of 360×360 pixels for each field. The kernel itself has a size comparable to the Abell radius of a cluster at $z \sim .15$, which is expected to be the median redshift of our cluster candidates. This combination of

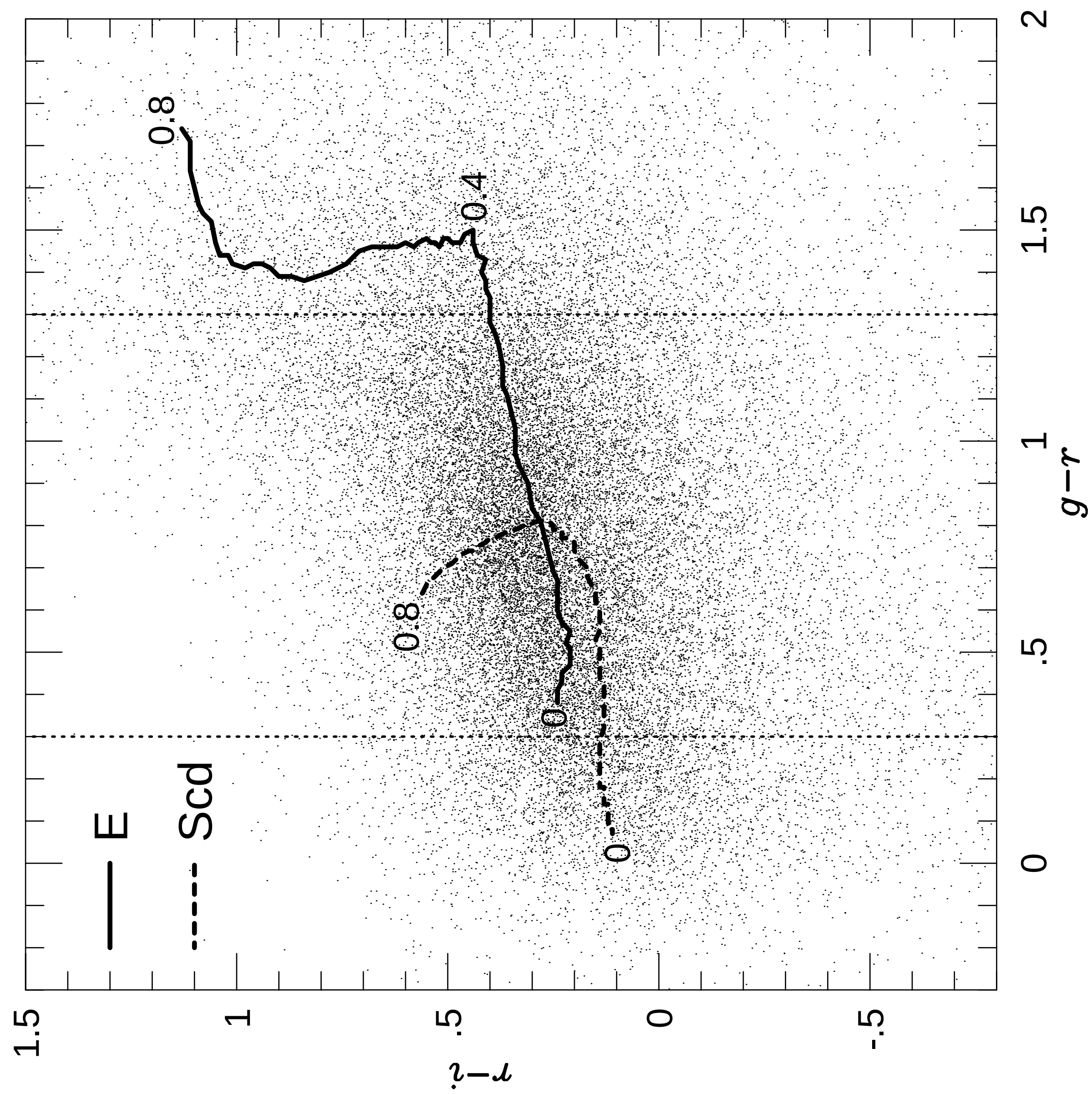

Figure 1: The $(g-r) vs.(r-i)$ diagram for galaxies in Field 475. The k–correction curve for E galaxies is shown as the solid line, and for Scd galaxies as the dashed line. The vertical dotted lines denote our color cuts.

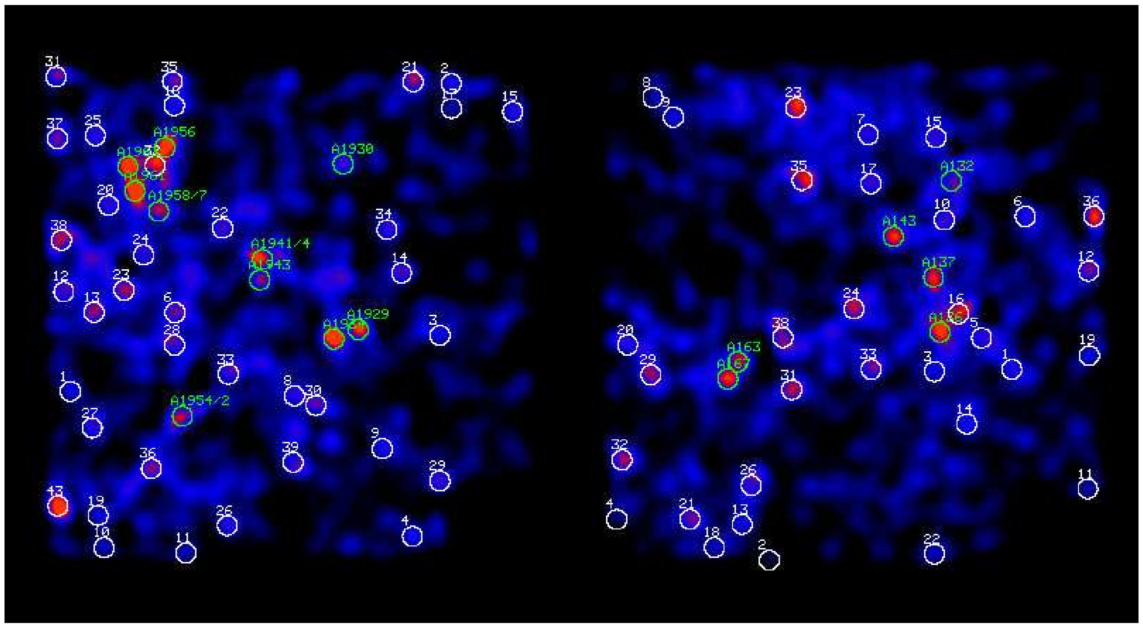

Figure 2: Density and significance maps for DPOSS Field 447. Abell clusters are marked in green, and new candidate clusters in white.

pixel and kernel size prevents us from over–resolving clusters into structural components, while maintaining a reasonable number of galaxies per pixel. Once a density map has been constructed, we use the *bootstrap* technique (Press *et al.* 1992) to build a significance map from 500 synthetic data sets. The density and significance maps for field 447 are shown in Figure 2. One can see that the structures seen in the significance map are also visible in the original density map.

We have also created a preliminary seamless catalog of galaxies over 2500 square degrees at the NGP. A density map of this region is shown in Figure 3. Both the Coma Cluster and the Corona Borealis Supercluster are visible in this map.

3.2 Selecting Cluster Candidates

The significance map constructed using the bootstrap technique is then used to identify peaks in the galaxy distribution, which we mark as candidate galaxy clusters. We utilize the FOCAS peak-finding algorithm (Jarvis & Tyson 1981) to detect enhancements with $\geq 3.0\sigma_F$ significance, where σ_F is the *rms* of the background in the significance map. In practice, this is a very liberal detection threshold, in the sense that we accept a larger number of false detections in return for both enhanced completeness at higher redshift and lower richness. We also impose a minimum area requirement of 66 pixels, which corresponds to a circular region of ~ 0.8 Mpc observed at $z = 0.15$. This yields 47 and 38 candidate clusters in fields 447 and 475, respectively. To quantify our false detection rate (which may be large at low significance), we have undertaken a spectroscopic follow–up campaign to measure redshifts for all the cluster candidates in these two fields, with results to be presented in a future paper.

The candidate clusters are marked in Figure 2, with Abell clusters shown in green, and new clusters in white. We successfully recover all known Abell clusters in these fields, and find a vastly larger number of new candidates. Clearly, Abell identified, in general, only the most significant density enhancements, although some rich clusters were still missed.

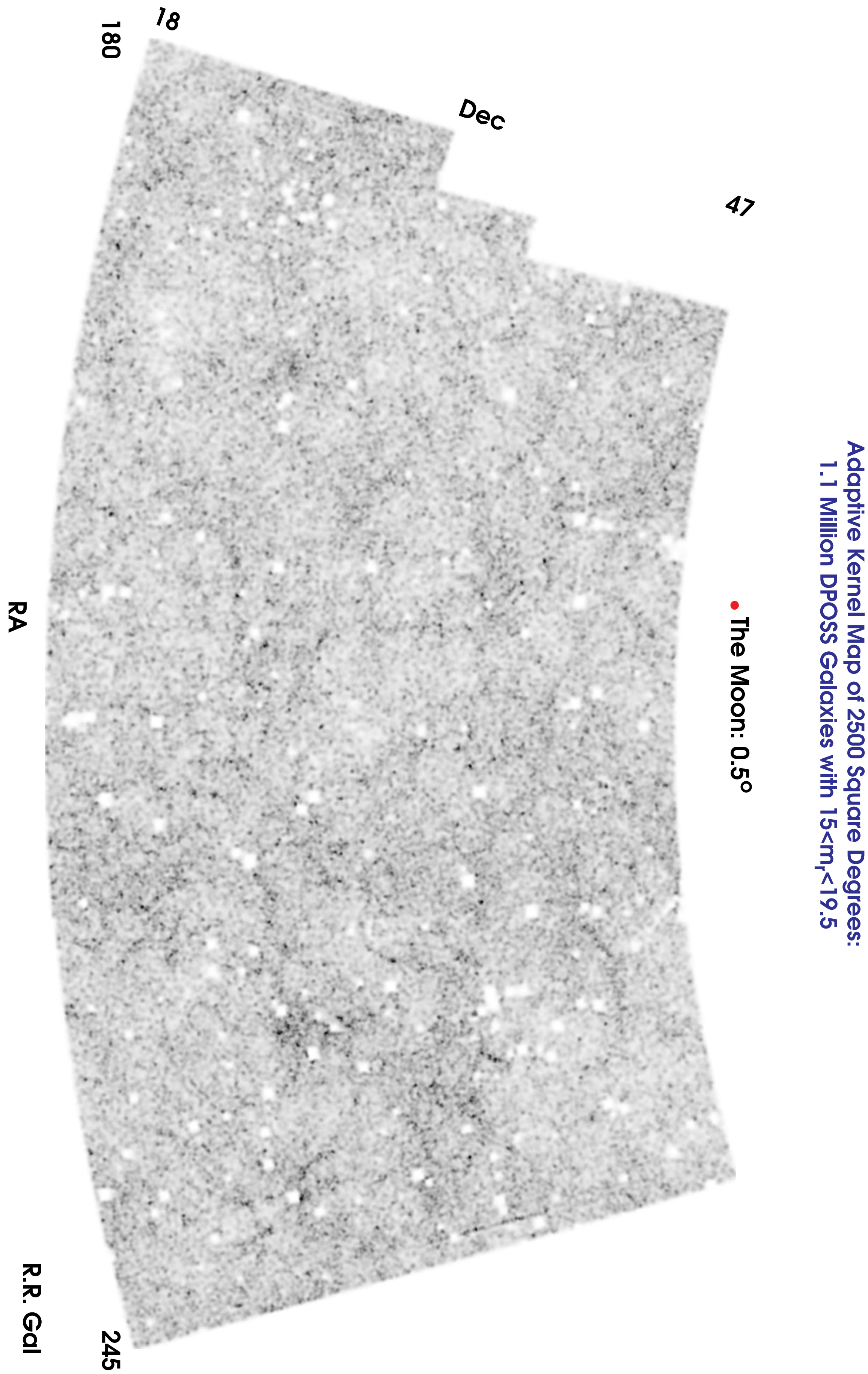

Figure 3: Density and significance maps for DPOSS Field 447. Abell clusters are marked in green, and new candidate clusters in white.

4 Estimating Cluster Properties

After detection from the significance map, some simple properties (such as size) of the clusters are measured from the original density map. Properties of the galaxy population of each cluster are derived from the input DPOSS galaxy catalogs. This helps to eliminate spurious detections, and allows us to compare our catalog with pre-existing catalogs. In this paper, we address properties measured directly from the plate data, while later papers will utilize additional CCD imaging and spectroscopic data obtained for the cluster candidates in these two fields for more detailed studies.

4.1 Photometric Redshift Estimation

From the plate data, we wish to measure the redshift and richness of each cluster candidate. First we estimate the redshift assuming that each cluster candidate contains a single cluster, at one redshift, and is dominated by early-type galaxies. We count the number of galaxies as a function of color, N_{g-r}, and the number as a function of r magnitude, N_r, inside the cluster area, as determined by FOCAS from our galaxy density maps. The background galaxy color and magnitude distributions, $N_{bg,g-r}$ and $N_{bg,r}$ are determined from a large (~ 146 square degrees) galaxy catalog around the North Galactic Pole, constructed from DPOSS data (Odewahn *et al.* 1999). This distribution, scaled to the appropriate area, is then subtracted from the color and magnitude distributions of each candidate cluster, and the median $g-r$ color and mean r magnitude of the remaining galaxies calculated.

This procedure was performed for 46 known Abell clusters with measured redshifts $z_{spect} > 0.05$, from 10 plates. However, in this case, we set the radius to the Abell radius ($R_{g-r} = 1.7'/z_{g-r}$) (Abell 1958) to measure the colors and magnitudes. We find that both the colors and magnitudes are equally well correlated with measured spectroscopic redshifts. We derive an empirical relation between redshift, median $g-r$ color, and mean r magnitude using a bivariate least–squares fit:

$$z_{phot} = 0.1074 \times r_{mean} + 0.1317 \times (g-r)_{med} - 1.9706 \tag{1}$$

The rms of $z_{spect} - z_{phot}$ is $\Delta z = 0.0165$. Figure 4 shows the photometrically estimated redshift against the spectroscopically measured redshift for all 46 clusters. Open circles show those clusters with only one galaxy with a spectroscopic redshift, and filled circles are clusters with two or more spectroscopic redshifts.

In practice, this must be an iterative procedure, because we do not initially know the redshift, and therefore the Abell radius, for our cluster candidates. Once a redshift is estimated, we then determine the Abell radius ($R_{phot} = 1.7'/z_{phot}$) for each candidate, and recenter on the median location of galaxies inside this radius. We then repeat the background correction and color and magnitude measurements inside R_{phot}, centered on the new location, and estimate the redshift anew. These steps are repeated until the redshift converges; $\Delta z_{phot} < 0.01$.

The complete iterative procedure was also performed on the 46 clusters with measured redshifts. As expected, starting the process with the incorrect Abell radius increases our errors. The final rms of $z_{spect} - z_{phot}$ is $\Delta z = 0.026$, with a mean offset of only -0.0016. This error includes all random and systematic errors, such as plate–to–plate photometry, classification, background correction, and even incorrectly measured spectroscopic redshifts. Approximately 50% of the 46 clusters used have only one galaxy with a measured redshift. In the future, as we gather more data, we will recalculate our photometric redshift estimator. Additionally, we have ignored clusters with $z_{spect} < 0.05$, as these are too extended for our technique to handle properly, and our photometric calibration for bright galaxies is currently too poor.

We find the small errors from this photometric redshift technique to be remarkable, particularly in light of the data we are using. The data is taken from 10 plates in various parts of

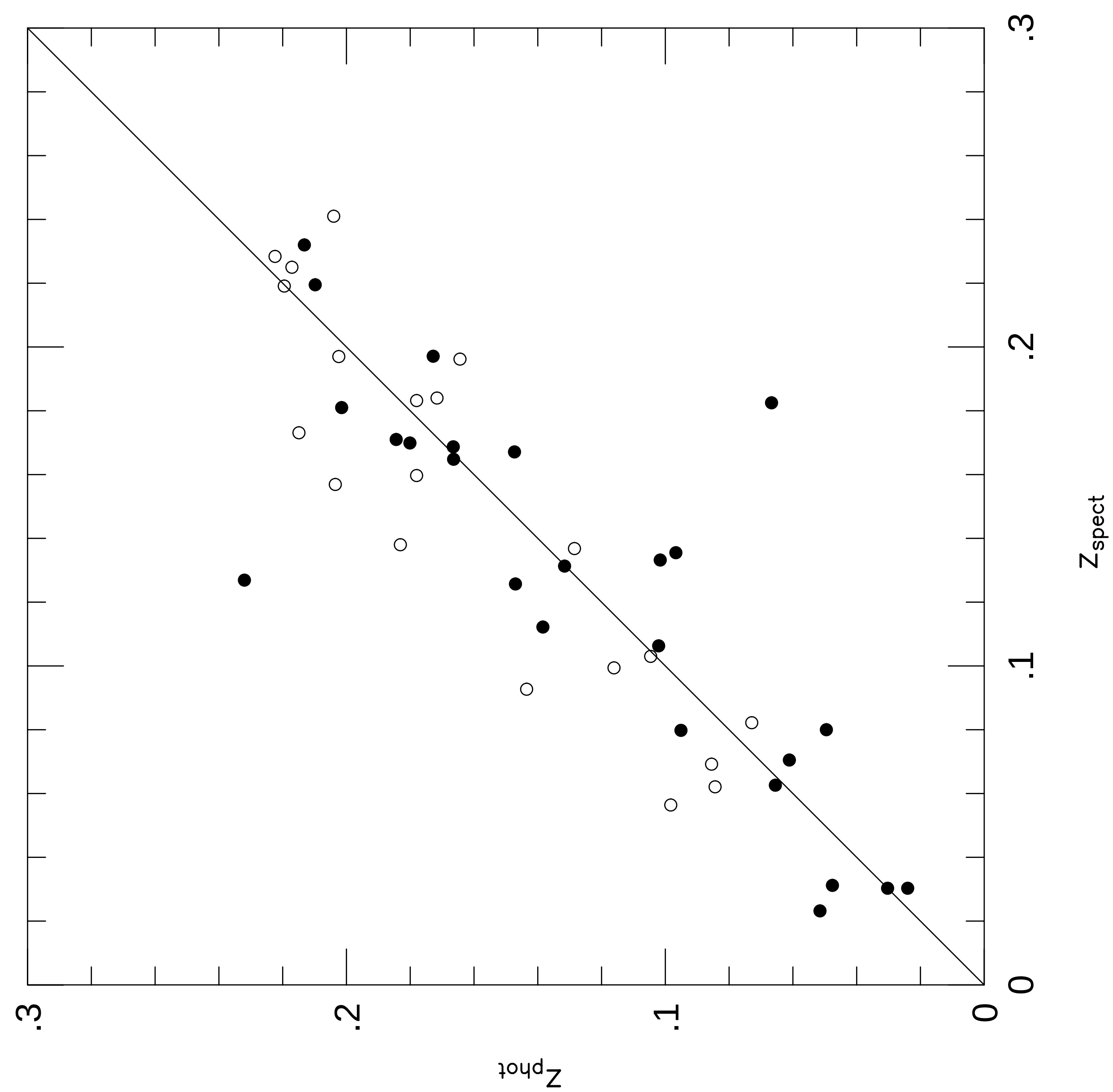

Figure 4: The photometrically estimated redshift against the spectroscopically measured redshift for all 46 clusters. Open circles show those clusters with only one galaxy with a spectroscopic redshift, and filled circles are clusters with two or more spectroscopic redshifts.

the sky, and therefore encompasses all of our photometric and classification errors, as well as varying amounts of extinction. The photometry is taken from photographic plates, where the quality is much lower than for CCDs. Additionally, the galaxy populations of these clusters must also vary; the fraction of bluer, late–type galaxies is not constant with richness or redshift. In comparison, the APM group, utilizing a richness-dependent, magnitude–based redshift estimator, achieved similar errors over only a much smaller redshift range ($0.04 < z < 0.1$, Dalton *et al.* 1994).

4.2 Extinction

An important consideration in the construction of an all–sky cluster catalog is extinction from dust in our Galaxy. Because the two fields considered in this paper are at high galactic latitude, the extinction correction is negligible. The mean extinction $\mathrm{E}(g-r)$ derived from the Schlegel *et al.* (1998) maps is only 0.01^m and 0.05^m for fields 447 and 475, respectively. Nevertheless, our final all-sky catalog will use extinction corrected magnitudes and colors to derive cluster properties. It is clear that galaxies are extincted out of our detection threshold, creating false large-scale structure. This cannot be avoided; it is impossible to correct the color of a galaxy that is not detected. This effect has been largely ignored by past cluster surveys, which only corrected magnitudes of detected galaxies , or were restricted to areas of low extinction. (Maddox, Efstathiou & Sutherland 1996). This can have dire consequences for measurements of large scale structure, introducing false power on all scales. Clusters are detected where the galactic extinction is low, and lost where the extinction is high. This non–uniform detection threshold across the sky must be accounted for when measuring the cluster–cluster correlation function, for instance.

5 Discussion and Conclusion

We have presented a simple but robust technique for generating an objective catalogue of galaxy clusters to $z \sim 0.3$ over the entire high–galactic–latitude Northern sky, using DPOSS data. With the multicolor nature of our data, and our demonstrated ability to measure redshifts photometrically, many scientific problems can be addressed. Future papers will present cluster catalogs over large areas of the sky, stitched together seamlessly from individual field catalogs (thus avoiding plate edge effects). Fundamental astrophysical questions will be addressed, including the cluster–cluster correlation function and the evolution of galaxy populations in clusters. Our observed cluster density, ~ 1 per square degree, is consistent with the detection of richness class 0 and higher clusters to $z \sim 0.3$. It is approximately 50% higher than the space density of clusters found by APM (Dalton *et al.* 1992), who only reached $z \sim 0.1$.

Additionally, we have undertaken an extensive follow-up campaign, including deep ($r_{lim} = 23.0^m$) 3-band (gri) CCD imaging at the Palomar $60''$ telescope, and multislit spectroscopy at the Palomar $200''$ Hale telescope. Every candidate in our two test fields will have both follow-up imaging and spectroscopy, with results to be presented in later papers. This data will allow us to provide an unprecedented statistical understanding of our cluster catalog. The spectroscopy will provide redshifts of ~ 80 candidate clusters in the two fields described in this paper. We will be able to directly measure our false detection rate and the redshift distribution of our clusters. The CCD imaging will provide detailed constraints on our photometric and classification errors, in addition to enabling studies of cluster luminosity functions, galaxy populations, etc. The data may also be used in searches for clusters at higher redshift.

Finally, we note that the next large–area galaxy cluster catalog will not be produced for a number of years. Until then, we hope that this modern catalog of galaxy clusters will be useful in broad–ranging, multi–wavelength studies by ourselves and other investigators. Upcoming and recent space missions, such as CXF and XMM, will also benefit greatly from this new resource.

References

1. Abell, G. O. 1958, ApJS, 3, 211
2. Abell, G. O., Corwin, H. G. & Olowin, R. P. 1989, ApJS, 70, 1
3. Bahcall, N. A. & Soneira, R. M. 1983, ApJ, 270, 20
4. Bahcall, N. A. & West, M. J. 1992, ApJ, 392, 419
5. Butcher, H. R. & Oemler, A. 1978, ApJ, 226, 559
6. Dalton, G. B., Efstathiou, G., Maddox, S. J. & Sutherland, W. J., 1992, ApJ, 390, L1
7. Dalton, G. B., Efstathiou, G., Maddox, S. J. & Sutherland, W. J., 1994 MNRAS, 269, 151
8. Djorgovski, S. G., Gal, R. R., Odewahn, S. C., DeCarvalho, R. R., Brunner, R., Longo, G. & Scaramella, R. 1999, in *Wide Field Surveys in Cosmology*, S. Colombi *et al.* ,eds., p. 89
9. Dressler, A. 1980, ApJ, 236, 351
10. Dressler, A. & Gunn, J.E. 1992, ApJ, 78, 1
11. Jarvis, J. F. & Tyson, J. A. 1981, AJ, 86, 476
12. Kennefick, J. 1996, *The Luminosity Function of Quasars at z¿4*, (Ph.D. Thesis)
13. Lasker, B. *et al.* 1996, *Astronomical Data Analysis Software and Systems V*, A.S.P. Conference Series, Vol. 101, 1996, George H. Jacoby and Jeannette Barnes, eds., p. 88.
14. Lumsden, S. L., Nichol, R. C., Collins, C. A. & Guzzo, L. 1992, MNRAS, 258, 1
15. Maddox, S. J., Efsthatiou, G. & Sutherland, W. J. 1996, MNRAS, 283, 1227
16. Odewahn, S. C. & Aldering, G. 1995, AJ, 110, 2009
17. Odewahn, S. C. *et al.* 1999, in preparation
18. Postman, M. *et al.* 1996, AJ, 111, 615
19. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. 1992, *Numerical recipes in FORTRAN, 2d. ed.*, (Cambridge: University Press)
20. Reid, I. N. *et al.* , 1991, PASP, 103, 661
21. Schlegel, D., Finkbeiner, D., & Davis, M. 1998, ApJ, 500, 525
22. Silverman, B.W. 1986, *Density Estimation for Statistics and Data Analysis*, (London: Chapman & Hall)
23. Struble, M. F. & Rood, H. J. 1991, ApJS, 77, 363
24. Weir, N., Fayyad, U., Djorgovski, S. 1995, AJ, 109, 2401
25. Weir, N., Djorgovski, S., Fayyad, U. 1995, AJ, 110, 1
26. Weir, N., Fayyad, U., Djorgovski, S. & Roden, J. 1995, PASP, 107, 1243
27. Worthey, G., 1994, ApJS, 95, 107