Ministério da Ciência e Tecnologia
Observatório Nacional

SELEÇÃO DE AGLOMERADOS DE GALÁXIAS A PARTIR DO SEGUNDO LEVANTAMENTO FOTOGRÁFICO DO OBSERVATÓRIO PALOMAR

Paulo Afrânio Augusto Lopes

Tese de doutorado
Orientador: Dr. Reinaldo Ramos de Carvalho

Apresentada como requisito para a obtenção
do grau de doutor em astrofísica
Rio de Janeiro - Junho de 2003

Para ver as meninas

Silêncio por favor
Enquanto esqueço um pouco
A dor do peito
Não diga nada
Sobre meus defeitos
Eu não me lembro mais
Quem me deixou assim
Hoje eu quero apenas
Uma pausa de mil compassos
Para ver as meninas
E nada mais nos braços
Só este amor
Assim descontraído
Quem sabe de tudo não fale
Quem não sabe nada se cale
Se for preciso eu repito
Porque hoje eu vou fazer
Ao meu jeito eu vou fazer
Um samba sobre o infinito

Paulinho da Viola

Agradecimentos

Gostaria de agradecer ao Dr. Reinaldo de Carvalho pelo apoio em todos os momentos ao longo dos últimos quatro anos. Apesar da distância, o Reinaldo sempre esteve disponível para dirimir dúvidas ou discutir novas questões relativas ao trabalho. Mais que isso, sempre mostrou entusiasmo com novos resultados e perspectivas. Durante minha estada em Caltech tive o prazer de trabalhar com o Dr. Djorgovski. Agradeço ao George pelo suporte nos mais de dois anos em que estive em Caltech. Pela ajuda com questões científicas, em resolver questões burocráticas e também pelo apoio financeiro.

Muito eu aprendi com os outros membros do DPOSS, principalmente com Roy Gal, que mostrou-se de incrível valia tanto para apontar os atalhos do céu do hemisfério norte como os atalhos de Pasadena e LA. Em inúmeras oportunidades tive o prazer de discutir projetos com Ashish Mahabal, S. Odewahn, R. Brunner e mais recentemente com J. Kohl. Ashish esteve sempre disponível para esclarecer dúvidas de programação, principalmente em *perl*.

Agradeco ao Dr. Hugo Capelato por nos ceder o programa do *kernel* adaptativo, utilizado nesta tese. Para detectarmos aglomerados distantes, além do *kernel* adaptativo utilizamos o programa de *voronoi tessclations*. Agradeço aos Drs. Massimo Ramella e Walter Boschin por cederem este programa e ajudarem na implementação do mesmo para os dados do DPOSS. Catarina Lobo foi muito gentil ao tornar disponível o programa conhecido como *matched filter*. Agradeço sua ajuda para testá-lo nos nossos dados e espero poder fazer melhor uso do mesmo no futuro. Infelizmente não tive tempo de empregar este programa nos estudos desenvolvidos nesta tese.

Quero agradecer a todos meus companheiros de pós-graduação no Observatório Nacional pelo excelente convívio nos últimos anos. À exceção do calor, o ambiente no ON é formidável, sendo as paradas para o café extremamente divertidas. Os pós-doutores, pesquisadores e alunos no terceiro andar que o digam ... Gostaria de agradecer à *Tia* Simone por sempre estar lá quando precisei.

Os dois anos em Caltech foram incrivelmente prazerosos. Infelizmente, no final de minha estada, minha figura tornou-se menos presente nas festas e bares, pois estava sendo sugado pelo meu computador ... Agradeço em especial ao Stan e Sandra por sempre tentarem me

arrastar para uma festa ou uma cerveja no *Amigos* ... As *TTs* não saem da memória ...

Agradeço à minha família pelo apoio em todos os momentos, em especial ao meu pai e minha mãe por toda a ajuda em Niterói e em Copa. Não sei se conseguiria chegar aonde cheguei sem seu apoio.

À ma princesse, merci pour tout. Tu es plus que super mon amour!

Resumo

Nesta tese abordamos a questão de seleção de aglomerados de galáxias com base nos dados digitalizados do segundo levantamento fotográfico do Observatório Palomar. Numa primeira etapa apresentamos o *Northern Sky Optical Cluster Survey* (NoSOCS), um catálogo de candidatos a aglomerados de galáxias cobrindo $\sim$ 11500 graus quadrados do hemisfério norte. Este é baseado nos catálogos de galáxias do DPOSS, completos até $r = 19.5$. A seleção de aglomerados é baseada num método de *kernel adaptativo* para geração de mapas de densidade para todos os campos do DPOSS. O programa SExtractor é utilizado para detecção de picos de densidade nestes mapas. O catálogo final contém $\sim$ 15000 candidatos, sendo $< z > \sim 0.16$ e $< N_{gals} > \sim 30$. Aglomerados ricos são detectados até $z \sim 0.3$.

Numa segunda etapa investigamos a possibilidade de estender o NoSOCS até $z \sim 0.5$. Para tal, utilizamos catálogos de galáxias até $r = 21.1$ e geramos um catálogo suplementar ao NoSOCS. Este novo catálogo é baseado em duas técnicas de detecção, o *kernel adaptativo* e o método conhecido como *voronoi tesselation*. O catálogo final contém $\sim$ 10000 candidatos numa área de $\sim$ 2700 graus quadrados, sendo $< z > \sim 0.30$ e $< N_{gals} > \sim 40$. Aglomerados ricos são detectados até $z \sim 0.5$.

Para ambos os catálogos, realizamos grande número de simulações para optimizarmos os algoritmos de detecção e estimarmos as taxas de contaminação e completeza. Estimativas de desvio para o vermelho e riqueza são fornecidas para os candidatos dos dois catálogos. Um levantamento espectroscópico foi realizado para candidatos do NoSOCS. Estes dados são utilizados para confirmar as estimativas fotométricas e as expectativas de contaminação provenientes das simulações.

Como subproduto, apresentamos uma comparação do DPOSS com o SDSS. Esta comparação é utilizada para estimarmos os limites de detecção do DPOSS. Alguns dos projetos a serem desenvolvidos no futuro são também discutidos.

Abstract

In this thesis we discuss the selection of galaxy cluster candidates based on the Digitized Second Palomar Observatory Sky Survey (DPOSS). First we present the Northern Sky Optical Cluster Survey (NoSOCS), a galaxy cluster catalog spanning $\sim$ 11500 square degrees of the northern sky. This catalog is based on the galaxy catalogs from DPOSS, which are complete down to $r = 19.5$. The cluster selection is based on a adaptative kernel technique to generate density maps for all DPOSS fields, with the SExtractor object detection algorithm being used to detect overdensities in these maps. The final catalog comprises $\sim$ 15000 cluster candidates, with $< z > \sim 0.16$ and $< N_{gals} > \sim 30$. Rich clusters are detected down to $z \sim 0.3$.

Later we investigate the possibility to extend the NoSOCS catalog out to $z \sim 0.5$. In order to generate a supplemental catalog to the NoSOCS we use galaxy catalogs down to $r = 21.1$. This new catalog is based on two techniques, the adaptive kernel and the voronoi tesselation method. The final catalog contains $\sim$ 10000 candidates over $\sim$ 2700 square degrees, with $< z > \sim 0.30$ and $< N_{gals} > \sim 40$. Rich clusters are detected down to $z \sim 0.5$.

In order to optimize the detection algorithms and estimate the contamination and completeness rates, we perform a large number of simulations for both catalogs. Redshift and richness estimates are also provided for all candidates in the two catalogs. Finally, an extensive spectroscopic survey of NoSOCS candidates is used to confirm the photometric estimates and the contamination rates derived from the simulations.

As a by-product we present a comparison between the DPOSS and SDSS surveys. This comparison is used to estimate the DPOSS detection limits. Some of the projects to be developed in the future are also discussed.

Índice

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

1.1 Uma Breve Revisão

Cosmologia é o estudo de modelos de universos. O termo tem origem grega (*kosmología*), significando o tratado das leis que regem o universo. Do ponto de vista da ciência moderna, cosmologia é o estudo da estrutura em grande escala do universo. Nossa definição e noção do universo evoluiu consideralvelmente ao longo da história, sendo nosso objeto de estudo cada vez mais ampliado, dos planetas no sistema solar ao estudo das estrelas em nossa Galáxia, até que finalmente nossos olhos voltaram-se para o universo extragaláctico. No século XVIII I. Kant já discutia a idéia de que algumas nebulosas vistas no céu eram objetos externos a nossa Galáxia. Naquela época já se imaginava que estas nebulosas constituiam sistemas estelares similares ao nosso. A Galáxia seria somente mais uma entre bilhões a povoar o universo. No entanto, até o início do século passado esta hipótese não tinha grande aceitação. Em realidade, existia um grande debate em curso, sendo uma das principais questões a natureza de nebulosas brilhantes observadas na época. H. Shapley, juntamente com muitos outros astrônomos, defendia que estas nebulosas eram parte de nossa própria Galáxia. H. Curtis tinha uma idéia oposta, sugerindo que estes sistemas encontravam-se além dos limites de nossa Galáxia. Observações realizadas por E. Hubble trouxeram luz a este problema. Em meados da década de 1920 Hubble observou estrelas variáveis em algumas destas nebulosas brilhantes, demonstrando que estas encontram-se a distâncias muito maiores do que previamente suposto, tornando possível a identificação destes sistemas como objetos extragaláticos. A cosmologia moderna ocupa-se quase que inteiramente do mundo extragaláctico, visando ao estudo da estrutura em grande escala, desde sua origem e

evolução até a época atual.

No trabalho aqui apresentado aglomerados de galáxias são nosso objeto de estudo. Estes são compostos por um grande número de galáxias de diferentes morfologias e características físicas. No que diz respeito a morfologia, galáxias podem ser divididas em quatro tipos básicos: elípticas, espirais, um tipo intermediário denominado S0 e galáxias irregulares. Em 1926, Hubble propôs um esquema de classificação de galáxias constituído destes quatro tipos. Galáxias elípticas (também conhecidas como *early-type*) possuem uma forma elipsoidal, baixa rotação e pequena quantidade de gás. São classificadas como E0, E1, ..., E7. Sendo a e b os eixos maior e menor da elipse, o índice n de uma galáxia En é dado por $n = 10(a-b)/a$. Galáxias espirais (conhecidas como *late-type*) são constituídas de duas componentes: uma região central (bojo), e um disco espiral. A componente disco apresenta um movimento de rotação e uma grande quantidade de gás, ao passo que o bojo tem propriedades similares a galáxias elípticas. De forma simplificada, estas galáxias são classificadas como Sa, Sb, Sc. Galáxias Sa são conhecidas como espirais *early-type*, enquanto que Sc são conhecidas como espirais *late-type*. A sequência está relacionada a razão de luminosidade entre as duas componentes (disco/bojo $\equiv$ D/B). A razão D/B é pequena para galáxias Sa, crescendo para galáxias Sc. Algumas espirais apresentam uma componente barra luminosa na região central, sendo denominadas espirais barradas e classificadas como SBa, SBb, SBc, etc. Galáxias lenticulares ou S0 têm uma razão D/B próxima de 1.0, mas não apresentam estrutura espiral, além de conterem pouco gás. Galáxias desprovidas de uma forma característica são classificadas como irregulares (Irr). Na Figura 1.1 apresentamos alguns exemplos de diferentes tipos de galáxias.

Observações de galáxias em diferentes bandas fotométricas tornam possível determinar a contribuição da distribuição de energia espectral (DEE) em diferentes intervalos de comprimento de onda. A diferença de magnitudes, ou a razão de fluxos nos correspondentes comprimentos de onda, entre duas bandas fotométricas fornece o indíce de cor ou simplesmente cor de um dado objeto. O tipo morfológico de galáxias é correlacionado com a cor. Galáxias Sc, por exemplo, apresentam uma fração de estrelas de alta massa da sequência principal maior do que galáxias Sa, o que torna galáxias Sc mais azuis do que Sa. As galáxias mais azuis da sequência de Hubble são irregulares. Galáxias elípticas são caracterizadas por uma população estelar mais velha. O espectro destas galáxias é dominado por características

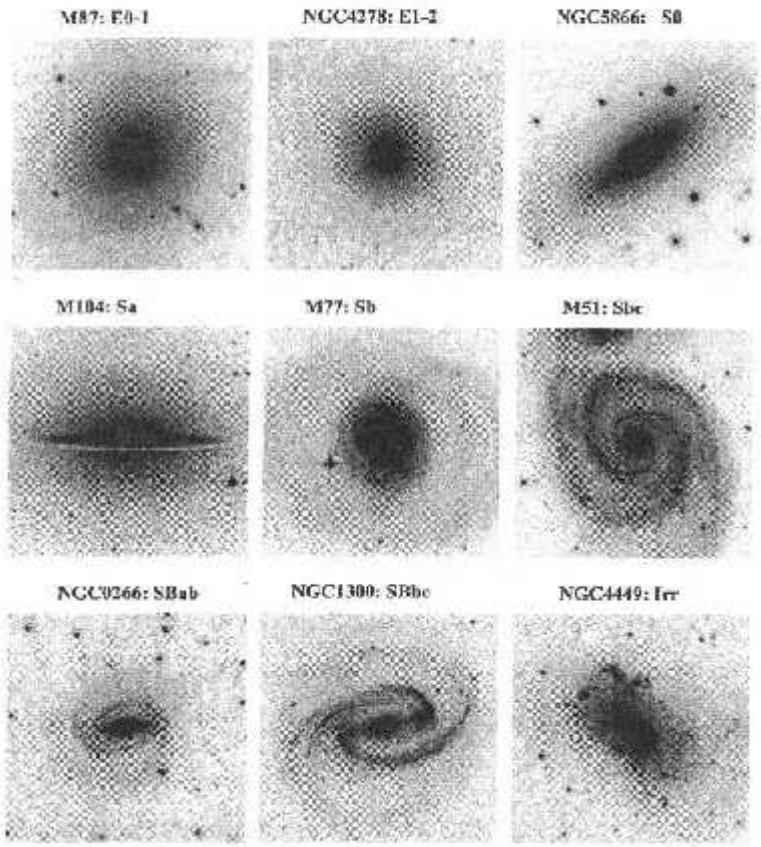


Figura 1.1: Exemplos de galáxias próximas com diferentes morfologias. Todas as imagens foram extraídas do *Nasa/IPAC Extragalactic Database* (NED).

de estrelas KOIII, o que justifica as cores mais vermelhas destes sistemas quando comparados com galáxias *early-type*.

A localização destes sistemas no universo também é bastante distinta. Galáxias elípticas tendem a se concentrar mais do que espirais, sendo a população de aglomerados de galáxias no universo local predominantemente composta por elípticas e S0. Em aglomerados mais distantes este efeito ainda é observado, no entanto o grau de concentração de elípticas é reduzido, sendo a fração de galáxias espirais (azuis) maior do que no universo local. O aumento da fração de galáxias azuis em aglomerados com o desvio para o vermelho é conhecido como efeito Butcher-Oemler (BO) (Butcher & Oemler 1978, 1984; Margoniner *et al.*

2001; Ellingson *et al.* 2001).

A origem das galáxias e da estrutura em grande escala do universo constitui um dos tópicos mais excitantes nos tempos atuais. Não somente pelo grande acúmulo de dados de alta qualidade, mas também pelo rápido progresso feito no campo de simulações cosmológicas nas últimas décadas.

No campo teórico a base da cosmologia moderna está na Teoria da Relatividade Geral (TRG) proposta por Einstein em 1915. A interação gravitacional é a única interação conhecida por ser de longo alcance e que não pode ser compensada (não existem massas negativas). Portanto, para grandes quantidades de matéria distribuídas em grandes regiões do espaço a força gravitacional representa a única interação relevante, sendo assim decisiva na determinação da evolução e dinâmica do universo. O aspecto mais original da TRG está em considerar a gravidade como uma propriedade intrínseca ao espaço-tempo. A teoria de Einstein pode portanto ser considerada uma *geometrização* da força gravitacional. O modelo mais simples do universo faz uso do *Princípio Cosmológico*, que assume homogeneidade e isotropia em grandes escalas. A métrica de Robertson-Walker fornece a descrição mais geral das propriedades geométricas de um universo em que o *Princípio Cosmológico* é válido. Esta métrica depende somente do fator de escala universal $a(t)$ e da curvatura do espaço-tempo (K) (Stephani 1990).

Em 1917 Einstein propôs um modelo isotrópico, homogêneo e estático para o universo[1]. No mesmo ano de Sitter propôs um modelo que ao contrário do universo de Einstein, era desprovido de matéria e não era estático. A ausência de matéria no universo de de Sitter gerava desconforto, no entanto a hipótese de um universo em expansão vinha ao encontro das observações de deslocamentos nas linhas espectrais de nebulosas, realizadas por Slipher na época (Slipher 1917). A interpretação destes deslocamentos como resultado da expansão do universo teve maior aceitação após inúmeras observações de nebulosas realizadas por Hubble (1929). Os modelos de Friedmann (1922-1924) combinavam as noções de expansão (de Sitter) e presença de matéria (Einstein). Estes modelos foram propostos alguns anos antes da descoberta de Hubble, e podem ser especificados com base em dois parâmetros $H(t) = \dot{a}/a$ e $\Omega(t) = \rho/\rho_c$ (onde o ponto representa a derivada com respeito ao tempo

[1]Em realidade as equações de Einstein não reproduziam um universo estático, sendo então modificadas pela adição da *constante cosmológica* (Λ) para tal.

próprio cosmológico t, e a densidade crítica é definida por $\rho_c = 3H(t)^2/8\pi G)$, conhecidos como parâmetros de Hubble e de densidade[2], respectivamente.

Após a descoberta da expansão, as questões mais importantes residiam na origem, evolução até a data presente, e futuro do universo. O modelo do *Big Bang* fornece uma explicação para origem do universo, segundo a qual o universo teria começado numa singularidade. Em outras palavras, o fator de escala $a(t)$ tenderia a zero quando $t \to 0$. Na década de 40 G. Gamow tentou descrever a origem da maioria dos elementos conhecidos com base no modelo do *Big Bang*. A formação dos núcleos seria baseada em fusão nuclear, num processo similar ao que ocorre em estrelas. As condições iniciais para tal processo são de densidade e temperatura muito altas (da ordem de 10^{10} K $1s$ após o *Big Bang*, Gamow 1946). Nesta época a densidade de energia de radiação seria muito maior do que a de matéria e o universo seria dominado por radiação. As expectativas de Gamow de que a maioria dos núcleos pudessem ser sintetizados no universo primordial foram confirmadas somente para elementos leves, como deutério e ^{4}He. Núcleos mais pesados seriam formados em estrelas. No entanto, o trabalho de Gamow teve outras previsões importantes. Em 1948 Gamow, em colaboração com R. Alpher e R. Herman, previu que os fótons extremamente quentes da era primordial teriam esfriado com a expansão do universo, dando origem na época atual a uma radiação de fundo na região de micro-ondas. Alpher & Herman (1948) estimaram que a temperatura atual desta radiação seria $\sim$ 5 K. Em 1965, como previsto com aproximadamente 20 anos de antecedência, este fato foi confirmado por Penzias & Wilson que detectaram uma radiação de fundo aproximadamente constante na frequência de micro-ondas. Assumindo que esta radiação representa uma radiação de corpo-negro, Penzias & Wilson determinaram a temperatura desta radiação de fundo como sendo $\sim$ 3.3 K. Observações do satélite COBE (*Cosmic Background Explorer*) indicam uma temperatura de $\sim$ 2.73 K (Mather *et al.* 1990). Apesar da radiação cósmica de fundo ser extremamente uniforme, o universo local apresenta

[2]O valor do parâmetro de densidade é fundamental para a determinação do destino do universo. Se $\Omega_0 \leq 1$ o universo expandirá para sempre, se for maior que 1 ele colapsará sobre si mesmo em algum momento (o índice zero refere-se a estimativa do parâmetro de densidade na época atual). Podemos definir o parâmetro total de densidade como $\Omega = \Omega_m + \Omega_\Lambda + \Omega_K$, onde o primeiro termo refere-se a contribuição de matéria (tanto bariônica como nâo-bariônica), o segundo está associado a constante cosmológica Λ e o último termo refere-se a densidade de energia associada a curvatura do universo (K = 0 para um universo de geometria plana).

grandes flutuações na distribuição de galáxias (vide Figura 1.2), sendo a origem e evolução destas estruturas áreas de grande interesse na cosmologia moderna. Desde a época de recombinação ($z \sim 1000$) até a época presente flutuações de temperatura na distribuição da radiação de fundo evoluiram sem interagir com a matéria. Portanto, detecções de flutuações de temperatura na radiação cósmica de fundo observada atualmente[3] carregam informação das flutuações de temperatura em $z \sim 1000$, e em consequência das flutuações de densidade de matéria nesta época. O conhecimento de $\delta\rho/\rho$ na época de recombinação permite-nos estimar como a matéria deveria estar aglomerada atualmente.

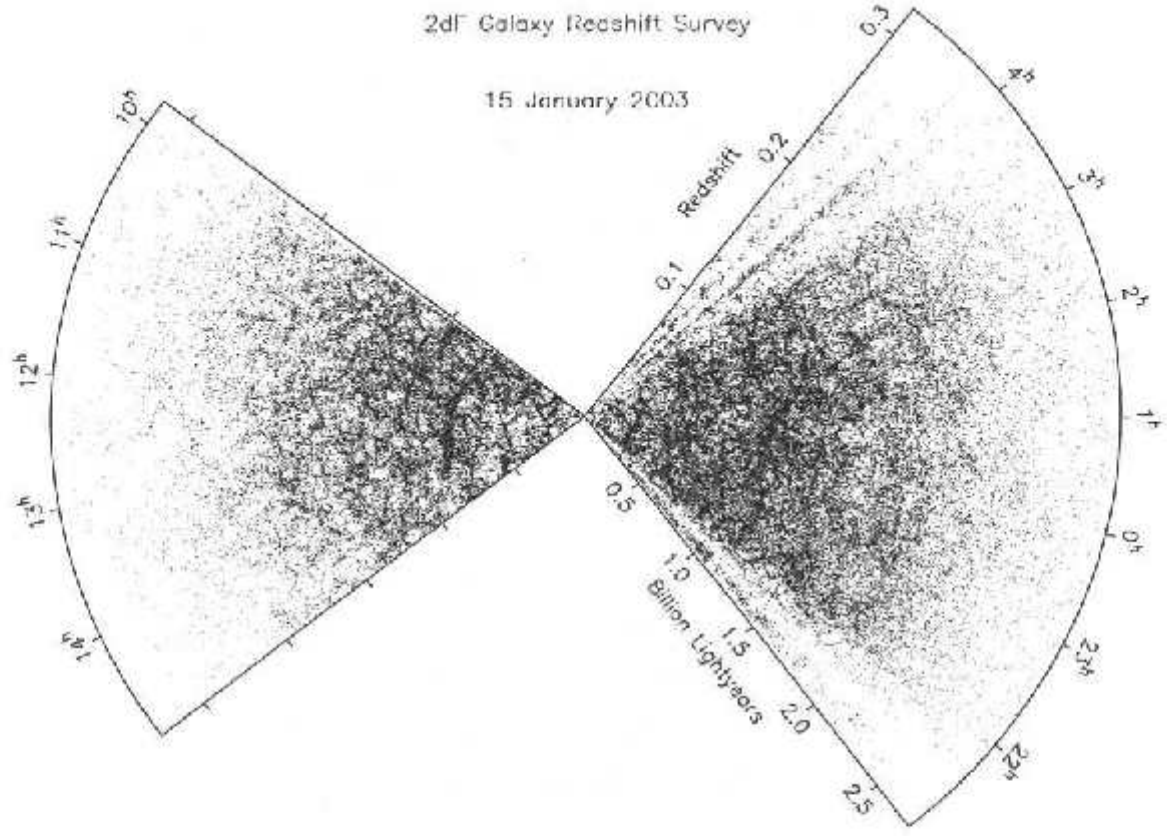


Figura 1.2: Mapa de distribuição de galáxias segundo o *Two-Degree Field Galaxy Redshift Survey*. Este levantamento apresenta espectroscopia para ~ 250000 galáxias abrangendo ~ 1700 graus quadrados do hemisfério sul, tendo uma magnitude limite $B = 19.7$, o que equivale a $< z >= 0.1$ (Colless *et al.* 1999).

[3]as primeiras detecções de anisotropia na radiação cósmica de fundo foram realizadas com dados do satélite COBE em 1992, apontando para $\Delta T/T \sim 10^{-5}$ (Smoot *et al.* 1992).

O modelo mais aceito para explicar a origem de galáxias e da estrutura em grande escala tem como base instabilidades gravitacionais, tendo sido proposto por *Sir* James Jeans no início do século passado (Jeans 1902). Esta teoria considera que pequenas flutuações de densidade no universo primordial teriam crescido com o tempo. Este efeito pode ocorrer se a força associada a pressão for desprezível em comparação a força gravitacional na região em questão. Neste caso, devido ao acúmulo de material de sua vizinhança, esta região ficará cada vez mais densa. O efeito final pode ser o colapso desta flutuação num sistema gravitacionalmente ligado.

Dois cenários foram inicialmente propostos para explicar a formação de estruturas. O primeiro é um modelo hierárquico (conhecido como *bottom-up* ou isotérmico) segundo o qual estruturas de baixa massa colapsam e se fundem, sendo a atração gravitacional responsável pela formação de objetos cada vez maiores. A segunda hipótese é baseada num modelo adiabático (conhecido como *top-down*) segundo o qual as primeiras estruturas formadas estão na escala de aglomerados e super-aglomerados de galáxias, e as estruturas menores são resultado da fragmentação das primeiras. No início dos anos 80, estas hipóteses saíram de cena, sendo substituídas por modelos que assumiam a existência (e predominância no universo) de matéria escura não-bariônica, podendo ser quente ou fria. Os dois principais modelos são baseados em matéria escura fria ou quente (CDM e HDM são as siglas em inglês para *Cold Dark Matter* e *Hot Dark Matter*). Em ambos os modelos as perturbações iniciais são adiabáticas, a diferença ficando no tipo de matéria escura admitida. Assumindo que $\Omega = 1$, CDM é capaz de reproduzir as propriedades de aglomeração na escala de galáxias e aglomerados de galáxias, sendo incapaz de gerar a estrutura em grande escala observada nos levantamentos mais recentes (Klypin *et al.* 1997; Brodbeck *et al.* 1998). O modelo HDM tem o problema inverso, ou seja, em reproduzir as propriedades de aglomeração na escala de galáxias. A estrutura em grande escala é bem reproduzida por este modelo, mas as estruturas levam muito tempo para serem formadas, o que representa um problema para justificar a existência de objetos com grandes desvios para o vermelho (Schaeffer & Silk 1988). Outros modelos consideram casos intermediários em que as partículas não são nem muito quentes nem muito frias (são os chamados modelos WDM, *Warm Dark matter* em inglês) (Schaeffer & Silk 1988). Outra possibilidade é a utilização de um modelo híbrido, misturando partículas quentes e frias. Estes modelos são normalmente denominados CHDM ou MDM (Davis *et al.*

1992) (*Mixed Dark Matter* em inglês).

O modelo do *Big-bang* encontra problemas para explicar o surgimento de pequenas não-homogeneidades no universo primordial. Uma solução para esta questão foi dada por A. Guth em 1981. A teoria da inflação, por ele proposta, prevê que nos seus primeiros instantes o universo teria sofrido uma grande expansão num período extremamente curto. Duas das principais previsões do modelo de inflação são: (i) o parâmetro de densidade seria igual a 1; (ii) o período inflacionário daria origem a pequenas flutuações de densidade, flutuações estas que seriam o ponto de partida para as estruturas observadas atualmente. As observações mais recentes apontam para um universo de baixa densidade, o que representaria um problema para um modelo de inflação que não assumisse a existência de uma constante cosmológica.

Existem diversas formas de testar modelos cosmológicos e tentar estabelecer limites para os principais parâmetros destes modelos. Algumas ferramentas que podem contribuir para este objetivo são: (i) observações de galáxias e aglomerados de galáxias; (ii) a determinação da anisotropia da radiação cósmica de fundo como função da escala angular; (iii) movimentos peculiares de galáxias; (iv) observações de quasares e galáxias com alto desvio para o vermelho.

1.2 A Importância de Aglomerados de Galáxias para a Cosmologia

Aglomerados de galáxias são os maiores sistemas colapsados encontrados no universo. Por serem raros, aglomerados são excelentes indicadores de como a matéria está distribuída em grande escala. Além disso, podemos utilizar aglomerados para determinação de parâmetros cosmológicos fundamentais (como Ω_m e σ_8). Considerando que as galáxias encontradas em aglomerados são provavelmente contemporâneas, aglomerados são também úteis no estudo da formação e evolução de galáxias.

Na seção 1.2.1 relacionamos algumas das principais características de aglomerados no universo local (Bahcall 1999). Nas seções seguintes fazemos uma breve revisão de alguns dos principais estudos desenvolvidos atualmente sobre aglomerados de galáxias.

1.2.1 Características de Aglomerados

Aglomerados de galáxias contém desde algumas dezenas até centenas de galáxias num raio típico de 1.5 h^{-1} Mpc, sendo que o raio central (R_c) é $\sim (0.1 - 0.25)h^{-1}$ Mpc. A densidade numérica de aglomerados é da ordem de 10^{-5} aglomerados h^3 Mpc $^{-3}$, sendo fortemente dependente da riqueza dos mesmos. A fração de galáxias no universo encontrada em aglomerados é $\sim 5\%$, dependendo do raio, da riqueza e luminosidades considerados. Se considerarmos somente aglomerados ricos, por exemplo, esta fração diminui. O perfil de densidade de galáxias em aglomerados, em geral, pode ser representado por $n_g(r) \propto r^{-2.4}$ (espacial) e $S_g(r) \propto r^{-1.4}$ (projetado). As galáxias num aglomerado possuem movimentos peculiares aleatórios da ordem de ~ 750 km s^{-1} dentro de 1.5 h^{-1} Mpc (este valor representa a dispersão de velocidade ao longo da linha de visada). Sendo assim, a massa dinâmica obtida através do teorema do virial considerando uma esfera de raio 1.5 h^{-1} Mpc é da ordem de (0.1-2)$\times 10^{15} h^{-1}$ $M_\odot$. A luminosidade na banda B é tipicamente (0.6-6)$\times 10^{12} h^{-2}$ $L_\odot$. Portanto, a razão massa-luminosidade típica de aglomerados no universo local é $(M/L_B)_{aglom.} \sim 300h$ $M_\odot/L_\odot$.

Aglomerados ricos (de alta massa) apresentam um meio intra-aglomerado composto por um plasma quente, abrangendo pelo menos a região onde a concentração de galáxias é significativa (raio de 1.5 h^{-1} Mpc). Este plasma quente é detectado através de emissão raio-X produzida por radiação bremsstrahlung térmica. A temperatura raio-X do gás intra-aglomerado é da ordem de 2-14 keV, sendo a luminosidade raio-X $L_X \sim 10^{44}$ erg s^{-1}. Estudos baseados em aglomerados próximos indicam que o gás encontrado em aglomerados está aproximadamente em equilíbrio hidrostático com o potencial do aglomerado. Estes resultados indicam que a distribuição de galáxias, em geral, delincia a distribuição de gás. Desta forma, a dispersão de velocidade de galáxias em aglomerados encontra uma boa correlação com a temperatura do gás intra-aglomerado (Lubin & Bahcall 1993; Yee & Ellingson 2003). Esta correlação sugere que a determinação de massa dinâmica é aproximadamente equivalente a determinação de massa baseada na temperatura do gás, pelo menos para aglomerados ricos e próximos.

1.2.2 Limites para a Densidade de Matéria (Ω_m)

Aglomerados de galáxias são úteis para estimar a densidade de matéria do universo e em consequência para estabelecer limites para modelos cosmológicos. Dois exemplos de como podemos estimar Ω_m são ilustrados a seguir. Estes são baseados na razão massa-luminosidade, e na fração de bárions em aglomerados.

A Razão Massa-Luminosidade

A matéria luminosa em galáxias não representa a massa total das mesmas, sendo a contribuição de matéria escura muitas das vezes especificada em termos da razão massa-luminosidade (M/L). Esta razão cresce com a escala ou raio dos sistemas em estudo, desde galáxias e grupos até aglomerados de galáxias. No entanto, espera-se que este crescimento diminua ou mesmo cesse na escala de aglomerados e super-aglomerados. Ou seja, para as maiores estruturas encontradas no universo a razão (M/L) tornaria-se constante. Aglomerados não apresentariam uma grande quantidade extra de matéria escura (ou halo escuro). Em aglomerados a matéria escura seria quase que totalmente justificada pelos grandes halos escuros das galáxias membro, além da contribuição do meio intra-aglomerado (Bahcall 1999).

Bahcall, Lubin & Dorman (1995) sugerem que os valores relativamente grandes da razão massa-luminosidade em aglomerados são devidos aos grandes valores de M/L encontrados para galáxias E-S0. Considerando que a razão M/L de E-S0 é cerca de 3-4 vezes maior do que a de espirais, e que aglomerados são essencialmente compostos de E-S0, grandes valores de M/L são esperados em aglomerados. Os autores ainda sugerem que o valor aproximadamente constante da razão M/L de aglomerados é consistente com um universo de baixa densidade ($\Omega_m \sim 0.2$). Este resultado assume que a quantidade de matéria escura em aglomerados é representativa do universo como um todo ($\Omega_m \sim (M/L) \times \rho_L/\rho_c$, onde ρ_c é a densidade crítica e ρ_L é a densidade de luminosidade típica do universo). Na Figura 1.3 (extraída de Bahcall *et al.* 2000) a razão massa-luminosidade é apresentada para diversos sistemas em função da escala. As razões M/L esperadas para $\Omega_m = 0.3$ e $\Omega_m = 1$ também são exibidas, da mesma forma que resultados de simulações para regiões de alta e baixa densidade. Estes resultados mais recentes também prevêem que a razão M/L fica aproximadamente constante a partir da escala de aglomerados. No entanto, nestas escalas o valor de M/L ainda depende

da densidade da região em questão, com regiões mais densas apresentando valores maiores de M/L.

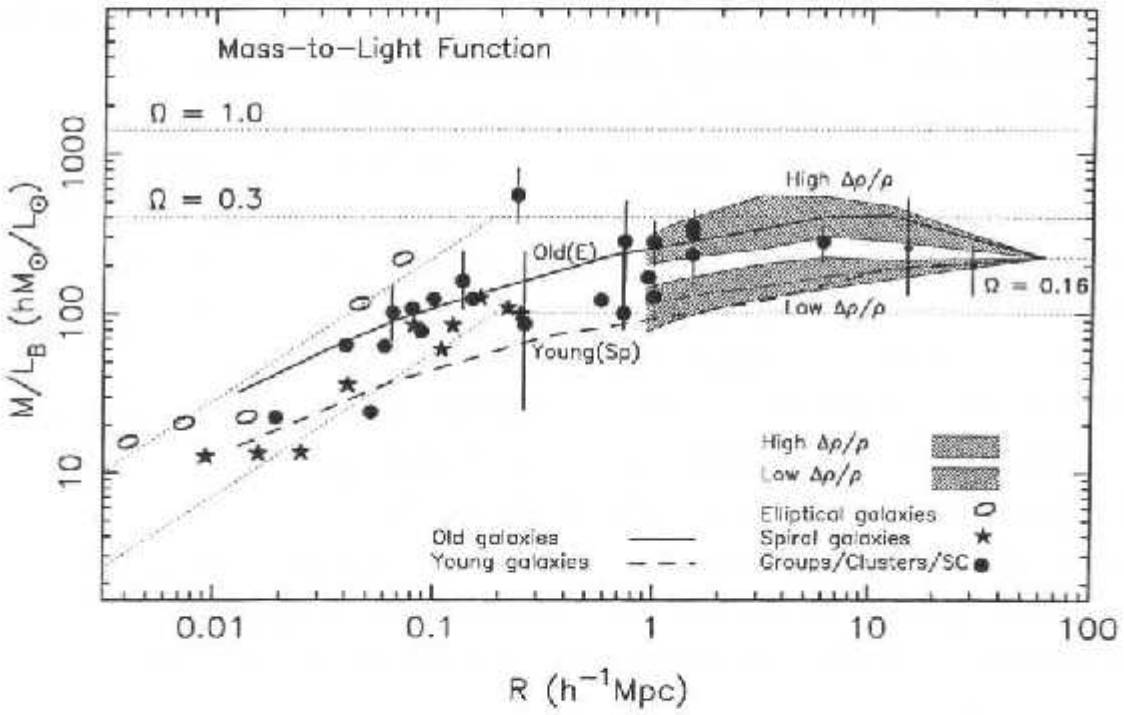


Figura 1.3: Razão massa-luminosidade para diferentes sistemas em função da escala. Os resultados para galáxias elípticas e espirais, grupos, aglomerados e super-aglomerados são exibidos. Também são mostradas as razões M/L para $\Omega_m = 0.3$ e $\Omega_m = 1$, além de resultados de simulações para regiões de alta e baixa densidade. Esta figura foi extraída de Bahcall *et al.* 2000, ApJ, 541, 1.

Girardi *et al.* (2002) e Hoekstra *et al.* (2002) encontram uma tendência similar a esta, com a razão M/L dependendo da massa do sistema (aglomerados tendo M/L maior do que grupos). No entanto, Hradecky *et al.* (2000) sugerem que a razão M/L de grupos não difere muito da de aglomerados.

A Fração de Bárions em Aglomerados

Em grandes escalas, aglomerados de galáxias são extremamente úteis na determinação de um limite superior da fração de bárions no universo. A matéria bariônica luminosa de aglomerados é composta de pelo menos duas componentes. Uma pequena contribuição é

proveniente de estrelas, enquanto que o gás intra-aglomerado é responsável por grande parte do material bariônico encontrado em aglomerados. Nucleosíntese primordial fornece limites precisos para a densidade bariônica, sendo esta $\Omega_b \sim 0.03 - 0.08 h_{50}^{-2}$ (Walker *et al.* 1991). Um limite superior para o parâmetro de densidade pode portanto ser obtido através da seguinte relação $\Omega_m < \Omega_b / f_b$, sendo a fração de bárions em aglomerados $f_b \sim 0.15 - 0.20 h_{50}^{-3/2}$. Para que $\Omega_m = 1$ a fração de bárions em aglomerados teria de ser muito menor do que o encontrado. Portanto, a densidade bariônica dada por nucleosíntese primordial e a alta fração de bárions observada em aglomerados sugerem que $\Omega_m \sim 0.2$. Em contrapartida existem argumentos na literatura de que as estimativas da fração de bárions em aglomerados podem estar super-estimadas, o que resultaria num valor maior para Ω_m. Sadat & Blanchard (2001) estimam a fração de bárions como sendo tipicamente a metade de outros estudos (David *et al.* 1995, Ettori & Fabian 1999). Eles ainda concluem questionando a confiabilidade da utilização da fração de bárions para determinar Ω_m. Segundo os autores, diversos efeitos podem contribuir para super-estimarmos f_b, como por exemplo o fato de que muitas das vezes a massa do gás é medida somente na região central de aglomerados sendo a medida extrapolada para as regiões mais externas (até o raio do virial, por exemplo).

1.2.3 Evolução na Densidade Numérica de Aglomerados

Os modelos mais aceitos atualmente para descrever a formação de estruturas são baseados em crescimento hierárquico, segundo o qual pequenas flutuações de densidade no universo primordial aumentam com o passar do tempo. O crescimento destas flutuações tem como produto final os sistemas virializados observados atualmente (desde galáxias até aglomerados). A época de formação destes sistemas é proporcional a massa dos mesmos. Num modelo que segue o formalismo proposto por Press & Schecter (PS, 1974) galáxias seriam formadas em desvios para o vermelho maior que 2, ao passo que aglomerados colapsariam após $z \sim 1$. A grande fração de aglomerados apresentando subestrutura no universo local sugere que os sistemas de mais alta massa ainda estão se formando (Solanes *et al.* 1999; Girardi *et al.* 1997; Kriessler & Beers 1997; Bird 1993). O formalismo PS é um dos modelos mais empregados para descrever a formação e evolução de estruturas. Em poucas palavras, a teoria PS prevê que estruturas virializadas com uma dada massa M são geradas quando flutuações de densidade atingem um valor superior a uma densidade limite (δ_c). Este formalismo pode então

ser empregado para investigar a evolução da densidade numérica comóvel de halos de massa M, segundo a expressão abaixo:

$$n(z,M) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{\rho}{M^2} \frac{\delta_c(z)}{\sigma(z,M)} \left|\frac{d\ln\sigma(z,M)}{d\ln M}\right| \exp\left(\frac{-\delta_c(z)^2}{2\sigma(z,M)^2}\right), \qquad (1.1)$$

onde $\overline{\rho} \equiv \Omega_m \rho_c$ é a densidade cósmica média; e $\delta_c(z) = \delta_c / D(z, \Omega_m, \Lambda)$, onde $\delta_c \sim 1.68$ é o contraste de densidade crítico necessário para um colapso e $D(z, \Omega_m, \Lambda)$ é o fator de crescimento de uma dada perturbação num regime linear. $\sigma(z, M)$ representa a flutuação de massa para diferentes valores de M, num dado desvio para o vermelho.

Uma simples estimativa de $n(0, M)$ no universo local é suficiente para vincular σ_8 e Ω_m por $\sigma_8 \Omega_m^{0.5} \sim 0.5$ (White, Efstathiou & Frenk 1993; Viana & Liddle 1996), onde σ_8 é a flutuação (rms) de massa medida em esferas de raio 8 h^{-1} Mpc. Apesar de extremamente poderosa, esta relação é degenerada (a determinação unívoca de um parâmetro depende de uma determinação precisa do outro). Uma forma de quebrar esta degenerescência é através da determinação da função de massa de aglomerados (dada pela equação 1.1 acima) até grandes valores de desvio para o vermelho. A evolução da abundância de aglomerados é extremamente sensível a como estes parâmetros são combinados. Num universo em que $\Omega_m = 1$ (com pequeno σ_8) esperamos encontrar uma densidade de aglomerados ricos muito pequena em $z \sim 1$ quando comparada com $z \sim 0$. Ou seja, a taxa de evolução na densidade de aglomerados é alta. Num universo em que Ω_m é pequeno (com grande σ_8) a formação destas estruturas é quase que interrompida em $z \sim 1$, representando somente uma pequena evolução na abundância de aglomerados. Fan *et al.* (1997) ainda sugerem que a taxa de evolução na abundânica de estruturas de mesma massa é muito mais sensível a σ_8 do que a qualquer outro parâmetro (inclusive Ω_m). Segundo estes autores a evolução da abundância de aglomerados é proporcional a σ_8^{-2} e $\Omega_m^{-2\alpha}$, onde α depende da inclinação (n) do espectro de potência $(-2.5 < n < -1)$. Portanto, a evolução da densidade numérica de aglomerados de alta massa representa um teste poderoso para a determinação unívoca de σ_8, e em consequência disto de Ω_m. Resultados recentes baseados no experimento *Boomerang* (Lange *et al.* 2001; Padmanabhan & Sethi 2001) sugerem que Ω total é igual a um, o que favorece um universo com geometria plana. Levando em consideração as estimativas de Ω_m e Λ baseadas em supernovas com grande desvio para o vermelho (Perlmutter *et al.* 1999), estes resultados implicam num baixo valor para Ω_m $(\sim 0.2 - 0.3)$, o que limita ainda mais a relação entre

Ω_m e σ_8.

É necessário aqui salientarmos a importância de aglomerados ricos para estimar $n(z, M)$. A equação 1.1 é válida para sistemas de qualquer massa, no entanto aglomerados ricos são os únicos sistemas com desvio para o vermelho de formação[4] não muito alto ($z < 1$). Sendo assim, catálogos de aglomerados ricos no universo local são *geralmente* completos e em mais alto z estas estruturas ainda são recuperadas sem grande custo observacional. Desta forma, podemos observar estes sistemas numa época próxima a que foram formados. Além disto aglomerados ricos podem ter a massa do virial determinada de forma *precisa* através de diferentes técnicas.

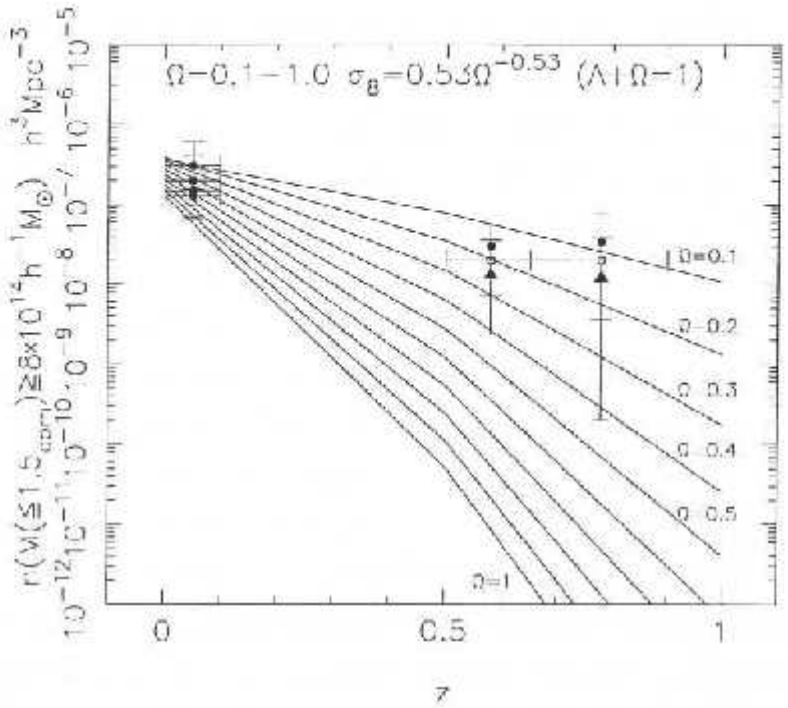


Figura 1.4: Evolução da abundância de aglomerados com desvio para o vermelho. As previsões do modelo PS são exibidas como curvas contínuas como função de $\Omega_m(\sigma_8)$. Os pontos em $z \sim 0.6$ e 0.8 representam um e dois aglomerados, respectivamente. Círculos cheios são para $\Omega_m = 1$, círculos vazios para $\Omega_m \sim 0.2, \Lambda = 0$ e triângulos para $\Omega_m \sim 0.2, \Lambda = 0.8$. Esta figura foi adaptada de Bahcall & Fan (1998).

Apesar do grande número de trabalhos buscando determinar a função de massa de aglom-

[4]por desvio para o vermelho de formação entenda-se a época em que o sistema tornou-se virializado.

erados (e assim obter σ_8 e Ω_m), os resultados ainda são bastante contraditórios. Eke *et al.* (1998) encontram que $\Omega_m \sim 0.45$. Mathiesen & Evrard (1998) e Kitayama & Suto (1997) encontram resultados similares. Bahcall *et al.* (2000), Bahcall & Fan (1998), Carlberg *et al.* (1997) e Donahue *et al.* (1998) sugerem um universo ainda menos denso, com $\Omega_m \sim 0.3$. Sadat *et al.* (1998), Reichart *et al.* (1999), Viana & Liddle (1999) e Blanchard & Bartlett (1998) encontram um valor mais alto para o parâmetro de densidade, sendo $\Omega_m \sim 0.8$. Na Figura 1.4 mostramos os resultados obtidos por Bahcall & Fan (1998). A figura apresenta o número de aglomerados com $M \geq 8 \times 10^{14} h^{-1}\ M_\odot$. As previsões do modelo PS são exibidas como curvas contínuas como função de $\Omega_m(\sigma_8)$. Os pontos em $z \sim 0.6$ e 0.8 representam um e dois aglomerados, respectivamente. Círculos cheios são para $\Omega_m = 1$, círculos vazios para $\Omega_m \sim 0.2, \Lambda = 0$ e triângulos para $\Omega_m \sim 0.2, \Lambda = 0.8$.

A Figura 1.4 exemplifica a grande eficiência da função de massa em discriminar modelos de baixa e alta densidade. Que fatores podem contribuir para resultados tão discrepantes na literatura? Segundo Eke *et al.* (1998) um dos principais fatores pode estar associado a própria determinação de massa. Incertezas na estimativa da temperatura do gás intra-aglomerado ou na estimativa da dispersão de velocidade, podem gerar erros sistemáticos na determinação da massa dos aglomerados. A escolha do perfil de densidade dos aglomerados também pode influir de maneira decisiva na determinação da massa dos mesmos. Um método alternativo é a determinação do perfil de densidade de massa utilizando lentes gravitacionais fracas. Um outro ponto para o qual devemos atentar refere-se aos seguidos melhoramentos que as simulações de N-corpos tem experimentado nos últimos anos. Simulações cosmológicas mais recentes (como por exemplo Jenkins *et al.* 2001) preveêm mais aglomerados de alta massa (especialmente em altos desvios para o vermelho, $z \sim 1$) do que o formalismo PS.

A existência de aglomerados de alta massa com grandes desvios para o vermelho pode também estar indicando que as condições iniciais não sejam gaussianas (uma premissa básica do formalismo PS) (Ribeiro *et al.* 2000; Willick 2000). Outro ponto de interesse é a normalização da abundância de aglomerados no universo local. Estudos mais recentes tem obtido valores tanto maiores como menores do que a normalização encontrada anteriormente (veja Viana, Nichol & Liddle 2002; Pierpaoli *et al.* 2001; Bahcall *et al.* 2003a). Como já mencionado, a normalização no universo local é de fundamental importância para a interpretação de resultados baseados em aglomerados mais distantes.

1.2.4 A Função de Correlação de Aglomerados

Uma das ferramentas mais utilizadas para descrever como sistemas em diferentes escalas (desde galáxias até super-aglomerados) encontram-se aglomerados tem como base funções de correlação. Estas funções descrevem as propriedades de aglomeração de um conjunto de pontos distribuídos numa dada região do espaço. Em particular, podemos definir que a probabilidade conjugada $\delta^2 P_2$ de encontrarmos uma dada galáxia num pequeno volume δV_1 e outra num volume δV_2, separadas por $\mathbf{r}_{12}$ é dada por

$$\delta^2 P_2 = n_V^2[1 + \xi(r_{12})]\delta V_1 \delta V_2, \tag{1.2}$$

onde os dois elementos de volume são selecionados aleatoriamente de um grande volume representativo do universo. Na equação acima n_V é o número médio de galáxias por unidade de volume e $\xi(r)$ é a função de correlação espacial de dois pontos. Assumindo isotropia e homogeneidade, ξ depende somente do módulo do vetor $\mathbf{r}_{12}$ e não da direção do mesmo. A função de correlação ξ representa a probabilidade de excesso (quando comparada a uma distribuição aleatória) de encontrarmos uma galáxia a uma distância r_{12} de qualquer outra galáxia dentro do volume em estudo. Este formalismo é aplicado a objetos em diferentes escalas, como galáxias, aglomerados e super-aglomerados. Mas pode ser aplicado para qualquer conjunto de pontos, como galáxias rádio, quasares, galáxias *Lyman-break* em altos *redhsifts*, etc. O formalismo descrito acima é válido para N=2 pontos, no entanto correlações de mais alta ordem (N > 2) também podem ser obtidas.

Em geral as estimativas de $\xi(r)$ utilizam-se de um conjunto poissônico de pontos gerados dentro dos mesmos limites e com a mesma função de seleção dos dados reais. Dois estimadores são exemplificados abaixo:

$$\xi(r) = \frac{n_{DD}(r)}{n_{RR}(r)} - 1 \tag{1.3}$$

$$\xi(r) = 4\frac{n_{DD}(r)n_{RR}(r)}{[n_{DR}(r)]^2} - 1, \tag{1.4}$$

onde $n_{DD}(r)$, $n_{RR}(r)$ e $n_{DR}(r)$ representam o número de pares com separação r nos dados reais, no catálogo poissônico, e com um membro no catálogo real e outro no catálogo poissônico, respectivamente. O estimador na equação 1.4 é conhecido como estimador de *Hamilton*, pouco afetado por incertezas na função de seleção para $\xi(r) < 1$ (Croft *et al.* 1997).

Quando estimadores como estes são aplicados a catálogos de galáxias, aglomerados ou super-aglomerados encontra-se que a função de correlação de dois pontos pode em geral ser descrita como uma lei de potência, com a forma:

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{1.5}$$

sendo r_0 a escala de comprimento de correlação, e o índice γ encontrado como $1.8 \leq \gamma \leq 2.2$. A pergunta básica em estudos da função de correlação é se os parâmetros r_0 e γ dependem da massa dos sistemas em estudo. Vários trabalhos sustentam a idéia de que a função de correlação ($\xi(r)$) é invariante de escala, sendo a forma da correlação e portanto (γ) a mesma, desde galáxias até super-aglomerados (Szalay & Schramm 1985; Bahcall & West 1992; Walter & Klypin 1996; Croft *et al.* 1997; Postman 1998; Colberg *et al.* 2000). O único fator de distinção é a amplitude das correlações.

Em virtude da sua grande separação espacial, aglomerados de galáxias são extremamente úteis para delinear a estrutura do universo em grande escala. Aglomerados estão mais fortemente correlacionados no espaço do que galáxias (veja Figuras 1.5 e 1.6). Desta forma, uma amostra pequena de aglomerados pode cobrir grandes distâncias, ao passo que um estudo baseado em galáxias torna-se muito mais caro (do ponto de vista observacional) para abranger o mesmo volume do espaço. Além disto, uma amostra de aglomerados pode ser dividida em classes de riqueza para testar a invariância de escala mencionada acima. Tendo em mente que a amplitude da correlação depende da riqueza ($\propto$ luminosidade ou massa) e da separação espacial média, Bahcall & West (1992) propuseram uma função de correlação *universal*. A função de correlação de aglomerados poderia ser representada por $\xi_i = A_i r^{-1.8}$, onde $A_i \propto N_i$ e $A_i \sim (0.4 d_i)^{1.8}$. A_i representa a amplitude da função de correlação, N_i a riqueza dos sistemas i em estudo (se aglomerados pobres ou ricos), e $d_i = n_i^{-1/3}$ é a separação espacial média dos mesmos (n_i é a densidade numérica espacial média). Com base na equação 1.5 podemos escrever a escala de comprimento de correlação (dos sistemas i em estudo) como $r_{0i} = A_i^{1/1.8} \sim (0.4 d_i)$. Na Figura 1.5, a dependência da amplitude A_i da função de correlação *universal* para aglomerados com a separação espacial média d_i é apresentada. A_i teria uma forte dependência com a riqueza ou massa dos sistemas (a Figura 1.5 é extraída de Bahcall & West 1992).

Esta função *universal* veio também justificar possíveis diferenças no valor de r_0 obtido

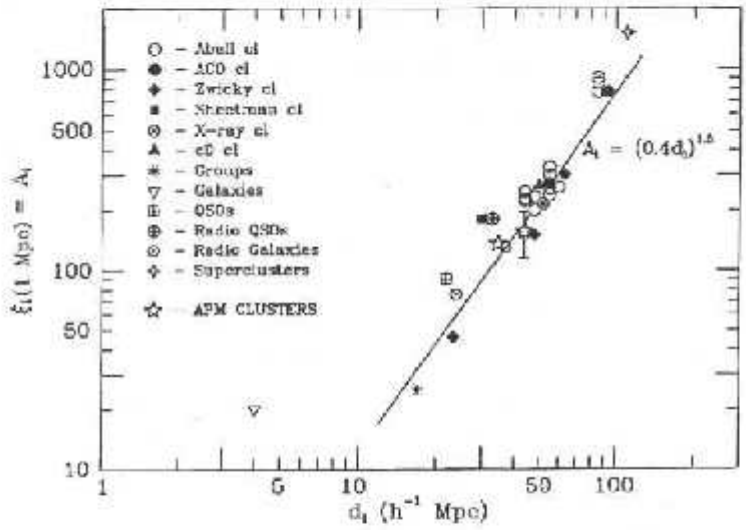


Figura 1.5: Correlações *universais* de aglomerados. A figura exibe a dependência da amplitude da correlação A_i com a separação espacial média dos sistemas. Esta figura foi adaptada de Bahcall & West (1992).

com aglomerados de Abell e com o catálogo APM. Usando o catálogo de Abell, Postman, Geller & Huchra (1986) e Bahcall (1988) apontam para um valor de r_0 de $20 - 24\ h^{-1}$ Mpc. Estudos baseados no catálogo APM sugerem que $r_0 \sim 13 - 14 h^{-1}$ Mpc (Dalton *et al.* 1992; Efstathiou *et al.* 1992; Dalton *et al.* 1994). Bahcall & West (1992) sugerem que as diferenças desaparecem quando se leva em conta a função universal por eles proposta. O menor valor de r_0 no catálogo APM deveria-se ao fato do mesmo ser mais completo do que o catálogo de Abell no regime de aglomerados pobres. Por outro lado, Efstathiou *et al.* (1992) sugerem que os grandes valores de r_0 com base no catálogo de Abell são devidos a aglomerações artificiais ao longo da linha de visada. O catálogo de Abell seria muito mais afetado por efeitos de contaminação e superposição do que o catálogo do APM, gerado de forma objetiva.

Trabalhos baseados em simulações e no catálogo APM, tentaram reproduzir a forte dependência com riqueza proposta por Bahcall & West (1992). Croft & Efstathiou (1994), Croft *et al.* (1997), Postman (1998) e Colberg *et al.* (2000) sugerem que esta dependência é menor do que anteriormente sugerido (a amplitude da correlação não cresceria tão rapidamente com a riqueza do sistema). Postman (1998) indica o valor assumido para γ como uma das possíveis causas de discrepâncias. Os resultados de Bahcall & West (1992) fixam γ em 1.8, ao passo que Croft *et al.* (1997) deixam este parâmetro livre, sendo em geral > 2.

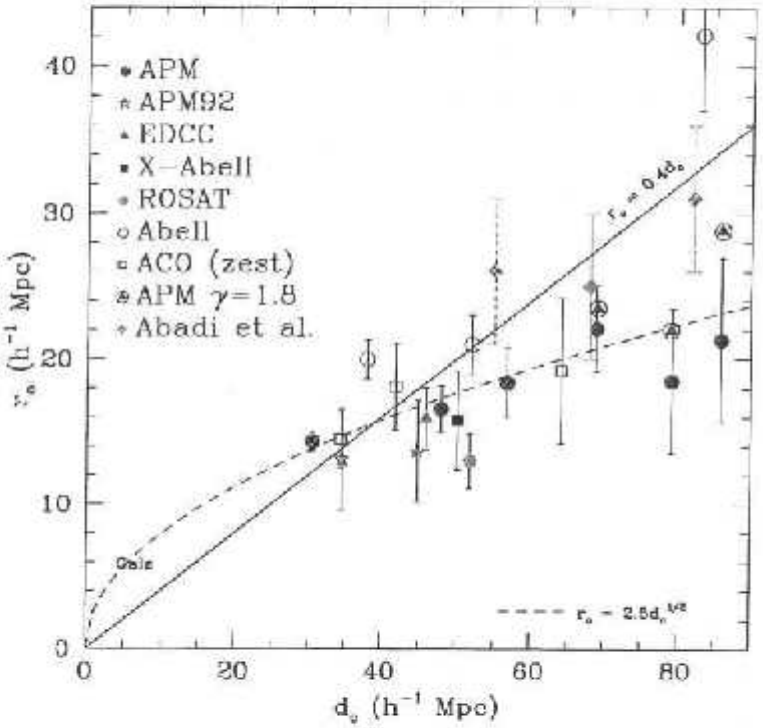


Figura 1.6: Dependência do comprimento de correlação com a separação média de aglomerados de diferentes classes de riqueza. Os resultados do APM são apresentados como obtidos originalmente e com $\gamma = 1.8$. A relação proposta por Bahcall & West (1992) ($r_0 = 0.4 d_c$) é exibida como linha sólida. Figura adaptada de Postman (1998).

Recalculando r_0 com base nos aglomerados do APM, mas fixando $\gamma = 1.8$, os resultados de Bahcall & West (1992) e Croft *et al.* (1997) são comparáveis, como indicado na Figura 1.6. Esta figura é análoga a figura anterior, mas com maior vínculo observacional. Outro fator que pode ser responsável pelas diferenças observadas é o procedimento de seleção dos aglomerados. Os resultados da Figura 1.6 sugerem que a dependência da função de correlação com a riqueza (ou de r_0 com d) é muito menor do que originalmente suposto somente com base no catálogo de Abell. Novos catálogos de aglomerados, selecionados de forma objetiva e com efeitos de seleção bem conhecidos, podem estabelecer de forma mais precisa este efeito para $d > 60h^{-1}$ Mpc. A Figura 1.7 apresenta resultados recentes do *Las Campanas Distant Cluster Survey* (LCDCS) (Gonzalez *et al.* 2002). O que chama atenção neste trabalho é o fato de ser a primeira estimativa fora do universo Local ($z > 0.2$) para função de correlação

de aglomerados. Os aglomerados utilizados cobrem o intervalo de $0.35 \leq z_{est} \leq 0.575$. Os autores encontram estimativas similares as obtidas no universo local (em particular aos resultados de Croft *et al.* 1997). A dependência de r_0 com d é semelhante a encontrada utilizando o catálogo APM. No entanto, estes resultados ainda não fornecem vínculos para as regiões de grandes separações ($d > 60h^{-1}$ Mpc).

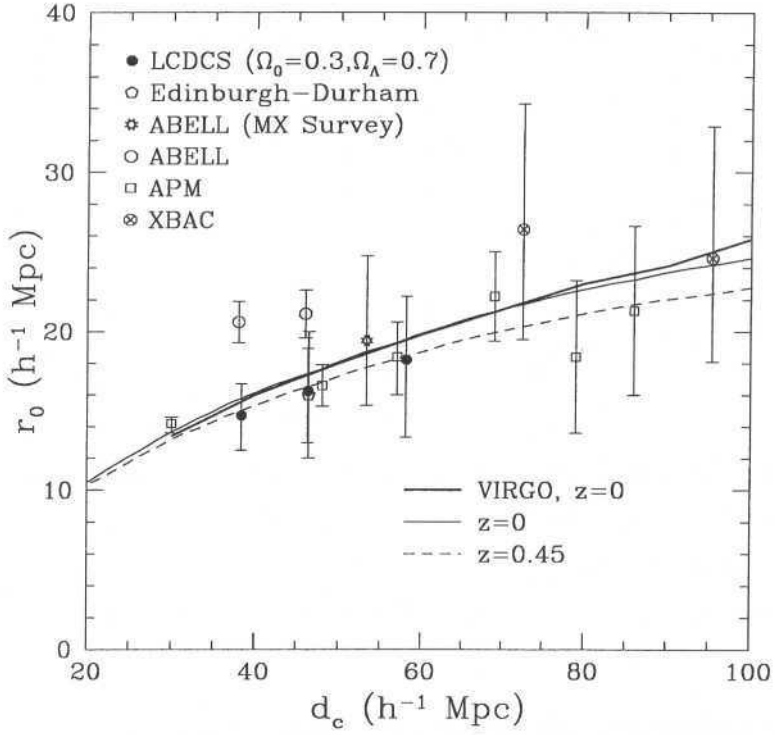


Figura 1.7: Dependência do comprimento de correlação com a separação média de aglomerados usando dados do LCDCS, Abell, EDCC e APM. Figura adaptada de Gonzalez *et al.* (2002).

1.3 Catálogos de Aglomerados

Nas seções 3.2 e 4.1 faremos uma revisão dos principais catálogos de aglomerados disponíveis para a comunidade científica. Nesta seção gostaríamos de resumir as principais características destes levantamentos.

A maioria destes catálogos são obtidos no óptico e infra-vermelho próximo, ou então através de levantamentos em raios-X. Os primeiros catálogos de aglomerados apresentavam grande cobertura angular, no entanto foram baseados na inspeção visual de placas fotográficas (Abell 1958; Abell, Corwin & Olowin 1989, Zwicky *et al.* 1961-1968). Desde a década passada um grande número de catálogos tem sido gerado de forma objetiva. Estes catálogos são provenientes de dados digitalizados de grandes levantamentos fotográficos, de levantamentos CCD profundos no óptico e infra-vermelho próximo (estes levantamentos alcancam $z \sim 1.4$, mas são limitados a regiões de algumas dezenas de graus quadrados), ou levantamentos raios-X. Apesar destes levantamentos em raios-X serem relativamente profundos, eles ainda não cobrem uma área tão ampla como a dos catálogos ópticos; ou quando cobrem, os catálogos são limitados a fontes brilhantes. Por exemplo, de Grandi *et al.* (1999) cobrem 8235 graus quadrados, apresentando um catálogo com somente 130 aglomerados, sendo que 78% destes são aglomerados de Abell.

Aglomerados detectados com imageamento óptico podem ser fortemente afetados por efeitos de projeção. No entanto, o uso de informação de cor tem sido extremamente útil para pré-seleção de galáxias *early-type*, típicas de aglomerados (Gladders & Yee 2000; Kim 2001). Levantamentos baseados em cor podem minimizar os efeitos de contaminação por superposição de objetos ao longo da linha de visada, além de fornecer estimativas precisas do desvio para o vermelho. Em raios-X efeitos de contaminação são minimizados, mas ainda presentes (por excesso de fluxo raios-X de AGNs por exemplo).

Um progresso muito grande também foi alcançado no que diz respeito aos métodos utilizados para detecção de aglomerados. Alguns destes métodos são baseados em diferentes algoritmos de percolação, estimadores de densidade como *kernels* adaptativos, *voronoi tesselation*, *wavelets*, etc. Outros métodos exploram propriedades intrínsecas de aglomerados como *matched filter* e diversos algoritmos baseados em informação de cor, etc. Obviamente, alguns destes métodos apresentam pré-disposição para detecção de alguns tipos de aglomerados, não sendo sensíveis a candidatos que não enquadram-se nas características básicas assumidas pelo método. A comparação dos resultados obtidos através de diferentes técnicas e em diferentes comprimentos de onda é extremamente importante para entender os efeitos seletivos de cada metodologia. Kim (2001) comparou catálogos baseados em diferentes algoritmos (*Matched Filter* e Voronoi Tesselation). Gilbank (2001) comparou os resultados do

Matched Filter com um método que utiliza a relação cor-magnitude para excluir galáxias com grande probabilidade de serem galáxias de campo. Além disto, Gilbank (2001) também comparou os efeitos seletivos de candidatos detectados em diferentes comprimentos de onda (óptico e raios-X). Diversos estudos apresentam comparações das propriedades de aglomerados em raios-X e no óptico, sendo estes de fundamental importância para um melhor entendimento de algumas propriedades de aglomerados. Por exemplo, a comparação de estimativas de massa obtidas com dados raios-X e no óptico pode esclarecer se existem efeitos sistemáticos nas estimativas de massa empregadas para o cálculo da função de massa (Lewis *et al.* 1999; Lubin *et al.* 2002).

Outros métodos que podemos citar são baseados em flutuações de brilho superficial, utilizando imagemento não tão profundo para detecção de aglomerados com alto desvio para o vermelho (Gonzalez *et al.* 2001); ou em *weak-lensing* (Wittman *et al.* 2001), o que fornece uma estimativa direta da massa do aglomerado; ou utilizam-se do efeito Sunyaev-Zeldovich para selecionar aglomerados por massa (Holder 2000).

1.4 Objetivos desta Tese

Um dos objetivos desta tese é a geração de um catálogo objetivo de candidatos a aglomerado no universo local ($< z > \sim 0.15$) para todo o hemisfério norte. Também pretendemos estender este catálogo para maiores valores de desvio para o vermelho ($< z > \sim 0.3$), além de compararmos o desempenho de dois métodos de detecção de aglomerados.

No capítulo 2 descrevemos a base de dados utilizada para geração dos catálogos apresentados nesta tese. Fazemos uso dos dados digitalizados do segundo levantamento fotográfico do Observatório Palomar (DPOSS). Uma breve comparação com o *Sloan Digital Sky Survey* é utilizada para estimativas dos limites de completeza do DPOSS. No capítulo 3 apresentamos o *Northern Sky Optical Cluster Survey* (NoSOCS). O catálogo aqui apresentado é um dos maiores catálogos selecionados objetivamente já obtido para o hemisfério norte, cobrindo 5800 graus quadrados (Gal *et al.* 2003a). Até o final deste ano esperamos publicar o restante do catálogo, completando $\sim$ 11500 graus quadrados (Gal *et al.* 2003b). Na Figura 1.8 apresentamos uma visão do céu inteiro do NoSOCS (pontos pretos), juntamente com o catálogo APM (pontos vermelhos). Os candidatos do NoSOCS cobrem $\sim$ 11500 graus quadrados, ao

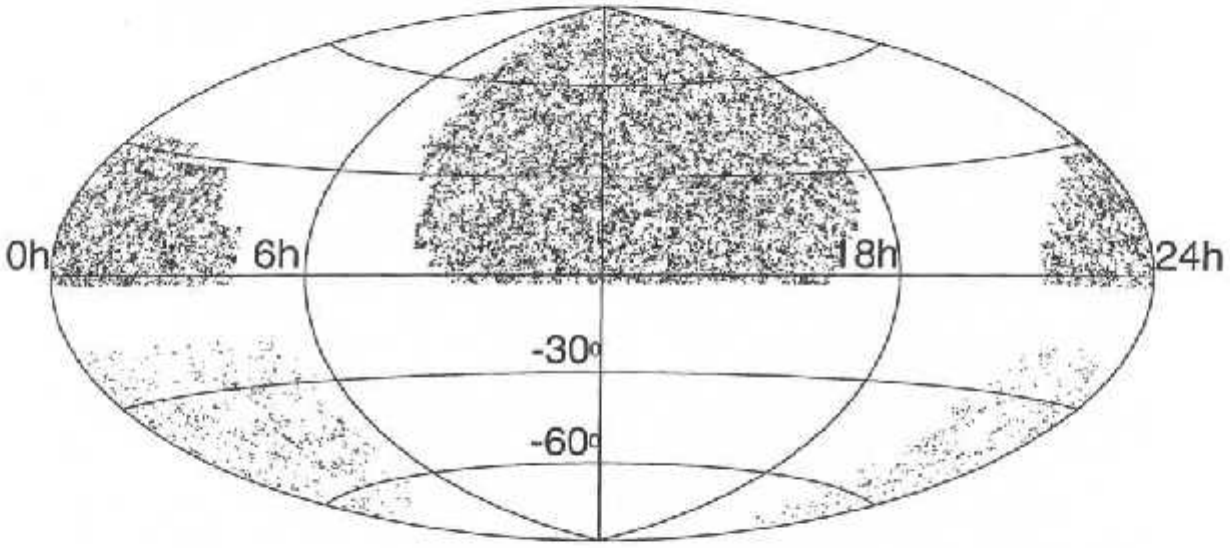


Figura 1.8: Distribuição de aglomerados no céu. As coordenadas da projeção Aitoff são equatoriais, com centro em $(\alpha, \delta) = (12^h, 0°)$. Candidatos do NoSOCS são apresentados como pontos pretos e do APM como pontos vermelhos no hemisfério sul.

passo que o catálogo do APM contém 957 aglomerados em cerca de 4300 graus quadrados (Dalton *et al.* 1997). Os dois catálogos são provenientes de dados de placa de qualidade semelhante. A razão principal pela qual o NoSOCS representa um catálogo muito mais denso do que o APM reside principalmente no fato da separação estrela-galáxia do DPOSS ser mais completa do que a do APM em limites de baixo fluxo (Weir *et al.* 1995; Odewahn *et al.* 2003).

Um extensão do NoSOCS é apresentada no capítulo 4, onde mostramos os primeiros resultados de um projeto para detectar aglomerados em cerca de 2700 graus quadrados até $z \sim 0.5$ a partir do DPOSS (Lopes *et al.* 2003a). Na figura 1.9 apresentamos uma imagem[5] de um dos candidatos do catálogo suplementar do NoSOCS apresentado no capítulo 4. No capítulo 5 apresentamos algumas possíveis aplicações dos catálogos aqui gerados e que constituirão projetos a serem desenvolvidos num futuro próximo. Esperamos que estes

[5]Esta imagem está disponível em
`http://www.astro.caltech.edu/~paal/thesis/colorimages/figure1.9.ps`

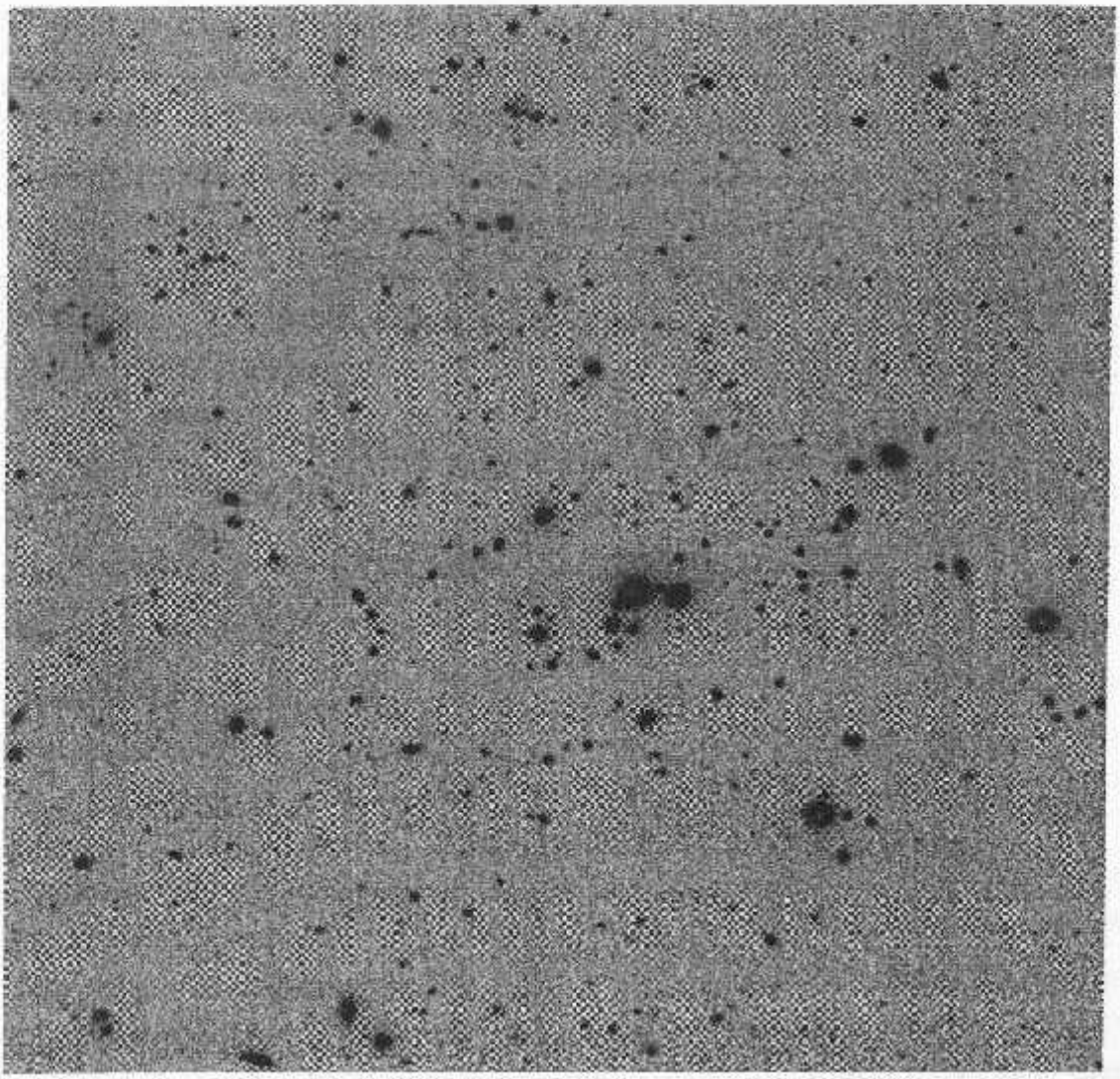


Figura 1.9: Exemplo de um aglomerado de galáxias detectado no catálgo suplementar do NoSOCS apresentado nesta tese (capítulo 4). O desvio para o vermelho estimado para este candidato é $z \sim 0.32$. Esta é uma imagem composta de observações feitas nas bandas *gri* utilizando o telescópio de 1.5m do Observatório Palomar. Outros exemplos de aglomerados são apresentados no apêndice A.

catálogos tornem-se referências importantes para a comunidade científica, possibilitando um melhor entendimento de algumas questões mencionadas nesta introdução.

Capítulo 2

O Segundo Levantamento Fotográfico do Observatório Palomar[1]

Resumo

Neste capítulo[2] apresentaremos as principais características do DPOSS, incluindo um resumo dos procedimentos adotados para detecção e medição dos objetos; geração do banco de dados associado ao DPOSS; separação estrela/galáxia; e calibração fotométrica do levantamento fotográfico. Também é apresentado um estudo dos limites de detecção do DPOSS. Para tal, fazemos uso das detecções em bandas múltiplas do próprio levantamento fotográfico, além de termos uma fonte de comparação externa dada pelo *Sloan Digital Sky Survey* (SDSS) (York *et al.* 2000). Os resultados apresentados nas seções 2.2 e 2.3 são úteis para verificarmos que $\sim 100\%$ dos objetos detectados na banda r, são também detectados na banda g, até $m_r \sim 19.5$. Isto corrobora a boa determinação de cor usada na estimativa de desvio para o vermelho fotométrico apresentada no capítulo 3. Também verificamos que o DPOSS é aproximadamente completo até $m_r \sim 21$, o que viabiliza a geração de um catálogo de aglomerados de galáxias até $z \sim 0.5$ (capítulo 4).

2.1 Um resumo do DPOSS

O segundo levantamento fotográfico do Observatório Palomar (POSS-II, Reid *et al.* 1991) é constituído por 894 placas distribuídas em todo o hemisfério norte ($\delta > -3^o$). Cada placa do

[1] *The Digitized Second Palomar Observatory Sky Survey* (DPOSS).

[2] Algumas figuras desto capítulo são provenientes de Gal *et al.* (2003c) e Lopes *et al.* (2003b).

POSS-II abrange uma área de 6.5° x 6.5° no céu, com espaçamento entre o centro das placas da ordem de 5 graus. As placas originais do levantamento fotográfico foram digitalizadas no *Space Telescope Science Institute (STScI)*. Cada pixel equivale a 1″, com o *seeing* típico dos dados digitalizados sendo da ordem de 2″. Cada placa representa $\sim$ 1Gb, o que equivale a $\sim$ 3Tb para o levantamento inteiro. A precisão astrométrica proveniente do STScI está na ordem de $rms \sim 0.5''$.

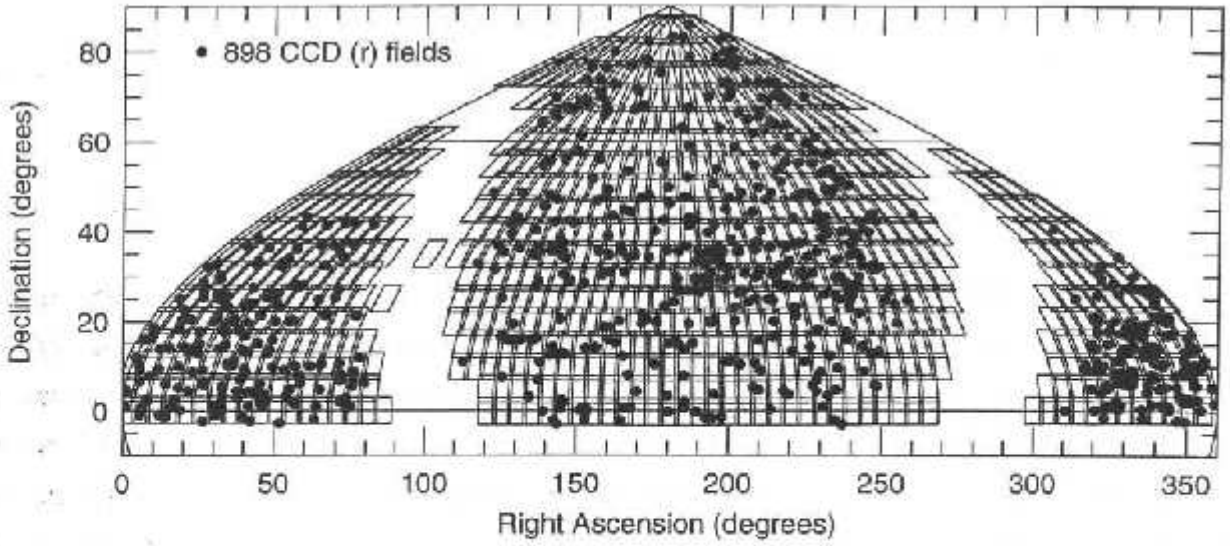


Figura 2.1: Distribuição das placas do POSS-II (retângulos) e cerca de 900 campos CCD (pontos).

As placas deste levantamento foram obtidas em três bandas (*JFN*): azul-verde (IIIa-J + GG395, $\lambda_{ef} \approx 4800$Å e FWHM ≈ 1600Å); vermelha (IIIa-F + RG610, $\lambda_{ef} \approx 6500$Å e FWHM ≈ 1300Å); infra-vermelho próximo (IV-N + RG9, $\lambda_{ef} \approx 8500$Å e FWHM ≈ 1700Å). A calibração fotométrica foi feita a partir de observações CCD de centenas de aglomerados de Abell com os filtros *gri* (com λ_{ef} de 5100Å, 6800Å, e 8100Å e FWHM de 900Å, 1000Å, e 1500Å, respectivamente) do sistema fotométrico de Thuan & Gunn (1976). As placas processadas dão origem a um catálogo final de todos os objetos detectados nos limites de completeza das placas ($B_J \sim 22.5^m$, $R_F \sim 20.8^m$, e $I_N \sim 19.5^m$). Os limites correspondentes dos dados de placa calibrados são $g \sim 21.5^m$, $r \sim 21^m$, e $i \sim 19.8^m$; sendo que os limites de detecção alcançam $r \sim 22^m$. As imagens CCD utilizadas na calibração fotométrica alcançam limites inferiores de fluxo, correspondentes a $\sim 1^m - 2^m$. Na figura 2.1 apresentamos a dis-

tribuição de todas as placas já processadas (retângulos), além dos campos CCD usados para a calibração fotométrica do levantamento fotográfico (pontos). Na figura 2.2 apresentamos a curvas de resposta das bandas do POSS-II (*JFN*), sobrepostas pelas bandas do sistema de Gunn-Thuan (*gri*).

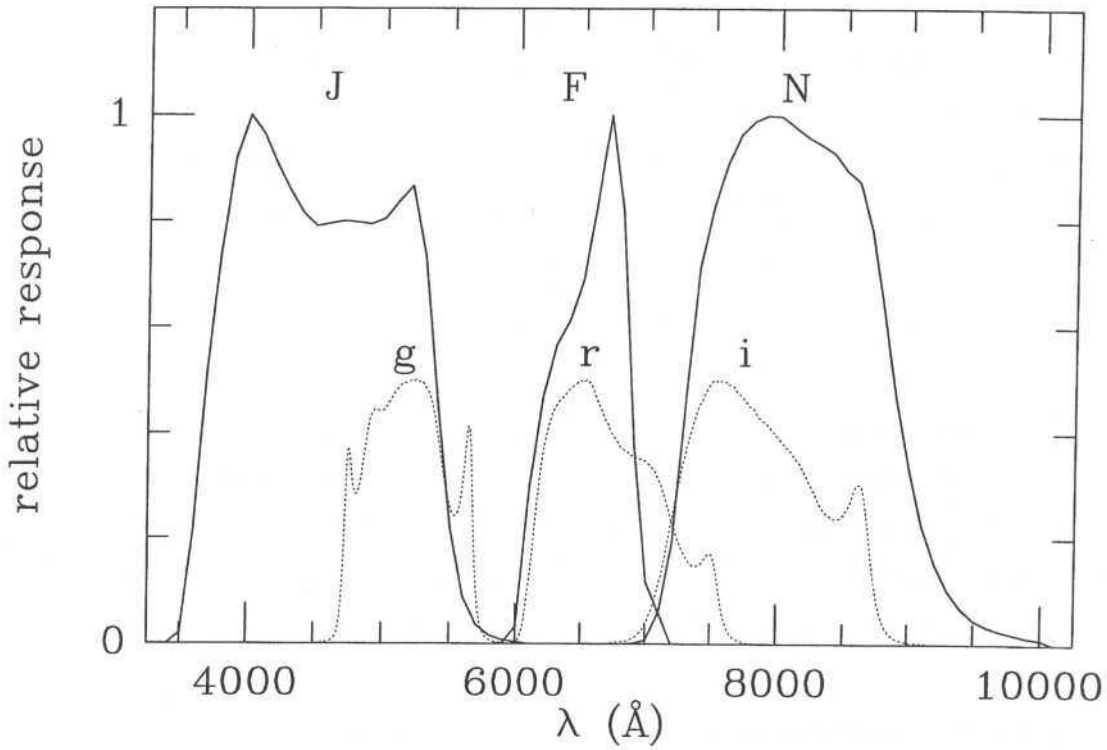


Figura 2.2: As bandas *JFN* do POSS-II (linhas cheias) e as curvas de resposta dos filtros *gri* de Gunn-Thuan (linhas pontilhadas).

O *Sky Image Cataloging and Analysis Tool (SKICAT)* (Weir *et al.* 1995abc) é um programa desenvolvido para facilitar a criação e utilização de catálogos a partir de grandes levantamentos de imageamento, em particular o DPOSS. Este programa utiliza-se de pacotes de processamento de imagens astronômicas, um banco de dados comercial (*Sybase*), além de técnicas de inteligência artificial (Weir 1995). Nós utilizamos o SKICAT para o processamento de cada placa do POSS-II, a partir do qual geramos catálogos com cerca de 60 atributos medidos por objeto, em cada banda fotométrica. Para cada campo, os catálogos resultantes de cada banda são correlacionados para geração de uma lista única de objetos por campo. Para cada banda, a classificação (estrela-galáxia) dos objetos é baseada em 2

métodos: árvore de decisão e redes neurais. Cada método fornece a classificação e a probabilidade associada a mesma. Esta probabilidade é utilizada para definir a classe de cada objeto em cada banda. A melhor classificação entre as três bandas (J, F, e N) é mantida no catálogo combinado para cada campo. Em geral a classificação tem melhor desempenho na banda J, e portanto a classe é proveniente desta banda, a não ser que o objeto não tenha sido encontrado na mesma. A classificação final tem precisão $> 90\%$ para objetos com r $\leq 19.5^m$ (Odewahn *et al.* 2003).

O produto final do DPOSS será dado pelo *Palomar Norris Sky Catalog* (PNSC) contendo $> 5 \times 10^7$ galaxies e $> 2 \times 10^9$ estrelas. Este catálogo e outros sub-produtos do DPOSS estão sendo disponibilizados em http://dposs.caltech.edu.

A extensa coleção de dados CCD é utilizada para dois propósitos muito bem definidos:

1. Classificações mais precisas são obtidas a partir dos dados CCD em comparação com os dados de placa, independente do regime de fluxo que se observa. Desta forma, os dados CCD são utilizados como um conjunto de treinamento para a classificação dos objetos detectados nas placas. Detalhes dos procedimentos adotados com relação a separação estrela/galáxia são dados em Odewahn *et al.* (2003).

2. Os objetos detectatos nos campos CCD são utilizados para a calibração fotométrica dos dados de placa, permitindo a transformação das magnitudes instrumentais para o sistema de Thuan-Gunn (1976). O procedimento completo da calibração fotométrica do DPOSS encontra-se descrito em Gal *et al.* (2003c). Na tabela 2.1 apresentamos os erros fotométricos em função da magnitude, para estrelas e galáxias, nas três bandas do DPOSS.

Fora estas aplicações , os campos CCD constituem um levantamento por si só, o *Palomar Abell Cluster Optical Survey* (PACOS), descrito em Gal *et al.* (2000b). Este levantamento tem diversas aplicações para estudos de propriedades fotométricas de aglomerados de galáxias no universo *local* ($z < 0.3$) (*e.g.*, determinação precisa de desvios para o vermelho fotométricos; efeito Butcher-Oemler; efeitos de alinhamento em aglomerados; subestrutura; evolução da relação cor-magnitude; etc).

Tabela 2.1: Erros fotométricos do DPOSS.

	g		r		i	
Mag	gal	star	gal	star	gal	star
17.0	0.25	0.22	0.13	0.15	0.17	0.22
18.0	0.17	0.24	0.16	0.17	0.24	0.25
18.5	0.19	0.25	0.17	0.22	0.29	0.32
19.0	0.21	0.24	0.19	0.21	0.32	0.31
19.5	0.25	0.26	0.24	0.24	0.36	0.34
20.0	0.29	0.29	0.29	0.27	0.44	0.35
20.5	0.37	0.32	0.34	0.33	0.42	0.39

2.2 Os Limites de Detecção do DPOSS

Antes de gerarmos um catálogo de aglomerados de galáxias é importante que tenhamos uma boa estimativa dos limites de completeza e da taxa de detecções espúrias presentes no DPOSS. Na figura 4.6 apresentamos a distribuição de magnitude (banda r) de um campo do DPOSS, além de mostrarmos a distribuição de magnitude limite de centenas de campos do levantamento. Pelo painel inferior desta figura concluimos que a magnitude limite típica dos campos do DPOSS é $r_{lim} \sim 21.5$. Uma análise mais detalhada do levantamento torna-se possível graças a disponibilidade de dados em mais de uma banda.

Na figura 2.3 apresentamos a fração de objetos que são detectados somente na banda r (linha contínua, #1), objetos detectados na banda r e somente mais uma banda, podendo ser g ou i (linha pontilhada, #2), objetos detectados na banda r e em mais uma ou duas bandas (linha tracejada, traços pequenos, #3) e objetos detectados nas três bandas (linha tracejada, traços grandes, #4). A primeira e terceira linhas são complementares. Da mesma forma que a soma da segunda e quarta linhas resulta na terceira.

A fração de objetos indicada pela linha 1 (somente na banda r) começa a crescer significativamente a partir de $m_r \sim 20$. Estes objetos tanto podem ser objetos reais, detectados somente nesta banda, como também podem ser espúrios. Nós assumimos que objetos detectados em mais de uma banda têm menor probabilidade de serem falsos. Ou seja, assumimos

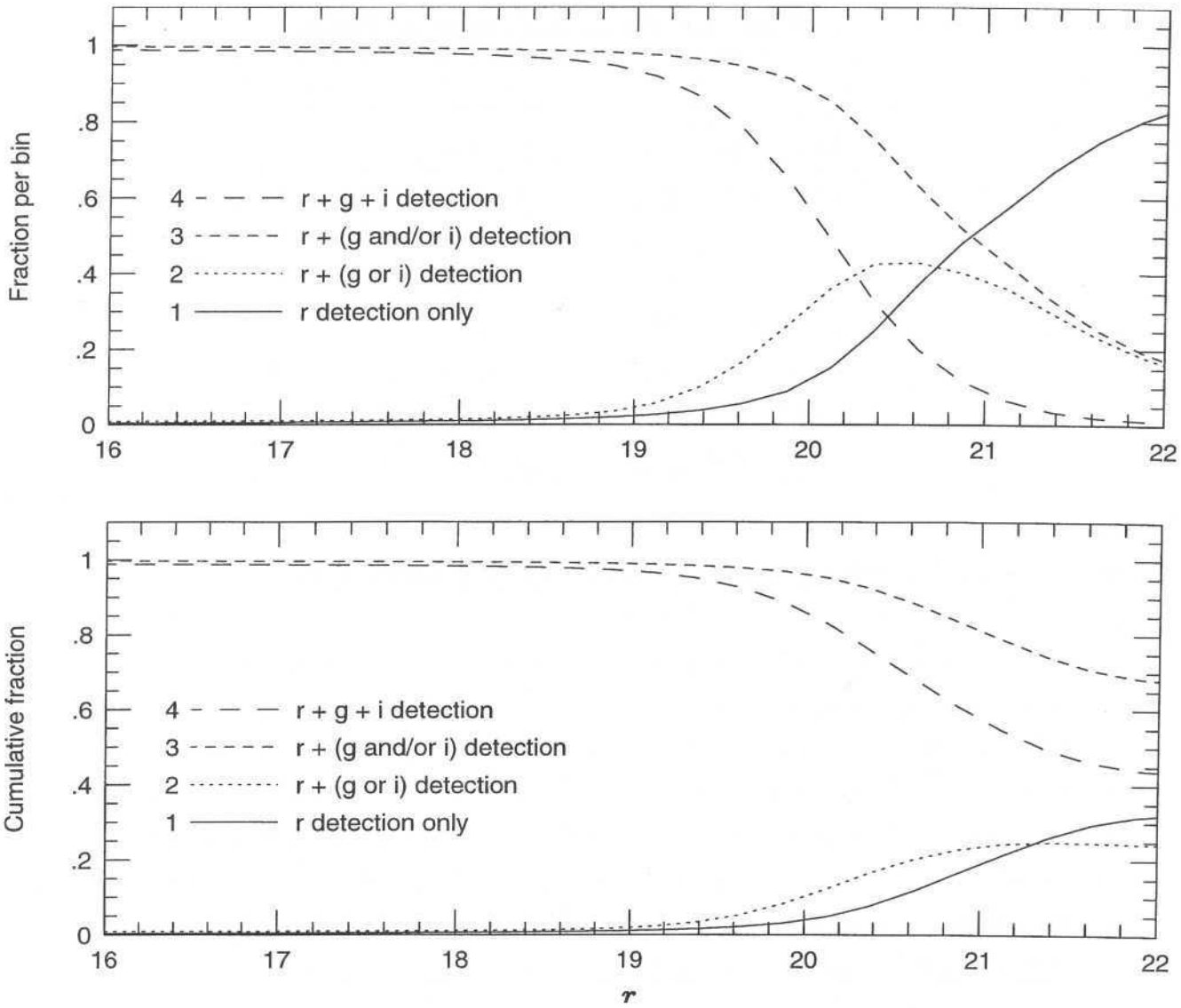


Figura 2.3: Análise interna do DPOSS com a banda r como referência.

que objetos espúrios são identificados numa banda somente; não significando porém que todos os objetos provenientes de uma só banda são espúrios. A linha 3 indica que o número de objetos detectados em duas ou mais bandas é $> 90\%$ até $m_r \sim 20$, decrescendo rapidamente a partir desta magnitude. Ou seja, até $m_r \sim 20$ praticamente todos os objetos detectados pelo DPOSS na banda r são provavelmente reais, pois apresentam uma contrapartida em pelo menos uma outra banda. Ao analisarmos a linha 1 percebemos que o número de

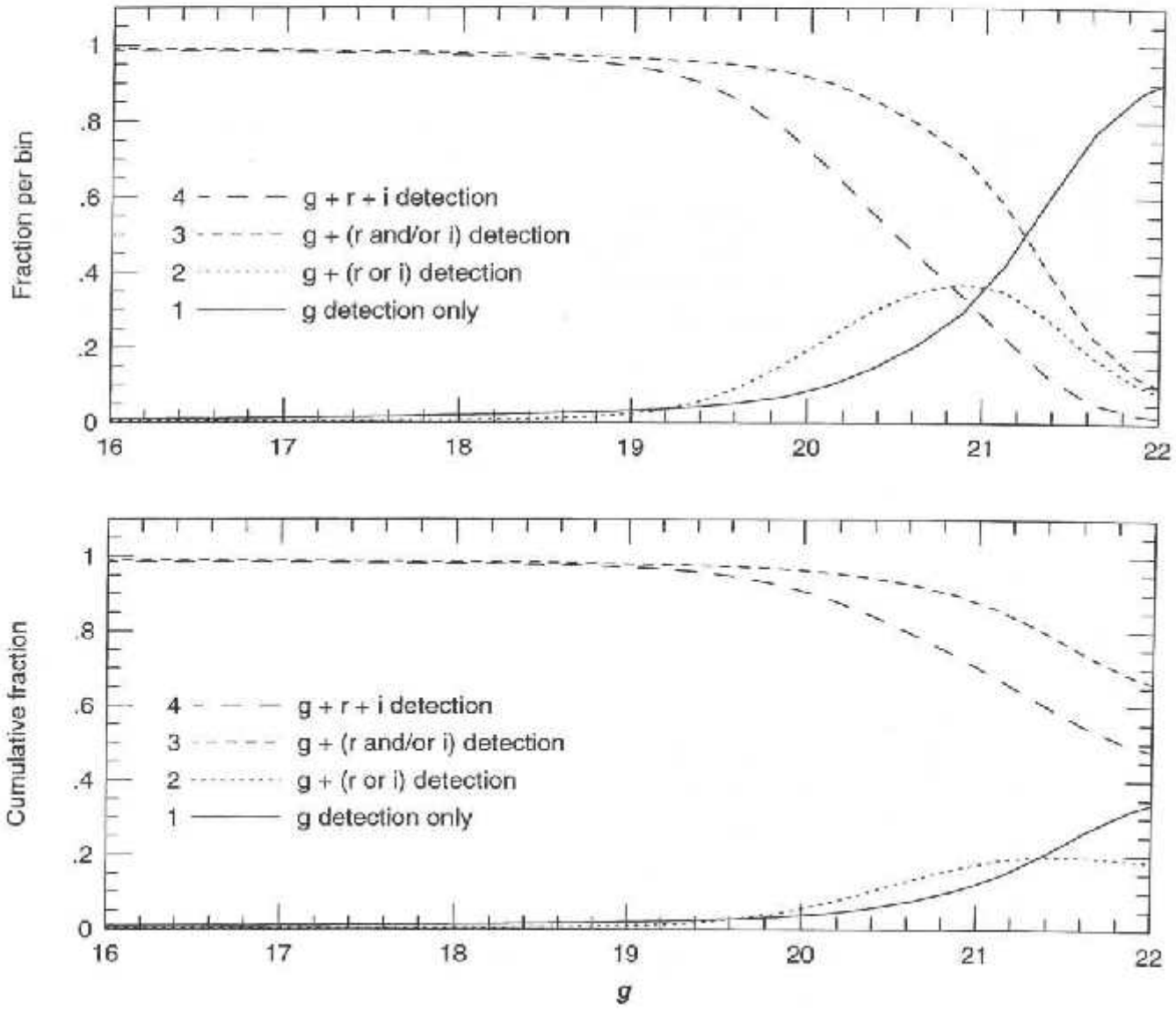


Figura 2.4: Análise interna do DPOSS adotando a banda g como referência.

detecções somente na banda r cresce rapidamente a partir de $m_r \sim 20$, ultrapassando o número de objetos em duas ou mais bandas (linha 3) em $m_r \sim 21$. Concluímos portanto que o limite máximo no qual podemos detectar aglomerados de galáxias com *redshifts* intermediários (capítulo 4) é $m_r \sim 21$. A partir deste limite, o número de objetos detectados somente numa banda cresce rapidamente, indicando um aumento na probabilidade de termos detecções espúrias. No entanto, ainda é importante que neste limite ($m_r \sim 21$) tenhamos

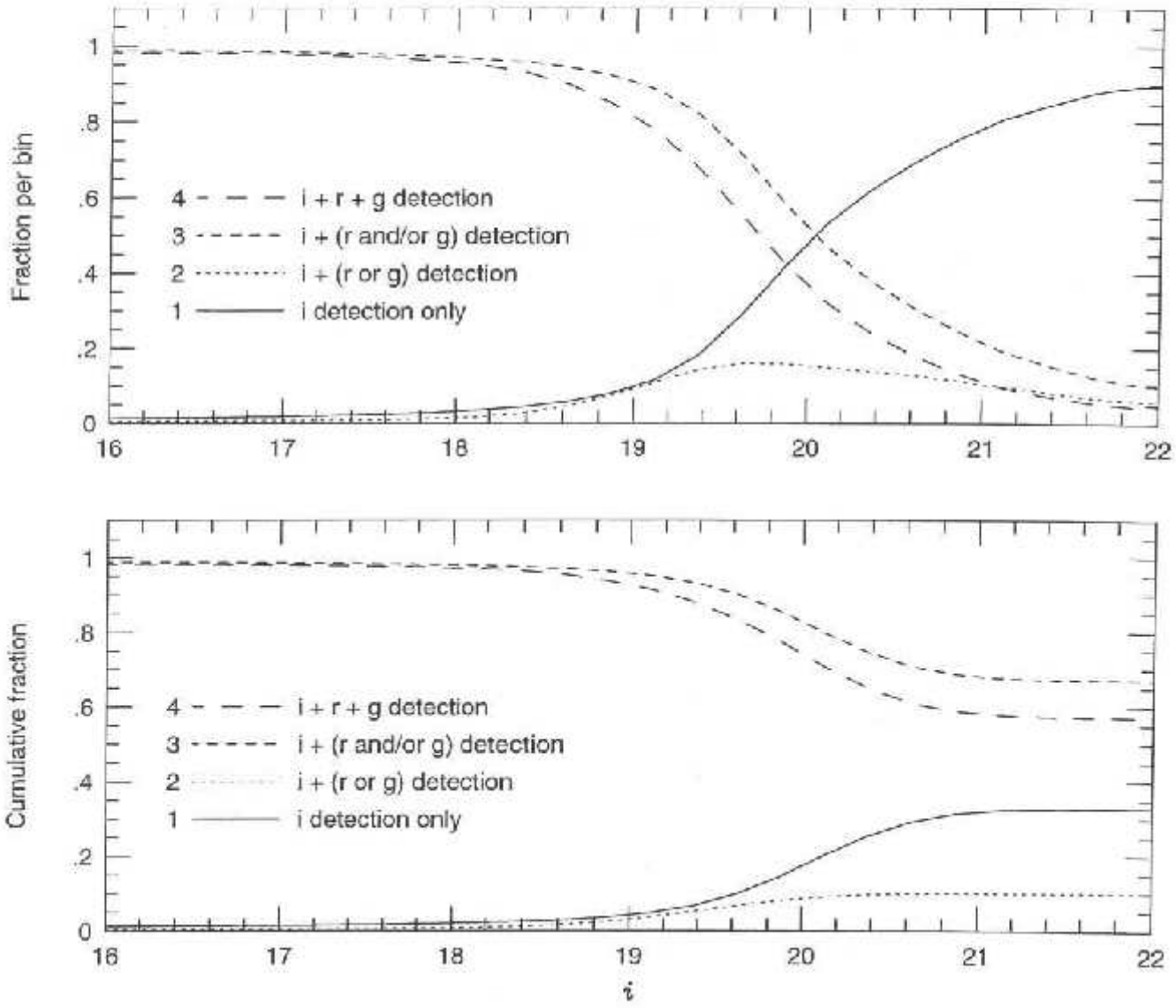


Figura 2.5: Análise interna do DPOSS tendo a banda i como referência.

uma estimativa do percentual de objetos detectados somente na banda r que são falsos e reais. Para tal, necessitamos de uma fonte externa para comparação. Na próxima seção apresentamos os resultados de uma comparação com o SDSS.

Nas figuras 2.4 e 2.5 apresentamos análises similares a da figura 2.3, exceto pelo fato de adotarmos as bandas g e i como bandas de referência, respectivamente. Fica evidente a baixa qualidade dos dados provenientes da banda i.

2.3 Comparação com o SDSS

Os dados de imageamento do *Sloan Digital Sky Survey* são obtidos em cinco bandas largas (u^*, g^*, r^*, i^*, z^*) centradas em 3561 Å, 4676 Å, 6176 Å, 7494 Å e 8873 Å, respectivamente. As observações são conduzidas em modo *drift-scan* num telescópio de 2.5 metros (York *et al.* 2000).

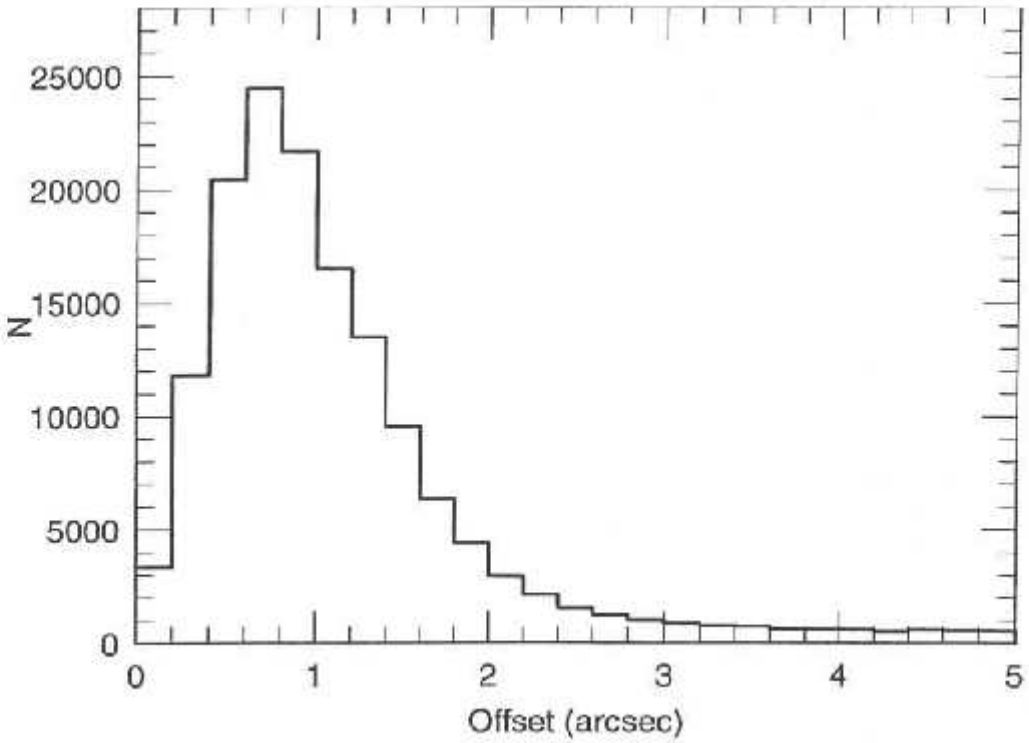


Figura 2.6: Distribuição das separações (em segundos de arco) de objetos provenientes dos dois catálogos (SDSS e DPOSS).

A comparação aqui efetuada utiliza-se de dados preliminares do SDSS cobrindo uma faixa equatorial em $163° \leq \alpha \leq 230°$ e $-0.65° \leq \delta \leq +0.65°$. Esta região é coberta por 14 campos do DPOSS. Nós ainda eliminamos um campo do DPOSS ($173° \leq \alpha < 178°$) para o qual não temos dados em três bandas, portanto também excluindo dados do SDSS nesta região. Os resultados apresentados abaixo utilizaram aproximadamente 2.6×10^6 e 1.2×10^6 objetos do SDSS e do DPOSS, respectivamente. Ao compararmos os dois catálogos adotamos um raio máximo de $5''$ e uma diferença máxima de 4 magnitudes. No entanto, para cada objeto de um

levantamento somente o objeto mais próximo no outro catálogo é mantido. A distribuição das separações (em segundos de arco) é exibida na figura 2.6, que apresenta um máximo em $\sim 0.7''$.

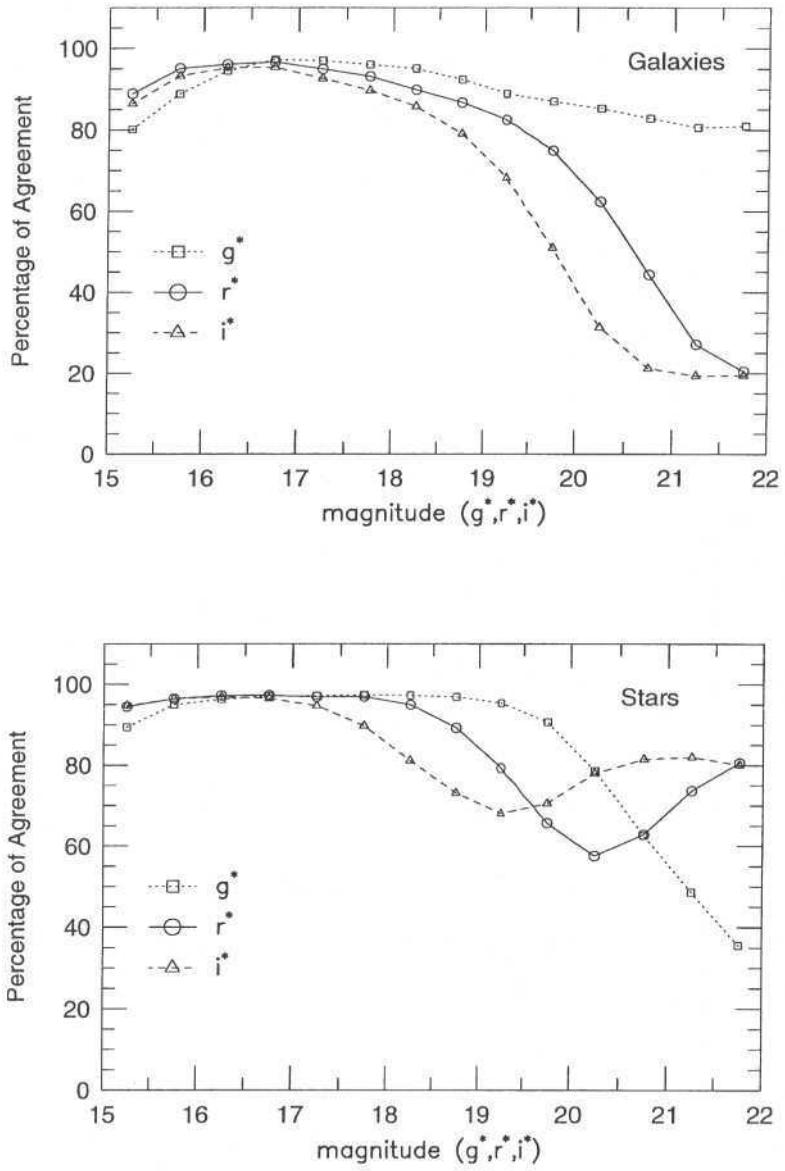


Figura 2.7: Eficiência em função da magnitude do classificador do DPOSS segundo o SDSS.

Primeiramente efetuamos uma comparação entre a separação estrela/galáxia no SDSS e no DPOSS. Detalhes da metodologia adotada para o DPOSS são encontrados em Odewahn

et al. (2003). Utilizamos o SDSS para avaliar a completeza do sistema de classificação adotado no DPOSS. As seguintes relações são utilizadas para estimarmos a taxa de sucesso do classificador adotado pelo DPOSS como função da magnitude:

$$S_{est} = \frac{N([SDSS = est] \bigcap [DPOSS = est])}{N(SDSS = est)} \tag{2.1}$$

$$S_{gal} = \frac{N([SDSS = gal] \bigcap [DPOSS = gal])}{N(SDSS = gal)} \tag{2.2}$$

Os resultados são apresentados na figura 2.7. Estes resultados confirmam a maior eficiência do classificador na banda azul (Odewahn *et al.* 2003), estando no nível de 95% em $g^* \sim 19$.

Na figura 2.8 é apresentada uma comparação das magnitudes do DPOSS na banda r e do SDSS (banda r^*). Para converter as magnitudes do SDSS para o sistema fotométrico do DPOSS, nós obtemos a seguinte transformação:

$$r = -0.20 + 1.02 \times r^* - 0.08 \times (g^* - r^*) \tag{2.3}$$

No diagrama de resíduos desta figura (painel inferior) podemos notar uma variação sistemática na transformação entre as bandas fotométricas do DPOSS e SDSS. Este efeito fica mais claro no regime de altos valores de fluxo, onde os resíduos são sistematicamente maior que zero. No entanto, a detecção de aglomerados (com qualquer desvio para o vermelho) em nada é afetada por este efeito.

Queremos agora utilizar o SDSS para confirmar os limites de detecção do DPOSS previstos na figura 2.3. Mais importante porém, é a utilização dos dados do SDSS para estimarmos a taxa de objetos espúrios nos limites do DPOSS. Na figura 2.9 apresentamos uma comparação entre as detecções do DPOSS e SDSS como função da magnitude (banda r). Nós assumimos que objetos detectados em duas bandas do SDSS são reais e portanto estes são usados como referência. Em ambos os painéis, de baixo para cima, a 1ª linha indica os objetos detectados somente na banda r do DPOSS sem contrapartida simultânea nas bandas g^* e r^* do SDSS. A segunda linha representa os objetos detectados na banda r e pelo menos uma outra banda do DPOSS sem par no SDSS. A 3ª linha mostra as detecções na banda r do DPOSS (podendo ocorrer ou não em outras bandas) sem contrapartida no SDSS. As linhas 4, 5 e 6 representam objetos selecionados da mesma maneira que nas linhas 1, 2 e 3, respectivamente; porém com uma contrapartida em duas bandas do SDSS (g^* e r^*). A soma

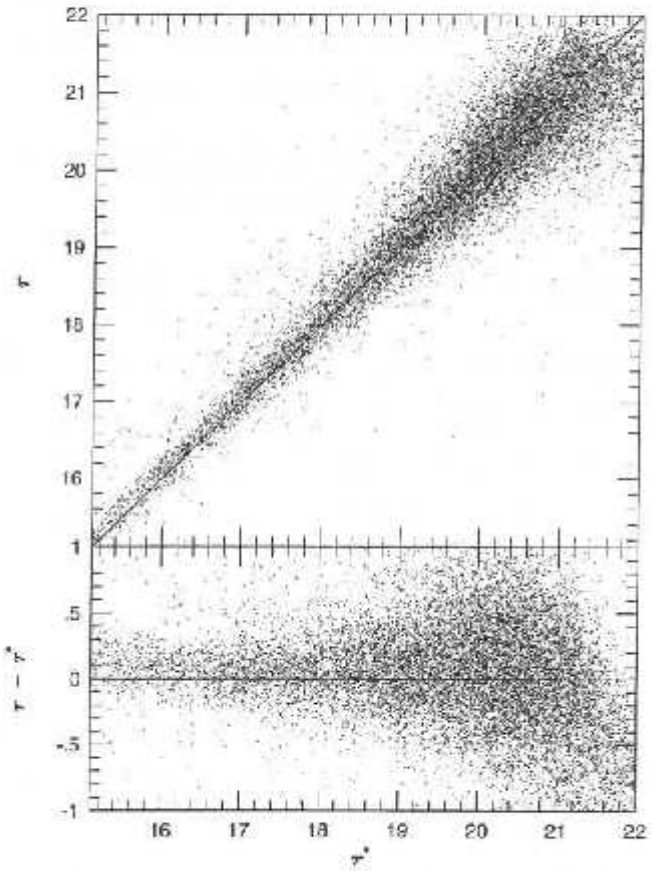


Figura 2.8: Comparação entre as magnitudes da banda r do DPOSS e da banda r^* do SDSS.

das linhas 1 e 2 resulta na 3ª, ao passo que a 6ª é o resultado da combinação das linhas 4 e 5.

A linha 6 indica que em torno de $m_r = 20$, cerca de 92% das detecções na banda r do DPOSS existem em duas bandas do SDSS, e portanto assumimos que sejam reais. Este resultado está em bom acordo com o que foi visto na figura 2.3 (linha 3). As linhas 1 e 4 somadas representam o total de detecções do DPOSS somente na banda r, sendo que a linha 1 indica objetos provavelmente espúrios (não encontrados no SDSS e a linha 4 representa objetos reais (identificados no SDSS). Em qualquer intervalo de magnitude (até o limite de $m_r = 21.5$) estas duas linhas são equivalentes, significando que aproximadamente metade dos objetos detectados somente na banda r são reais, sendo a outra metade possivelmente constituída de objetos falsos. No entanto, para a detecção de aglomerados de galáxias usando

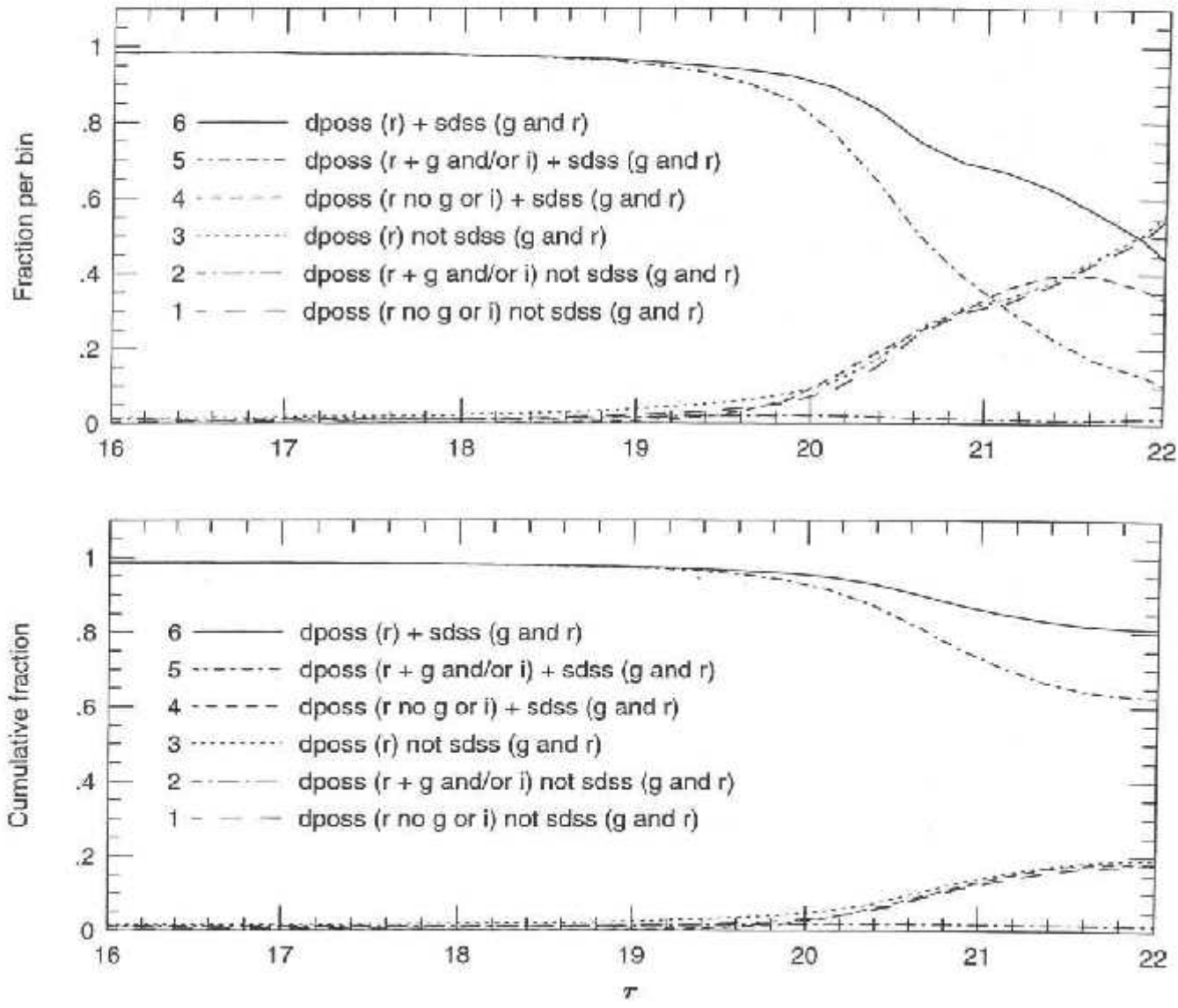


Figura 2.9: Comparação com o SDSS usando a banda r como referência.

objetos até $m_r = 21$ (capítulo 4) não podemos simplesmente ignorar detecções em uma única banda. Ao decompormos a linha 6 nas linhas 4 e 5 (em $m_r \sim 21$), percebemos que cerca de metade das detecções reais na banda r ocorrem somente neste filtro.

Uma outra forma de compararmos os dois levantamentos é começarmos a análise pelo SDSS, determinando a fração de objetos recuperados pelo DPOSS em função da magnitude. Os resultados desta comparação são exibidos na figura 2.10. A linha 4 (soma da 2ª e 3ª)

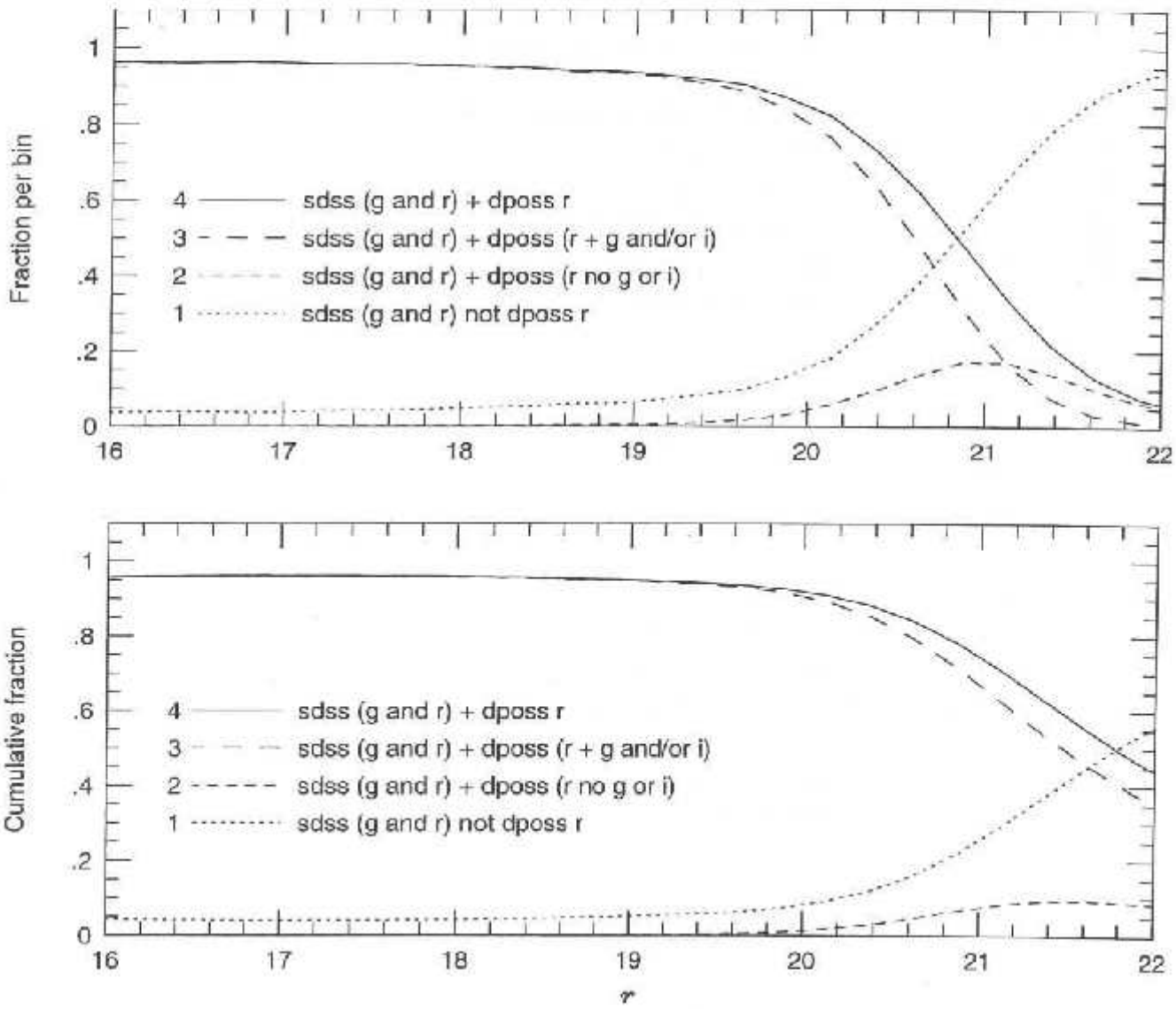


Figura 2.10: Comparação entre o SDSS e o DPOSS usando a banda r como referência.

mostra a fração de objetos comuns como função da magnitude, ao passo que a linha 1 exibe a fração de objetos detectados somente pelo SDSS.

Um fato que chama a atenção é a taxa aproximadamente constante ($\sim 5\%$) de objetos detectados somente pelo SDSS em $16 < m_r < 19$. Neste intervalo de magnitude seria de se esperar $\sim 100\%$ de acordo entre os dois levantamentos. Decidimos então investigar estes objetos em detalhe. Selecionamos todos os objetos que a princípio foram detectados somente

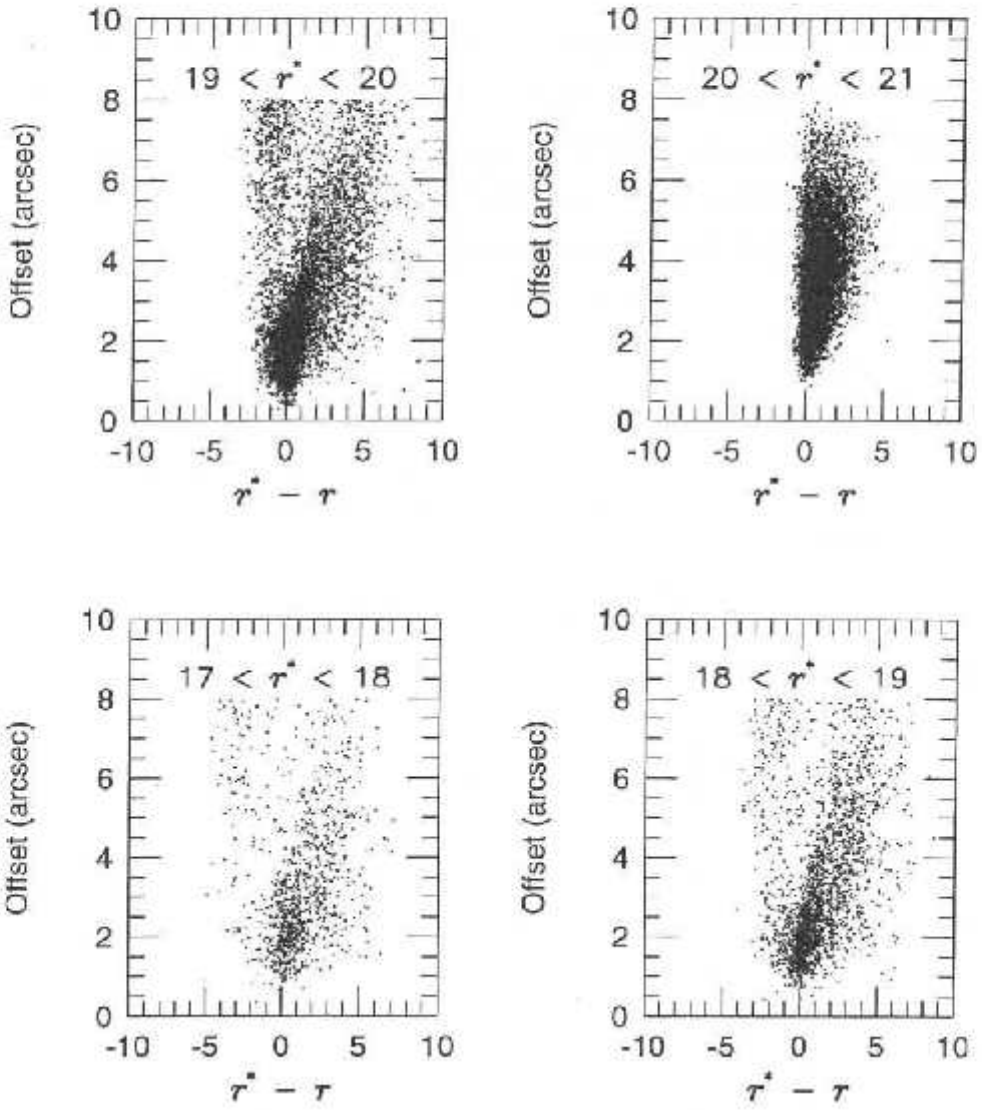


Figura 2.11: Objetos a princípio detectados somente pelo SDSS, mas com um companheiro brilhante no DPOSS. A figura apresenta a distribuição das separações (segundos de arco) *vs* diferença de magnitude, em quatro intervalos de magnitude.

pelo SDSS e os comparamos novamente com o DPOSS, procurando ainda um possível par. O resultado desta nova comparação é exibido na figura 2.11 para quatro intervalos de magnitude diferentes. Para estes objetos do SDSS, nós mostramos a separação (em segundos de arco) até o objeto mais próximo no DPOSS (em geral, mais brilhante), em função da diferença de magnitude. Fica evidente que a maioria destes objetos possui um companheiro brilhante no

DPOSS, com uma separação tipicamente $> 1''$. Neste caso em que temos dois objetos muito próximos (sendo um muito mais brilhante do que o outro), o DPOSS falha na distinção dos dois objetos, recuperando somente o mais brilhante. O SDSS, dotado de imagens de mais alta resolução consegue separá-los. Obviamente o *software* empregado para detecção e separação dos objetos é de fundmental importância para este resultado.

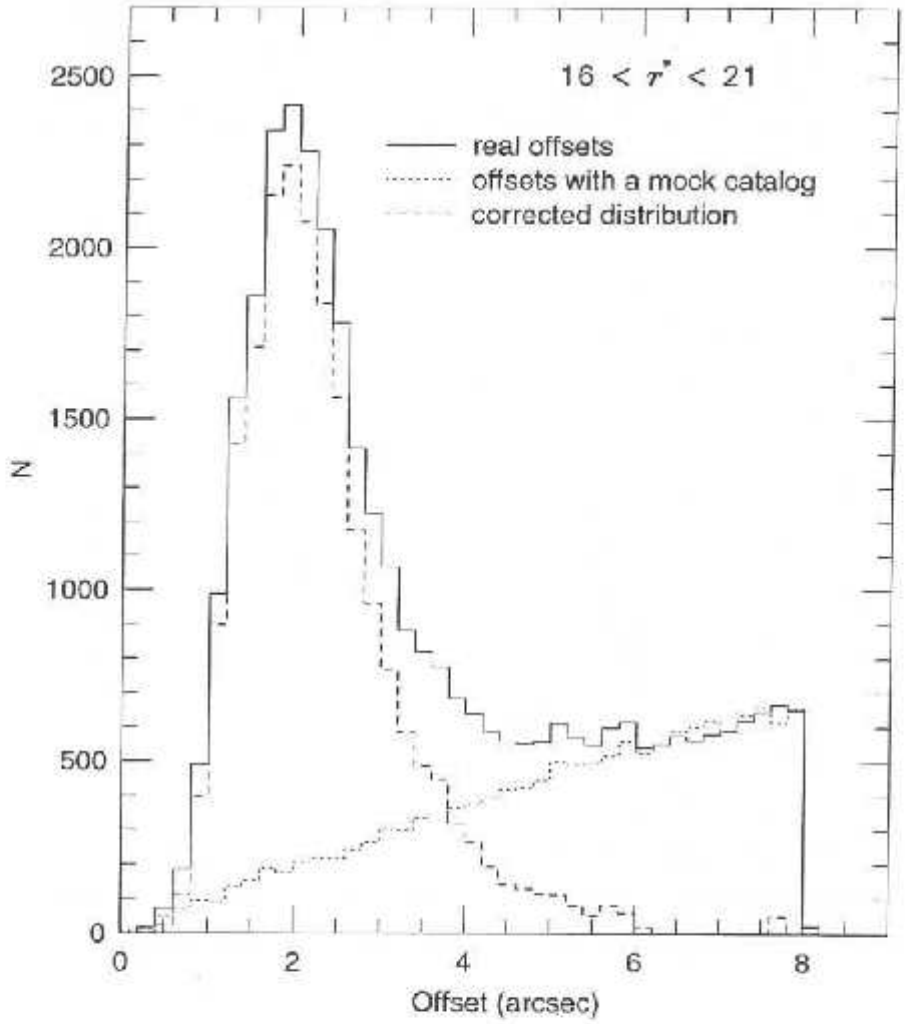


Figura 2.12: Distribuição das separações entre objetos do SDSS e objetos brilhantes do DPOSS (linha contínua). A linha pontilhada mostra a distribuição utilizando um catálogo falso, enquanto a linha tracejada representa a distribuição corrigida.

Na figura 2.12 exibimos a distribuição das separações (em segundos de arco) entre os objetos do SDSS e os companheiros brilhantes no DPOSS. O resultado cobre o intervalo

de $16 < r^* < 21$. Esta distribuição é exibida como linha contínua, ao passo que uma comparação do SDSS com um catálogo falso (simplesmente transladamos o DPOSS por 30″ em α e δ) é exibido pela linha pontilhada. A distribuição corrigida é apresentada pela linha tracejada. Esta distribuição apresenta um máximo em $\sim 2''$. Pelo menos cerca de 50% dos objetos detectados somente pelo SDSS em $16 < r^* < 19$ são justificados por este efeito de duplicidade. É interessante ressaltar que este problema afeta qualquer levantamento, e não somente o DPOSS. Uma possível comparação do SDSS com dados do telescópio espacial *Hubble*, revelaria a inoperânica do primeiro em separar objetos muito próximos.

Capítulo 3

Aglomerados de Galáxias no Universo *Local* ($< z > \sim 0.15$)[1]

Resumo

Apresentamos neste trabalho um novo catálogo de candidatos a aglomerados para o hemisfério norte, contendo $\sim$ 8000 sistemas. Este catálogo é construído com critérios de seleção objetivos, tendo como base catálogos de galáxias do segundo levantamento fotográfico do Observatório Palomar. Os resultados aqui apresentados são derivados dos campos com melhor calibração fotométrica do DPOSS, limitados por latitude galáctica ($b > 30^{o}$), totalizando uma área de 5800 graus quadrados em torno do pólo norte galáctico.

Candidatos a aglomerados são selecionados em regiões de alta densidade de galáxias. O processo de seleção é baseado num méTodo de *kernel* adaptativo para geração de mapas de densidade. Utiliza-se o programa SExtractor para identificação dos picos de densidade presentes nestes mapas. Para cada campo do DPOSS o programa de *kernel* adaptativo é utilizado para construção em duas etapas de um mapa de densidade de galáxias. Na primeira etapa, utilizamos um *kernel* de tamanho fixo para obtermos uma estimativa da densidade de galáxias em cada ponto do campo. Na segunda etapa, um *kernel* de tamanho variável é empregado para gerar o mapa de densidade final. O tamanho deste *kernel* variável depende da densidade local, dada pela estimativa obtida na primeira etapa. Em seguida, utilizamos o SExtractor para detecção de regiões de alta densidade de galáxias em cada mapa.

[1]Os resultados exibidos neste capítulo são provenientes do artigo "The Northern Sky Optical Cluster Survey II: An Objective Cluster Catalog for 5800 Square Degrees", Gal, R. R., de Carvalho, R. R., Lopes, P. A. A., Djorgovski, S. G., Brunner, R. J., Mahabal, A., Odewahn, S. C. 2003a, AJ, 125, 2064.

Os dois parâmetros que buscamos optimizar no programa SExtractor são a área mínima e o nível de detecção (*threshold level* em inglês). O segundo parâmetro reflete o número mínimo de galáxias que um excesso de densidade deve possuir acima da flutuação (*rms*) da distribuição de fundo para ser considerado como um candidato a aglomerado. O processo de optimização é baseado em simulações da distribuição de galáxias de fundo. Estas mesmas simulações fornecem uma estimativa da taxa de contaminação presente em nosso catálogo. Nós ainda fazemos uso de aglomerados de galáxias artificiais para estimarmos a taxa de completeza. A optimização e estimativas de contaminação e completeza são efetuadas de maneira indepentente para cada campo de nossa amostra. Este fato é extremamente importante para geração de um catálogo para diversos campos espalhados no céu, pois campos de diferentes densidade serão optimizados em diferentes níveis.

Também fazemos uso de informações de cor e magnitude para estimarmos o desvio para o vermelho e a riqueza de cada candidato. Nós obtemos uma relação empírica para estimarmos o desvio para o vermelho com base na magnitude média e na cor mediana de galáxias contidas[2] numa região de 1.0 h^{-1} Mpc em torno de cada candidato. A comparação dos desvios para o vermelho fotométricos com espectroscópicos para 369 aglomerados resulta numa dispersão de $\Delta z = 0.033$. O desvio para vermelho mediano da amostra total é $z_{med} \sim 0.16$. As estimativas de riqueza são baseadas na contagem de galáxias entre $M^* - 1$ e $M^* + 2$ num raio de 1.0 h^{-1} Mpc. Estas estimativas apresentam bom acordo com medidas apresentadas em catálogos recentes. A riqueza mediana de nossa amostra é $N_{gals,med} = 31$, correspondendo a R = 0 (classe de riqueza de Abell).

Finalmente, um grande levantamento espectroscópico foi realizado para confirmarmos as estimativas da taxa de contaminação provenientes das simulações. Este levantamento é realizado em dois campos do DPOSS, contendo 1245 galáxias com desvio para o vermelho medido, em 79 candidatos e 20 regiões de densidade moderada que não constituem candidatos de nossa lista. O desvio para o vermelho mediano desta amostra é $z_{med} \sim 0.2$, em bom acordo com as estimativas fotométricas da amostra total de candidatos. Sendo conservadores, nós estimamos que **pelo menos** 80% dos candidatos com dados espectroscópicos são reais, o que está em bom acordo com as estimativas de contaminação de $\sim 10\%$ baseadas nas simulações.

[2]Estas estimativas levam em conta uma correção para as distribuições de cor e magnitude de galáxias de fundo.

3.1 Introduction

Over the past two decades, large numbers of astronomers have expended a great deal of effort on the study of galaxy clusters. Motivations for these works cover a wide range of astrophysical topics, from the study of galaxy evolution in dense environments, through the search for dark matter, the measurement of the cluster mass function and its comparison to theoretical predictions, and on to the characterization of large scale structure in the universe. Despite this labor, understanding the underlying physics has often proved enigmatic, due to the complex physical nature of clusters, which is only compounded by the lack of an objectively selected, well-characterized statistical sample.

To help remedy this situation, and provide a basis for future studies, we have undertaken the creation of a new cluster catalog which fulfills a number of fundamental criteria:

1. Cluster detection should be performed by an objective, automated algorithm to minimize human biases and fatigue.

2. The algorithm utilized should impose minimal constraints on the physical properties of the clusters, to avoid selection biases.

3. The sample selection function must be well-understood, in terms of both completeness and contamination, as a function of both redshift and richness. The effects of varying the cluster model on the determination of these functions must also be known.

4. The catalog should provide basic physical properties for all the detected clusters, such that specific subsamples can be selected for future study.

In this paper, we describe how we generate a cluster sample that meets the above four criteria, and provide the final catalog for 5800□°. We discuss past cluster surveys and their limitations in §3.2. In §3.3 we briefly review the DPOSS survey (Djorgovski *et al.* 2003) which provides the basis for our catalog, while §3.4 describes the cluster detection technique (a modified version of that presented in Gal *et al.* 2000a, hereafter Paper I). §3.5 describes simulations used both to optimize the detection algorithm, and define its statistical properties, including selection functions in richness and redshift for each plate. The results of these simulations are compared to extensive spectroscopic follow-up observations. Our

photometric redshift estimator and richness measure are described in §3.6. An overview of the final catalog is provided in §3.7. We conclude with a brief discussion of future extensions of this survey (to cover the South Galactic Cap area, and the remaining areas of the NGP), and discuss some projects we have undertaken utilizing this sample. We use a cosmology with $H_0 = 67$ km s^{-1} Mpc^{-1} and $q_0 = 0.5$ throughout; other choices for these parameters have no significant effects on our results except for the poorest and most distant clusters.

3.2 A Brief History of Cluster Surveys

Surveys for galaxy clusters have only recently benefited from the automation afforded by computers. The earliest surveys relied on visual inspection of vast numbers of photographic plates, usually by a single astronomer. As early as 1938, Zwicky discussed such a survey based on plates from the 18″ Schmidt telescope at Palomar, and results were presented in 1942 by both Zwicky and Katz & Mulders. Even then, Zwicky, with his typical prescience, noted that elliptical galaxies are much more strongly clustered than late-type galaxies. However, the true pioneering work in this field did not come until 1957, upon the publication of a catalog of galaxy clusters produced by George Abell as his Caltech Ph.D. thesis, which appeared in the literature the following year (Abell 1958). Zwicky and collaborators followed suit a decade later, with their voluminous Catalogue of Galaxies and of Clusters of Galaxies (Zwicky, Herzog & Wild 1968).

Abell used the red plates of the first National Geographic-Palomar Observatory Sky Survey. In selecting clusters, Abell applied a number of criteria in an attempt to produce a fairly homogeneous catalog. He required a minimum number of galaxies within two magnitudes of the third brightest galaxy in a cluster, a fixed physical area within which galaxies were to be counted, a maximum and minimum distance (redshift) to the clusters, and a minimum galactic latitude to avoid obscuration by interstellar dust. The resulting catalog, consisting of 1,682 clusters in the statistical sample, remained the predominant such resource until 1989. In that year, Abell, Corwin & Olowin (hereafter ACO) published an improved and expanded catalog, now including the Southern sky. These catalogs have been the foundation for many cosmological studies over the last four decades, despite serious questions about their reliability (which are addressed later in this section). Some other catalogs based on

plate material have also been produced, such as Shectman (1985), from the galaxy counts of Shane & Wirtanen (1954), and a search for more distant clusters carried out on plates from the Palomar 200″ by Gunn, Hoessel & Oke (1986).

Only in the past ten years has it become possible to utilize the objectivity of computational algorithms in the search for galaxy clusters. These more modern studies required that plates be digitized, so that the data are in machine readable form. Alternatively, the data had to be digital in origin, coming from CCD cameras. Unfortunately, this latter option provided only small area coverage, so the hybrid technology of digitized plate surveys blossomed into a cottage industry, with numerous catalogs being produced in the past decade. All such catalogs relied on two fundamental data sets: the Southern Sky Survey plates, scanned with the Automatic Plate Measuring (APM) machine (Maddox, Efstathiou & Sutherland 1990) or COSMOS scanner (to produce the Edinburgh/Durham Southern Galaxy Catalog / EDSGC, Heydon-Dumbleton, Collins & MacGillivray 1989), and the POSS-I, scanned by the APS group (Pennington *et al.* 1993). The first objective catalog produced was the Edinburgh/Durham Cluster Catalog (EDCC, Lumsden, Nichol, Collins & Guzzo 1992), which covered 0.5 steradians ($\sim 1,600$ square degrees) around the South Galactic Pole (SGP). Later, the APM cluster catalog was created by applying Abell-like criteria to select overdensities from the galaxy catalogs, and is discussed in detail in Dalton, Maddox, Sutherland & Efstathiou (1997). The work by Odewahn & Aldering (1995), based on the POSS-I, provided a Northern sky example of such a catalog, while utilizing additional information (namely galaxy morphology). Some initial work on this problem, using higher quality POSS-II data, was performed by Picard (1991) in his thesis.

In addition to these hybrid photo-digital surveys, smaller areas, to much higher redshift, have been covered by numerous deep CCD imaging surveys. Notable examples include the Palomar Distant Cluster Survey (PDCS, Postman *et al.* 1996), the ESO Imaging Survey (EIS, Olsen *et al.* 1999; Lobo *et al.* 2000), Gonzalez, Zaritsky, Dalcanton, & Nelson (2001), KPNO/Deeprange (Postman, Lauer, Oegerle, & Donahue 2002), and many others. None of these surveys provide the angular coverage necessary for large-scale structure and cosmology studies, and are typically designed to find rich clusters at high redshift. Only the Sloan Digital Sky Survey (SDSS, York *et al.* 2000) will provide large area, moderately deep CCD coverage. Cluster surveys from the SDSS, including those of Annis *et al.* (1999), Kim *et al.*

(2002) and Goto *et al.* (2002), are only now appearing, and will cover a much smaller area until the photometric survey is completed.

Despite these efforts, one thing is still missing: a catalog of galaxy clusters, produced by objective means, that is at least as deep as the Southern surveys, but which covers the Northern sky. This paper is intended to provide such a catalog.

We note that there have also been many surveys for galaxy clusters at other wavelengths, most notably in the X-ray. All-sky surveys, such as RASS (Collins *et al.* 2000; Ebeling *et al.* 2000; Böhringer *et al.* 2000) and EMSS (Gioia & Luppino 1994), as well as pointed ROSAT observations (Scharf *et al.* 1997; omer *et al.* 2000), have been used to produce cluster catalogs at X-ray wavelengths. Future Sunyaev- Zel'dovich effect (SZ) surveys will also generate extremely important datasets. However, the relationship between optically and X-ray selected clusters is still not fully understood, so an optically selected catalog is essential.

3.2.1 Limitations of Existing Catalogs

Most of the optical studies to date have been limited by the statistical quality of the available cluster samples. For instance, the Abell catalog suffers from a non-objective selection process, poorer plate material, a bias toward centrally-concentrated clusters (especially those with cD galaxies), a relatively low redshift cutoff ($z \sim 0.15$; Bahcall & Soneira 1983), and strong plate-to-plate sensitivity variations. Still, many far-reaching cosmological conclusions have been drawn from it (*i.e.,* Bahcall & West 1992), although later studies have sometimes shown these to be flawed. Photometric errors and other inhomogeneities in the Abell catalog (Sutherland & Efstathiou 1991; Efstathiou *et al.* 1992), as well as projection effects (Lucey 1983; Katgert *et al.* 1996) are serious and difficult-to-quantify issues. These effects have resulted in discrepant results on the correlation function (Bahcall & Soneira 1983; Dalton, Maddox, Sutherland & Efstathiou 1997; Miller *et al.* 1999), and later attempts to disentangle these issues relied on models to decontaminate the catalog (Sutherland 1988; Olivier *et al.* 1990). Also, Sutherland & Efstathiou (1991) find a discrepancy between the angular and spatial correlation function of the Abell catalog, which is not found in the APM catalog (Dalton, Maddox, Sutherland & Efstathiou 1997). The extent of these effects is also surprisingly unknown; measures of completeness and contamination in the Abell catalog disagree

by factors of a few. For instance, Miller *et al.* (1999) claim that under- or over-estimation of richness is not a significant problem, whereas van Haarlem, Frenk & White (1997) suggest that one-third of Abell clusters have incorrect richnesses, and that one-third of rich ($R \geq 1$) clusters are missed. The largest study of $R \geq 1$ clusters (Miller *et al.* 1999) suggests projection effects are not of great concern for the Abell catalog; however, completeness cannot be gauged without deeper samples. Unfortunately, some of these problems will plague any optically selected cluster sample, including our own, but objective selection criteria and a strong statistical understanding of the catalog can mitigate their effects.

Even some of the other objective catalogs preceding ours have their drawbacks. 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 in the redshift range probed ($z < 0.3$). Comparable surveys, such as the EDCC, already find a factor of two higher space density of clusters than Abell, and more sensitive CCD surveys find as many as five times more, although these results may be due to strong detection efficiency differences at lower richnesses. Other more recent surveys, such as the EDCCII (Bramel, Nichol & Pope 2000) have not yet achieved the area coverage of DPOSS. Additionally, the survey presented here utilizes at least one color (two filters) for photometric redshifts, and a significantly increased amount of CCD photometric calibration data.

3.3 DPOSS: A Brief Overview

The POSS-II (Reid *et al.* 1991) covers the entire northern sky ($\delta > -3°$) with 894 overlapping fields (each 6.5° square, with 5° spacings), and, unlike the old POSS-I, has no gaps in the coverage. Approximately half of the survey area is covered at least twice in each band, due to plate overlaps. Plates are taken in three bands: blue-green, IIIa-J + GG395, $\lambda_{\rm eff} \sim 480nm$; red, IIIa-$F$ + RG610, $\lambda_{\rm eff} \sim 650nm$; and very near-IR, IV-$N$ + RG9, $\lambda_{\rm eff} \sim 850nm$. Typical limiting magnitudes reached are $g_J \sim 21.5$, $r_F \sim 21.0$, and $i_N \sim 20.3$, *i.e.*, $\sim 1^m - 1.5^m$ deeper than POSS-I. The image quality is improved relative to POSS-I, and is comparable to the Southern photographic sky surveys.

The original survey plates are digitized at STScI, using modified PDS scanners (Lasker

et al. 1996). The plates are scanned with 15μ (1.0″) pixels, in rasters of 23,040 square, giving $\sim$ 1 GB/plate, or $\sim$ 3 TB of pixel data total for the entire digital survey (DPOSS). Currently, astrometric solutions, provided by STScI, are good to $rms \sim 0.5''$.

An extensive effort to process, calibrate, and catalog the scans, with the detection of all objects down to the survey limit, and star/galaxy classifications accurate to 90% or better down to $\sim 1^m$ above the detection limit, has been undertaken over the past several years. Object detection and photometry is performed by SKICAT, a software system developed for this purpose (Weir *et al.* 1995a,b,c), incorporating standard astronomical image processing packages, a commercial Sybase database, as well as a number of artificial intelligence and machine learning modules. Using this code, we measure $\sim$ 60 attributes per object on each plate. Nearly all plates at $|b| > 10^\circ$ have been processed into catalogs. The catalogs are photometrically calibrated using extensive CCD sequences, with typical rms magnitude errors of 0.25^m at $m_r = 19.5$; see Gal *et al.* (2003c) for details of the calibration procedure. A fraction of this error is due to systematic offsets in the photometric zero points between plates. Gal *et al.* (2003c) show that the mean zero-point error is negligible, but has a 1σ scatter of 0.07^m in the r-band, which can produce significant plate-to-plate depth variance. Our solution to this potential problem is discussed in §3.4. Star-galaxy separation is performed using a combination of FOCAS, neural networks, and decision trees, maintaining an accuracy of $> 90\%$ at $m_r < 19.5$ (Odewahn *et al.* 2003).

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 the $g - r$ colors of our galaxies; this also reduces the likelihood that any object is a spurious detection. Areas on the plate containing saturated objects are masked. These areas often contain large numbers of falsely identified galaxies, as the plate processing software, tuned to find faint objects, handles large, bright objects improperly.

In this paper we utilize a total of 237 plates with good calibration ($> 1,000$ calibrating galaxies per plate from the CCD fields), at $b > 30^\circ$, in the North Galactic Cap region. The distribution of DPOSS fields (as well as the detected clusters) can be seen in Figure 3.14. The total area coverage is 5800 square degrees. In addition, we have run our procedures on one field (475) from the SGP region of DPOSS, which is used to obtain additional spectroscopic

follow-up (see §3.5.3).

3.4 The Detection Algorithm

For this survey we use a modified version of the detection strategy first described in Paper I. We urge the reader to review Paper I for details of the adaptive kernel technique; we do not repeat the details here. In that earlier work, a color selection was first applied to the galaxy catalogs, after which we used an adaptive kernel technique (Silverman 1986) to produce galaxy density maps, and a bootstrap technique was used to create significance maps. The FOCAS object detection algorithm (Jarvis & Tyson 1981) was then used to detect density enhancements in the significance maps, which we identified as candidate clusters. We generated cluster catalogs for only two sky survey fields, for which we performed extensive follow-up imaging and spectroscopy. Since the publication of that work, we have gained a greater understanding of the plate-to-plate variance in DPOSS, significantly modified our photometric calibration techniques, and learned a great deal about our detection algorithms as a result of the additional data obtained. This has led us to significantly modify our procedures, while maintaining the underlying principles and methodology.

In Paper I, we utilized galaxies with magnitudes $m_r < 20.0$. We have found that the random photometric errors, plate-to-plate zero-point offsets, and classification accuracy, can all reach unacceptable levels at that magnitude limit. Although some plates perform well to this limit, we found that imposing a slightly brighter magnitude limit, $m_r \leq 19.5$, and requiring a g detection, produces a significantly more uniform galaxy catalog, without greatly sacrificing depth. Our intent with the current survey is to produce a catalog with good uniformity; therefore, we have elected to be rather conservative in the data utilized. Future work will use fainter objects to create a higher redshift cluster catalog which may not be suitable for large scale structure work, but which will provide a useful sample for cluster studies over a larger distance/time baseline. Additionally, in Paper I we imposed a liberal selection in $g - r$ color; we no longer apply this cut for the same reasons that we adopt the shallower magnitude limit. Because photometry in the g and r bands is completely independent, the use of a color cut effectively increases the pistoning due to zero-point offsets among plates by a factor $\sqrt{2}$. Finally, star/galaxy separation for DPOSS has been

improved from the version used to produce the input data for Paper I, resulting in purer and more uniform galaxy samples being generated from the individual plate catalogs. The final catalog of objects now used in cluster detection therefore consists of all galaxies, with detections in both the g and r bands, having $15.0 \leq m_r \leq 19.5$. A typical galaxy catalog for a single plate contains $\sim 27,000$ galaxies over an area of $\sim 34\square^\circ$, resulting in a mean galaxy density of 0.79×10^3 galaxies per square degree.

The resulting galaxy catalog is used as input to the adaptive kernel (AK) density mapping algorithm. As described in Paper I, this technique uses a two-stage process to produce a density map. First, it produces an initial estimate of the galaxy density at each point in the map, which is then used to apply a smoothing kernel whose size changes as a function of the local density, with a smaller kernel at higher density. We refer the reader to Paper I for a more detailed description. As before, we generate our maps with 1′ pixels. However, rather than the 900″ kernel used earlier, we now use a significantly smaller kernel of 500″ radius. Based on the simulations discussed in the next section, we found that this smaller kernel prevents over-smoothing the cores of higher-redshift ($z \sim 0.3$) clusters, while avoiding fragmentation of most low redshift ($z \sim 0.08$) clusters. The effect of varying the initial smoothing window is demonstrated in Figure 3.1. In this figure, we have placed four simulated clusters into a simulated background, representing the expected range of detectability in our survey. There are two clusters at low-z (0.08), and two at high-z (0.24), with one poor and one rich cluster at each redshift (100 and 333 total members, respectively). Of the 100 (333) total cluster galaxies, only 54 (186) are brighter than our magnitude limit ($m_r \leq 19.5$) at $z = 0.08$, and 9 (17) at $z = 0.24$. The corresponding richnesses (N_{gals}, see §3.6.2) are 80 and 25. The initial smoothing window is varied from 300″ to 800″ in 100″ steps. This figure clearly shows the segmentation of rich, low-z clusters by small kernels, and the smoothing away of high-z clusters by large kernels.

Unlike Paper I, we do not perform bootstrap tests to produce significance maps. Cluster candidates are detected directly from the AK map of the actual data. In the current work, we use SExtractor (Bertin & Arnouts 1996) to perform object detection (rather than FOCAS, as we did originally). We have found that SExtractor is both faster and more easily configured to meet our needs. For each plate, we find the set of SExtractor detection parameters (a pair of threshold level and minimum area) which produces an acceptable level of contamination

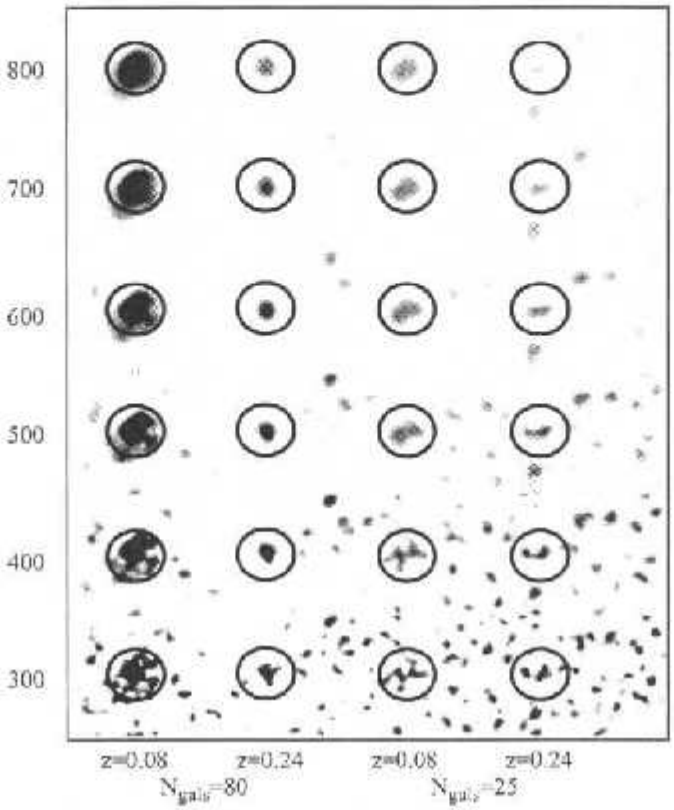


Figure 3.1: The effect of varying the initial smoothing window on cluster appearance. Each panel contains a simulated background with four simulated clusters, as described in the text. The smoothing kernel ranges in size from 300″ for the upper panel, to 800″ for the lower panel, in 100″ increments.

by false clusters (10%). The simulations used to determine these parameters are discussed in §3.5.1. Once the set of optimal detection parameters for each plate is determined, further simulations are used to assess the catalog completeness as a function of redshift and richness, as discussed in §3.5.2. We have opted to vary detection parameters between plates to maintain a constant level of contamination, even though the completeness may change. Due to the inhomogeneous nature of plate data, as well as large scale structure, setting a fixed threshold (in galaxy surface density) for all plates can lead to significant variance in both completeness *and* contamination between plates. By adjusting the parameters for each plate, we can minimize the variance of one of these quantities, and estimate the impact on the other. Because contamination is more easily quantified (as an integrated rate, independ-

Table 3.1: Percentage of Clusters Detected vs. Number of Maps

Percent	N(maps)	Percent	N(maps)
99.9	1	92.8	6
99.7	2	89.7	7
98.9	3	68.6	8
98.5	4	66.3	9
96.3	5	62.1	10

ent of richness and redshift) as the fraction of detections that are false, we have chosen to adjust the parameters to maintain a fixed contamination rate. To allow large scale structure work, we provide a table (see Table 3.2) for each plate, including the RA/Dec boundaries used and the completeness function, as described in §3.5.2.

Additionally, to assess the impact of photometric zero-point errors, and remove those cluster candidates whose detection is sensitive to these errors, we generate a set of ten galaxy catalogs for each plate with random zero-point offsets added to the r-band magnitude, drawn from the known photometric error distribution for DPOSS given in Gal *et al.* (2003c). Like the original data, these catalogs are clipped at $15.0 \leq m_r \leq 19.5$, and an AK map generated for each one. Cluster detection is then performed using the same parameters as for the original map, and the resulting lists of cluster candidates are compared. Using the photometric redshifts (described in §3.6) for each candidate from the original map, we match clusters within a 300kpc projected radius. In Table 3.1 we present the fraction of total candidates that appear in a given number of maps. Only those candidates which appear in the original catalog, and more than 7 of the 10 zero-point-error-added AK maps are kept in the final catalog; the number of simulated maps a cluster is found in (out of a maximum of ten) is provided in the final cluster catalog, and can be used to select increasingly certain subsamples. We use seven detections among the Monte-Carlo maps as our limit because the fraction of clusters detected in fewer maps is constant, whereas it drops steeply when more detections are required.

The requirement of ≥ 7 detections from the Monte-Carlo maps removes $< 10\%$ of the cluster candidates. Because the detection of candidates that appear in fewer maps is very sensitive to small photometric errors, these candidates are likely to be either false, extremely

poor, or at high redshift. For the latter two cases, our detection efficiency is very low (as described below), and such clusters should not be used in statistical studies. Thus, we expect that this requirement imposes no significant bias on the final catalog, and merely serves to reduce the catalog inhomogeneity at the limits of our detection procedure.

3.5 Contamination, Completeness and Optimizing Detection

The weakest aspect of optical imaging surveys for galaxy clusters is the lack of distance information. The two-dimensional projection of the galaxy distribution can produce many apparent overdensities which may be identified as clusters but are not actual physical associations. This issue has been discussed in some detail for the Abell catalog (see the overview in §3.2), where rates of false clusters due to projection effects as high as 40% and as low as 10% have been claimed. Other types of surveys (*e.g.*, X-ray, SZ) are not as strongly affected by projection, but suffer from other selection effects or observational difficulties. Therefore, redshifts are needed to validate the existence of a given cluster.

The number of cluster candidates varies significantly with the detection threshold (Kim *et al.* 2002). Obviously, this dependence is also related to the number of false positive detections. As we try to achieve high completeness we may suffer from high contamination. Our goal is to minimize N_{false}/N_{cand}, while maximizing N_{cand}, where N_{false} is the number of false positives and N_{cand} is the number of candidates for each field. We make use of a simulated background distribution in order to evaluate N_{false}, while N_{cand} is derived from the real galaxy catalogs. Because each plate suffers from slightly different systematic errors in both photometry and star/galaxy separation, our procedure to optimize detection and evaluate completeness and contamination is done for each plate individually. This is preferable to selecting a single region of the sky, optimizing the algorithm for that region, evaluating the contamination rate, and assuming that is true for the rest of the sky.

3.5.1 The Rayleigh-Levy Distribution and Contamination

Initially, we attempted to estimate contamination rates by producing random catalogs with the same mean density as our survey fields. This resulted in obvious underestimation of the contamination rate. As pointed out by Postman *et al.* (2002, hereafter P02), the basic sources for false positives in cluster catalogs based only on photometric data are random fluctuations and superpositions of poor galaxy groups. For this survey, in which we effectively use a single photometric band for cluster detection, the susceptibility to false detections due to projection effects is high. Therefore, we adopt a methodology similar to that of P02, using the Rayleigh-Levy (RL) distribution to generate simulated galaxy positions. The RL random walk creates a linear ordering among the distributed galaxies, implying that galaxies interact only with the ones in their immediate vicinity, and each small set of neighbors interacts independently of any other small set. The resulting galaxy distribution, as shown by Mandelbrot (1975), follows the same two- and three-point correlation functions as described by Peebles & Groth(1975).

For a given separation θ, the RL distribution gives the probability of finding a pair with a larger separation:

$$P(>\theta) = (\theta/\theta_0)^{-d} \quad if \quad \theta \geq \theta_0$$

and

$$P(>\theta) = 1 \quad if \quad \theta < \theta_0$$

We ran a suite of RL simulations with different values for the parameters θ_0 and d, and chose values for these such that the resulting pair distribution matches the real galaxy distribution for the DPOSS data. We use $\theta_0 = 500''$ and $d = 0.6$. This value of θ_0 also corresponds to the diameter of the smoothing kernel; angular correlations are thus removed at smaller scales. We find that the values of θ_0 and d do not vary significantly among plates, and therefore adopt the values above for all plates.

For each plate, we create the simulated RL distribution by choosing a random position within the plate limits that does not fall into a bad area, and then selecting a separation from the above distribution. From this new point in the frame we repeat the same procedure until seven galaxies are selected. We then choose a new random location and select another

new point in the frame, repeating the same procedure until seven galaxies are selected. This process goes on until we generate a number of positions equal to the total number of galaxies for that particular plate. The resulting distributions are very insensitive to the number of galaxies we choose before restarting.

We also tested the full random walk distribution, where we start from one location, and continue relocating to new points with step sizes following the $P(>\theta)$ distribution (as opposed to choosing a new random point, distributing a fixed number of points around it, and repeating). Again, the final distribution is very similar to the restricted random walk; this is because the clustering scale we are interested in is well below the cutoff in the two-point correlation function introduced by the restricted random walk. By comparing the angular separation distribution for these different sets we find them to be indistinguishable up to scales of 2°. We therefore use the restricted RL in our simulations, as it is significantly less computationally intensive than the full random walk.

For each plate in the survey, we produce an RL distribution with the same number of objects as the galaxy catalog for that plate, with the same bad areas excised. A density map is generated from this distribution, and SExtractor is run with a range of threshold/minimum area pairs on this map, along with the map generated from the real galaxy catalog. We use detection thresholds ranging from 900 to 2200 galaxies per square degree, in increments of 20, and minimum areas of 30, 40, 50 and 100 square pixels (or arcmin). Initial tests using a broader range of parameters indicated that the optimal pair would always be found within these ranges. Once detection is completed on both the RL and real maps, we measure the contamination rate as $C = N_{false}/N_{cand}$, and select the parameter set for which C is closest to 10%, while maximizing N_{cand}; typically, this results in $8\% < C < 12\%$. We performed tests using other values for the contamination rate: we found that completeness is not substantially improved by allowing higher C, while lower rates, such as $C = 5\%$, produced more widely varying parameter sets due to small fluctuations in C. Importantly, enforcing a fixed contamination rate does not produce large plate-to-plate variations in the completeness; this can be seen by comparing the completeness functions for a large set of plates. Nevertheless, it would be inappropriate to use a mean completeness function for the entire survey.

We then examined the optimal parameters derived for each plate. We found that min-

imum areas of 30 or 40 square arcminutes produced very similar results, in terms of the distribution of both the detection thresholds and final number of candidates per plate. We also examined the completeness functions (described below), and found that they were extremely similar for both values of the minimum area. Using a minimum area of 40 square arcminutes produced slightly lower variance in the number of candidates per plate; we have therefore used this minimum detection area for all plates. Physically, this is sensible: we know that the depth and classification accuracy vary from plate to plate, which will affect the threshold in galaxies per square degree, but there is no reason to expect the sizes of clusters to vary from plate to plate. The larger minimum areas (50 and 100 square arcminutes) produce radically fewer cluster candidates.

We stress that the entire optimization procedure is performed separately for each field. The optimal parameters for a given field will not be appropriate for another. This is not due only to zero point differences or plate variations, but also to the large scale structure (LSS) in the universe (Bramel, Nichol & Pope 2000). Our final cluster catalog is designed to maintain a constant contamination level across the whole sky. One must also note that the RL distribution does not take into account inhomogeneities at large scales; this test is specifically designed to estimate the contamination rate due to chance projections of small groups or physically unassociated galaxies. Because it is impossible to know *a priori* which overdensities are true physical associations contributing to large scale structure, there is no way to use the actual catalogs to measure contamination rates. Only complete spectroscopic follow-up can determine the accuracy of our estimates; as described below in §3.5, our own spectroscopy suggests low contamination rates, consistent with the 10% threshold imposed. Additionally, the results of P02 demonstrate that the RL estimates of false positive rates is verified by such spectroscopy.

An alternative method to estimate contamination rates is to shuffle the galaxy positions in a plate catalog, while maintaining the object magnitudes. This method preserves some very large-scale correlations, while removing those at smaller scales. As shown in Lopes *et al.* (2003a), this technique produces contamination rates (and the corresponding optimized detection parameters) which are very similar to those derived using the RL distribution.

3.5.2 Completeness From Simulations

Perhaps the most important feature of this cluster catalog is the detailed assessment of the selection function. In order to properly assess this, it is fundamental that each plate be treated individually. Because the detection parameters for each DPOSS field are determined separately, we assess the completeness as a function of redshift and richness individually for each plate, and provide this information for each plate. When generating mock catalogs from simulations (or from other observations), one can then apply these selection functions to mimic potential systematic effects in our catalog.

To assess the selection function for each plate, we use simulated galaxy clusters which are placed in the background field derived from the respective plate. We then use these galaxy catalogs consisting of a background plus artificial clusters to generate density maps on which we run SExtractor. When running SExtractor, we use the same detection parameters obtained in the previous section, which are also used for the real cluster detections.

Artificial clusters are typically defined using a Schecter luminosity function with its choice of α and M^*; a surface density profile; a given cluster composition; a maximum radius and a core radius. Lobo *et al.* (2000) tested the variation of the completeness rate as a function of the spatial profile slope, while P02 also tested different cluster compositions. As expected, steeper profiles provide the highest completeness rates. We discuss the effects of varying different cluster parameters at the end of this section.

An important issue is the background into which we insert the artificial clusters. While the RL distribution is well suited to estimate the contamination rate and optimize the detection parameters, it only represents an ideal case, being unable to reproduce all the inhomogeneities due to large scale structure. As we discuss below, the RL distribution overestimates the completeness rate. Thus, for each plate, we use the actual galaxy catalog from that plate to provide the background field in which we insert the artificial clusters

The simulated clusters follow a Schecter luminosity function with parameters given by Paolillo *et al.* (2001) ($\alpha = -1.1$ and $M_r^* = -21.53$). The constituent galaxies have $-23.4 \leq M_r \leq -16.4$, such that our observed magnitude range is fully covered at all redshifts of interest. The surface profile adopted is a power law in radius of the form r^β, where $\beta = -1.3$. This is in the middle of the observed range, $-1.6 \leq \beta \leq -1.0$, as discussed in Tyson & Fischer (1995) and Squires *et al.* (1996). The galaxies are placed within a maximum radius

of $r_{max} = 1.5h^{-1}$ Mpc with a core radius of $r_{core} = 0.15h^{-1}$ Mpc. The clusters are composed of 60% elliptical galaxies and 40% Sbc galaxies, with SEDs taken from Coleman, Wu, & Weedman (1980), convolved with the DPOSS r filter. Each galaxy has the appropriate k-correction applied, as well as a random photometric error from Gal et al.(2003c) added.

We proceed as follows:

1. We generate clusters with six different richnesses, which are shifted to seven redshifts, with the catalogs trimmed at $15.0 \leq m_r \leq 19.5$. The richness classes are $N_{gals} = [15, 25, 35, 55, 80, 120]$ galaxies, while the redshifts adopted are $z = [0.08, 0.12, 0.16, 0.20$ $0.24, 0.28, 0.32]$.

2. For each of the 42 richness/redshift combinations we generate 5 simulated clusters, which are placed at random positions in the real background galaxy distribution, avoiding both excised areas and cluster candidates detected in the real fields. This is repeated ten times, resulting in ten catalogs containing a total of 50 simulated clusters for each richness/redshift combination. Thus, there are 420 total simulated catalogs generated for each plate (6 richnesses × 7 redshifts × 10 realizations). An example is shown in Figure 3.2, where we plot Plate 389, with the bad areas marked, and five simulated clusters at $z = 0.16$ and $N_{gals} = 80$ inserted.

3. We then use the AK to produce density maps for each of these 420 background plus cluster galaxy catalogs. This is performed individually for each plate, and SExtractor run with the optimal parameters determined from the contamination tests.

4. Next, we compare the candidate positions in these maps with the initial input positions. We use a matching radius of $400h^{-1}$ kpc, which is small compared to the typical size of a cluster.

RL *vs.* Real Backgrounds

For one DPOSS field (389) we repeated the entire procedure twice, first using the real background and then the RL background distribution. The purpose of this test is to check how well the RL distribution reflects the real background. We show the results of this

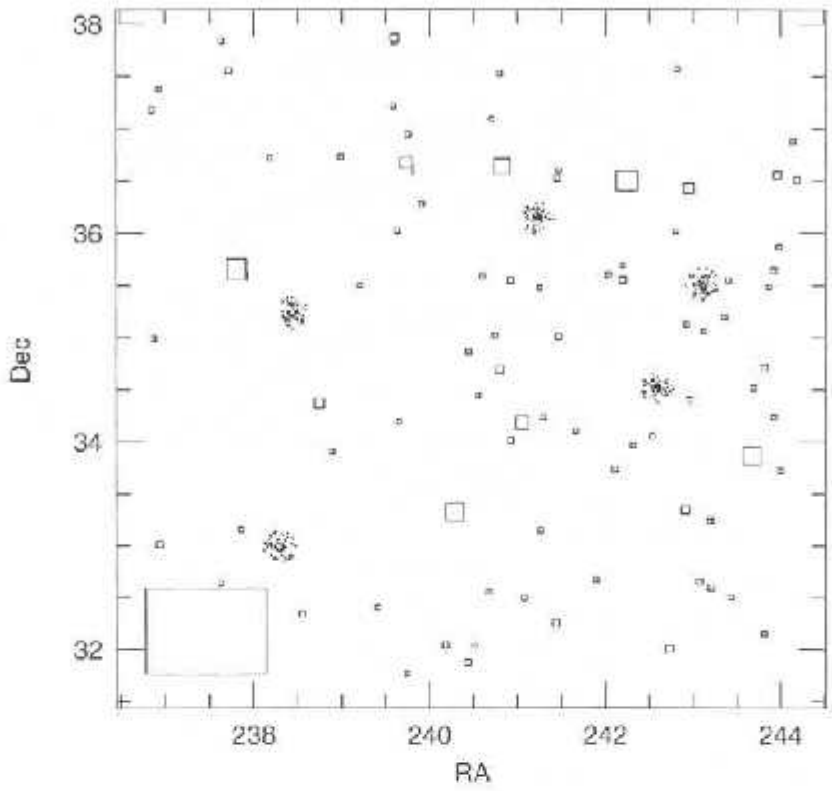


Figure 3.2: Plate 389, with the bad areas marked and five simulated clusters at $z = 0.16$ and $N_{gals} = 80$ inserted.

test in Figure 3.3, where the completeness rates obtained using the RL distribution are the dashed lines, while the solid lines use the real background. Each panel shows a different richness class, with richness increasing from bottom to top. We see that the dashed lines (RL background) are always higher than the solid lines, for all richness classes and at all redshifts. At $z = 0.16$ (the median redshift of the DPOSS sample) and $N_{gals} = 35$ (close to the median richness value), the recovery rate is 94% using the RL background, but only 72% using the real background. This suggests that using the RL distribution likely overestimates the completeness. An interesting comparison can be made with Figure 3.11. For each richness class shown there, the median redshift is similar to the redshift where the completeness estimate for that subsample starts to decrease significantly.

We note that Kim *et al.* (2002) modified their detection limits after comparing the recovery rate of artificial clusters using uniform and real backgrounds. Their goal was to improve the completeness rate; however, this necessarily has an adverse impact on the contamination.

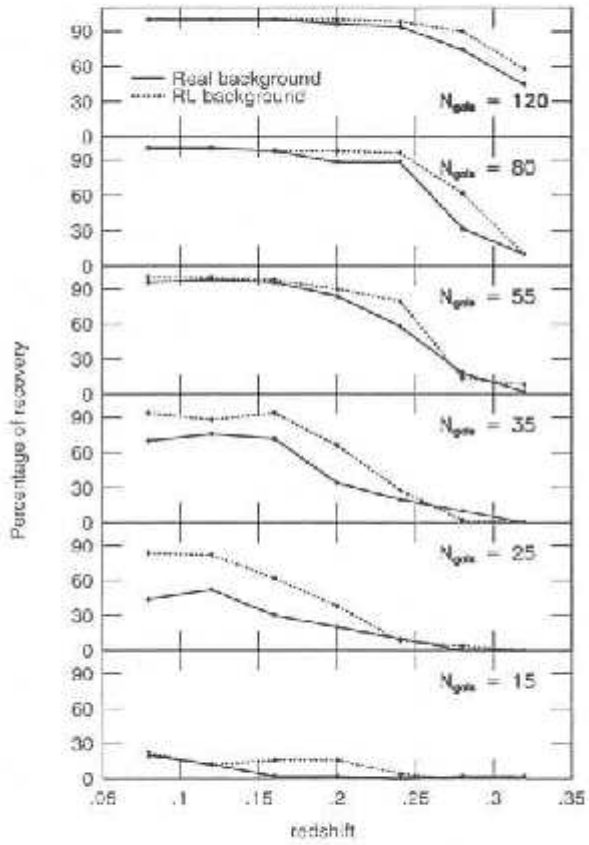


Figure 3.3: The completeness rate evaluated with two different backgrounds: a RL distribution (dashed line) and a real galaxy catalog from DPOSS (full line). The richness, increasing from bottom to top, is indicated on each panel.

We do not apply such corrections, as our goal is to maintain a constant contamination rate.

Dependence on Cluster Model

We also wish to test the variation of completeness with the cluster composition, surface density profile slope, luminosity function slope, and maximum and core radii. We use a single plate, varying the properties of the simulated clusters and repeating our completeness tests. Figure 3.4 shows the dependence of the completeness on the cluster composition and spatial profile slope (β). Each row represents a different value for the slope (β = -1.0, -1.3, -1.6, from bottom to top), while each column shows the results for different compositions (from right to left: 100% E, 60% E + 40% Sbc, and 20% E + 80% Sbc). The improvement when

we adopt steeper profiles is evident; as expected, tighter cores are easier to detect. There is little if any correlation with the cluster composition. This is expected as the difference in the k-corrections used (for E and Sbc galaxies) is only 0.2^m at the highest redshift of the simulations ($z = 0.32$). For instance, for $N_{gals} = 35$ and $z = 0.16$, using 60% E + 40% Sbc, the recovery rates for β = (-1.0,-1.3,-1.6) are (62%,72%,80%), respectively. However, when we fix $\beta = -1.3$, $N_{gals} = 35$ and $z = 0.16$, the recovery rates are (74%,72%,70%) for cluster compositions of (100% E, 60% E + 40% Sbc, 20% E + 80% Sbc).

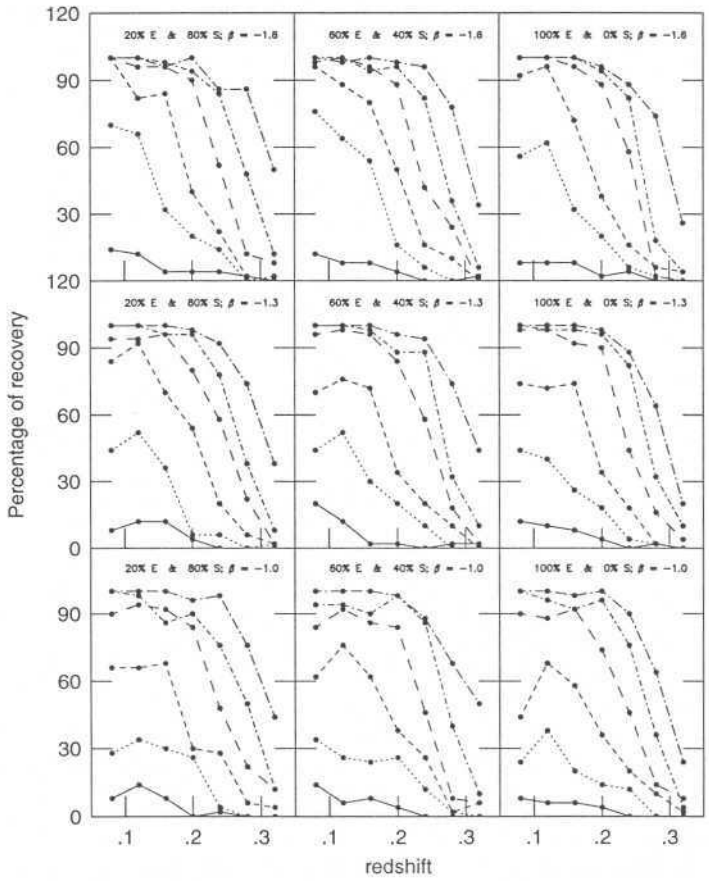


Figure 3.4: The selection function evaluated with different cluster compositions and spatial profile slopes β. From bottom to top β assumes the values -1.0, -1.3 and -1.6. The cluster composition, from right to left panels is 20% E + 80% Sbc, 60% E + 40% Sbc, and 100% E. Richness classes are the same as Figure 3.3.

The effects of changing the luminosity function slope, core radius and maximum radius are shown in Figure 3.5. Each row shows the results for one of these 3 parameters. For the

richest clusters, steepening the LF slope increases completeness, while poor clusters show an opposite trend, albeit with large scatter. The variation with core radius shows a clear trend to higher completeness for smaller values of this parameter; this is commensurate with the results of the radial profile test. Finally, tests with the cutoff radius show the worst recovery rate for $r_{max} = 1.0h^{-1}$ Mpc, while the results are about the same for the other two values tested.

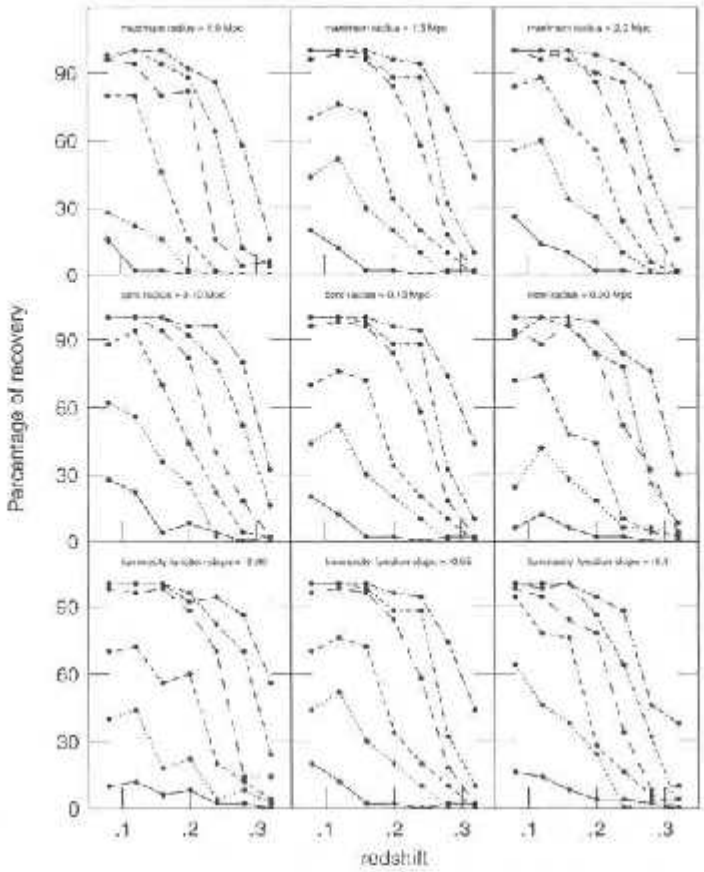


Figure 3.5: Dependence of the selection function with the luminosity function slope (bottom panels), core radius (middle panels) and cutoff radius (top panels). The values tested, from left to right, are $\alpha = -0.8, -0.95, -1.1$ for the LF slope, $r_{core} = 0.1, 0.15, 0.2h^{-1}$ Mpc for the core radius, and $r_{max} = 1.0, 1.5, 2.0\ h^{-1}$ Mpc for the maximum radius. Richness classes are the same as shown in Figure 3.3.

We conclude that our canonical cluster model ($\alpha = -1.1, M_r^* = -21.53, \beta = -1.3, r_{max} = 1.5h^{-1}$ Mpc, $r_{core} = 0.15h^{-1}$ Mpc, 60%E + 40%S) provides a realistic estimate of the catalog

Table 3.2: Example Completeness Function: Plate 389

N/z	0.08	0.12	0.16	0.20	0.24	0.28	0.32
15	0.16	0.14	0.08	0.04	0.04	0.02	0.00
25	0.64	0.46	0.38	0.24	0.00	0.04	0.04
35	0.94	0.78	0.76	0.28	0.16	0.06	0.00
55	0.98	0.94	0.84	0.78	0.34	0.08	0.10
80	1.00	1.00	1.00	0.86	0.64	0.32	0.00
120	1.00	0.98	1.00	0.94	0.88	0.46	0.38

completeness. For each candidate, we provide a table of the completeness as a function of richness and redshift. An example is shown in Table 3.2 and Figure 3.6; a similar table is provided online (at `http://dposs.caltech.edu/dataproducts/`) for each plate used in this survey. Additionally, a list of the areas excised due to bright stars, airplanes, etc. is provided for each plate at the same location.

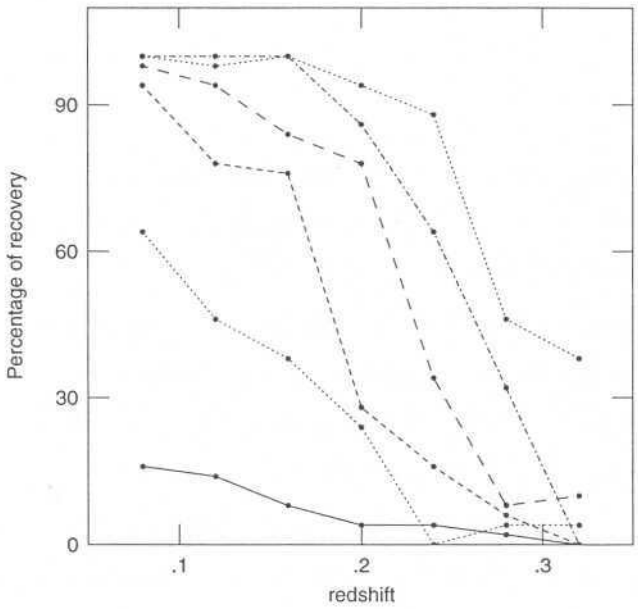


Figure 3.6: An example completeness function, for Plate 389. From top to bottom, the functions correspond to clusters of richnesses $N_{gals} = [120, 80, 55, 35, 25, 15]$.

3.5.3 Spectroscopic Confirmation

To assess the validity of the above results, as well as to calibrate our photometric redshift estimator (see the next section and Paper I) and examine the redshift distribution of our clusters, we undertook a complete spectroscopic survey of candidates in two DPOSS 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, and because scans in all three bands were available when this project was started. However, these maps were generated, and cluster detection performed, using the methodology discussed in Paper I, which (as discussed earlier) differs significantly from the final technique used here. The original list of cluster candidates in these two fields was drawn from the individually calibrated plates, using galaxies to $m_r = 20$. Therefore, the sample with follow-up spectroscopy is only a subset of the final cluster catalog in these areas, and also contains targets which are no longer candidates. We remind the reader that the catalog for Field 475, which is in the SGP, is not a part of the NGP, and thus not in the area covered by the catalog presented in this paper.

Observations

Spectra of the cluster candidates were obtained between April 1996 and April 2000 with the COSMIC instrument (Kells *et al.* 1998) in re-imaged mode, mounted at the prime focus of the Hale 5m Telescope at Palomar Observatory. Targets for each mask were chosen from the DPOSS catalog for the appropriate field, selecting only galaxies with $m_r \leq 20$ to maintain reasonable exposure times. No color selection was applied; this was done to maximize the number of possible objects which could have slits placed on them. Multi-slit masks, made of photographic film negatives, with $1.5''$ slit widths, were mounted at the focal plane. The re-imaged pixel scale is $0.399''$/pixel; the slits therefore correspond to ~ 3.75 pixels. The available field for spectroscopy is theoretically $8' \times 12'$, with the long dimension corresponding to the spatial axis. The field is further limited by the lack of an accurate distortion map in the slitmask production software; this requires that slits be placed within $\sim 3'$ of the central axis in the direction perpendicular to the slits. Nevertheless, the $8' \times 3'$ field of the masks is larger than the core radius of clusters at all but the lowest redshifts in our sample.

Slitmasks were designed using the *cosmicslitmask* program, written by M. Pahre. This software simply takes an input list of coordinates, and allows the user to interactively select objects for slit assignment. There is no automated slit optimization algorithm, so objects were selected visually for each mask, to maximize the number of targets observed per mask. This often resulted in densely packed slitmasks, with as many as 45 slits on a single mask, some as short as 10″.

Spectra were taken using a grism with 300 l/mm, blazed at 5500Å, which produced a dispersion of 3.03Å pixel^{-1}. The spectra typically cover a wavelength range $\sim$3000Å to $\sim$9000Å, with a central wavelength of $\sim$6000Å. Since the dispersing element is a grism and cannot be tilted, the wavelength range is fixed and varies only as a result of the vertical displacement of the slits from the mask center. Therefore, nearly all spectra are observed from blueward of the 4000Å break (for $z = 0$), through the redshifted Na lines (5890Å rest frame, 7950Å at our highest expected $z = 0.35$). The COSMIC CCD has a gain of $3.1e^-/DN$ and a relatively high read noise of $13e^-$. Exposure times varied greatly from night to night, depending largely on the seeing (which ranged from 0.8″ to as high as 2.0″) and spectrograph focus changes (due to temperature fluctuations). Exposures ranged from a single 1800s exposure up to two 3600s exposures per slitmask, with most consisting of two 1800s exposures.

Data Reduction

All data reduction was performed using the IRAF package. The object spectra were overscan subtracted, and flattened using spectra of a halogen lamp reflected off the dome interior. Cosmic rays were removed using the *szap* task (written by M. Dickinson), and night sky emission removed by subtraction in the spatial direction on the two-dimensional images, using the *background* task to perform median filtering and sigma clipping. Because the slits were typically very narrow, we used only a second order fit along the spatial direction. After sky subtraction, each spectrum was traced and optimally extracted in the individual exposures. The extracted spectra were wavelength calibrated using spectra of an arc lamp, taken immediately before or after the object spectra, to minimize the effects of instrument flexure. The individual extracted, wavelength-calibrated spectra were then combined to produce the final spectrum for each object, used to measure the redshifts. All redshifts were measured by RRG using the IRAF task *redshift*, written by T. Small. This program allows the user to visually

mark specific spectroscopic features, and input their rest wavelengths. The object's redshift is then calculated, and the positions of other common absorption and emission lines overplotted. In this way, the user can check the redshift assignments. Such visual measurement was necessitated by the poor S/N of the majority of our spectra; automated techniques (such as cross-correlations) were strongly affected by residuals from poor sky subtraction. To enable completion of the survey in a reasonable time, a trade-off in exposure times *vs.* number of masks observed was made, since we were only interested in measuring redshifts and not any specific galaxy properties (such as velocity dispersions or line strengths). Comparison of redshifts obtained by different reducers and at different times suggests the redshifts are accurate to $\sim \Delta z = 0.002$, or approximately 600 km s^{-1}.

In addition, over the course of the survey, our list of candidate galaxy clusters changed as the photometric calibration and sample selection improved. This resulted in a significant number of targets being observed which are not included in the final cluster sample. This extraneous data is actually useful in assessing the rate at which we miss real clusters and is discussed later in this section. A total of 3249 individual spectra were visually inspected; of these, 1655 (51%) were identified as galaxies, and 326 (10%) as stars. Thus, a total of 1981 spectra were identifiable, a 61% success rate. Of the galaxies, 1245 (75%) had securely measured redshifts; the remainder had insufficient S/N. We also note that the 10% stellar contamination is consistent with the results of Odewahn *et al.* (2003), and should be taken as an upper limit on misclassification, as extended sources (*i.e.*, galaxies) tend to have lower S/N spectra, making identification more difficult. This large spectroscopic sample constitutes a significant survey in its own right, with other applications in addition to those discussed here.

In Field 447, there are 64 cluster candidates. Of these, only 24 have usable spectroscopy, with all 24 showing evidence for clusters. There are an additional 4 masks targeted at areas with no current candidate; these nevertheless show spectroscopic evidence for physical galaxy associations. However, they all have $z \geq 0.22$ and $N_{gals} < 30$, a regime where our completeness is typically only $\sim 30\%$. There are unfortunately 40 current candidates in this field without spectroscopy.

The situation for Field 475 is much better, with 55 candidates in our current sample, of which 37 have corresponding slitmasks. These all show evidence for physical clustering (as

Table 3.3: Spectroscopic Survey Results

Field	N(cands)	N(spec) [frac]	N(extra)
447	64	24 [0.38]	4
475	55	37 [0.67]	16

discussed below). There are 16 masks not associated with any new candidates. Of these sixteen, four are near the edge of the plate or the area used for the densitometry spots. There remain twelve masks which show evidence for real galaxy associations. Nevertheless, these twelve areas appear to be only moderate overdensities; even lowering our detection threshold to allow 30% contamination recovers only three of these. Running our richness estimator on these areas, using the spectroscopic redshifts, shows that indeed these possible clusters all have $N_{gals} < 30$. Additionally, four of these 12 masks have clusters at $z \geq 0.22$. Finally, eighteen current candidates have no spectroscopy; these are all of similar overdensity to those candidates with spectra, and therefore likely to be real.

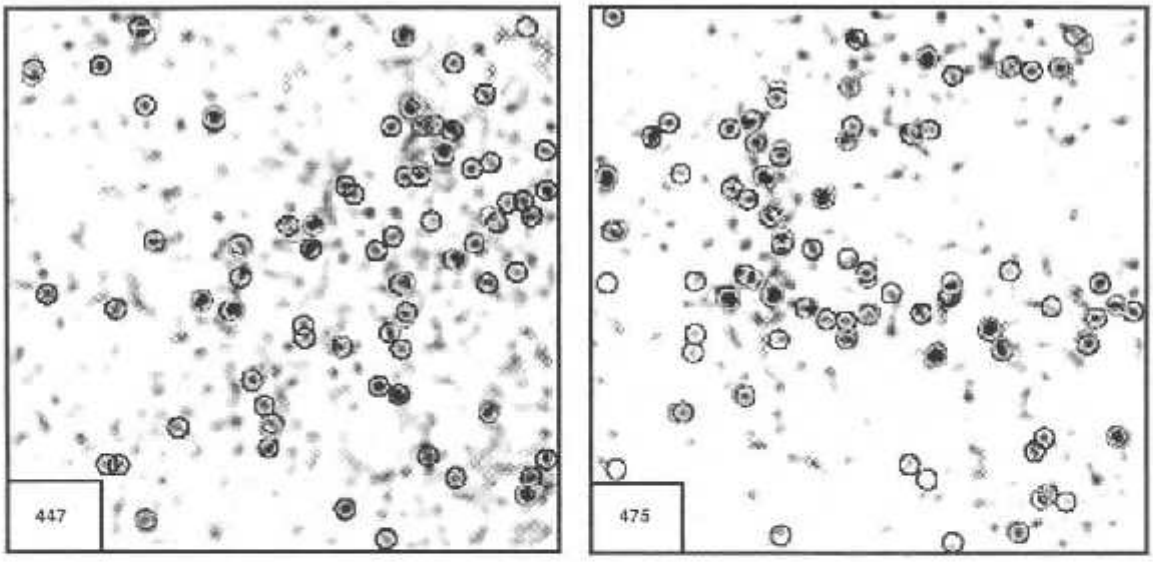


Figure 3.7: AK galaxy density maps for Fields 447 (left) and 475 (right), with blue circles marking the locations of current candidate clusters, green circles showing slitmasks associated with current candidates, and red circles showing locations where spectra were taken but there is no current candidate.

Table 3.4: Field 447 Spectroscopy

Candidate/Mask	z_{spec}	$N(z)$	Comment
NSCJ142311+320840	0.200	10	
NSCJ142841+323859	0.127	5	
NSCJ142920+270609	0.268	6	
NSCJ142937+301403	0.103	4	
NSCJ143203+293404	0.221	14	
NSCJ143237+313532	0.131	13	
NSCJ143330+292738	0.219	12	
NSCJ143400+301222	0.222	3	
NSCJ143437+284024	0.205	11	
NSCJ143539+281143	0.203	3	
NSCJ143737+300923	0.338	9	
NSCJ143744+302547	0.160	13	
NSCJ143910+290229	0.253	11	
NSCJ144210+294444	0.216	4	
NSCJ144229+292545	0.224	8	
NSCJ144231+323227	0.243	11	
NSCJ144250+314342	0.243	9	
NSCJ144315+305758	0.227	4	
NSCJ144432+311149	0.233	14	
NSCJ144457+300112	0.177	3	
NSCJ144603+301148	0.109	5	
NSCJ144635+281740	0.229	4	
NSCJ144713+302554	0.170	4	
NSCJ144820+272134	0.233	20	
J144328+313136	0.240	11	High z; $N_{gals} = 36.3$
J144352+302724	0.320	4	High z; $N_{gals} = 22.5$
J144404+313214	0.233	6	High z; $N_{gals} = 25.3$
J144902+323713	0.200	3	$N_{gals} = 13.2$

The results of the spectroscopic survey are summarized in Table 3.3. We give the total number of candidates, the number (and fraction) with and without spectroscopy, and the number of extraneous masks. Details for each candidate with spectroscopic observations are provided in Tables 3.4 (for Field 447) and 3.5 (for Field 475). Column (1) provides the candidate name, columns (2) and (3) provide the spectroscopic redshift and the number of galaxies at that redshift (if spectra were obtained), and column (4) provides comments, if any.

The above results are shown in Figure 3.7, for Field 447 (left panel) and 475 (right panel). We show the AK galaxy density maps, with blue circles marking the locations of current candidate clusters. These clearly correspond to areas with the highest galaxy density. Green

Table 3.5: Field 475 Spectroscopy

Candidate/Mask	z_{spec}	$N(z)$	Comment
NSCJ005559+262442	.194	24	
NSCJ005618+254729	.150/.245	3/3	
NSCJ005957+234739	.240/.302	10/6	
NSCJ010201+250504	.273	11	
NSCJ010211+261816	.239	7	
NSCJ010251+252028	.189	15	
NSCJ010255+235859	.266	6	
NSCJ010310+270349	.166	12	
NSCJ010319+264850	.127::	3	
NSCJ010348+262628	.242	12	
NSCJ010403+255906	.245	7	
NSCJ010408+250654	.160	17	
NSCJ010420+271828	.239	8	
NSCJ010434+263908	.168	7	
NSCJ010439+254013	.158	15	
NSCJ010546+245803	.241	17	
NSCJ010641+261142	.164	15	
NSCJ010748+243721	.238	7	
NSCJ010749+265059	.192	8	
NSCJ010758+272626	.117/.238	4/3	
NSCJ010847+252214	.200	12	
NSCJ010853+245311	.197	10	
NSCJ011059+265458	.113	7	
NSCJ011152+274612	.115	8	
NSCJ011210+242646	.196	17	
NSCJ011251+250601	.183	5	
NSCJ011444+244250	.178	8	
NSCJ011521+242925	.189	6	
NSCJ011601+222736	.248	7	
NSCJ011601+273938	.120	8	
NSCJ011726+225257	.123	6	
NSCJ011825+273800	.177	10	
NSCJ011932+243213	.142	7	
NSCJ011954+244954	.207	5	
NSCJ012049+233053	.117	21	
NSCJ012057+245751	.190	9	
J010030+243949	.082	7	$N_{gals} = 17.1$
J011252+222124	.140	6	$N_{gals} = 16.0$
J010432+243635	.266	9	$N_{gals} = 20.4$
J005940+262717	.193	14	$N_{gals} = 22.8$
J005645+230911	.180	6	$N_{gals} = 18.9$
J010441+222618	.250	5	High z; $N_{gals} = 22.9$
J010752+253143	.200	6	$N_{gals} = 29.4$
J011142+230337	.195	7	$N_{gals} = 9.8$
J011818+224753	.267	5	High z; $N_{gals} = 27.63$
J010025+242744	.125	10	$N_{gals} = 20.3$
J011000+250756	.200	6	$N_{gals} = 26.7$
J011050+231344	.115	5	$N_{gals} = 23.4$
J010027+251510	.227	5	High z; $N_{gals} = 28.6$
J011548+252229	.185	9	$N_{gals} = 22.8$
J011747+245810	.145	11	$N_{gals} = 23.2$
J005608+251240	.175	8	$N_{gals} = 16.80$

circles show the locations of slitmasks associated with current candidates, while red circles mark locations where spectra were taken but there is no current candidate.

We show example spectroscopic redshift histograms for some cluster candidates in Figure 3.8. From these redshift distributions, we can clearly see strong clustering in redshift space, providing evidence for true galaxy clusters in most of the slitmasks. The median redshift of our spectroscopically confirmed cluster candidates is $z_{med} \sim 0.2$. This value is in keeping with the photometric redshifts presented in the following section, and with our earlier estimates of the depth of our galaxy catalogs.

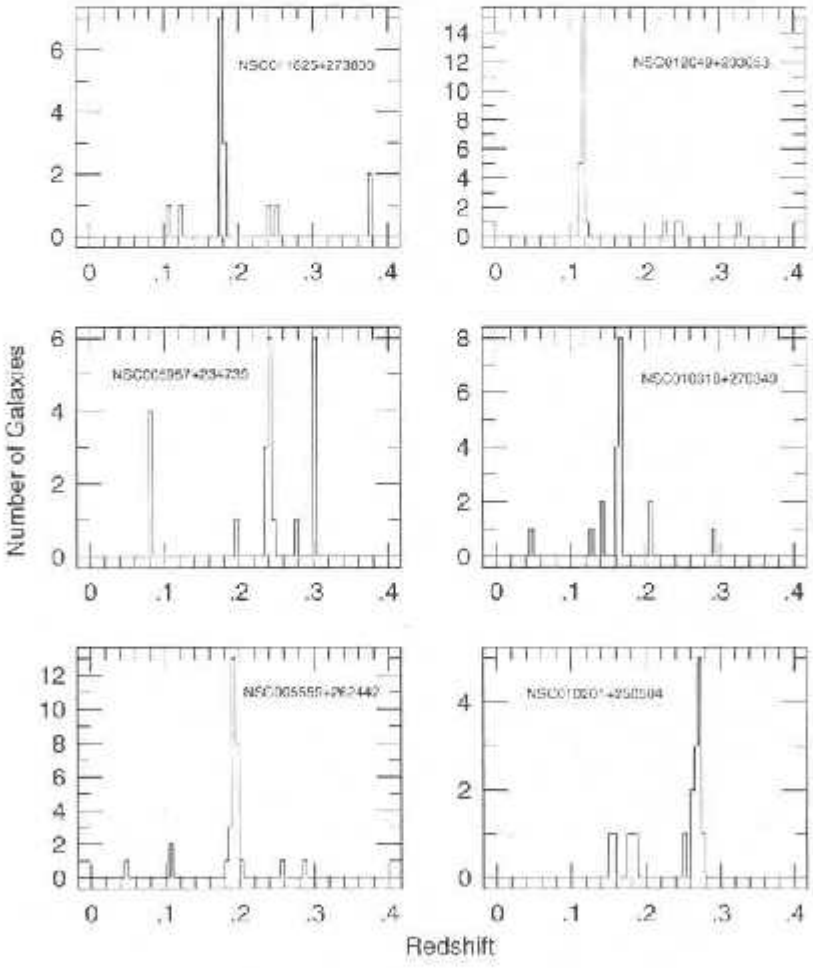


Figure 3.8: Spectroscopic redshift histograms for selected candidates in Field 475.

Are Redshift Peaks Real Clusters?

Although the peaks in the redshift distributions provide evidence for clustering, we would like to quantify the likelihood that a given peak is a real structure. Certainly, large peaks (as in candidate NSC005559+262442, with 24 galaxies) are undoubtedly clusters. However, many of the masks show peaks with only 3 or 4 members. Individual galaxies are assigned as members of a peak if they fall within $\Delta z = 0.005$ (± 1500 km s^{-1}) of the peak center. Increasing this range does not change the results significantly, since most redshift peaks are well isolated.

A rough estimate of the likelihood for such small peaks being real can be made following the argument of Zaritsky *et al.* (1997). If the underlying redshift distribution of galaxies were smooth, the formal probability of a pair of galaxies within $\Delta z = 0.005$ would be negligible, much less for three or more galaxies. However, because galaxies are correlated, the likelihood of pairs, triplets, etc., is increased. If we assume that *all* slitmasks where the richest clump has only three or four members correspond to spurious detections (there are 18 such masks out of 95 total masks), the probability of finding a single false peak is 18/95, or $\sim 20\%$. We might then expect a comparable fraction of our histograms to show a second redshift peak with similarly few members, but there are only eight ($\sim 8\%$). In fact, some of the distributions with multiple peaks may be projections of multiple poorer clusters which we detect as a single candidate. Additionally, there is only one mask with a secondary peak containing five members. An additional argument in favor of the reality of three-or-more galaxy clumps is that if peaks with three members were often spurious, we expect to find many more such peaks than those containing four galaxies. However, we find an almost identical number of peaks with three and four galaxies, supporting the idea that even three-galaxy peaks are often real. Therefore, we conservatively identify all peaks with four or more members as real clusters, and peaks with only three members as tentative. The number of clumps with two galaxies is much higher (nearly every mask contains one); we therefore ignore those completely. Furthermore, while Zaritsky et al. follow similar arguments for only two galaxies, their spectroscopy covers a significantly larger magnitude range ($m_r < 22$ compared to our $m_r < 20$), thereby greatly increasing their likely contamination rate. Conservatively, we estimate that at least 80% of our robust candidates are true clusters; this is roughly consistent with our attempt to generate catalogs with $\sim 10\%$ contamination.

3.6 Photometric Redshifts and Richnesses

Two of the most fundamental properties of galaxy clusters are their distance (or redshift), and mass. Because the latter is difficult to measure accurately, and impossible from our photometric data alone, we instead measure the richness. The details of our redshift and richness estimation techniques are given below. It is important to note that both are measured in a completely model independent way; this is possible because, these properties are measured *after* detection is completed, and are not an inherent part of the procedure. We note that this is inherently different from many other techniques; for instance, the Matched Filter simultaneously performs detection and redshift/richness estimation, while others, such as the Cut & Enhance method (Goto *et al.* 2002) or the RCS (Gladders & Yee 2000) output the redshift, based on galaxy colors used as part of the detection procedure.

3.6.1 Redshift Estimates

Paper I presented an extremely simple yet effective photometric redshift estimator for DPOSS cluster candidates, based on the assumptions that each candidate is a single cluster, at one redshift, and the cluster galaxy population is dominated by early-type galaxies. The estimator is an empirical relation between spectroscopic redshift and the median $g-r$ color and mean r magnitude of the background-corrected galaxy population for each candidate. We use both colors and magnitudes as we find both to be equally well correlated with the spectroscopic redshift. 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 a radius of 1 Mpc ($0.67 \times R_{Abell}$). The background galaxy color and magnitude distributions ($N_{bg,g-r}$ and $N_{bg,r}$) are determined independently for each plate, scaled to the appropriate area, and subtracted from the color and magnitude distributions of each candidate cluster. The median $g-r$ color and mean r magnitude of the remaining galaxies is then calculated. This differs slightly from the methodology in Paper I, where universal background distributions, taken from a large contiguous area, were utilized. 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. We therefore start with an initial radius corresponding to 1 Mpc at $z = 0.05$, within which we measure the above quantities and derive the initial redshift estimate. The Abell radius

is adjusted using this new redshift estimate, and the procedure repeated until the redshift converges.

Spectroscopic redshifts for the clusters used to derive our empirical photometric redshift estimator were taken from Struble & Rood (1999, hereafter SR99). We do not use our own spectroscopic data, as they cover only two plates, and we do not want to bias our estimator. The SR99 data are culled from diverse sources in the literature, and are of highly variable quality. A majority of the cluster redshifts, especially those at $z > 0.1$, are based on very few (one or two) individual galaxy redshifts, and may therefore be incorrect. Nevertheless, this catalog is the largest, somewhat homogenized resource currently available. The larger area utilized in this paper results in a final training sample comprising 369 clusters (compared to only 46 in Paper I), with a median redshift of $z_{med} = 0.138$. Because we now have many more clusters, and their redshift distribution is heavily weighted with low-redshift clusters, we bin the colors and magnitudes into spectroscopic redshift bins. We divided the range $0.02 \leq z_{spec} \leq 0.35$ into ten bins, and calculate the median z_{spec}, $(g-r)_{med}$, and r_{mean} for each bin. These binned values are used to derive an empirical relation between redshift, median $g-r$ color, and mean r magnitude using a bivariate least-squares fit:

$$z_{phot} = 0.5694 \times (g-r)_{med} - 0.00215 \times r_{mean} - 0.057243 \tag{3.1}$$

The upper panel of Figure 3.9 shows the photometric redshift against the spectroscopically measured redshift for these 369 Abell clusters, with the lower panel showing the residual. Performing the complete iterative procedure, where we begin without assuming the spectroscopic redshift to define a starting radius, yields a Q_{sigma} of $(z_{spec} - z_{phot})/(1+z_{spec}) = \Delta z = 0.033$.

In addition, we tested the effect of varying the starting cluster radius from $5'$ (corresponding to ~ 1.0Mpc at $z = 0.2$) to $15'$ (~ 1.0Mpc at $z = 0.05$). In principle this could lead to different final photometric redshifts if the convergence procedure is unstable. The results are shown in Figure 3.10; the Q_{sigma} for $\Delta z_{phot} = 0.004$. There are a very few outliers, which demonstrates the robustness of our simple technique. This also suggests that estimating z_{phot} with two different initial radii and selecting those with varying results can be used to test for clusters where the estimated redshift is questionable. The starting redshift also does not affect the accuracy of the estimator.

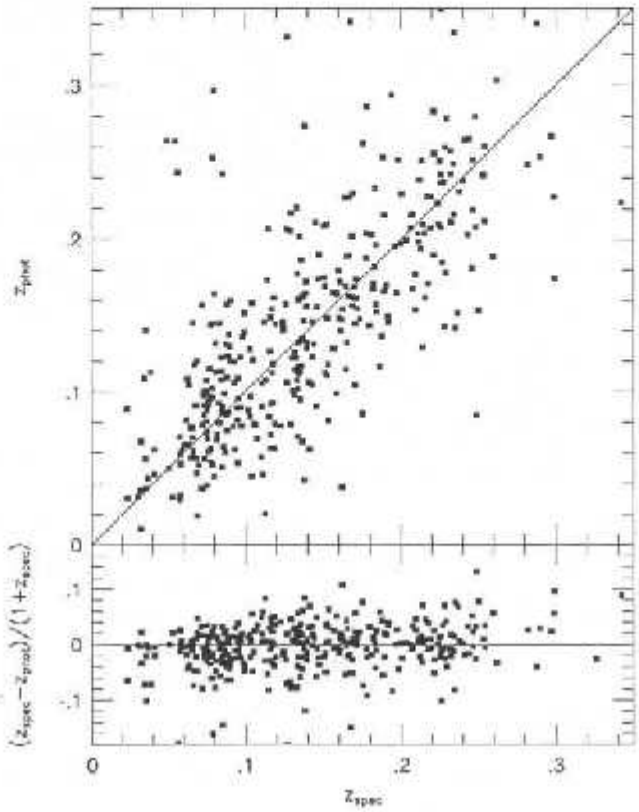


Figure 3.9: The photometrically estimated redshift *vs.* the spectroscopically measured redshift for 369 Abell clusters. Residuals as a function of redshift are shown in the bottom panel.

The photometric redshift estimates for the candidate clusters are provided in the fourth column of Table 3.6. Those clusters where the photometric redshifts using the two starting redshifts disagree by more than $\sqrt{2}$ times the redshift error are marked with a colon; in cases where the estimator failed entirely, we set $z = 0.0$. The mean redshift for our sample is $z_{phot,med} = 0.1579$, which is comparable to our original estimate of $< z > \sim 0.15$, based on the magnitude range covered. Our highest redshift clusters are at $z = 0.3$, with many at $z > 0.2$. The redshift distribution of our robust sample is shown in Figure 3.11. The entire sample is shown in the top panel, with subsamples of varying richnesses in the lower bins. The median redshift for each sample is marked. As expected, poor clusters are found only at low redshift, while richer clusters can be seen to larger distances.

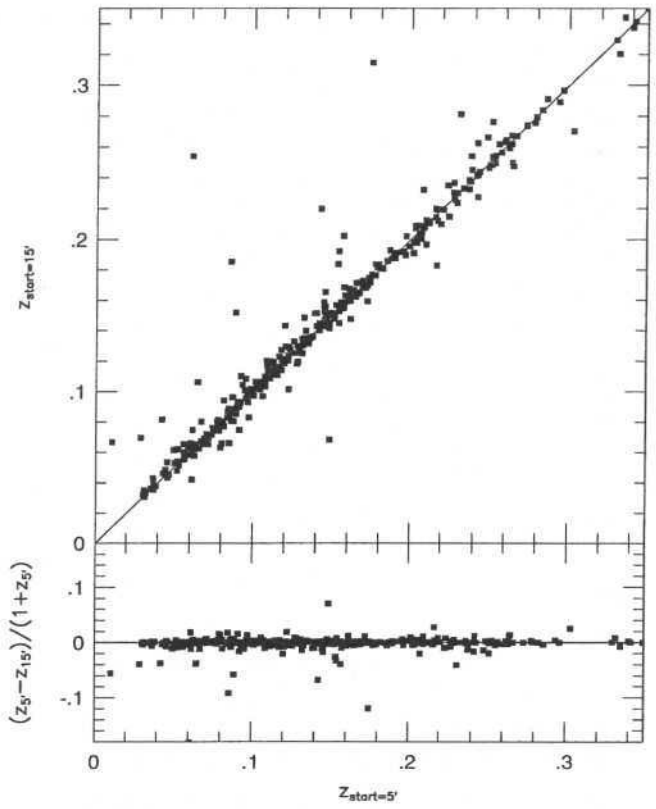


Figure 3.10: The photometrically estimated redshift using starting radii of 5 'and 15'. Differences between the estimates are shown in the bottom panel.

3.6.2 Richness

A fundamental physical property of galaxy clusters is their mass. Mass can be measured using X-ray data, velocity dispersions, lensing (strong and weak); all of these methods require data which is difficult to obtain for extremely large cluster samples, and often impossible for low-mass systems. One expects that the number of galaxies in a cluster (the cluster richness) should be correlated with the total cluster mass. Unfortunately, cluster richnesses are notoriously difficult to measure, and there are a variety of different estimators. Abell's richness, for instance, is extremely poorly correlated with mass; other estimators fare somewhat better (Yee & López-Cruz 1999). Because our catalog covers a large fraction of the sky, we can measure optical richnesses for many clusters which have spectroscopic data, as well as prepare specific subsamples for future spectroscopic study.

Given a redshift estimate for each cluster, we can measure a richness that samples the

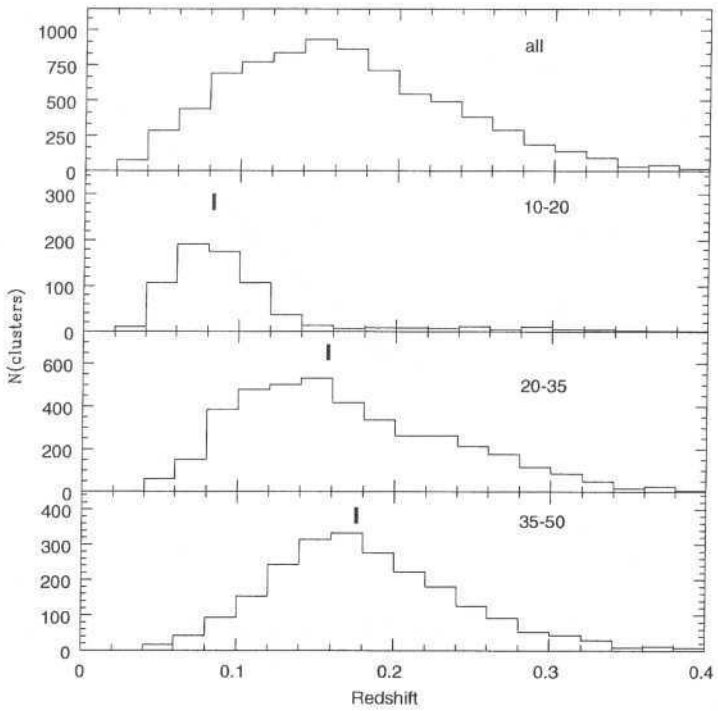


Figure 3.11: The redshift distribution of our candidate clusters. The distribution for all clusters is shown at top, with lower panels dividing the sample by N_{gals}.

same absolute magnitude range. We compute the richness in a fixed absolute magnitude interval, $M_r^* - 1 < M < M_r^* + 2$, using the r-band data, and assuming $M_r^* = -21.53$, taken from the cluster luminosity functions derived by Paolillo *et al.* (2001). Galaxies are counted within a radius of $1h^{-1}$ Mpc. For clusters at low redshift ($z<0.17$), we measure the richness by directly summing the background-corrected number of galaxies in the appropriate magnitude interval. Only at these low redshifts is this entire magnitude range contained within our data. For more distant clusters ($z > 0.17$), we directly sum the galaxies with $M_r^* - 1 < M < M_{20}$, where M_{20} is the absolute magnitude limit corresponding to our catalog limit $m_r = 20.0$. We then calculate a correction factor γ for the richness, defined as

$$\gamma = \frac{\int_{M_r^*-1}^{M_r^*+2} \Phi(M)dM}{\int_{M_r^*-1}^{M_{20}} \Phi(M)dM} \tag{3.2}$$

We set the exponent in the luminosity function to $\alpha = -1.1$, as we do for our simulated

clusters; the correction factor is slightly larger for lower α (by $\sim 5-10\%$ for $\alpha=-0.87$). Typical values of γ are 1.125 for $z=0.2$ and 1.50 at $z=0.3$.

We use a background determined individually for each plate. We simply take the distribution of galaxies as a function of magnitude for the entire plate, and subtract this (scaled to the appropriate area) from the same distribution for each cluster. This ensures that any systematic errors in an individual plate (calibration, classification) are maintained when performing the background subtraction. Our methodology lies between completely local background estimation (such as taking an annulus near the cluster), and completely global estimates (which would use the entire survey area). Tests using median filtering to remove overdense regions from the background measurement show that this method does not change the final richnesses appreciably; the mean richness is increased by $\Delta N_{gals}=3$ if this alternative method is used.

The final richness (N_{gals}) for each cluster is listed in the fifth column of Table 3.6. We show the richness distribution of our robust cluster candidates in Figure 3.12, for the whole sample and in different redshift bins. At high z, where our completeness drops, we find only richer clusters, while at low z, where the volume is smaller, we find mostly poor clusters. The median richness for our clusters is $N_{gals,med}=31$, which corresponds to the poorest end of Abell's richness class 0. We therefore expect that our catalog contains many very poor clusters (or even groups).

Richness Errors

Errors in our photometric redshift estimator will introduce an error in the measured richness for each cluster. We investigated the magnitude of this effect by calculating richnesses for the Abell clusters with spectroscopic redshifts. We calculate the richness using both the spectroscopic redshift and our photometric redshift. The results are shown in the upper left panel of Figure 3.13. The two estimates are very well correlated, with no systematic offset, and larger scatter at lower richness, where the redshift estimates are likely to be worse.

We also tested various cluster radii for measuring richness. The upper right panel of Figure 3.13 shows the ratio of richnesses using 1.5Mpc and 1Mpc radii, versus the 1.5Mpc radius richness. We see that the ratio is approximately constant regardless of richness, and that the ratio is ~ 1.3, less than the ratio of the areas. This is clearly because the outer

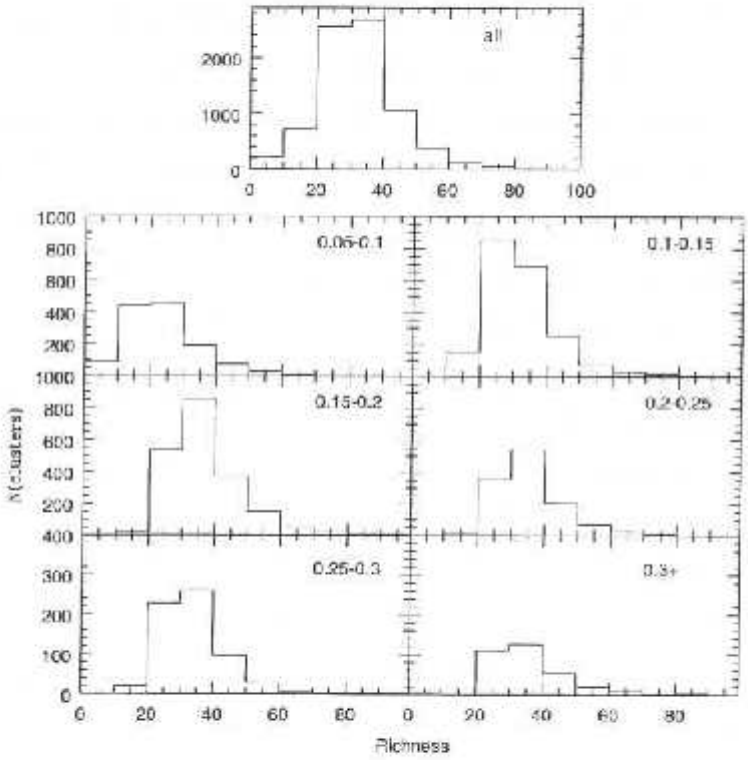


Figure 3.12: Richness distributions for the entire sample (top panel) and in different redshift bins (lower panels).

parts of the clusters have low density. We therefore use the 1Mpc radius richnesses, as the background subtraction over a larger area will introduce more noise.

There are many other potential sources of error, both random and systematic. A future paper will address many of these issues in detail, including the choice of background area, cosmology, k-corrections, the radius used, etc. We will measure cluster richnesses using a variety of techniques; this will allow detailed tests to find which methodology provides the best surrogate for a mass measurement. Nevertheless, we are confident that our *ad hoc* estimator provides a useful measurement. We have compared our richness estimator to the A_{cl} measurement of Kim (2001) for clusters detected by both surveys. The results are shown in Figure 3.14, where solid circles are clusters where the redshift estimates agree (within errors), while open circles are clusters where they do not. The two richness measures are extremely well correlated, suggesting that they are both measuring similar properties of the clusters. The agreement is quite remarkable, as they are based on different surveys, and

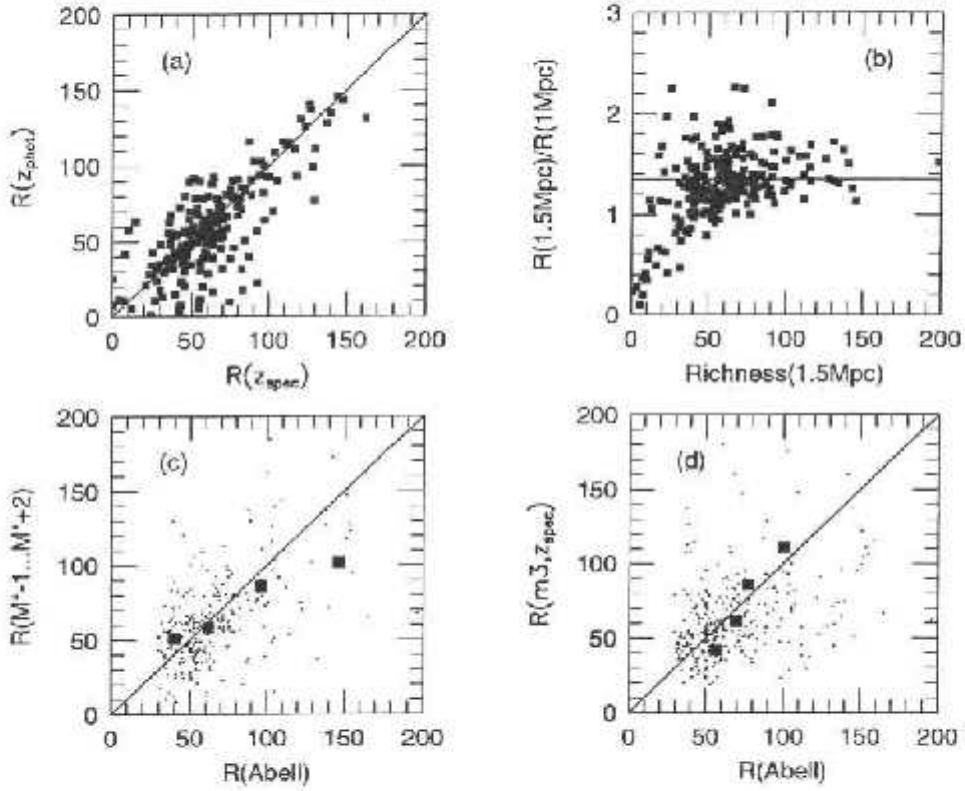


Figure 3.13: Richness estimator tests using Abell clusters with spectroscopic redshifts. Panel (a) shows our richness measure using the spectroscopic redshift from SR99 compared to the richness using our photo-z. Panel (b) is the ratio of richnesses using 1.5Mpc and 1Mpc radii. Panel (c) shows our richness (using z_{spec}) against Abell's. Panel (d) compares Abell's richness to our measurement using the Kim *et al.* (2001) technique to count galaxies with $m_3 < m < m_3 + 2$. In the lower panels, each small dot is a single cluster, with the large squares representing binned data.

measure different properties. One method (N_{gals}) simply counts galaxies, while the other considers the luminosity of the galaxies (Λ_{gl}).

Comparison to Abell

Finally, we attempted to compare our richness estimate to those of Abell. The results are shown in the lower left panel of Figure 3.13, with each comparison cluster as a small dot, and the boxes representing binned data. Although a correlation is evident, the scatter is large; significantly larger than the errors due to our measure alone. Abell's richness

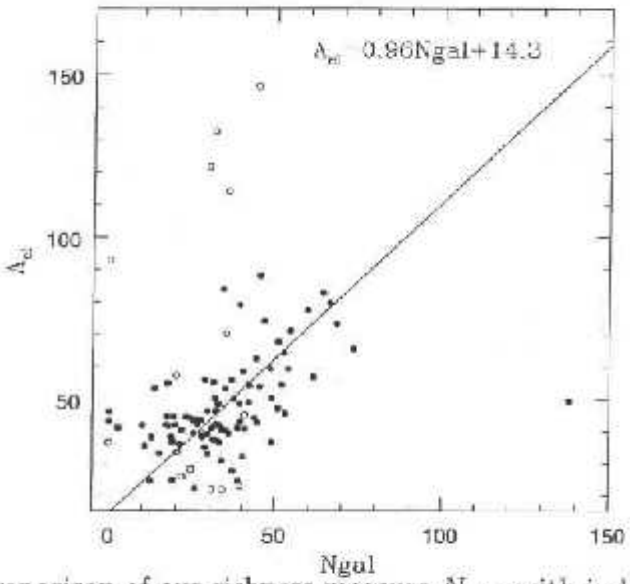


Figure 3.14: Comparison of our richness measure N_{gals} with independent measurements of Λ_{cl} from Kim *et al.* (2001).

estimate, counting from the third brightest galaxy to two magnitudes fainter, suffers from many drawbacks. First, his redshift estimates are often incorrect, causing errors in the radius used to count galaxies. The use of an apparent magnitude, m_3, to set the magnitude limits means that different clusters have their richnesses measured in different absolute magnitude ranges, and introduces random errors due to bright foreground galaxies. This effect is shown in the bottom right panel of Figure 3.13, where we plot Abell's richness against our estimate of the cluster richness measure using a version of his technique (detailed in Kim 2001). A similarly poor, but extant, correlation can be seen. The difficulties with Abell's richness measures are well documented, as discussed in §3.1.

3.7 The Cluster Catalog

The first installment of the Northern Sky Optical Cluster Survey catalog presented here covers 5834□°, and contains 8,155 cluster candidates, yielding 1.4 clusters per square degree. In comparison, Abell's 1958 survey covered a much larger area, yet contains only 2,712 clusters. Our catalogue thus represents an increase of roughly one order of magnitude in

cluster counts. More importantly, these clusters have been selected using an automated, objective algorithm, with extremely well characterized contamination rates and selection functions. Each cluster also has a consistently measured richness and photometric redshift. Figure 3.15 shows the sky distribution of all our cluster candidates. The median redshift of our sample is $z_{med} = 0.1579$, with a median richness of $N_{gals,med} = 31$. Thus, our sample is both somewhat deeper than Abell, and extends to significantly poorer systems. As mentioned earlier, the redshift distribution is shown in Figure 3.11, and the richness distribution in Figure 3.12. DPOSS F-plate images (500″ diameter) of four new, rich clusters are shown in Figure 3.16. Two were not previously known, and two (NSC172013+264028 = RXC J1720.1+2637 and NSC122906+473720 = RXC J1229.0+4737) were detected only as X-ray clusters.

We provide the complete cluster catalog for our survey area in Table 3.6, as well as at `http://dposs.caltech.edu/dataproducts/`, sorted by increasing RA.

The table provides:

1. The cluster name. The naming convention is NSChhmmss+ddmmss. Coordinates are J2000.

2. The cluster RA and

3. The cluster Dec, in decimal coordinates, also J2000.

4. The photometric redshift. If this is zero, the estimator failed to converge.

5. The measured richness. If $z_{phot} > 0.17$, the correction factor γ, as described in §3.6.2, is applied.

6. The number of simulated maps in which the cluster was detected. Only clusters detected in seven or more maps are included here.

7. The plate number from which the cluster is drawn.

Table 3.6: The Northern Sky Cluster Catalog: Excerpt

Name	RA (J2000.0)	Dec (J2000.0)	z_{phot}	N_{gals}	N_{maps}	Plate
NSCJ080140+623236	120.41708	62.54239	0.0732	28.1	10	124
NSCJ080142+575030	120.42386	57.83877	0.1356	20.6	7	121
NSCJ080205+634700	120.52274	63.78326	0.0726	45.9	10	089
NSCJ080217+633248	120.57204	63.54685	0.0626	22.9	10	089
NSCJ080218+604112	120.57571	60.68660	0.1080	23.6	10	124
NSCJ080244+642523	120.68704	64.42315	0.1107	28.0	7	089
NSCJ080245+613414	120.69027	61.40387	0.1149	29.8	10	124
NSCJ080300+637419	120.75356	63.23854	0.1434	35.3	8	089
NSCJ080320+652141	120.83549	65.36151	0.1479	29.5	7	089
NSCJ080327+640531	120.86491	64.09182	0.1341	28.6	7	089
NSCJ080337+585755	120.90653	58.96515	0.2260	20.4	10	124
NSCJ080347+633326	120.94792	63.55546	0.0874	30.0	9	089
NSCJ080401+625402	121.00716	62.90080	0.0850	33.0	8	089
NSCJ080435+393027	121.14700	39.42425	0.1914	27.7	10	312
NSCJ080542+670818	121.42574	67.13827	0.0942	32.2	10	089
NSCJ080608+374711	121.53357	37.75649	0.2353	30.2	10	312
NSCJ080615+381832	121.56289	38.30891	0.2618	60.7	10	312
NSCJ080621+514129	121.59971	51.69127	0.1091	33.1	10	124
NSCJ080650+520358	121.71079	52.06607	0.2518	37.6	10	124
NSCJ080717+393521	121.82319	39.58903	0.0939	15.6	10	312
NSCJ080745+605029	121.94125	60.84157	0.2797	29.7	10	124

3.7.1 Consistency with Previous Surveys

Although our data and methodology is different from those of prior surveys, we expect that optically-based cluster searches will find similar objects. Thus, we may ask if our results are consistent with other recent surveys.

The obvious survey to compare with is that of Abell. In the area covered in this paper, there are 1090 Abell clusters. Of these, $\sim 93\%$ are recovered by us; an exact number is difficult to determine due to centroiding errors, varying matching radii, unknown redshifts, proximity to excised areas, and other effects; some of these will be treated in a full comparison when the second (and final) portion of our catalog is completed. Nevertheless, the high recovery rate suggests that the Abell catalog consists mostly of real clusters (as argued by Miller *et al.* 1999), but is incomplete, as our sample contains eight times more clusters. Unfortunately, it is difficult to assess the Abell catalog incompleteness in the regime where it is used for statistical studies ($N_{Abell} > 50$, $z < 0.2$) because of the poor correlation between Abell's richness and our N_{gals}. If we assume N_{gals} and N_{Abell} are equivalent, then the two catalogs contain very similar numbers of clusters in the aforementioned range. However, a small shift in the limits used for comparison (for instance, setting $N_{gals} \geq 45$) would lead one to conclude that the statistical Abell catalog is $\sim 35\%$ incomplete. These discrepant results

are consistent with the difficulties encountered by other authors in assessing the completeness of the Abell catalog.

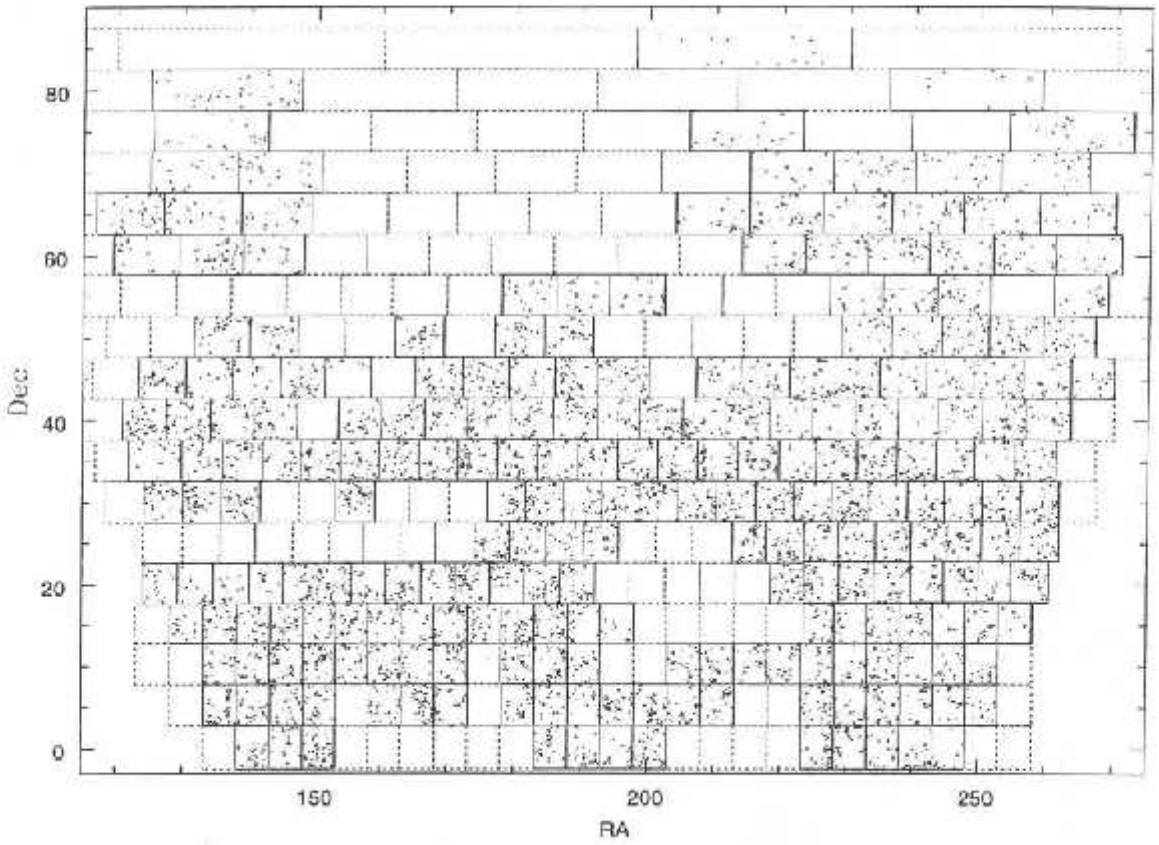


Figure 3.15: The sky distribution of our candidate clusters. Black lines show the boundaries of plates used in this paper; dotted boundaries are unused plates with poorer calibration.

An excellent alternative comparison sample is that of P02, who discuss the surface density of clusters as a function of redshift. They find 1.19 clusters per square degree for $z \leq 0.2$ and $\Lambda_{cl} \geq 40$. In our 5800 square degree sample, we therefore expect 6902 clusters for the same range of parameters. Unfortunately, we do not measure Λ_{cl}, but P02 provide a conversion to Abell richness, $\Lambda_{cl} = 1.24 \times R_{Abell}$; their Λ_{cl} cut thus corresponds to $R_{Abell} \geq 30$. At these low richnesses, the Abell richness corresponds roughly to our N_{gals} (see Figure 3.13). We therefore apply these cuts ($z_{phot} \leq 0.2$, $N_{gals} \geq 30$) to our catalog, resulting in a sample of 2888 clusters. We then apply our selection function (shown in Figure 3.3); at the higher end of

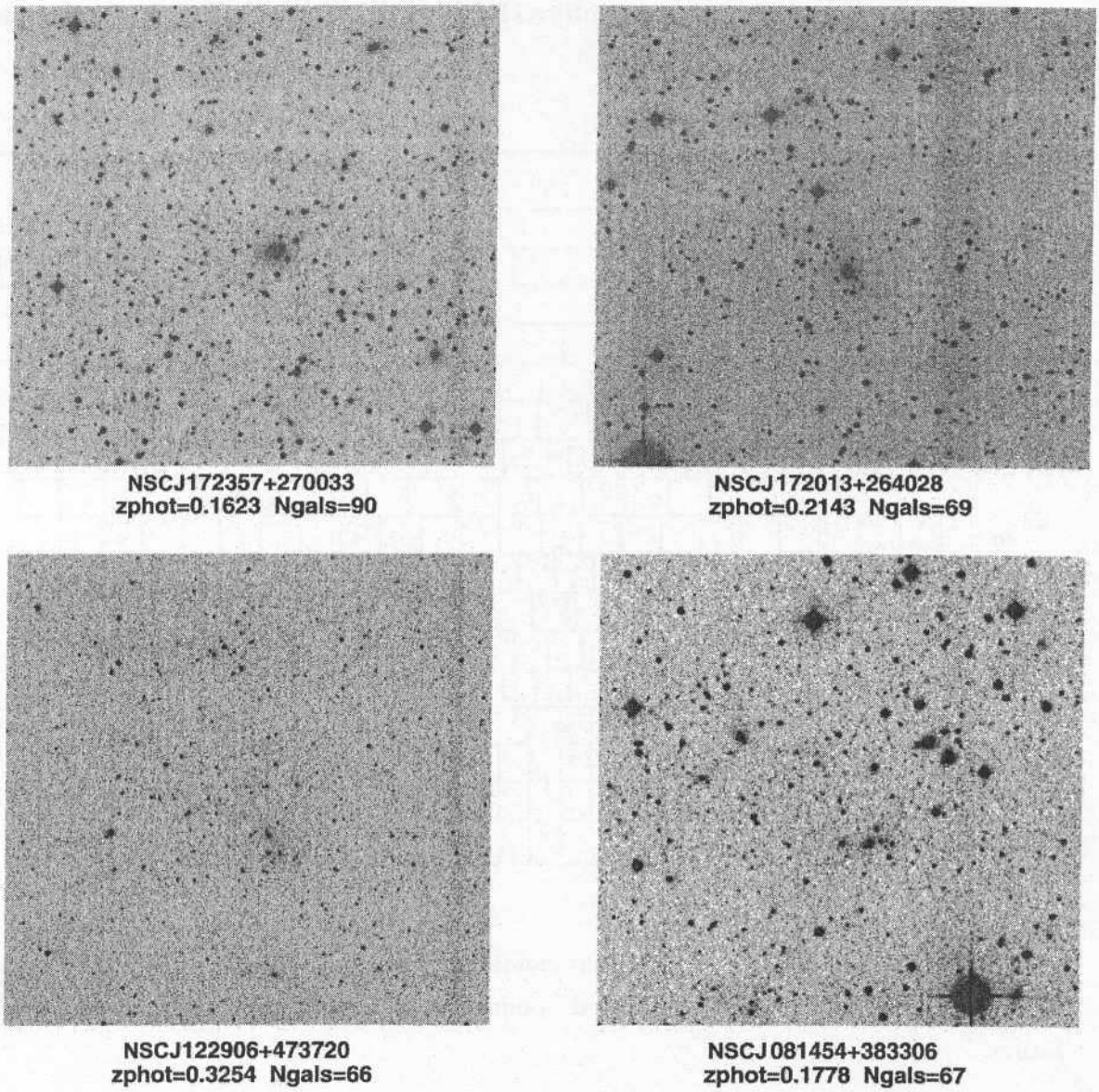


Figure 3.16: DPOSS F-plate images of four rich clusters. Two were not previously known, and two (NSCJ172013+264028 = RXC J1720.1+2637 and NSCJ122906+473720 = RXC J1229.0+4737) were detected only as X-ray clusters.

the redshift range ($z \sim 0.15$) our detection efficiency is quite low for the most common, poorer clusters. The estimated actual number of clusters expected in our survey area, after applying this correction, is 6461 clusters, compared to 6902 predicted by the constant comoving space density given by P02. This is only a 6% difference, which is remarkable given that both

surveys use different techniques, and have to apply estimated selection functions to arrive at actual space densities.

This excellent agreement suggests that both our survey and the smaller but deeper survey of P02 detect similar objects in the overlapping redshift range. In addition, the selection functions presented by the two surveys provide realistic estimates of their completeness.

We note also that P02 find a space density of Abell-like clusters that is a factor of 1.5 higher than in the Abell catalog. This result relies on a conversion of Λ_{cl} to N_{Abell}, which, like our own conversion, is somewhat uncertain. Nevertheless, the agreement in density between our survey and P02 does suggest significant incompleteness in the Abell catalog. Because the number of detected clusters is strongly increasing at the poor end of Abell's $R = 1$ bin, these results are very sensitive to small changes in the richness cuts applied.

A preliminary comparison to the NORAS X-ray cluster survey (Böhringer *et al.* 2000) shown in Gal (2001) demonstrates that nearly 90% of NORAS clusters at $z < 0.2$ are detected in DPOSS. In contrast, the DPOSS cluster catalog contains over an order of magnitude more clusters than NORAS; in most cases these are poor systems. To understand the biases in these catalogs, we are undertaking a full comparison in which the same properties (f_X, N_{gals}, z_{phot}) are consistently measured for both NORAS and DPOSS clusters.

3.8 Discussion and Future Directions

We have presented the community a large, new cluster catalog, which meets the original objectives outlined at the beginning of this paper. This catalog includes accurate photometric redshifts and richnesses for a sample of 8,155 cluster candidates. Additionally, we provide detailed selection functions in both redshift and richness, on an individual plate basis, which can be used in the generation of mock catalogs for testing cosmological models, measuring the correlation function, and other large scale structure studies.

Certainly, this will be superseded in the future by deeper surveys, with more accurate photometry in more bandpasses, and with superior spectroscopic or photometric redshift information. Our catalog is still limited by the use of the projected galaxy distribution to detect clusters, and the poor photometric accuracy in comparison to modern digital surveys, such as the SDSS.

In the future, we will publish an additional catalog, covering the Southern Galactic Cap region, and including the less well calibrated plates from the NGP. The final Northern Sky Cluster catalog will then cover over 10,000 square degrees.

The scientific uses for this catalog are numerous. Measurements of statistical properties of the cluster population, such as the cluster-cluster correlation function ξ_{cc} and the cluster mass function are all avenues of further research. Comparisons to the relatively large X-ray selected samples from the RASS are forthcoming, with the potential to find many poorer systems by performing joint optical/X-ray detections. The existence of independent samples of thousands of clusters will allow us to evaluate the biases present in the different detection methods.

We also plan to examine the multiplicity function (Puddu *et al.* 2003), from small groups to rich clusters. This requires a separate catalog of poor groups, which is being constructed using a different algorithm by Iovino *et al.* (2003). The catalog presented here and the compact group catalog are expected to have significant overlap, which will also provide important cross-checks in the difficult domain of small galaxy associations.

This sample is also fertile ground for the selection of specific, well-defined subsamples for follow-up observational studies. Using these cluster locations and the now public DPOSS data, we can search for clusters with substructure, excess blue galaxy populations, unusual optical/X-ray flux ratios, or any of a variety of interesting properties. Other methods for cluster detection have also been applied to our galaxy catalogs (Puddu *et al.* 2001), which can be used for further testing and comparisons.

We remind the reader that all of the data for this survey (including future installments) can be found at `http://dposs.caltech.edu/dataproducts/`.

We thank the Norris Foundation and other private donors for their generous support of the DPOSS project. RRG was supported in part by an NSF Fellowship, NASA GSRP NGT5-50215, a Kingsley Fellowship, as well as a KDI grant to A. Szalay. This work would have been impossible without the POSS-II photographic team and the STScI digitization team. We also thank the Palomar TAC and Directors for generous time allocations for the DPOSS calibration effort. The DPOSS survey would have been impossible without the hard work of the POSS-II observing staff, and numerous undergraduates who assisted with the calibration. RRG would like to thank Sandra Castro for assistance with the spectroscopy

reductions, and RdC thanks Hugo Capelato, Angela Iovino and Gary Mamon for discussion on several issues in this project. This work was made possible in part through the NPACI sponsored Digital Sky project and a generous equipment grant from SUN Microsystems. Access to the POSS-II image data stored on the HPSS, located at the California Institute of Technology, was provided by the Center for Advanced Computing Research.

Capítulo 4

Aglomerados de Galáxias com Desvio para o Vermelho Intermediário ($< z > \sim 0.30$)[1]

Resumo

Apresentamos neste capítulo um catálogo com 9982 candidatos a aglomerados de galáxias selecionados em $\sim$ 2700 graus quadrados dos dados digitalizados do segundo levantamento fotográfico do Observatório Palomar. Este trabalho é uma extensão do *Northern Sky Optical Cluster Survey* (NoSOCS), apresentado no capítulo anterior. Nosso principal objetivo neste capítulo é explorar a utilização do DPOSS no seu limite. Descrevemos aqui a estratégia para detectar aglomerados no intervalo $0.1 < z < 0.5$. Para tal, nós utilizamos as melhores placas (na banda F) do DPOSS, selecionadas de acordo com a latitude galáctica ($|b| > 50°$), *seeing* e magnitude limite.

A detecção de aglomerados é baseada em catálogos de galáxias do DPOSS limitados por $r = 21.1$, em comparação com $r = 19.5$ adotado no capítulo 3. Neste limite a separação estrela-galáxia não é mais confiável. Portanto, utilizamos somente os campos com $|b| > 50°$, onde esperamos que a contaminação estelar seja aproximadamente uniforme. Além disso, uma seleção estatística é realizada para converter galáxias erroneamente classificadas como estrelas nos limites do DPOSS (seção 4.2).

Utilizamos dois métodos para a geração do catálogo de aglomerados. O primeiro é baseado numa técnica de *kernel* adaptativo (KA, capítulo anterior), e o segundo é baseado no método

[1]Os resultados exibidos neste capítulo são provenientes do artigo "The Northern Sky Optical Cluster Survey IV: An Intermediate Redshift Galaxy Cluster Catalog and the Comparison of Two Detection Algorithms", Lopes, P. A. A., de Carvalho, R. R., Gal, R. R., Djorgovski, S. G., Odewahn, S. C., Mahabal, A. A., Brunner, R. J. 2003a, AJ, *em preparação*.

conhecido como *voronoi tesselation* (VT) (seção 4.3). Os dois métodos são optimizados de maneira similar a apresentada no capítulo 3, no entanto, a taxa de contaminação aqui adotada é de $\sim 5\%$. Apresentamos também (seção 4.4) uma comparação das estimativas do número de falsas detecções para diferentes tipos de distribuições de galáxias de fundo.

Simulações de aglomerados são utilizadas para obter a função de seleção dos dois métodos (seção 4.6). Diferentemente do capítulo anterior, a função de seleção não é obtida para todos os campos. As estimativas para alguns campos são dadas como exemplo dos limites de completeza de cada método. Utilizamos ainda a função de seleção para (i) verificar se as distribuições de desvio para o vermelho (estimado) e riqueza são compatíveis com nossas previsões; (ii) para confirmar se os dois métodos apresentam bom acordo no intervalo onde ambos esperam alta completeza; e (iii) para testar um possível ganho em completeza quando permitindo uma taxa de contaminação mais alta (10% ao invés de 5%). No entanto, uma taxa de contaminação de $\sim 5\%$ foi utilizada neste trabalho, como será discutido adiante.

Estimativas do desvio para o vermelho e da riqueza dos candidatos a aglomerados são apresentadas (seção 4.5). Estas estimativas são úteis para seleções de sub-amostras para estudos posteriores, além de serem de fundamental importância para compararmos o desempenho dos dois métodos para detecção de aglomerados. Como descrito na seção 4.7, o catálogo baseado em VT é mais completo no regime de baixos desvios para o vermelho e sistemas pobres, ao passo que o KA apresenta melhor desempenho para mais altos valores de desvio para o vermelho. Estes resultados estão em bom acordo com as estimativas da função de seleção.

Uma comparação do catálogo aqui obtido com os catálogos originais do NoSOCS é apresentada na seção 4.8. Encontramos um excelente acordo no regime em que ambos os catálogos tem expectativas de alta completeza ($0.1 < z < 0.2$ e $N_{gals} > 65$). Nossos resultados são também comparados com os catálogos preliminares do SDSS. Nossas estimativas de riqueza apresentam boa correlação com as de Kim (2001), e o catálogo proveniente do DPOSS apresenta alto nível de coincidência com os catálogos do SDSS (Kim 2001 e Goto *et al.* 2002) no regime de $0.2 < z < 0.3$ e $N_{gals} > 65$.

4.1 Introduction

Clusters of galaxies constitute the largest bound structures in the universe. Hence, clusters have been widely used to trace the mass distribution and its evolution (Bahcall & Soneira 1983; Postman, Huchra & Geller 1992; Guzzo *et al.* 1992), as well as to place constraints on cosmological models (Bahcall *et al.* 1997; Evrard 1989; Viana & Liddle 1996; Carlberg *et al.* 1997). Galaxy clusters are also likely composed of coeval stellar systems, and are therefore well suited for studying the formation and evolution of galaxies in dense environments (Aragón-Salamanca *et al.* 1993; Butcher & Oemler 1984; Dressler *et al.* 1997; Margoniner *et al.* 2001). Studies of this nature require the use of a well defined and understood sample of galaxy clusters. Although whole sky cluster catalogs exist, they typically are subjective in nature or do not span a large redshift range. On the other hand, deep cluster catalogs are limited to small regions of the sky.

In the last decade many authors have used different wavelengths and techniques to identify clusters of galaxies. Each method has its own biases, and an ideal sample would derive from a combination of multiple data sets and techniques. Optical cluster catalogs use galaxy overdensities as a proxy for mass overdensities. Clusters lacking a significant galaxy population are not likely to be recovered by optical surveys, although the existence of many such clusters is questionable. Optical catalogues are observationally cheap, but have projection effects as a critical drawback. However, such effects can be minimized with the aid of colors. X-ray catalogues do not suffer from projection effects, but are biased against clusters with an unresolved gas distribution or lower gas content. Some other methods include the weak-lensing technique (Wittman *et al.* 2001) and the search for distortions in the cosmic microwave background radio emission, where clusters may scatter the microwave background radiation via the Sunyaev-Z'eldovich effect (Sunyaev & Zeldovich 1980).

Most cluster studies to date have made use of the Abell catalog (Abell 1958; Abell, Corwin & Olowin 1989), which is the result of visual inspection of photographic plates. Other examples of subjective catalogs generated from plate data can be found in Zwicky, Herzog & Wild (1968) and Gunn, Hoessel & Oke (1986). The main drawbacks of human-based cluster searches are incompleteness, reproducibility and generation time, as well as posing difficulty in quantifying the selection effects associated with the resulting catalog.

Sheetman (1985) was the first to use an automated method to search for clusters, with much progress since that time. Many automated cluster finding techniques have been developed, with some applied to wide field data. A few examples are the APM Cluster Survey (Dalton, Maddox, Sutherland & Efstathiou 1997), the EDSGC (Lumsden, Nichol, Collins & Guzzo 1992) and the EDCCII (Bramel, Nichol & Pope 2000). Except for the work of Gunn, Hoessel & Oke (1986) all the catalogs mentioned above sample the nearby universe. Automated or not, all of these surveys are based on plate data, including that of Gal *et al.* (2003a), which contains rich clusters to $z \sim 0.3$ in comparison to the $z \sim 0.15 - 0.20$ limit of previous surveys.

A large number of deep optical/near-IR surveys have recently become available, reaching as deep as $z \sim 1.4$, but the sky coverage is typically a few tenths to a few square degrees. The pioneering work is that of Postman *et al.* (1996), who developed and utilized a matched filter technique. Following this work, a large number of cluster catalogs have became available both in the northern and southern hemispheres (Olsen *et al.* 1999; Lobo *et al.* 2000; Gladders & Yee 2000; Gonzalez, Zaritsky, Dalcanton, & Nelson 2001; Postman, Lauer, Oegerle, & Donahue (2002). At $0 < z < 0.6$ there are also preliminary results from the Sloan Digital Sky Survey (SDSS), such as those of Kim (2001), Annis *et al.* (1999) and Goto *et al.* (2002).

A necessary byproduct of these surveys was the development of various techniques for cluster detection, which utilize different properties of the clusters. Some of these algorithms and techniques include the matched filter (MF) (Postman *et al.* 1996) and all its variants (Kawasaki *et al.* 1998; Kepner *et al.* 1999; Lobo *et al.* 2000; Schuecker & Böhringer 1998; Kim *et al.* 2002); Voronoi Tessellation (Ramella *et al.* 2001; Kim *et al.* 2002); the Adaptive Kernel (Gal et al. 2000a, 2003a, 2003b - hereafter Papers I, II and III respectively); Surface Brightness Fluctuations in shallow images (Gonzalez, Zaritsky, Dalcanton, & Nelson 2001) and methods based on the color and magnitude of the cluster galaxy population (Gladders & Yee 2000; Goto *et al.* 2002; Annis *et al.* 1999). Nonetheless, only a few authors have compared the performance of different cluster search techniques on the same data. Kim (2001) and Kim *et al.* (2002) applied a matched filter, an adaptive matched filter (AMF) and a Voronoi Tessellation technique (VT) with a color cut to $\sim 152\ \square^{\circ}$ of SDSS commissioning data. The final comparison merged the MF and AMF in a hybrid matched filter (HMF) which uses the MF to create likelihood maps and detect cluster candidates, while the AMF determines

richness and redshift more precisely, resulting in a final comparison with two cluster detection algorithms (HMF and VT). Goto *et al.* (2002) compared their results with Kim (2001) and Annis *et al.* (1999). Bahcall *et al.* (2003b) are currently working on comparing cluster catalogs based on a variety of cluster finding techniques applied to SDSS data.

This paper presents a galaxy cluster catalog covering 2700 $\square^{\circ}$ of DPOSS data. The catalog is expected to be complete for rich clusters out to $z \sim 0.30$, with clusters still detected to $z \sim 0.5$. In the regime of poor systems this catalog is shallower than the SDSS preliminary cluster catalogs (Kim *et al.* 2002; Goto *et al.* 2002; Annis *et al.* 1999), but the sky coverage is larger. The main goal of this project is to provide a catalog of rich structures to $z \sim 0.5$ covering the northern hemisphere. We also aim for a low contamination rate at the level of $\sim 5\%$. As no other similar resource is available to date, this catalog represents a valuable reference for follow-up studies. Our catalog will also serve as a valuable cross reference for the final SDSS cluster catalog as well as X-ray based catalogs.

We also demonstrate the feasibility of using photometric survey data for cluster detection in the regime where star/galaxy separation is not reliable. At the faint end of DPOSS ($m_r > 19.5$) misclassified stars can contaminate galaxy catalogs. As in Postman, Lauer, Oegerle, & Donahue (2002) (hereafter P02), we assess the probability that a faint object classified as a star is indeed a galaxy. We apply this statistical correction as a function of magnitude to obtain an object catalog in agreement with the galaxy counts to $m_r = 21.1$. In this way, we minimize stellar contamination effects which would otherwise reach 25% of the galaxy counts. We also restrict the search to fields at high galactic latitude, where the star counts are approximately uniform. Finally, as a byproduct, this work presents a performance comparison of two different cluster search algorithms applied to DPOSS data.

Since in this work we do not impose a uniform completeness limit over the entire surveyed area, we do not recommend this catalog for studies such as the cluster angular correlation function or estimates of the cluster space density and mass function. At the faint end of DPOSS photometric errors are quite large (Gal *et al.* 2003c), playing a key role in the generation of a cluster catalog. Poor and/or distant clusters are extremely sensitive to the input galaxy catalog, which is severely affected by our poor photometry. Such studies can be addressed using the catalogs from Papers II and III. However, this paper provides a valuable source of rich, relatively distant optical clusters. This sample can be used to select candidates

Observatório Nacional

Ministério da Ciência e Tecnologia

BRASIL UM PAÍS DE TODOS GOVERNO FEDERAL

December 10, 2009.

Dear Dr. Dimitrios N. Argyriou,

As you will be in Brazil at the end of the year I would like to invite you to visit the Geophysics Coordination of the National Observatory, at Rio de Janeiro (ON/MCT) to discuss on possible collaborations.

We are aware that your time in Rio is short, and with the Christmas holidays you only have three days available: 21 and 22/12/2009 and 5/1/2010. For us these dates are perfect.

Unfortunately we will not be able to cover any of your expenses.

Looking forward to see you soon.

Andres R. R. Papa
Chefe da Divisão de Programas de Pós-Graduação
Observatório Nacional/MCT
P.O. 846/2007

Andres R. R. Papa
General Coordinator of Master and PhD Programs
Observatório Nacional (www.on.br)

Rua General José Cristino, 77 – Rua General Bruce, 586 – São Cristóvão – Rio de Janeiro – RJ - Brasil * CEP 20.921-400 *
Tel.: (21) 2580-6087 – Fax.: (21) 2580-6041 – dir@on.br - www.on.br

4.2.1 How far can we see with DPOSS?

To detect clusters at a given redshift, a photometric survey should be deep enough to sample sources significantly fainter than the characteristic luminosity at that redshift. Figure 4.1 shows the apparent magnitude-redshift relation for elliptical galaxies at different ranges around m^*. We use $M_r^* = -21.52$ (Paolillo *et al.* 2001) and assume clusters are dominated by early-type galaxies. The K-correction for elliptical galaxies is derived with the use of the spectral energy distributions from Coleman, Wu, & Weedman (1980), convolved with the DPOSS r-band filter. We adopt $m_r = 21.1$ as the limiting magnitude of the survey, as indicated by the vertical line in Figure 4.1. If we assume that we need to go as deep as $m^* + 1$ to detect a significant number of elliptical galaxies to identify a cluster over the background, then the survey limit lies below $z = 0.4$. At $z \sim 0.5$ we are able to identify only the richest systems, which have a large number of galaxies brighter than m^*. We should keep in mind that these limits are valid for clusters entirely composed of early-type galaxies. The presence of late-type galaxies would render distant cluster members brighter as the K-correction effects are less strong.

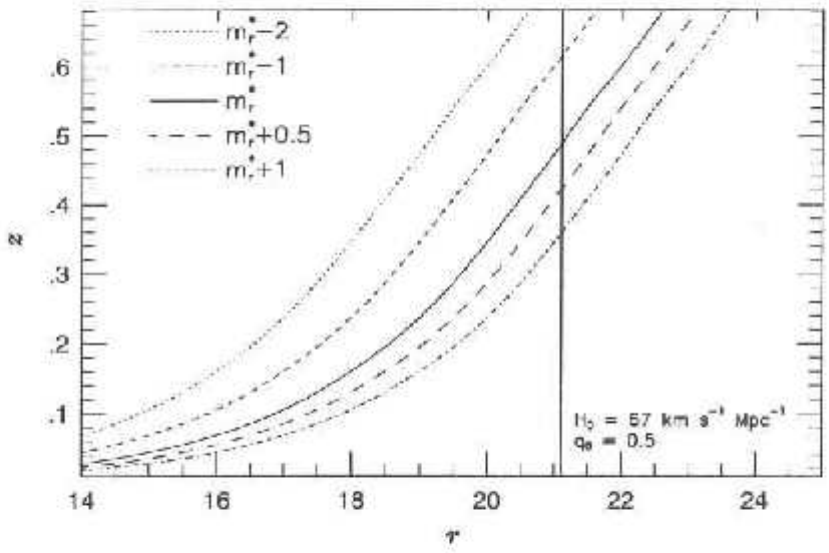


Figure 4.1: Magnitude-redshift relation for elliptical galaxies with $M_r^* = -21.52$. The survey limit is indicated by the vertical line at $m_r = 21.1$.

Figure 4.2 shows the differential magnitude distribution of galaxies within the projected area of four rich galaxy clusters (solid lines) with known spectroscopic redshifts and richnesses

given by N_{gals} (§4.5.2). Field counts are also shown (dotted lines). All counts are normalized to one square degree. The shift of the luminosity function to fainter apparent magnitudes as we go to higher redshifts is obvious. However, it is clear from this figure that the cluster profile is easily differentiated from the background counts for these rich clusters to at least $z \sim 0.35$.

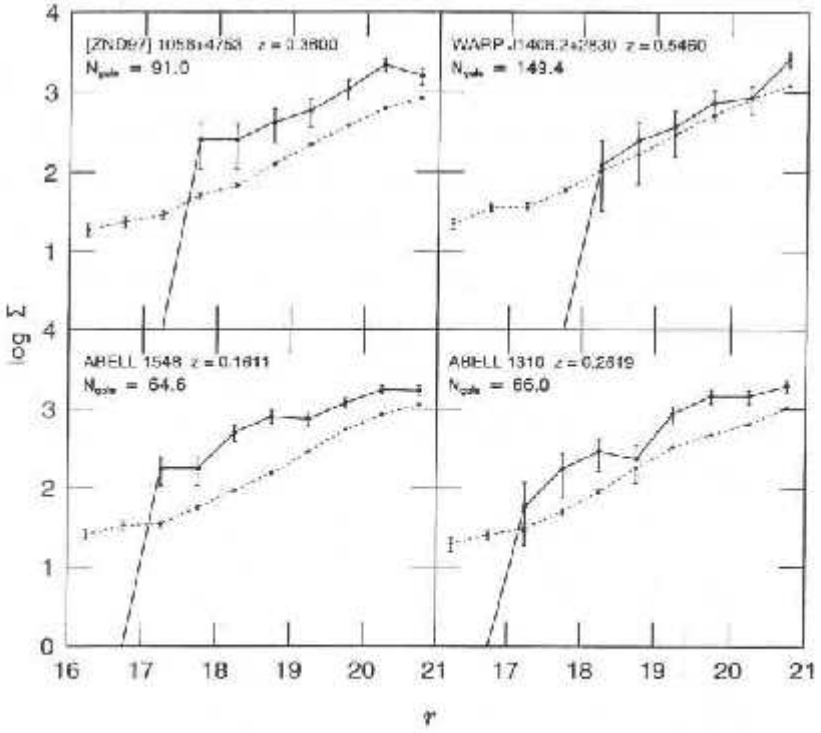


Figure 4.2: Differential magnitude distributions for four rich galaxy clusters with known spectroscopic redshifts. Cluster counts are shown as solid lines, while dotted lines show the local background counts. All counts are normalized to an area of one square degree.

4.2.2 Field Selection

As mentioned above, the star-galaxy separation accuracy drops below the 90% level at $m_r > 19.5$. Photometric errors and observing conditions also play a key role in the generation of an input catalog for our cluster search. To mitigate these effects, we apply a galactic latitude cut when selecting the DPOSS fields to be used. Additionally, we exclude fields based on their limiting magnitudes and seeing, as outlined below.

The plate catalogs used in this paper contain galaxies at $16 \leq m_r \leq 21.1$. Initially we were simply ignoring classification for the faint magnitudes ($m_r > 19.5$). We assumed stars would add a constant background to the galaxy distribution at high galactic latitudes. However, this contribution is not expected to be mild. We therefore adopt a statistical approach to check the probability that a faint star is indeed a galaxy. This procedure is explained below.

The first criterion to select DPOSS fields is the galactic latitude ($|b|$). We look for regions on the sky where density gradients as a function of $|b|$ are minimized. In these areas we can assume that the stellar contamination (regardless of the actual level) represents a uniform background added to the galaxy distribution. Figure 4.3 shows the stellar (left panel) and galaxy (right panel) number counts for different magnitude bins as a function of galactic latitude ($|b|$). The magnitude bins are indicated in the right panel. All lines represent spline fits to the mean density from 375 DPOSS plates. There is a strong dependence on galactic latitude, mainly for low-$|b|$ fields ($|b| < 45°$), due to misclassified stars close to the galactic plane, with the increase in density due to stellar contamination in these low-$|b|$ fields.

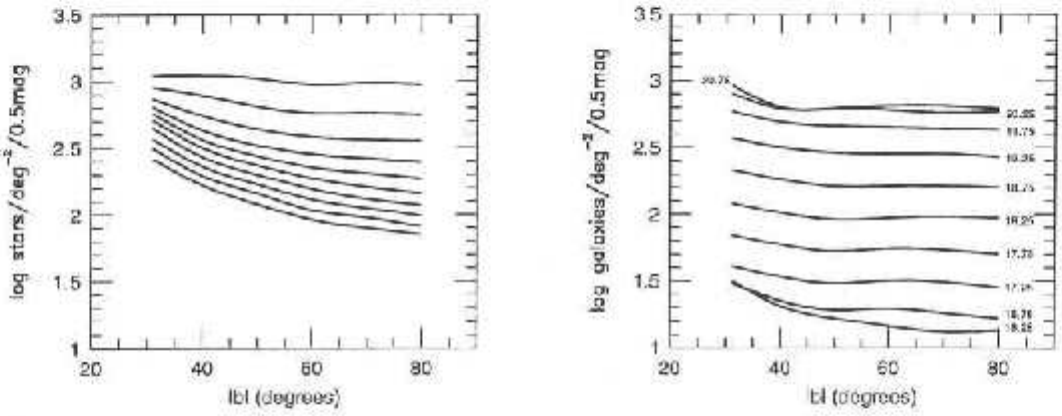


Figure 4.3: Star and galaxy counts as a function of galactic latitude for different magnitude bins. The left panel shows the star counts, while the right panel shows the galaxy counts. The magnitude bins are indicated in the right panel.

Another issue is whether the Poissonian fluctuations in the number of stars misclassified as galaxies could affect the cluster detection. In order to estimate the magnitude of this effect, we computed the expected $\sqrt{N}$ fluctuations in the number of such misclassified stars within a typical cluster aperture, as functions of both magnitude and galactic latitude. We

find that for $m_r < 19$, the contamination is less than 0.1 star per cluster aperture, increasing to a level of ~ 2 stars per cluster aperture at $m_r = 20.5$. Thus, we conclude that the effect of such fluctuations is negligible for our purposes.

Based on Figure 4.3, we conservatively select only fields with $|b| > 50^{\circ}$. This is motivated by the fact that in this latitude range we expect any contamination caused by stars to be approximately uniform. However, we should still avoid high contamination levels. Initially we decided to simply ignore classification for magnitudes fainter than $m_r = 19.5$ (the star-galaxy classification limit). Unfortunately, in this magnitude range galaxy counts do not exceed star counts by a large fraction. At $m_r = 19.5$ the ratio of galaxy to star counts is expected to be only 1.3, increasing to 2.5 at $m_r = 20.5$. If we ignore classification for objects with $m_r > 19.5$, 30% of our catalog would consist of stars (misclassified as galaxies) at these magnitudes. Since $\sim 84\%$ of all objects with $m_r < 21.1$ have $m_r > 19.5$, the overall expected stellar contamination would be $\sim 25\%$. To reduce this effect and generate object catalogs with statistically reliable galaxy counts, we opted for a statistical approach to assess the probability that faint stars are actually galaxies. This methodology was previously employed by P02 and Postman *et al.* (1998). The procedure consists of the extrapolation of the bright star counts ($m_r < 18.5$) to fainter magnitudes. We can then compare the number of stars that should be in each magnitude bin to the actual value found in the DPOSS survey, computing the probability that a given star in a given magnitude bin is actually a misclassified galaxy. This function is applied to the DPOSS faint stars ($m_r > 18.5$), having as a final product an object catalog with statistically reliable galaxy counts. The probability function (described below) is given by

$$P = 5.16 - 3.2 \times 10^{-3} r^2 + 4.97 \times 10^{-5} r^4 \qquad (4.1)$$

This effect is illustrated in Figure 4.4, where we plot the number counts for stars (squares) and galaxies (triangles) as a function of magnitude, to $m_r = 20$, in bins of 0.5^m. The dotted line is the best fit to the bright galaxy counts, extrapolated to $m_r = 21$. The dashed line is the relation used to assess the probability given by equation (4.1); it is the best fit to the bright star counts ($m_r < 18.5$) extrapolated to $m_r = 21$. Equation (1) is derived from the comparison of the number of faint stars found in DPOSS to the expected value given by the dashed line relation. The corrected galaxy counts are shown as circles. These are only slightly higher than the observed counts (triangles) at $m_r = 19.5$. This is in agreement with

the 90% galaxy success rate at this magnitude. The circles are also well fit by the dotted line at faint magnitudes, demonstrating that the corrected counts are in agreement with the extrapolated galaxy counts from the brighter bins.

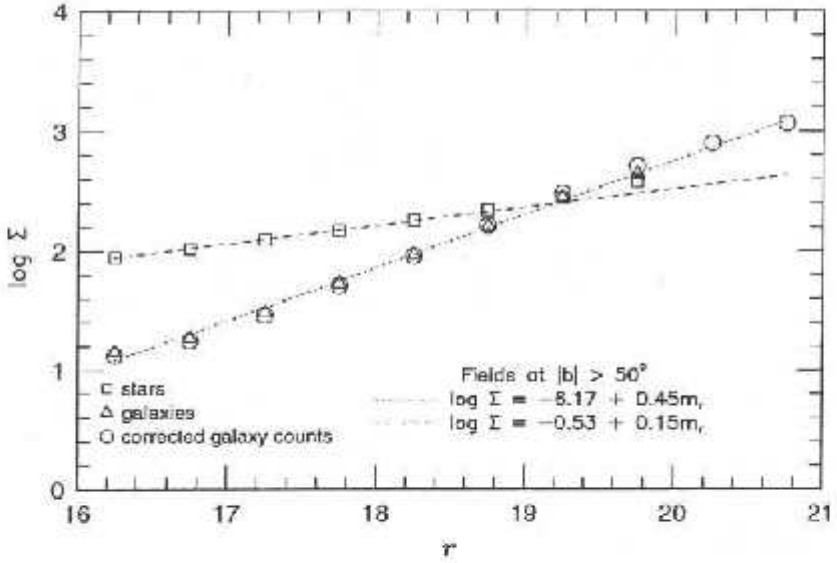


Figure 4.4: Star and galaxy counts at $|b| > 50°$. Stars are shown as squares and galaxies as triangles. The dotted line is the best fit to the bright galaxy counts extrapolated to $m_r = 21$. The dashed line represents the bright star counts ($m_r \leq 18.5$) extrapolated to $m_r = 21$. Circles represent the corrected galaxy counts out to $m_r = 21$. Error bars are equivalent to the error estimates for field counts in Figure 4.2.

We apply two additional criteria to select DPOSS fields for this survey. First, we restrict the cluster search to fields with good seeing. We use the intensity-weighted second moment of the light distribution for stellar objects ($IR2_r$) as an indicator of the quality of the observations. Figure 4.5 shows the $IR2_r$ distribution for $|b| > 50°$ fields. We use only fields with $IR2_r \leq 1.98''$, (close to the median value) as indicated by the vertical line in the plot. This is equivalent to seeing measures varying from $2.0''$ to $2.5''$.

Finally, we exclude fields with limiting magnitudes brighter than $m_r = 21$. The top panel of Figure 4.6 shows the magnitude (r band) distribution for a typical DPOSS field. As in Picard (1991), we consider the magnitude limit to be the magnitude where the distribution starts to drop steeply (as indicated by the arrow). The bottom panel of the same figure shows the distribution of magnitude limits, with the chosen cut indicated by a vertical line

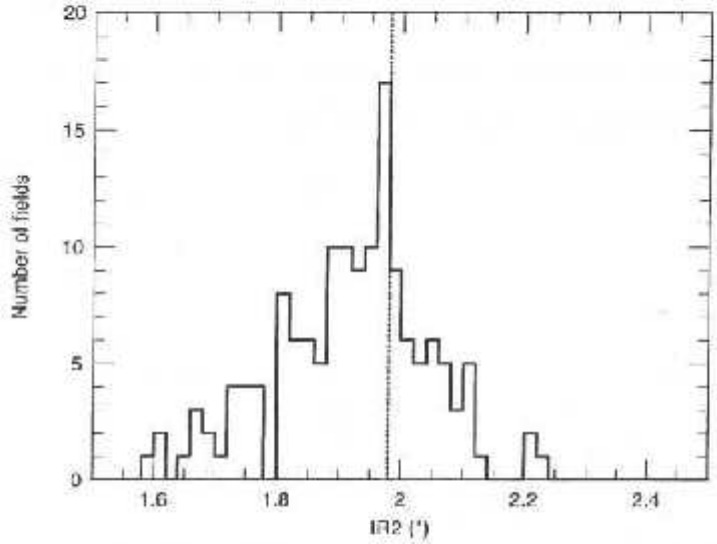


Figure 4.5: The intensity-weighted second moment of the light distribution for stellar objects (IR2) in the r-band is given as an indication of the quality of DPOSS fields. Its distribution is shown above for the well calibrated and high latitude DPOSS fields. We select fields with IR2 $< 1.98''$.

at $m_r = 21$.

To summarize, we select the best fields for this project based on three criteria. Initially we begin with 375 fields. Excluding fields with poor photometry and at $|b| < 50°$ leaves 146 fields, which is reduced to 109 after selecting those with IR2 $\leq 1.98''$. Finally we eliminated one field which has a bright magnitude limit, $m_r < 21$. The final sample used for the cluster search thus comprises 108 fields distributed over the northern sky, providing an area coverage of ~ 2700 square degrees.

4.3 Cluster Detection Algorithms

In Papers I, II and III we detected galaxy clusters using the DPOSS galaxy catalogs brighter than $m_r = 19.5$. At this limit, all real objects are detectable in both the r and g bands, so we require that all objects to $m_r = 19.5$ must also have a counterpart in the g band, to avoid spurious detections associated with satellite or airplane trails and plate defects. For the deeper survey presented here, this requirement is no longer practical. At the current limit of $m_r = 21.1$ there are J plates which are not as deep as the F plates, resulting in real objects

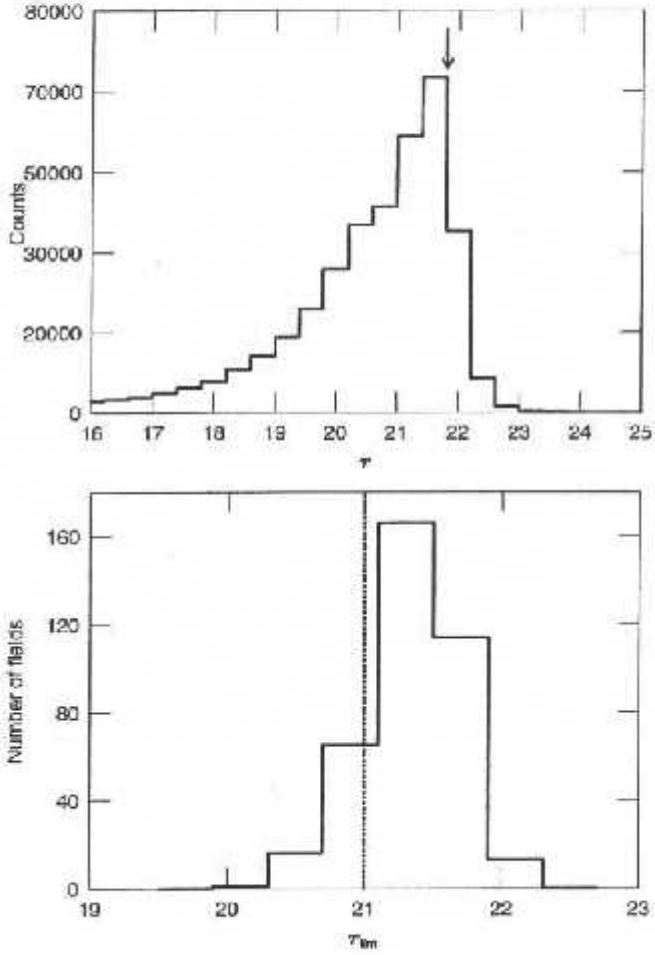


Figure 4.6: Typical magnitude distribution of a DPOSS field (top). The limiting magnitude is indicated by the arrow. In the bottom panel we show the distribution of limiting magnitudes for 375 DPOSS plates. We exclude fields with $m_{r,lim} < 21$.

detected only in the r-band which have no counterpart in the shallower g-band catalogs. We have therefore opted to use only the r-band when preparing the galaxy catalogs. As described in Djorgovski *et al.* (2003) this gives rise to $\sim 10\%$ of r-band only detections which may be spurious; these are mainly associated with meteor and aircraft trails. We exclude such objects (detected only in the r-band) by fitting a linear relation to trails present in our galaxy catalogs and removing detections in the corresponding areas. Approximately 20% of plates require this special treatment. An example galaxy catalog for a plate with a cleaned trail (in the bottom right corner) is plotted in Figure 4.7. As in the previous papers, detections

in the vicinity of bright objects (where our photometry fails) are also excised.

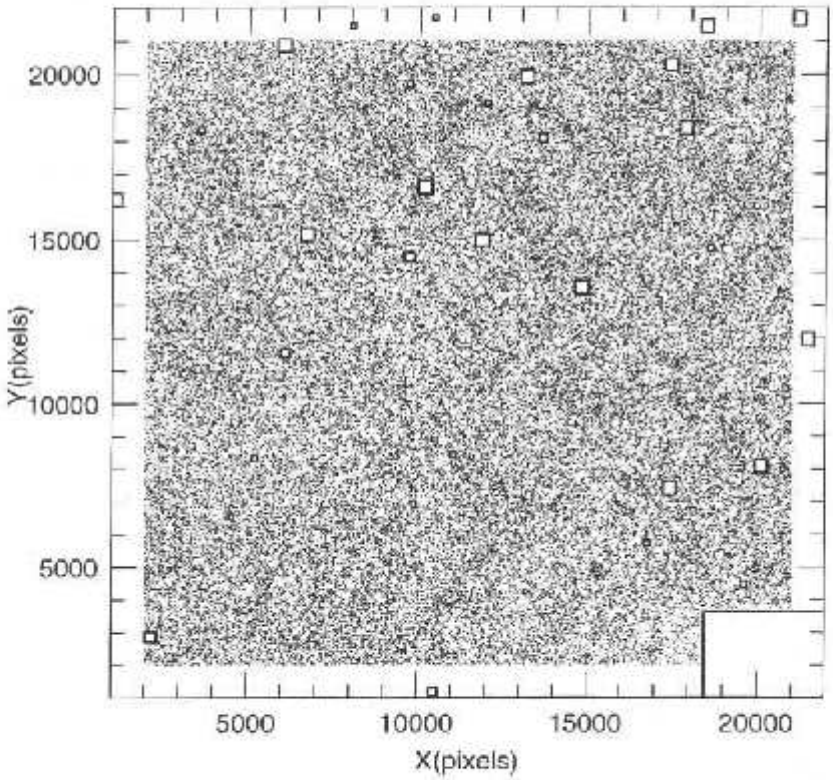


Figure 4.7: Galaxy distribution in DPOSS Field 444 ($\alpha = 201.64°$, $\delta = 29.74°$) for $16.0 \leq m_r \leq 21.1$. Excised areas due to bright objects are indicated by rectangles. A cleaned path due to a satellite trail is visible in the lower right corner.

Another difference from the previous papers is the use of only the central part of each plate. We apply cuts in pixel coordinates, using only objects with $2000 < X, Y \leq 21000$ (out of the maximum $0 \leq X, Y \leq 23040$), resulting in catalogs of typically ~ 116000 galaxies over an area of ~ 25 square degrees (compared to 34 square degrees from Papers II and III). This is necessary to avoid the heavily vignetted areas near the edge of the plates. The mean galaxy density is 4.6×10^3 galaxies per square degree (approximately six times the number density from the previous DPOSS papers). The survey covers $\sim 2,708\square°$, containing $\sim 1.3 \times 10^7$ galaxies.

In this paper two independent techniques are employed to detect galaxy clusters. One, the adaptive kernel technique (AK) (Silverman 1986) was already used by our group (Papers I, II and III), while the second technique uses Voronoi Tessellation (VT). The main differences

between the applications of these two methods to DPOSS data are:

(i) The AK is applied to the projected distribution of galaxies only, while the VT method also incorporates magnitude information.

(ii) As our goal in the present work is the detection of intermediate-z rich clusters, the galaxy catalogs used by the AK have objects at $19 \leq m_r \leq 21.1$ only. We expect to find clusters with luminosity functions spanning these fainter magnitudes and thus avoid low-z clusters. Initially we planned to employ the VT technique using the same galaxy catalogs, but because we can use magnitudes to minimize background/foreground contamination, we opted for a brighter cut at $m_r = 16$, instead of $m_r = 19$, as detailed in section 4.3.1.

The next two sections present a brief description of both techniques applied to DPOSS data. More details can be found in the references provided below.

4.3.1 The Voronoi Tessellation Technique

Considering a homogeneous distribution of particles it is possible to define, on average, a characteristic volume associated to each particle. This is known as the Voronoi volume, whose radius is of the order of the mean particle separation. Voronoi Tessellation has been applied to a variety of astronomical problems. A few examples are found in Ikeuchi & Turner (1991), Zaninetti(1995), El-Ad *et al.* (1996), Doroshkevich *et al.* (1997). Ebeling & Wiedenmann (1993) used Voronoi Tessellation to identify X-ray sources as overdensities in X-ray photon counts. Kim *et al.* (2002) and Ramella *et al.* (2001) looked for galaxy clusters using Voronoi Tessellation (VT). As pointed out by Ramella *et al.* (2001) one of the main advantages of employing VT to look for galaxy clusters is that this technique does not distribute the data in bins, nor does it assume a particular source geometry intrinsic to the detection process. The algorithm is thus sensitive to irregular and elongated structures.

The parameter of interest in our case is the galaxy density. When applying VT to a galaxy catalog each galaxy is considered as a seed and has a Voronoi cell associated to it. The area of this cell is interpreted as the effective area a galaxy occupies in the plane. The inverse of this area gives the local density at that point. Galaxy clusters are identified at

high density regions, composed by small adjacent cells. In other words, cells small enough to give a density value higher than the chosen density threshold. An example of Voronoi Tessellation applied to the same DPOSS field as shown in Figure 4.7 is presented in Figure 4.8. For clarity, we show only galaxies at $17.0 \leq m_r \leq 18.5$ in this plot.

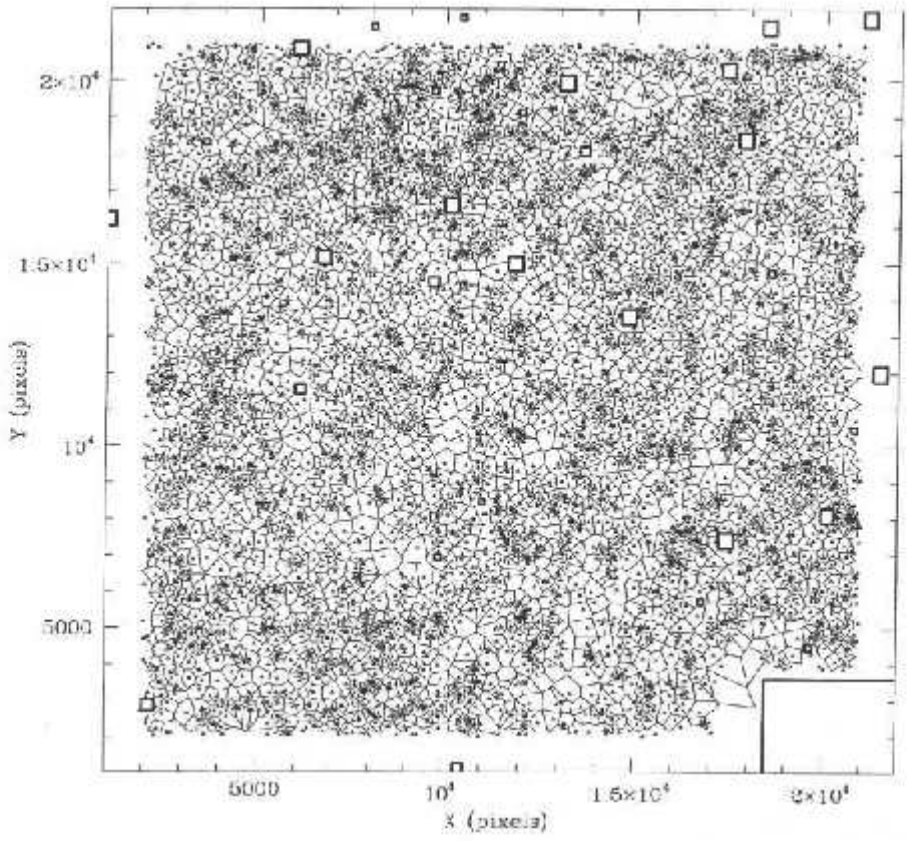


Figure 4.8: Voronoi Tessellation of galaxies with $17.0 \leq m_r \leq 18.5$ in Field 444. Each triangle represents a galaxy surrounded by its associated Voronoi cell (indicated by the polyhedrals). Excised areas (due to bright objects) are shown as rectangles.

In order to detect galaxy clusters using Voronoi Tessellation we use the code employed by Ramella *et al.* (2001). It uses the *triangle* C code by Shewchuk (1996) to generate the Tessellation. The code is designed to avoid the borders of the field, as well as the excised areas around saturated objects. The algorithm identifies cluster candidates based on two primary criteria. The first is the density threshold itself, which is used to identify

fluctuations as significant overdensities over the background distribution, and is termed the search confidence level (**scl**). The second criteria rejects candidates from the preliminary list using statistics of Voronoi Tesselation for a poissonian distribution of particles (Kiang 1996), by computing the probability that an overdensity is a random fluctuation. This is called the rejection confidence level (**rcl**). More details can be found in Ramella *et al.* (2001).

Kim *et al.* (2002) used a color-magnitude relation to divide the galaxy catalog into separate redshift bins, and ran the VT code on each bin. The candidates originating in different bins are cross-correlated to filter out significant overlaps and produce the final catalog.

Ramella *et al.* (2001) follow a different approach, as they do not have color information. Instead, they use the object magnitudes to minimize background/foreground contamination and enhance the cluster contrast, as follows:

(i) The galaxy catalog is divided into different magnitude bins, starting at the bright limit of the sample and shifting to progressively fainter bins. The step size adopted is derived from the photometric errors of the catalog.

(ii) The VT code is run using the galaxy catalog for each bin. Hence, each magnitude slice will have a catalog of cluster candidates associated with it.

(iii) The centroid of a cluster candidate detected in different bins will change due to the statistical noise of the foreground/background galaxy distribution. Thus, the cluster catalogs from all bins are cross-matched, and overdensities are merged according to a given criteria (described below), producing a combined catalog.

(iv) A minimum number (N_{min}) of detections in different bins is required in order to consider a given fluctuation as a cluster candidate. N_{min} acts as a final threshold for the whole procedure. After this step, the final cluster catalog is complete.

Ramella *et al.* (2001) applied their algorithm to one of the Palomar Distant Cluster Survey (PDCS) galaxy catalogs. They divided this deep data into bins of two magnitudes, starting with $18.0 \leq V_4 \leq 20.0$ and shifting to fainter bins in steps of 0.1 mag down to the detection limit ($V_4 = 23.8$). This procedure resulted in 39 bins. They adopt **scl** = 0.80 and **rcl** = 0.05 to select fluctuations in each bin. They then merged candidates whose centers had a projected distance equivalent to $d_{12} \leq 0.3min(R_1, R_2)$, where R_1 and R_2 are the radii

of any two candidates being compared. Finally, they kept only candidates with $N_{min} = 5$. While not stated in their paper, they further merged clusters whose radii overlap.

When applying their algorithm to DPOSS data we tested different bins and step sizes. As with the PDCS galaxy catalog, we span five magnitudes, but in a brighter regime and with larger photometric errors. We adopt a bin size of 1.5 magnitude and a step size of 0.2 magnitudes. This step size is comparable to the photometric error at $m_r = 19.5$. A wider magnitude bin could be used, but would significantly decrease the number of steps. A narrower bin would result in a low number of cluster counts per bin. We thus have 19 bins spanning the range $16 \leq m_r \leq 21.1$. As seen in Figure 4.2, 1.5 magnitudes is a reasonable range to see clusters over the background.

The threshold as given by **scl** selects overdensities above a given galaxy surface density. Due to the non-uniform nature of plate data and the effects of large scale structure, we do not use a fixed threshold for the 108 different plates used in this project. Instead, we maintain **rcl** fixed at 5%, but allow **scl** to vary. Thus, the VT code is run for each DPOSS field varying **scl** from 0.78 to 0.92 (in steps of 0.02), instead of adopting a fixed value of 0.80 as Ramella *et al.* (2001) did.

The percolation analysis we apply to the cluster candidate catalogs from different bins is similar to that of Ramella *et al.* (2001). However, the optimal parameters found for DPOSS data are slightly different. We chose $d_{12} \leq 0.8 min(R_1, R_2)$ and $N_{vmin} \geq 3$, plus a final merging of structures whose separation is smaller than the radius of the largest neighboring candidate. These values are selected based on the redshift range over which we are detecting clusters, and the nature of our data. The PDCS data used by Ramella *et al.* (2001) has no nearby large clusters, as the data covers only 1 $\square^{\circ}$ and was designed to avoid bright objects and low-z clusters. Tests done with a smaller matching radius show that many nearby clusters would be broken into subcomponents. Larger radii would incorrectly associate adjacent clusters into a single candidate. The choice of $N_{vmin} \geq 3$ is explained in the next section, as it depends on the number of fake clusters (the contamination rate) produced for each field. It is a sensible choice given that we have half the bins (19 compared to 39) of Ramella *et al.* (2001).

As with the AK, the VT is optimized based on simulations of the background galaxy distribution, where we apply the VT code to both these simulated fields and real DPOSS

fields. These techniques are described in §4.4.

4.3.2 The Adaptive Kernel Technique

As described in Papers I and II, the adaptive kernel technique works in a two stage process. The galaxy catalogs ($19 \leq m_r \leq 21.1$) are used as input to the AK. Initially, the code uses a fixed size kernel to evaluate the galaxy density at each point within the catalog. It then applies a smoothing kernel whose size changes as a function of the local density estimated from the first step, with smaller kernels at higher densities. Finally, a density map is generated for each field. SExtractor is used to locate overdensities on this map. In the previous NoSOCS papers the pixel scale adopted for the density maps was 1 pixel = 60″. For the current project, we aim to detect distant (generally low contrast) structures. We therefore require higher resolution for the density maps, which are here generated with a scale of 1 pixel = 10″.

Gladders & Yee (2000) utilize a more powerful technique to detect clusters, with a color-magnitude relation as a filter to minimize contamination from foreground sources. However, the surface density of objects is evaluated in a similar way to ours. The main difference is that they use a simple fixed kernel. They argue that our estimator might not be the best choice as the AK has a tendency to resolve high density regions in the data. Based on cluster simulations, we show (see Figure 1 in Paper II) that with the proper choice of the initial kernel size we do not break up nearby rich clusters into subcomponents, while remaining sensitive to low contrast structures (poor and close, or rich and distant). A simple fixed kernel technique might not be appropriate in our case, as it is not simultaneously sensitive to clusters of different richness classes at different distances. Having color information it is possible to divide the data in color slices, and then apply a fixed kernel estimator to each slice, as done by Gladders & Yee (2000).

In Papers II and III we adopted a kernel of 500″ radius to look for *low-z* clusters ($z < 0.3$) with our brighter galaxy catalogs. As before, we use simulated clusters to test the kernel size for the fainter catalogs at $19 \leq m_r \leq 21.1$. The optimal value found is 260″. Figure 4.9 shows a density map generated with a 260″ radius initial kernel. 48 artificial clusters of six richness classes ($N_{gals} = 15, 25, 35, 55, 80, 120$) at 8 different redshifts (z =0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50) are inserted. Clusters are marked with circles and each

column represents a different richness class (increasing from left to right), while each row is a different redshift (increasing from bottom to top). We can see that our choice for the initial kernel radius does not break up the nearby clusters, while it is still sensitive to distant rich structures.

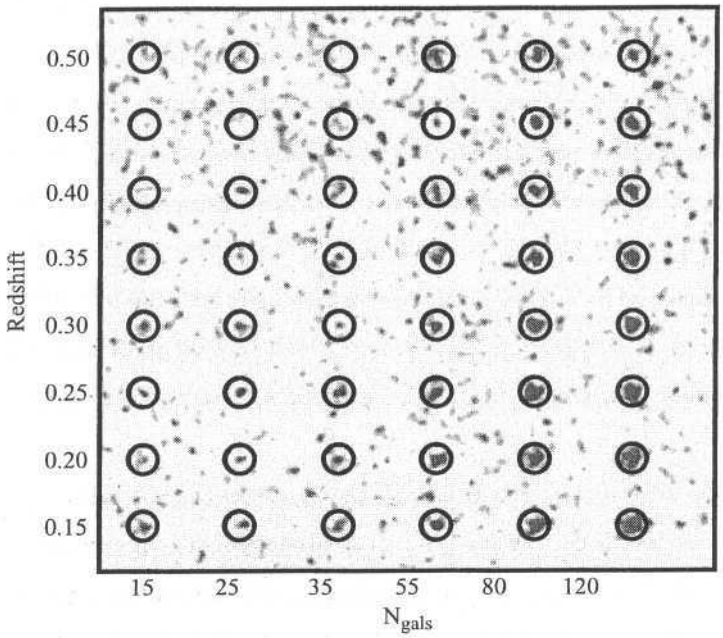


Figure 4.9: Density map of a simulated background galaxy distribution with 48 artificial clusters inserted. The initial kernel adopted has a 260″ radius. Artificial clusters are marked by circles. Each row represents a different redshift (z =0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50), increasing from bottom to top. Richness increases from left to right ($N_{gals} = 15, 25, 35, 55, 80, 120$).

After generating a density map for each field we then run SExtractor to detect high density peaks, which we associate with cluster candidates. Simulated background fields are used to optimize SExtractor parameters (see section 4.4). As in Papers II and III, each field is optimized independently.

4.4 Contamination and Optimization of the Algorithms

As stated in Paper II, contamination by random fluctuations in the galaxy distribution and chance alignments of galaxy groups can seriously affect any cluster catalog. This problem

is exacerbated in a survey like this one, which lacks color information to minimize background/foreground contamination (Gladders & Yee 2000; Goto *et al.* 2002).

The most common methods to estimate how much a given catalog is affected by such effects use simulated background distributions. A simple random distribution with the same number density as the original field would underestimate contamination. A background constructed in this way reproduces random fluctuations, but not the large scale structure present in a given field. Goto *et al.* (2002) tested three different types of background, where they: (i) shuffled the positions without touching colors; (ii) shuffled the colors, but did not touch the positions; (iii) shuffled the colors and smeared the positions. That was specifically done for their algorithm, which makes use of the extensive multicolor information from SDSS. Algorithms which do not make use of color usually have their contamination estimate based on a shuffled background (Lobo *et al.* 2000), a Poissonian galaxy distribution (Ramella *et al.* 2001), or on a field generated with the same angular correlation properties as the real field (P02; Paper II; Gilbank 2001).

For this project we have tested four types of simulated backgrounds: (i) a Raleigh-Levi (RL) distribution (P02; Paper II); (ii) a shuffled background; (iii) a randomized galaxy distribution; and (iv) a smeared background. In the latter case galaxy magnitudes are maintained, while the positions are randomly redistributed within 7′ of their original ones. Goto *et al.* (2002) adopted 5′ for the smearing scale. They found the contamination rate evaluated with this background to be extremely high in comparison to the other ones tested. As described below, we test the optimization of our algorithms at the 5% and 10% contamination levels. Commensurate with the results of Goto *et al.* (2002), we find that it is not possible to optimize the AK or VT at these levels with the smeared background, and we have therefore abandoned it in the tests described below. The RL distribution is intended to simulate fields with an angular two point correlation function similar to the original fields. As described below, we see no significant difference between the RL and the shuffled estimates at the 5% contamination level. However, the random background produces fewer fake clusters in comparison with the other backgrounds tested. We finally selected the shuffled background to optimize our algorithms and evaluate the contamination rate; this choice is further justified in §4.4.1.

4.4.1 Voronoi Tessellation

A variety of methods have been used to minimize the number of false detections (N_{fake}) in a simulated field when optimizing cluster detection algorithms. The number of fake clusters per square degree has been presented as an indicator of how low contamination effects could be for a given catalog. However, as the photometric depth of catalogs varies, so will the absolute number of false detections. Thus, the contamination rate ($C = N_{fake}/N_{real}$) has been more commonly used to optimize cluster detection algorithms. N_{fake} gives the number of false detections in a simulated field, while N_{real} is the number of detections in the real galaxy distribution. Optimal detection thresholds are achieved for the highest recovery rate allowed at a given contamination level.

We simulate field distributions and run our algorithm on both the mock and real catalogs. The algorithm is optimized to achieve a given value of C. The contamination rate and the number of candidates change rapidly with the optimization threshold. The threshold for the VT algorithm is given by a combination of **scl** and N_{min} (see section 4.3.1).

We performed tests on 20 DPOSS fields and compared the results obtained at the 5% and 10% contamination levels. We chose not to fix the values of **scl** or N_{min}. These should be optimized independently for each field. For a given field the best option could consist of a high value of **scl** with a low number of minimum detections, while the opposite could be true in other cases.

Figure 4.10 compares the variation of N_{fake} with N_{min} for the RL, randomized and shuffled distributions. Each dashed line represents the variation of N_{fake} for one of the 20 DPOSS plates tested. The solid line shows the mean variation. Figure 4.11 shows the variation of N_{fake} with **scl**. In order to compare the performance of the three different backgrounds we use N_{fake}, as it is the only parameter to vary when estimating C. N_{real} is the same regardless of which background we use, as it is measured off the real data. In Figure 4.10, the dispersion is basically the same for all distributions, being slightly higher when the shuffled background is used. Figure 4.11 indicates the contamination estimate is lowest, and least sensitive to, the randomized background, while the RL and shuffled distributions show similar results to each other (except for two outliers in the latter case). This suggests that the randomized background underestimates the contamination. The RL is a model representation of the angular distribution in a galaxy field. N_{fake} smoothly decreases

as we move to higher thresholds (given by N_{min} and scl). However, the RL distribution might not properly represent variations in high density fields. We expect the correlation properties to vary for different DPOSS fields as they have different galaxy densities (Maddox, Efstathiou & Sutherland 1996). The photometric errors at $m_r = 21.1$ are $\sim 0.4^m$, which could seriously affect the galaxy density in DPOSS fields, and therefore the modeling of the angular correlation given by the RL distribution. As a full analysis of the correlation properties of all DPOSS fields is beyond the scope of this project, we decided not to use the RL distribution for this paper. Instead, the shuffled background is adopted for optimizing our algorithms. It shows no large differences relative to the RL distribution, and is a model-independent representation of the background.

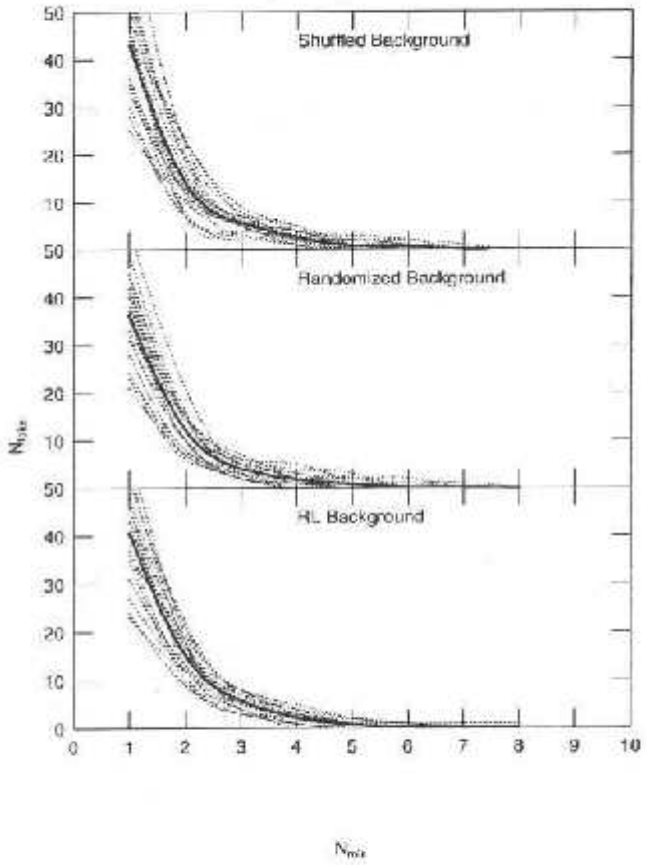


Figure 4.10: Comparison of the number of false detections using an RL distribution, a randomized background, and a shuffled background. The dotted lines show the variation of N_{fake} with N_{min} for 20 different fields, while the solid line shows the mean variation.

We then generate a shuffled catalog (simply repositioning all galaxies within a field and keeping the magnitudes) and run the VT on this list, which gives N_{fake}. The VT is also run on the real catalog, yielding N_{real}. Each field is optimized for the values of **scl** and N_{min} which gives the highest value of N_{real}, while keeping the contamination rate fixed at $C \sim 5\%$. Note that 5% is a lower limit for contamination effects. A full assessment requires a spectroscopic follow-up. In section 4.6 we discuss the improvement in the selection function as the allowed contamination rate C is increased from 5% to 10%.

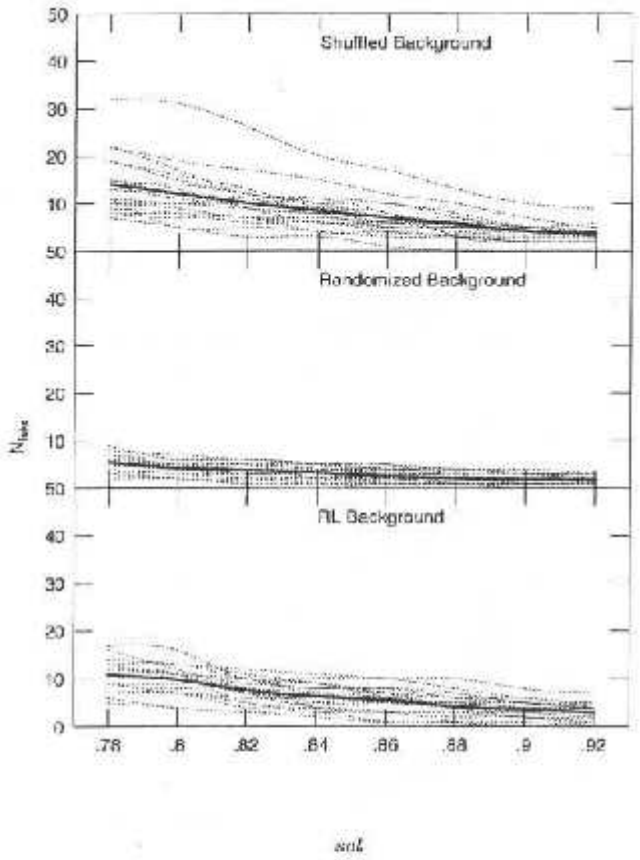


Figure 4.11: Comparison of the number of false detections using an RL distribution, a randomized background, and a shuffled background. The dotted lines show the variation of N_{fake} with **scl** for 20 different fields, while the solid line shows the mean variation.

4.4.2 The Adaptive Kernel

The strategy adopted to optimize the AK algorithm is basically the same as for the VT. We take the same shuffled catalogs generated for the VT optimization and trim them to the appropriate magnitude limit ($19 < m_r < 21.1$) adopted for the AK. We run the AK on these catalogs as well as the originals to generate density maps for all of them. SExtractor is used to detect overdensities on these maps.

The parameters that give the final optimization of the AK are the minimum area and the threshold required for a detection when running SExtractor. The minimum area gives the minimum number of pixels a candidate shall have, while the threshold gives the minimum number of galaxies per square degree. We vary the threshold from 3000 to 9000 galaxies per $\square^{\circ}$. Tests in a broader range for 20 plates showed the optimal threshold is always found within these values. Four different values are tested for the minimum area: 200, 300, 500 and 900 pixels^2. In order to choose the optimal value for the minimum area we evaluate the selection function with these four different values (see Figure 4.12). We find that 200 and 300 pixels^2 produce similar results, with a slightly higher recovery rate in the former case. Going to 100 pixels^2 does not represent an improvement. Thus we adopted 200 pixels^2 as the optimal minimum area for all fields when running SExtractor. As stated above the threshold is optimized for each plate.

4.5 The Cluster Catalogs

The main goal of this project is to provide a catalog with rich galaxy clusters to intermediate redshift for follow-up studies. In order to be able to select subsamples from the main catalog we provide some basic properties, such as redshift and richness, for the cluster candidates. When estimating these two quantities we use similar, but not identical, techniques to those employed in Papers II and III. However, we face a variety of challenges for the current project. For instance, the photometric errors are large in the magnitude range utilized here, which would also affect color measures; the number of clusters with known spectroscopic redshift to be used as a training sample is small when we go out to $z \sim 0.6$; the $(g-r)$ color is not useful for $z \geq 0.4$ clusters (as the 4000 Å break shifts from $g-r$ to $r-i$); and the blue plates do not necessarily go as deep as the red ones.

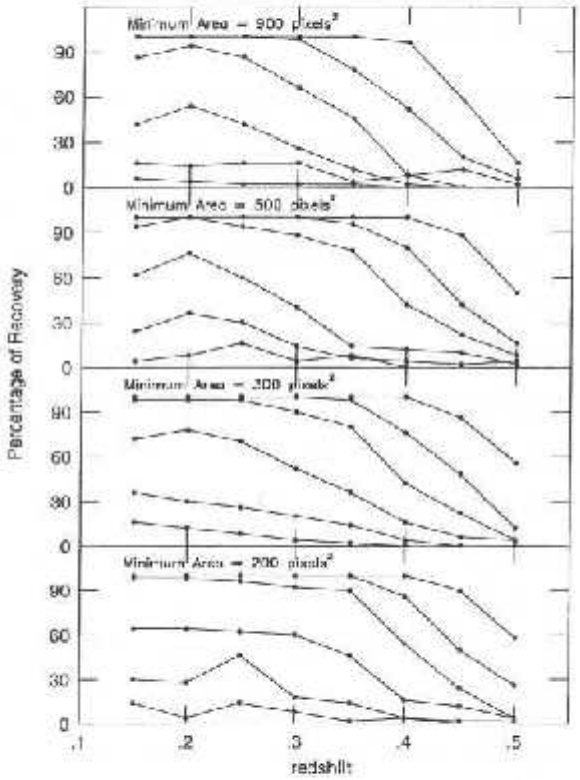


Figure 4.12: The selection function with the AK code for 4 different values of the minimum area: 200, 300, 500 and 900 pixels^2. In each panel, from bottom to top, each line represents a different richness ($N_{gals} = 15, 25, 35, 55, 80, 120$).

In §4.5.1 and §4.5.2 we describe our efforts to estimate redshift and richness for the cluster candidates. A visual inspection is employed in §4.5.3 to eliminate obvious fake clusters associated to bright stars, nearby galaxies and groups. A combined version of the VT and AK cluster catalogs is presented in section 4.5.4.

4.5.1 Photometric Redshifts

An empirical relation based on the mean r magnitude and median $(g-r)$ color was successfully used to estimate redshifts for the NoSOCS clusters (Papers I, II and III). Here, we assume these same properties are strong indicators of the redshift of a cluster. However, this paper employs significantly different methodologies, and it is not clear *a priori* if useful photometric redshift estimation is possible using the faintest DPOSS data.

The first step towards the determination of photometric redshifts is the compilation of a list of clusters with known spectroscopic redshift. Unfortunately, there are rather few clusters with spectra taken at $z > 0.3$, which bias our sample to low-z clusters. The list we generated has clusters from Struble & Rood (1999), Holden *et al.* (1999), Vikhlinin *et al.* (1998), containing 212 clusters over the $\sim 2,700$ square degrees of this survey. In comparison, the training sample for Paper II had 369 clusters.

The main differences with respect to the photometric redshift technique employed in Paper II are: (i) The galaxy catalogs used here have a statistical correction applied to them (section 4.2.2), which gives rise to a $\sim 15\%$ difference in number counts around $m_r = 20$; (ii) counts were previously done at $15 < m_r < 20$, while now we count galaxies at $16 < m_r < 21.1$; (iii) we now adopt local estimates for the background counts, instead of taking these estimates from the whole area of each plate. Additionally, we tested different counting radii (0.50, 0.75, 1.0 and 1.5 h^{-1} Mpc), and found that 1.0 h^{-1} Mpc produces the fewest outliers and minimizes the overall dispersion.

We proceed as follows: each of the 212 clusters with known spectroscopic redshifts have the number of galaxies as a function of magnitude (N_r) and color ($N_{(g-r)}$) determined. These counts have the local background counts subtracted, resulting in a determination of the net cluster counts as a function of both color and magnitude. The mean m_r magnitude (r_{mean}) and median color ($(g-r)_{median}$) is then computed for each cluster. Then we bin the colors and magnitudes in redshift bins of $\Delta z = 0.05$, calculating the mean z_{spec}, r_{mean} and $(g-r)_{median}$ for each bin. These values are then used to derive two empirical relations for redshift estimation. The first uses r_{mean} only, and the second uses r_{mean} and $(g-r)_{median}$.

Actually, if we do not have any idea of the cluster redshift, we should evaluate it in a iterative manner. We choose an initial guess of the redshift ($z_{start} = 0.15$), and compute r_{mean} and $(g-r)_{median}$ based on the cluster corrected counts within 1.0 h^{-1} Mpc, apply the empirical relations derived above and derive a photometric estimate of z. We then repeat the same procedure until it converges.

In Figure 4.13 we compare the photometric and spectroscopic redshifts using $z_{start} = 0.15$, for the two relations mentioned above. We find the dispersions to be similar, being slightly smaller when we make use of color to evaluate z_{phot}. However, it seems that the relation based in both color and magnitude fails to recover the redshift of clusters with $z \geq 0.4$, as the

photometric redshifts are severely underestimated. Most cluster members of those distant systems lie at the faint end of DPOSS, where the photometric errors induce large color errors. The 4000 Å break is also shifted from the $(g-r)$ color to the $(r-i)$. However, maybe the most important factor for underestimation of the $(g-r)$ color lies in the incompleteness of the J (blue) plates. As shown in §4.2.2, the field selection is based only in the F (red) plates. We select only F plates going as deep as $m_r = 21$, but the blue plates might not go as deep as the red selected ones. In Figure 4.14 we show, as function of magnitude, the fraction of r and g (as well as r and i) detections to the r-band detections. These values are calculated from the 108 fields used for this survey. We see that almost every r detection has a g-band counterpart down to $m_r \sim 19.5$. Fainter than that the number of two band detections drops steeply, reaching $\sim 70\%$ at $m_r = 21$. For Papers I, II and III the magnitude limit of the galaxy catalogs ($m_r = 19.5$) allowed us to impose a g-band detection when generating a galaxy catalog for cluster search. That is not possible for the current project.

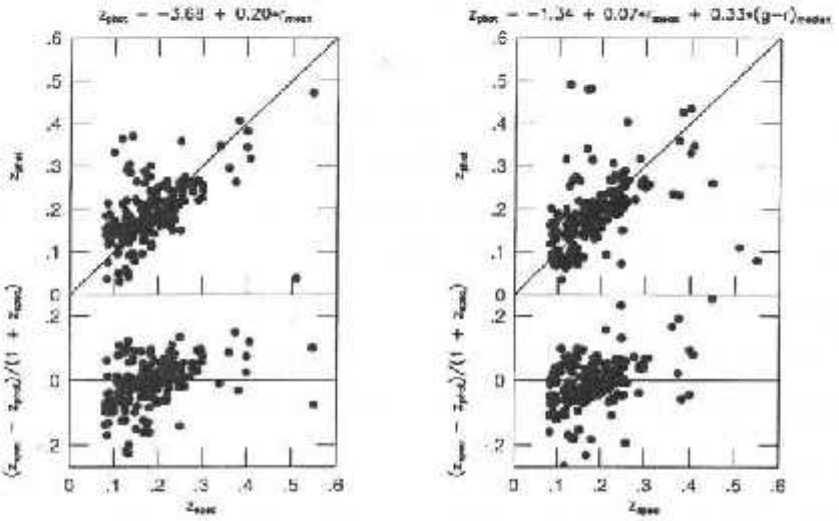


Figure 4.13: The photometric redshift estimate *vs* spectroscopically measured redshift for 212 clusters. Residuals as function of magnitude are shown in the bottom panel of each plot. We show photometric estimates based on the mean magnitude (left panel), as well as the mean magnitude and median color $(g-r)$ (right panel).

As we are mainly looking for rich and distant clusters, we then prefer not to use colors when estimating redshifts. For this project we chose an overall less accurate redshift estimate (using only magnitudes) instead of a more precise estimate but which may fail for the most

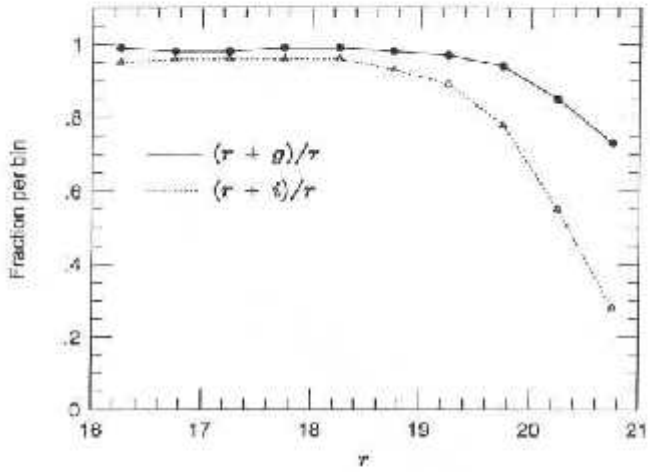


Figure 4.14: The fraction, as a function of magnitude, of r and g (solid line) and r and i (dotted line) detections to the r-band detections. All the 108 plates selected for this survey were used for generating this plot.

distant systems (if including color for the estimates). We then opted for the relation that uses only magnitude, which results in a redshift dispersion of $\Delta z = 0.053$. Those estimates are at the same level of other methods which use only magnitude for redshift estimation (like the matched filter). The fitted relation is

$$z_{phot} = -3.68 + 0.20 \times r_{mean} \tag{4.2}$$

We also want to test the dependence of the photometric redshift estimator on the starting radius. Figure 4.15 compares the redshift estimates obtained with $z_{start} = 0.15$ and $z_{start} = 0.05$. We find the relation to have almost no dependence on the starting guess redshift. Q_{zsigma} is $(z_{0.15} - z_{0.05})/(1 + z_{0.15}) - \Delta z_{phot} - 0.006$.

For the 212 clusters with known z_{spec} we tested four different starting redshifts ($z_{start} = 0.05, 0.10, 0.15, 0.20$). As we see from Figure 4.15 the results are not very sensitive to the starting redshift. When estimating redshifts for the cluster candidates we also tested these four starting redshifts. As before, we did not see any significant dependence on the initial guess redshift. We chose $z_{start} = 0.15$ as it has fewer outliers. In Figure 4.16 we compare the redshift distributions for these four initial redshifts, for the VT and AK. Using $z_{start} = 0.15$ recovers more redshifts than the other cases, with the main difference being for the nearest clusters ($0.10 < z < 0.25$).

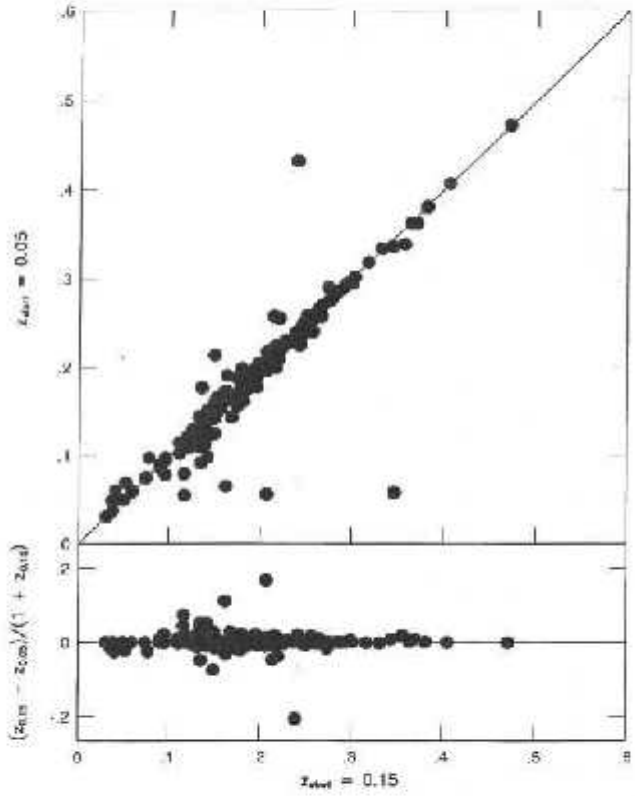


Figure 4.15: The dependence of the photometric z estimator on the initial redshift guess. We compare the estimates for two different starting redshifts ($z_{start} = 0.15$ and $z_{start} = 0.05$).

Figure 4.17 shows the redshift distribution for different richness class for both methods. We can see the cutoff on each case is generally in agreement from what is seen in the selection functions (§4.6).

4.5.2 Richness Estimates

Richness is evaluated in a similar manner to Papers II and III. We describe below all the steps in the richness determination, pointing out the minor differences from the previous procedure. We proceed as follows:

1. We count the number of galaxies at $16.0 \leq m_r \leq 21.0$ within 1.0 h^{-1} Mpc of the cluster center. The background counts in the same range are evaluated locally, scaled to the cluster area, and subtracted, yielding the background-corrected cluster counts

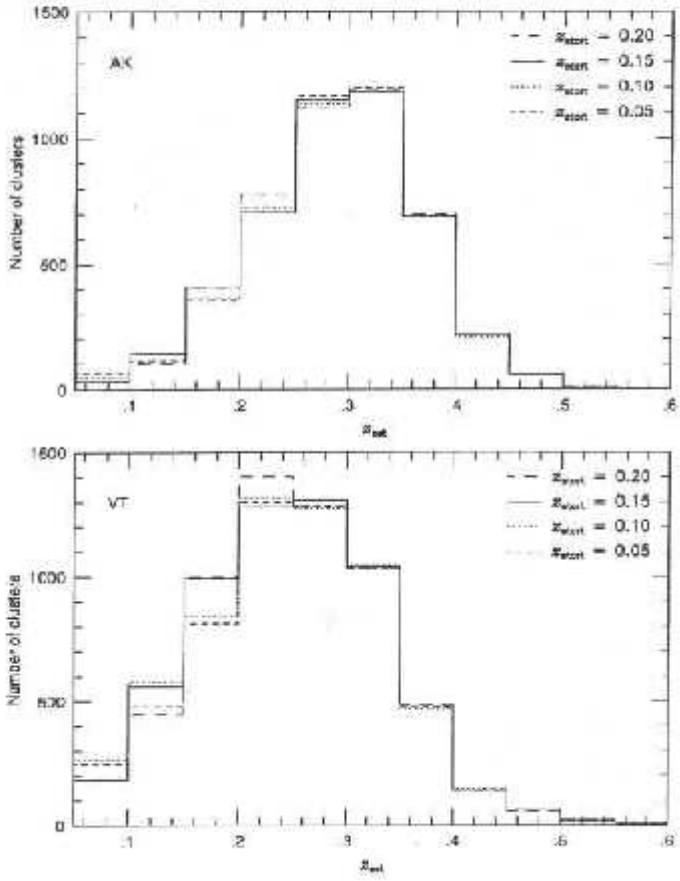


Figure 4.16: The estimated redshift distributions for the AK & VT candidates, with different starting redshifts ($z_{start} = 0.05, 0.10, 0.15, 0.20$).

(hereafter N_{corr}). In the previous papers the magnitude range was $15.0 \leq m_r \leq 20.0$ mag and the background was estimated from the whole plate.

2. We run a bootstrap procedure with 100 iterations. In each iteration, we randomly select N_{corr} galaxies at $16.0 \leq m_r \leq 21.0$, within 1.0 h^{-1} Mpc. Each galaxy has its apparent magnitude converted to an absolute magnitude. As with the artificial clusters (§4.6), we consider the clusters to be composed of 60% early-type and 40% late-type galaxies. When transforming to absolute magnitudes, we simply use elliptical K-corrections for 60% of the objects, and Sbc K-corrections for the remainder. We then count the number of galaxies within M_r^*-1 and M_r^*+2, where $M_r^* = -21.52$ (Paolillo *et*

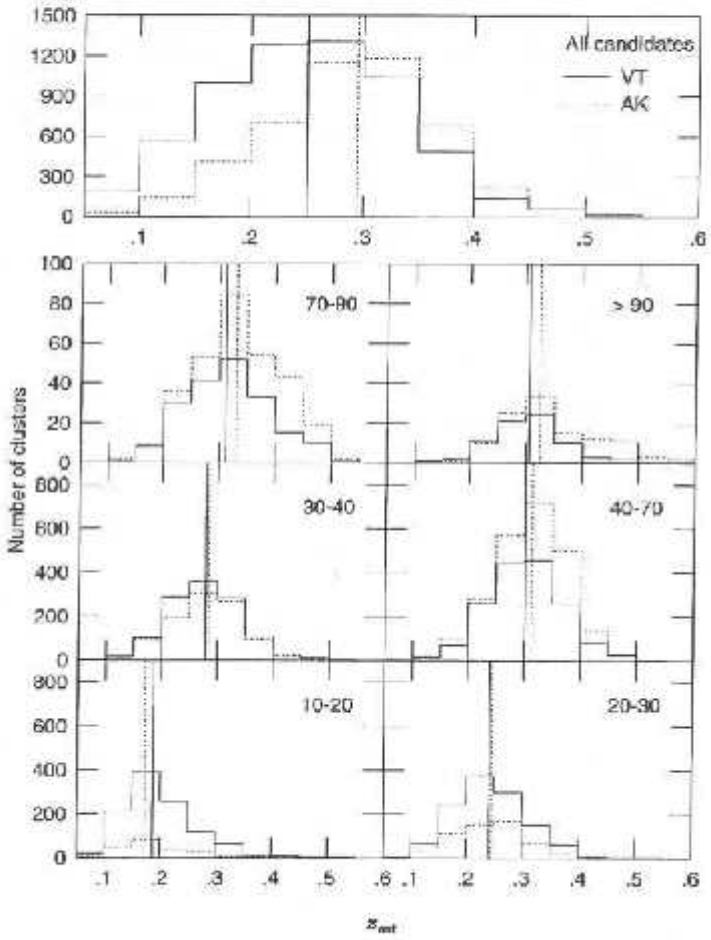


Figure 4.17: The estimated redshift distribution in different richness ranges and for the whole sample (top panel). The richness (N_{gals}) range is indicated in all panels. The median redshift is also indicated for the VT (solid line) and AK (dotted line) candidates.

al. 2001). The mean value from the 100 iterations is computed, yielding the cluster richness N_{gals}, which can be decomposed into 40% Sbc galaxies ($N_{gals-Sbc}$) and 60% ellipticals (N_{gals-E}). If the cluster has a redshift such that M*-1 to M*+2 is fully sampled in the range $16.0 \leq m_{z'} < 21.0$, then the procedure is finished here. Otherwise we have to apply a correction factor to the richness estimate (N_{gals}), as described in step 3.

3. If the cluster is too close or too distant, then M*-1 < M_{16} or M*+2 > M_{21}, respectively; where M_{16} and M_{21} are the corresponding bright and faint absolute magnitude limits

to $m_r = 16.0$ and $m_r = 21.0$. In practice these limits are different if a given galaxy is considered to be elliptical or spiral, due to the K-correction factor. Whenever M^*-1 to M^*+2 does not lie within M_{16} and M_{21} we have to apply a correction factor for the richness estimate. This factor is defined by:

$$\gamma_1 = \frac{\int_{M_r^*-1}^{M_r^*+2} \Phi(M)dM}{\int_{M_{16}}^{M_r^*+2} \Phi(M)dM} \tag{4.3}$$

$$\gamma_2 = \frac{\int_{M_r^*-1}^{M_r^*+2} \Phi(M)dM}{\int_{M_r^*-1}^{M_{21}} \Phi(M)dM} \tag{4.4}$$

We call γ_1 and γ_2 the low and high magnitude limit correction factors. Actually, each of these correction factors are evaluated twice, once for the magnitude limit for elliptical galaxies and the other for late-types. Whenever necessary, one of the above factors (as we span 5 magnitudes it is impossible to miss both the bright and faint end) is multiplied by N_{gals-E} and/or $N_{gals-Sbc}$.

The main differences in steps 2 and 3 to the procedure adopted for the previous NoSOCS papers lies in the assumption that the cluster is not totally composed of elliptical galaxies and also in the fact that we consider a correction factor to the lower magnitude limit.

In Figure 4.18 we show the richness distribution for the VT and AK cluster candidates for different redshift bins. The top panel shows the richness distribution for the entire sample.

Further considerations related to redshift and richness errors are given in section 4.8, where we compare the current catalog to the *shallow* NoSOCS cluster catalog and to other surveys.

4.5.3 Elimination of Spurious Cluster Candidates

Before combining the VT and AK cluster catalogs we perform a visual inspection of the cluster candidates to eliminate obviously false clusters. Such a procedure was previously adopted by Kim (2001) and Goto *et al.* (2002). While inspecting the candidates we realized that many of them were associated to bright stars and a few to nearby bright galaxies and groups, globular or open clusters, plate defects, trails, etc. Thus, we decided to try to eliminate most of these fake clusters in an automatic way.

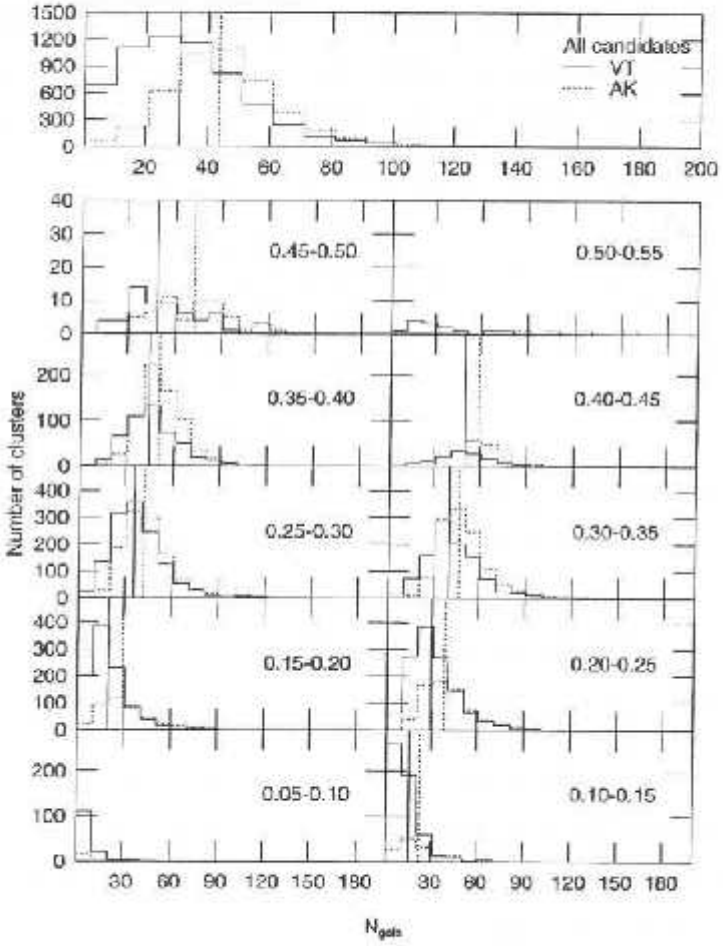


Figure 4.18: The richness distribution in different estimated redshift bins and for the whole sample (top panel). The redshift bins are indicated in each panel, as well as the median richness, which is indicated by the vertical lines (solid for the VT candidates and dotted for the AK).

The major source of spurious detections is bright objects which were missed when generating the list of bad areas to be excised. In some cases, the object was removed, but the excised area was not big enough. We therefore decided to compare the cluster candidates positions to a bright star catalog. We used the Tycho-2 catalog (Høg *et al.* 2000), which is $\sim 90\%$ complete to $V \sim 11.5$.

Figure 4.19 shows for 4 different magnitude bins the offset distribution (in arcsec) of Tycho-2 stars from the nearest cluster candidate. The solid lines show the real offset distri-

butions, while the dotted lines show the offset distribution between a mock star catalog and the cluster catalog. At $8 < V < 10$, almost all stars are associated with a cluster candidate within $\leq 120''$. Our photometry fails at these bright stars, which gives rise to a large number of faint spurious detections in the halos of these objects. These spurious objects generate overdensities detected as cluster candidates. The catalogs presented in Papers I, II & III did not face this problem as most of these spurious detections are fainter than $m_{z'} = 19.5$, and do not have counterparts in the g-band. For the single band galaxy catalog utilized here, the elimination of bright objects turns out to be a serious issue.

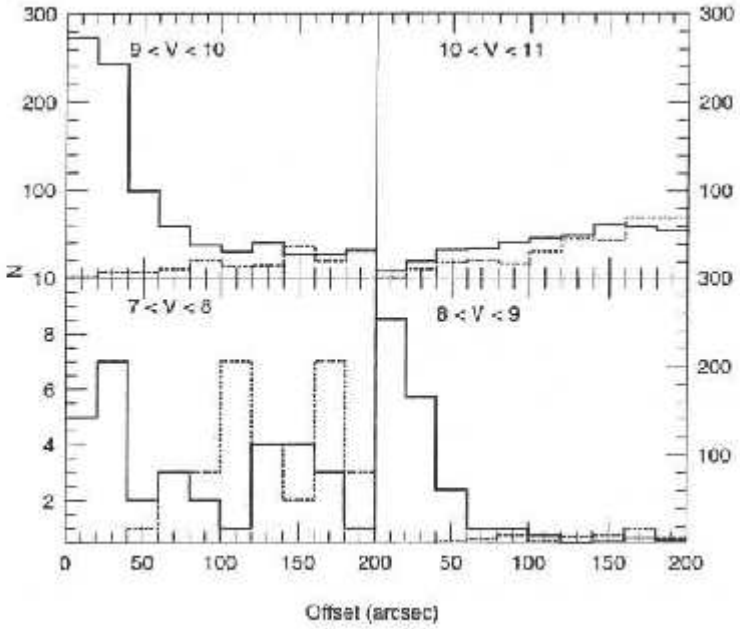


Figure 4.19: Offset distribution between Tycho-2 stars and cluster candidates (solid lines). The dotted lines represent the offset distribution to a mock star catalogue. The magnitude ranges are indicated on each panel.

We use the above information to clean our cluster catalog in a semi-automated fashion. From the inspection of offsets distributions in different magnitude bins, we selected a minimum offset between a star and a cluster candidate (in each magnitude bin) to exclude the cluster as a fake detection, and a function is interpolated to these minimum offsets as function of magnitude.

$$\Delta\theta_{min} = 185 + 0.37 \times (V^2) - 0.01 \times (V^4) \tag{4.5}$$

This function gives the minimum separation (in arcsec) a given cluster candidate can have from a V magnitude star. If the separation is smaller than this value, the candidate is excluded. Adopting this procedure, we eliminated 1031 VT and 1303 AK candidates. We further compared the cluster candidates to lists of bright galaxies, and globular and open clusters, which resulted in the elimination of a few more candidates. Finally, we visually inspect all the remaining candidates to exclude any associated to plate defects or trails. The total number of eliminated cluster candidates is 1602 ($\sim 17\%$) and 1748 ($\sim 27\%$) for the VT and AK catalogs, respectively. Note that as the detection codes were optimized to produce 5% contamination, reducing the number of real candidates by 20% results in an increase of the contamination rate to $C = \mathrm{N}_{fake}/\mathrm{N}_{real} \sim 6\%$.

4.5.4 The Northern Sky Optical Cluster Survey (NoSOCS) High-Redshift Catalog

The NoSOCS supplemental catalog presented here covers $\sim 2{,}708\ \square^{\circ}$ and originates from two cluster detection algorithms. The catalog contains 9,982 cluster candidates, which translates into a surface density of ~ 3.7 clusters per square degree. The mean estimated redshift of the AK clusters is 0.30, and is 0.25 for the VT catalog. The mean richness is $\mathrm{N}_{gals} = 43.7$ for the AK and $\mathrm{N}_{gals} = 30.6$ for the VT. The sky distribution for the combined AK-VT cluster candidates is shown in Figure 4.20, for the regions in the northern galactic hemisphere (NGH) and southern galactic hemisphere (SGH). Examples of rich and distant clusters detected by the VT and AK codes are shown in Figure 4.21. The stamps are taken from the DPOSS F-plate images and sample 250″ on a side. However, the low quality of the plates makes it difficult to visually identify most of these distant clusters. Figure 4.22 compares a CCD image taken for one candidate to the DPOSS-F plate. The CCD field was taken at the Palomar 60″ telescope using the r-band filter.

The combined cluster catalog with 9,982 candidates is presented in Table 4.1, which is sorted by RA. An online complete version of this table can be found at `http://dposs.caltech.edu/dataproducts`. The table is organized as follows: column 1 gives the cluster name, where the convention is NSCS Jhhmmss+ddmmss (the second "S" in the name stands for *supplemental*); columns 2 and 3 give the RA and Dec in decimal degrees (J2000); the redshift estimates are shown in columns 4 (AK) and 5 (VT). Entries are set

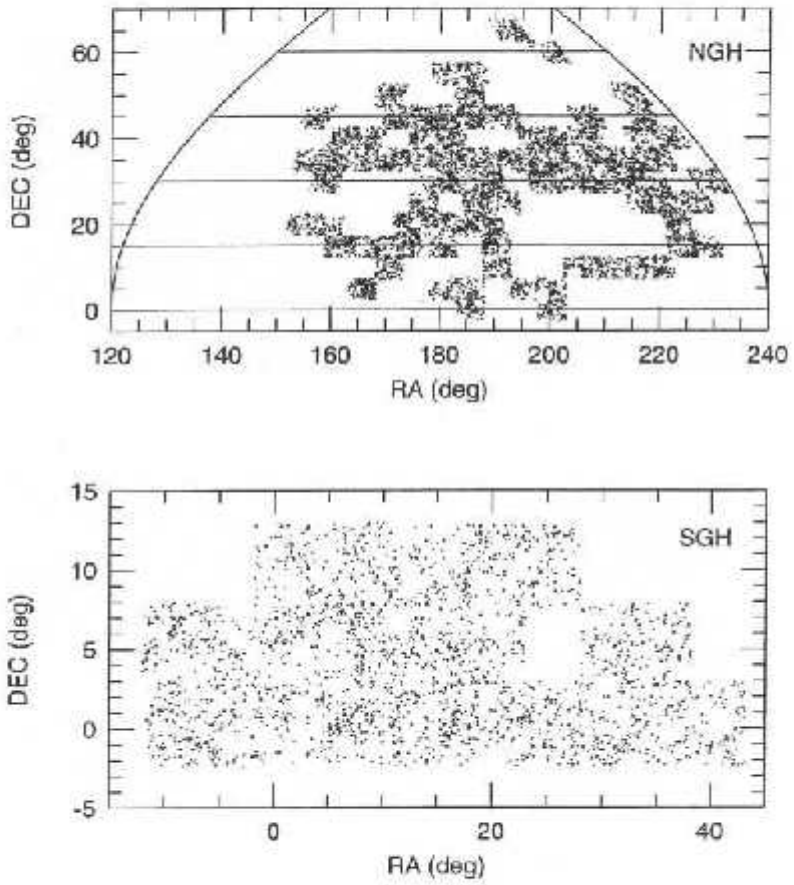


Figure 4.20: The sky distribution in equatorial coordinates for the combined AK-VT catalog (9,982 candidates). The region covering the northern galactic hemisphere is shown in the top panel, while the bottom panel shows the southern galactic hemisphere.

to zero whenever the corresponding estimator failed or if the candidate was not detected by that algorithm. Columns 6 and 7 present the richness estimates. Column 8 has the plate number and the codes that retrieved the cluster are indicated in column 9. This column has "A" for the AK code, "V" for the VT and "AV" for both.

4.6 The Selection Function

In order to estimate the completeness of this catalog we evaluated the selection function (SF) for a small number of fields. Unlike Papers II and III, we do not perform a complete field-by-field analysis of the SF, as this catalog is not intended to provide a statistical sample.

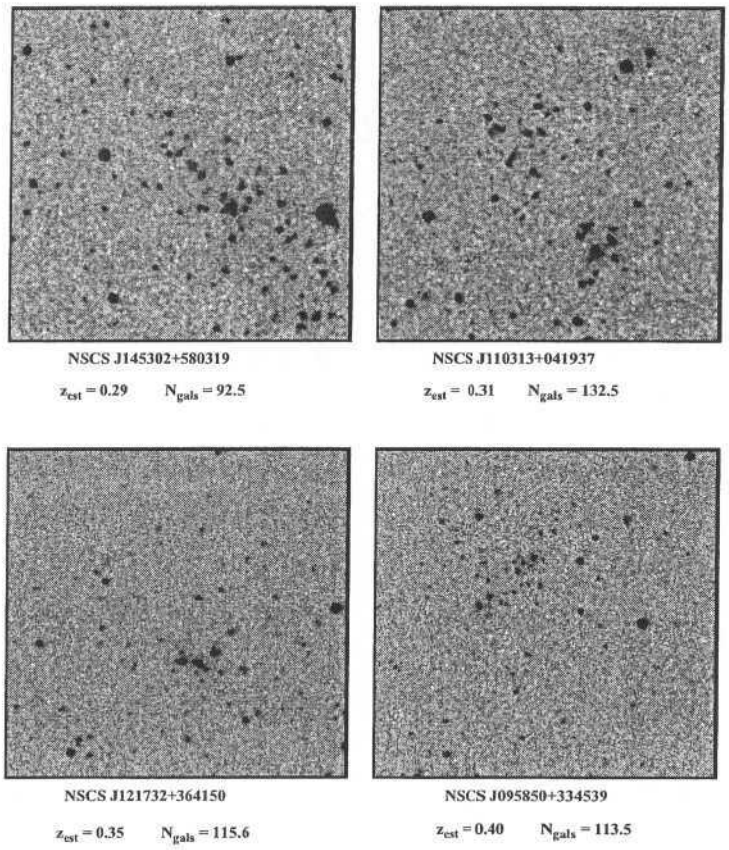


Figure 4.21: Examples of intermediate redshift rich clusters candidates detected at DPOSS. Images are 250″ on a side. NSCS J145302+580319 = Abell 1995 at $z_{spec} = 0.32$; NSCS J110313+041937 was not previously catalogued; NSCS J121732+364150 = RX J1217.5+3641; NSCS J095850+334539 = ZwCl 0955.8+3401. Those matches were found at NED adopting a search radius of 3 arcmin.

Instead, the SF measurements provide an approximate completeness estimate, and are used to check for possible improvement in the recovery rate for rich clusters when allowing a higher contamination level. In addition, the SF is necessary to determine if the estimated redshift distributions shown in Figure 4.17 are realistic. Finally, in the richness and redshift regime where we predict our catalog to be nearly complete, we should find good agreement with other catalogs of rich clusters spanning the same volume. Comparisons to such catalogs are presented in §4.8.

The most common technique to evaluate the selection function makes use of artificial

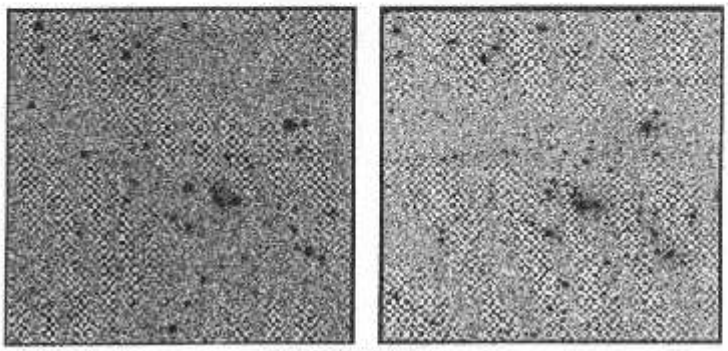


Figure 4.22: Candidate NSCS J234730-000853, also found by the CE method (Goto *et al.* 2002), with an estimated redshift from SDSS data of $z_{est} = 0.24$. The left panel shows a DPOSS F-plate image and the right one shows a CCD image taken at the Palomar 60″ telescope. Images are 250″ on a side. This figure demonstrates the difficulty of visually confirming distant rich clusters from DPOSS plates.

clusters. These are inserted in a background distribution and their recovery rate as a function of richness and redshift gives the SF. However, there are three major complicating factors:

1. The background used for completeness tests should not be the same as the one used for the optimization of the algorithm. As shown by Goto *et al.* (2002) and in Paper II, a simulated background overestimates the completeness rate. The real field should be used instead.

2. When randomly placing artificial clusters in the background distribution we must avoid bad areas (due to saturated objects), as well as the positions of cluster candidates, as done by P02 and in Paper II. A simulated cluster located in the vicinity of a real cluster has its recovery probability changed from what it would be in a region devoid of clusters (which is more likely to be a background region). In this way we avoid the bias in the detection rates discussed by Goto *et al.* (2002).

3. The background plus artificial clusters galaxy catalog must have the same number density as the real field, as the optimal threshold was obtained with the original number density. Galaxies from rich artificial clusters can increase the number density of the whole field by a few percent.

Table 4.1: The Northern Sky Cluster Catalog Supplemental: Excerpt

Name	RA (J2000.0)	Dec (J2000.0)	$z_{phot-AK}$	$z_{phot-VT}$	$N_{gals-AK}$	$N_{gals-VT}$	Plate	codes
NSCSJ124108+260838	190.283890	26.143654	0.33	0.00	46.1	0.0	506	A
NSCSJ124108+203645	190.284241	20.445881	0.30	0.27	38.9	36.8	574	AV
NSCSJ124109+201638	190.285614	20.277033	0.35	0.00	53.5	0.0	574	A
NSCSJ124109+345240	190.289682	34.877716	0.30	0.30	45.7	61.4	381	AV
NSCSJ124112+200102	190.300283	20.017195	0.30	0.29	50.9	52.0	574	AV
NSCSJ124114+441253	190.309921	44.214638	0.32	0.32	50.2	59.0	268	AV
NSCSJ124115−524407	190.311127	52.735004	0.17	0.19	18.6	30.8	172	AV
NSCSJ124115+360354	190.312140	36.048342	0.00	0.24	0.0	27.9	381	V
NSCSJ124115+180058	190.313309	18.014511	0.38	0.35	76.7	73.0	574	AV
NSCSJ124117+503547	190.320557	50.596100	0.29	0.00	29.8	0.0	318	A
NSCSJ124118+183052	190.323275	18.564341	0.09	0.09	34.2	31.8	574	AV
NSCSJ124118+261817	190.326447	26.304502	0.48	0.00	77.4	0.0	442	A
NSCSJ124121+251031	190.337302	25.175100	0.33	0.00	46.4	0.0	507	A
NSCSJ124126+570034	190.359171	57.009380	0.29	0.00	29.9	0.0	172	A
NSCSJ124127+455630	190.363968	45.941415	0.00	0.00	0.0	0.0	388	V
NSCSJ124128+434030	190.366119	43.637419	0.33	0.00	44.0	0.0	268	A
NSCSJ124132+265817	190.382889	26.971125	0.40	0.00	35.6	0.0	506	A
NSCSJ124134+151039	190.386332	15.277313	0.35	0.00	44.0	0.0	645	A
NSCSJ124134+540220	190.390579	54.038685	0.00	0.25	0.0	24.9	172	V
NSCSJ124135+232315	190.393692	23.387304	0.43	0.00	53.8	0.0	506	A
NSCSJ124135+273913	190.397263	27.003359	0.00	0.23	0.0	51.9	442	V
NSCSJ124135+435248	190.397753	43.879933	0.00	0.23	0.0	26.4	388	V
NSCSJ124136−325055	190.398514	32.848547	0.29	0.29	72.5	77.2	381	AV
NSCSJ124141−201639	190.421219	20.277275	0.00	0.41	0.0	38.6	574	V
NSCSJ124141−565018	190.422526	56.838177	0.00	0.31	0.0	11.7	172	V
NSCSJ124146+294460	190.440155	29.747133	0.36	0.37	58.4	58.2	442	AV
NSCSJ124146+705833	190.440583	70.975756	0.30	0.16	0.0	25.0	717	V

We generate sets of 48 clusters with 6 different richness classes ($N_{gals} = 15, 25, 35, 55, 80, 120$) and placed at 8 different redshifts ($z = 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50$). The simulated clusters are placed at random locations within a field; if it falls on a bad area or cluster candidate, a new position is chosen. We also preserve the same number of objects as in the original field by randomly removing the same number of galaxies from the original catalog as are inserted with the artificial clusters. This procedure is repeated 50 times.

For both the AK and VT methods we use a 0.75 h^{-1} Mpc radius when comparing the input and output positions (this choice is explained below). An example of the selection function obtained for a DPOSS field is shown in the middle panel of Figure 4.23.

Artificial clusters were generated in the same way as in Paper II. They follow a Schecter luminosity function, with parameters given by Paolillo *et al.* (2001). The characteristic magnitude is $M^* = -21.52$, while $\alpha = -1.1$. Cluster galaxies lie at $-23.4 \leq M_r \leq -16.4$ and are composed of 60% elliptical galaxies and 40% Sbc galaxies. K-corrections are obtained through the convolution of SEDs taken from Coleman, Wu & Weedman (1980), with the DPOSS r filter. We adopt a power law in radius for the surface profile, r^{β}, where $\beta = -1.3$

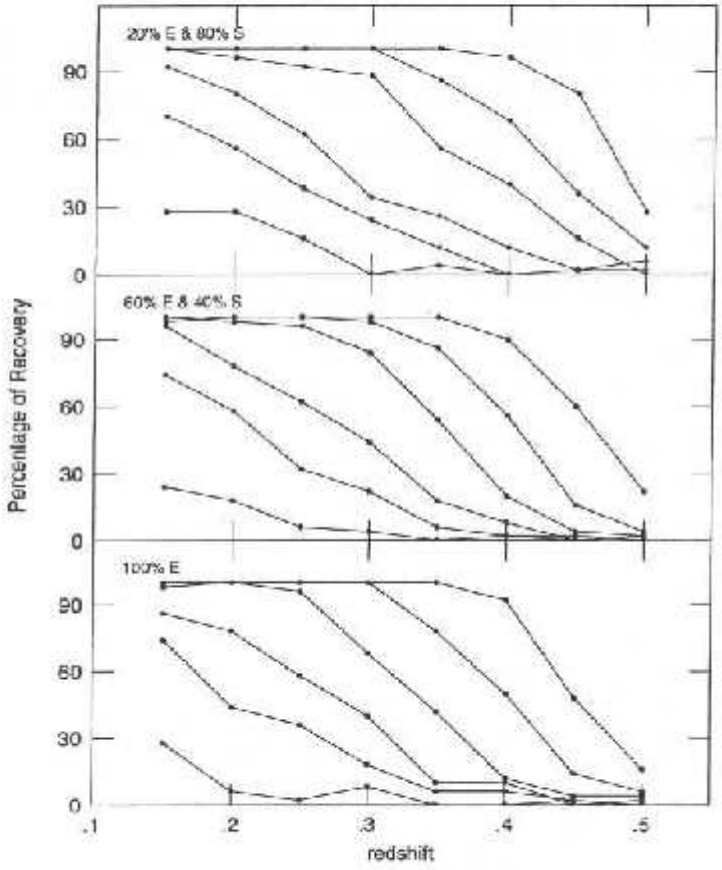


Figure 4.23: Comparison of the selection function evaluated with three different cluster compositions. The top panel shows the results for 20% E & 80% Sbc, the middle panel is for 60% E & 40% Sbc, while the bottom panel is for 100% ellipticals. The richness classes are as indicated in Figure 4.12.

lies in the middle of the observed range (Squires *et al.* 1996; Tyson & Fischer 1995). Clusters also have a cut-off radius at $r_{max} = 1.5$ h^{-1} Mpc and a core radius $r_{core} = 0.15$ h^{-1} Mpc.

In Paper II all of these five basic parameters (luminosity profile slope, r_{max}, r_{core}, spatial profile slope (β) and the cluster composition) had their effects on the evaluation of the SF tested. The strongest variation is with the value of β. Large fluctuations in the SF were not seen when varying the other parameters. We then decided to adopt the canonical values above. In this work, significant differences could arise from testing different cluster compositions. At higher redshifts, the K-correction effects could be more significant when testing clusters with different compositions. Figure 4.23 shows a comparison of the SF for

the VT code evaluated with three different cluster compositions (20% E & 80% Sbc; 60% E & 40% Sbc; 100% E). There is a clear trend toward a higher recovery rate when using clusters mainly composed of late-type galaxies, although the effect is weak. If we fix the comparison at $z = 0.35$ and $N_{gals} = 80$, the recovery rate decreases from 88% to 84% from the top to the middle panel and finally to 80% for the bottom panel.

We also test for improvement in the SF when we change the required value of C from 5% to 10%. This is shown in Figure 4.24 for the VT and AK. As there is no significant improvement for the most distant and rich clusters when allowing higher contamination, we keep the contamination rate fixed at $C = 5\%$.

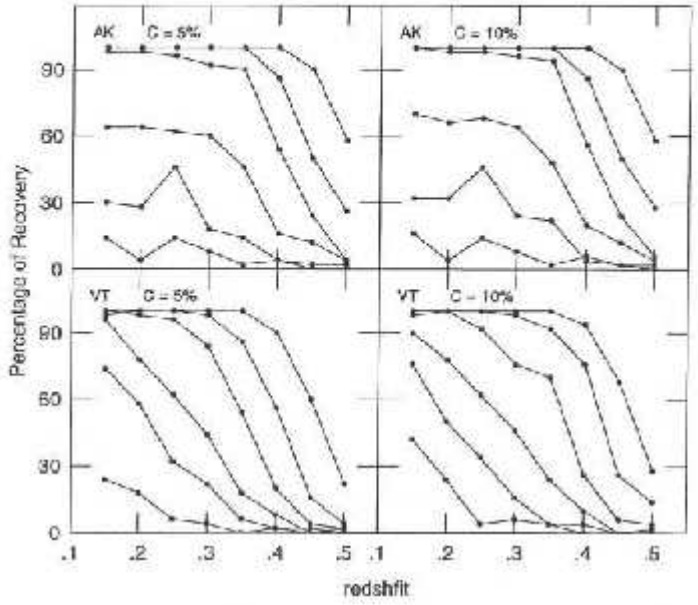


Figure 4.24: Comparison of the SFs obtained for the VT (bottom) & AK (top) codes when fixing C = 5% (left) and C = 10% (right). The richness classes are as indicated in Figure 4.12.

When assessing the SF of a given cluster catalog, it is important to pay attention to items (1), (2) and (3) listed above, as well as the matching radius used to compare input and output positions of artificial clusters. Goto *et al.* (2002) checked the dependence of the position deviation according to redshift and richness. They used a small radius of 1.2 arcmin to match input and output catalogs, and found a negligible variation to $z \sim 0.4$. Instead of using an apparent radius for the match, we use a physical radius. We perform a test

with 1.5 h^{-1} Mpc (the size of the input clusters), checking the dependence of the offsets on redshift and richness. Our final choice is 0.75 h^{-1} Mpc for comparing the original positions of artificial clusters with the output of both our algorithms.

Figure 4.25 shows the SF evaluated with four different matching radii for the VT code: 0.4, 0.75, 1.0 and 1.5 h^{-1} Mpc. There is a great improvement when we go from 0.4 h^{-1} Mpc to 0.75 h^{-1} Mpc, while the SF does not change significantly for larger matching radii.

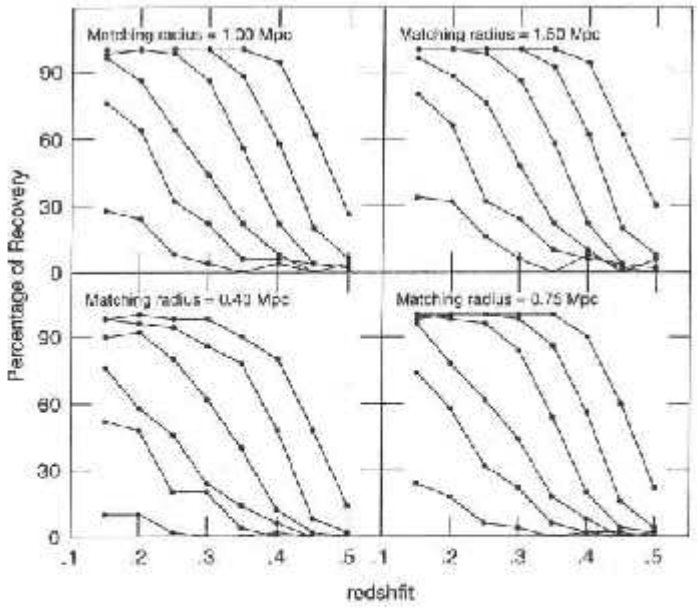


Figure 4.25: The SF evaluated with four different matching radii (0.40, 0.50, 0.75 & 1.5 Mpc). The richness classes are as indicated in Figure 4.12.

Figure 4.26 shows (for both the AK & VT) the positional offset (in Mpc) according to redshift for different richness classes. Each line represents a different richness class. Thicker lines represent richer clusters. For the poorer clusters the offset goes to infinity for higher redshifts ($z > 0.40$). This simply means the recovery rate goes to zero for these clusters. We note that the offset is slightly larger and noisier for the VT code as compared to the AK. This is due to the percolation analysis needed for the VT, where we begin with candidates in different magnitude bins. An artificial cluster can have its recovered position slightly changed if it percolates with a nearby fluctuation.

Finally, it is important to stress that the simulations used to assess the contamination rate (§4.4) only provide a lower limit. Similarly, the clusters simulations employed provide

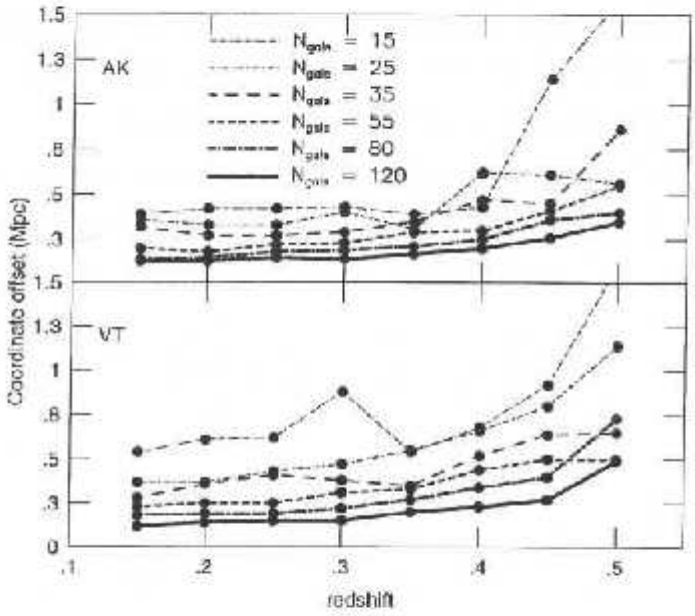


Figure 4.26: Positional offsets (in Mpc) for the VT and AK codes as a function of redshift. Each line represent a different richness class ($N_{gals} = 15, 25, 35, 55, 80, 120$), with the thickest representing the richest clusters.

only an upper limit for the completeness rate of the catalog.

4.7 Performance comparison of the two algorithms

The AK cluster catalog contains 4,813 candidates, whereas the VT contains 7,962. The combined catalog shown in Figure 4.20 contains 9,982 cluster candidates. From the analysis of Figure 4.24 we would expect the AK catalog to have more rich clusters, and to go deeper, as compared to the VT catalog. On the other hand, the VT catalog is more complete in the regime of poor/nearby systems. The redshift and richness distributions of the cluster candidates confirm the expectations from the simulations. The cutoffs in every panel of Figures 4.17 and 4.18 show that the cluster distributions are in good agreement with the SFs.

It is important to stress that both methods applied here are simple density estimators, which detect fluctuations in the projected galaxy distribution above a given threshold. Both methods make no assumptions whatsoever on any physical properties of galaxy clusters

(e.g., the luminosity function, colors of galaxy members, spatial profile, etc). The AK code is sensitive to an initial smoothing scale, while the VT does not distribute the data in bins. However, with the proper choice of the initial kernel size, we expect the AK to be as sensitive to irregular structures as the VT. We believe the differences shown here arise from the way we minimize the background in each case. The VT is applied to galaxies spanning a wide range of magnitudes ($16.0 \leq m_r \leq 21.1$), while the AK code is applied only to the faint data of DPOSS (the bright cut is at $m_r = 19.0$). Thus, the AK naturally avoids nearby fluctuations and enhance the contrast of distant systems, while we employ a binning scheme to minimize contamination for the VT algorithm.

Obviously, these same density estimators when applied to multi-wavelength data provide a much more powerful tool to minimize background/foreground contamination (Kim *et al.* 2002; Goto *et al.* 2002; Gladders & Yee2000). It is not a straightforward task to compare catalogs generated by these different techniques. As noted by Kim (2001), clusters are missed by one method and not by another due to the fact that they are *seen* with different *eyes*. Some cluster finding techniques will be biased to clusters with some specific properties, thus having a low efficiency for overdensities that do not match these properties. Furthermore, clusters which are recovered by different techniques might have large differences in the measurement of properties such as richness, redshift and projected density profile.

Kim (2001) used a 0.7 h^{-1} Mpc radius when matching candidates from different algorithms. However, she finds that, in some cases, the cluster centers from different algorithms can differ by as much as the extent of the cluster. This could be especially true for poor and irregular clusters.

Large differences in the redshift estimates for candidates from the two codes would make this comparison more difficult, or even impossible. However, we do not expect the photometric redshifts to have a large variation in our case. So, we decided to adopt a physical radius, instead of an apparent radius when comparing the AK and VT catalogs. We adopted a 0.75 h^{-1} Mpc maximum radius, as done for the evaluation of the selection functions.

Out of 4,813 (AK) and 7,962 (VT) cluster candidates, 2793 (58% of AK and 35% of the VT) are common sources. The VT catalog is denser than the AK catalog because it is more complete in the regime of poor systems, which are more abundant. Figure 4.27 shows the offset distribution for the matched clusters in Mpc (solid line) and arcseconds (dotted line).

Most clusters show no large offsets (less than 100″ or 0.40 h^{-1} Mpc). The combined catalog is presented in Table 4.1 and Figure 4.20.

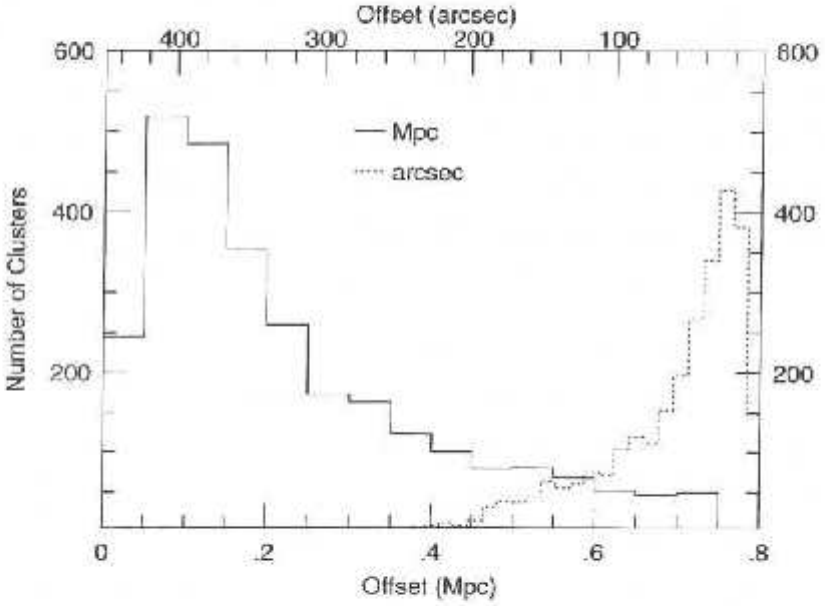


Figure 4.27: Centroid offset distribution for common clusters from the VT and AK codes. Offsets in Mpc are indicated by the solid line, while the offsets in arcsec are represented by the dotted line.

Redshift and richness estimates are obtained post-detection. They are however sensitive to the cluster center, as the cluster corrected counts will depend on this choice. We then divide the common clusters into bins of 40″ in centroid offsets, from 0″ to 240″. The residuals as a function of redshift for different offsets are shown in Figure 4.28, with a clear trend to large residuals for large offsets. Figure 4.29 shows the same effect, but for richness. Finally, we investigate the dependence of the richness evaluation on the adopted redshift. Figure 4.30 is similar to the previous one, but the bins are in redshift residuals (from $\Delta z = 0.01$ to 0.06, in steps of 0.01), instead of cluster center offsets. The top left panel shows the results when cutting the sample at offset $< 40''$ and $\Delta z < 0.01$. There is an evident improvement when comparing this panel to the bottom left panels of this Figure and Figure 4.29.

We show in Figure 4.31 richness (top) and redshift (bottom) distributions for VT only detections (dashed-dotted line), AK only clusters (dotted line) and common candidates (heavy solid line). This plot fully confirms the expectations from the selection function estimates.

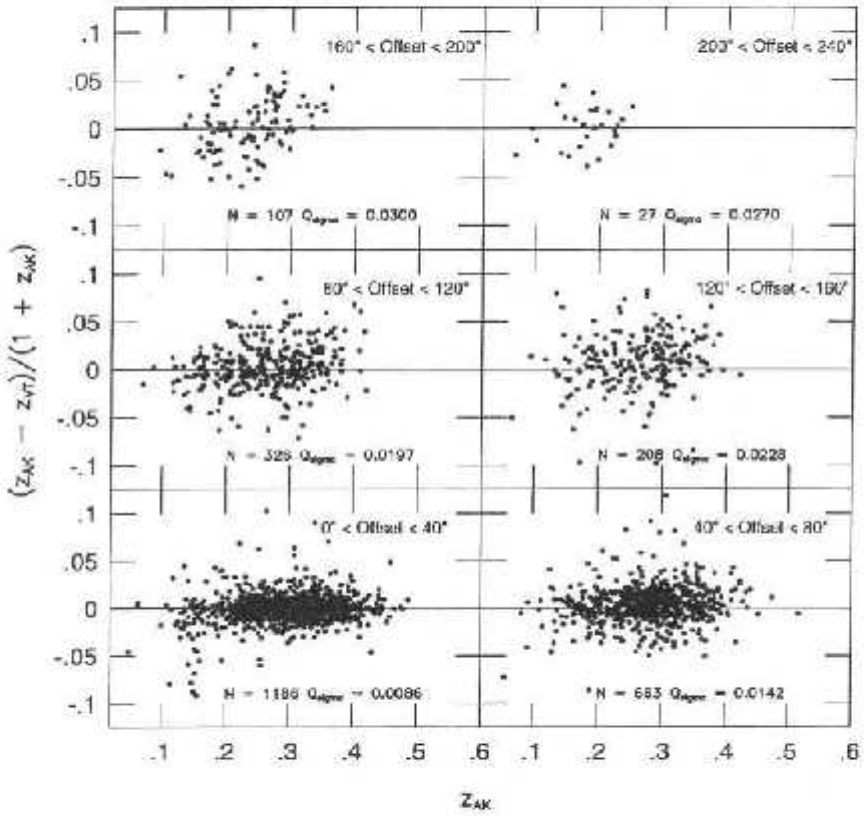


Figure 4.28: The dependence of the redshift estimate on the cluster centroid. Redshift residuals as a function of redshift are shown for different offset ranges, which are indicated on each panel. The number of clusters (N) is also indicated.

The VT code performs better for poor nearby clusters, while the AK goes deeper when detecting rich systems. This plot also illustrates the main difference in the performance of the algorithms. Most of the VT only detections are nearby poor systems, while the AK only clusters are mainly distant and rich. The two catalogs overlap at intermediate richness and redshift ranges.

Figure 4.32 shows the distribution of N_{gals} with respect to the estimated redshift for both AK and VT. A similar relation was previously found by Kim (2001). The richer systems are rarely found at low redshifts where we probe less physical volume per unit redshift interval. On the other hand the poor clusters are easier to be detected at low-z, where they are still seen as high-contrast systems. One surprising result is that only the VT code apparently finds relatively poor systems ($N_{gals} \leq 60$) at $z > 0.4$. For these faint systems the redshift

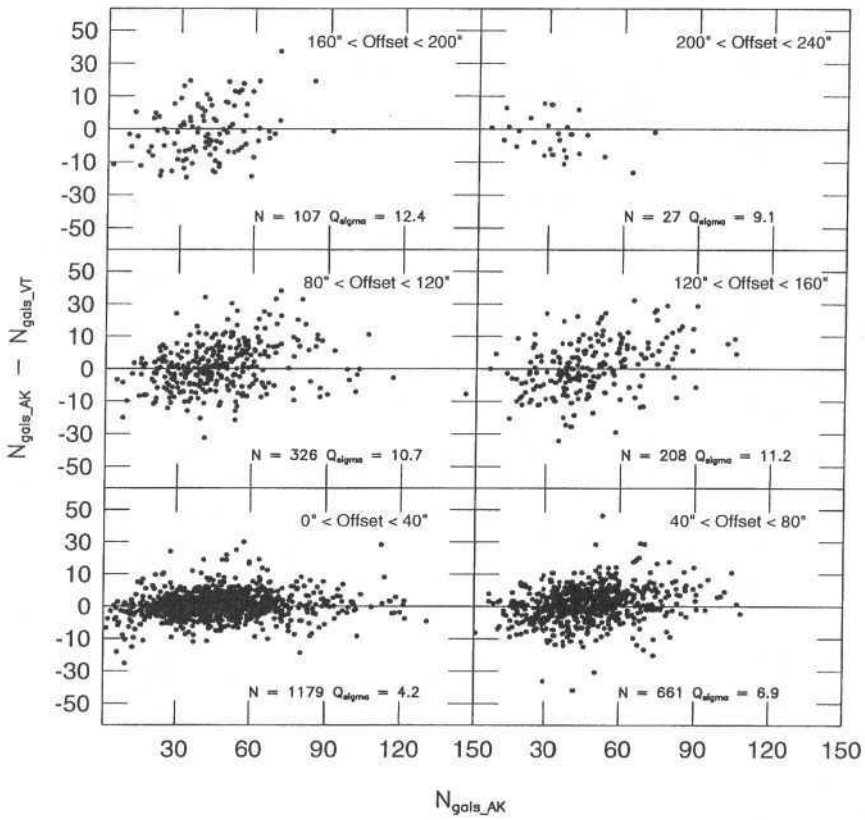


Figure 4.29: Richness residuals as a function of richness are shown for different offset ranges, which are indicated on each panel. The number of clusters (N) is also indicated.

and richness estimates are more difficult to measure as the number of galaxies used for these estimates is typically low. The VT code is also less accurate than the AK algorithm in determining the centroid of a cluster (Figure 4.24). For these low contrast systems that will represent a great impact to the redshift and richness estimates, and could result in underestimation of the richness value. Note also that Kim (2001) found the same trend for the VT candidates when using SDSS data.

Finally, we test the overlap of both catalogs in the regime where they are expected to have high completeness. The SF predicts that both cluster catalogs are nearly complete for rich clusters ($N_{gals} \geq 65$) at $0.2 < z < 0.3$ (Figures 4.24 and 4.25). To be conservative we assume a completeness level $> 90\%$ and select all cluster candidates from both catalogs for these richness classes at this redshift range. We find 140 VT candidates and 177 in the AK catalog. All VT candidates have a match in the AK catalog, which represents an overlap of

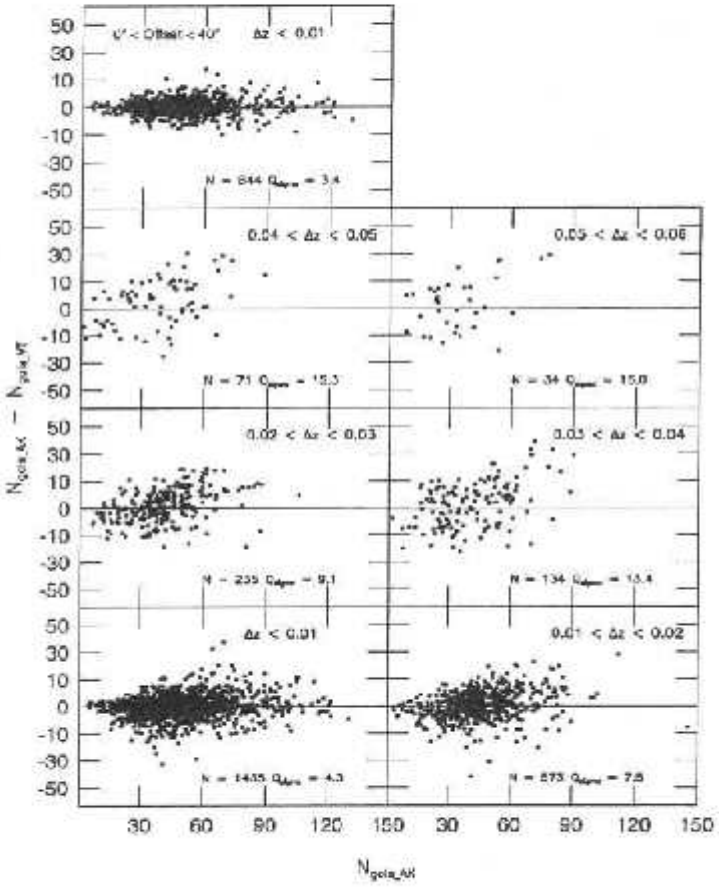


Figure 4.30: Richness residuals as a function of richness are shown for different bins of redshift residuals (indicated on each panel). The number of clusters (N) is also indicated. The upper left panel has the results when selecting only clusters with offsets $< 40''$ and $\Delta z < 0.01$.

$\sim 80\%$, in agreement with the expectations from the SF. For this comparison we adopted a matching radius of 6′, as done by Goto *et al.* (2002).

4.8 Comparison with other cluster surveys

A comparison of the NoSOCS and the Abell cluster catalog was previously done in Papers II and III. We would like to compare the supplemental NoSOCS catalog presented here to other intermediate redshift cluster catalogs. The best sources of comparison are given by the preliminary SDSS catalogs (Kim 2001; Goto *et al.* 2002), which are then used as a reference

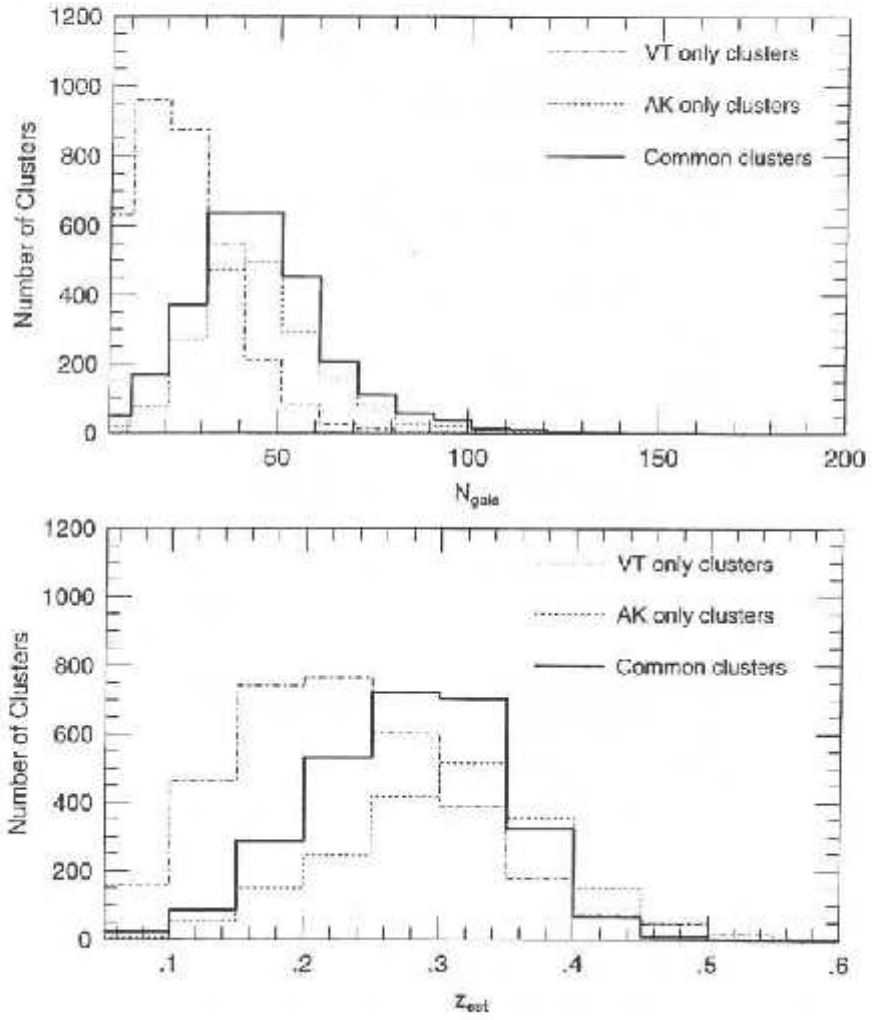


Figure 4.31: Richness (top) and estimated redshift (bottom) distributions for clusters detected only by the VT code (dashed-dotted line), only by the AK code (dotted line) and by both methods (heavy solid line).

for the comparison shown in the end of this section. Additionally, as the DPOSS data used here is different than that used in Papers II and III (the current galaxy catalog is deeper and has a statistical correction applied to it), we also compare to those catalogs. We would also like to investigate differences in the redshift and richness measurements. As stated before, there are some fundamental differences in the estimates presented here and in the previous DPOSS papers.

From the $>$ 12,000 NoSOCS (Papers II & III) clusters candidates over the whole high

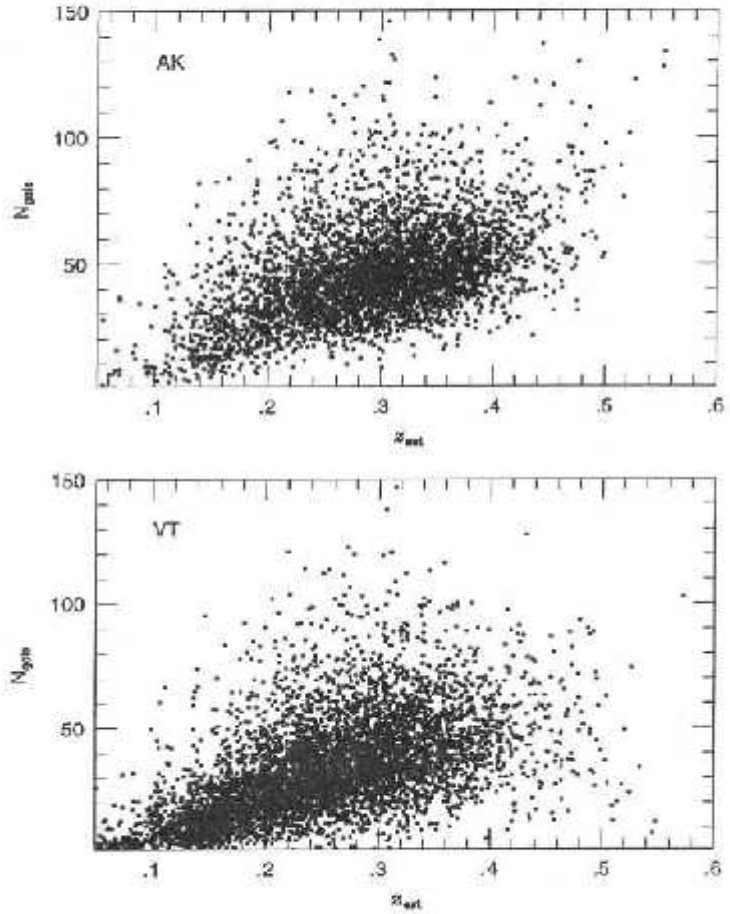


Figure 4.32: Richness *vs* estimated redshift for the AK (top) and VT (bottom) candidates.

latitude ($|b| > 30°$) northern sky, we find that 4,211 are within the 2,700 square degrees sampled by the current project. Due to large differences in the redshift estimates we could not use a physical radius when comparing NoSOCS to the deeper catalog presented here. We therefore adopt a 200″ radius for this comparison. We found that 2,417 clusters are common to both catalogs. Figure 4.33 shows the estimated redshift distribution (bottom panel) of NoSOCS only clusters (Papers II and III) as dotted lines, while the dashed lines represents the distribution of clusters present only in the current survey. The distribution of common clusters is represented by the heavy solid line. It is clear that most of new candidates found here are at $z_{est} \geq 0.2$. On the top panel of the same Figure we show the ratio of common clusters to each of the catalogs. The dotted line represents the ratio to the catalog presented in Papers II & III, while the dashed line is the ratio of common clusters to the supplemental

catalog presented here. Down to $z_{est} \sim 0.2$ approximately 50% of NoSOCS clusters are also common to the new catalog. The supplemental catalog has almost all of its common detections at low-z.

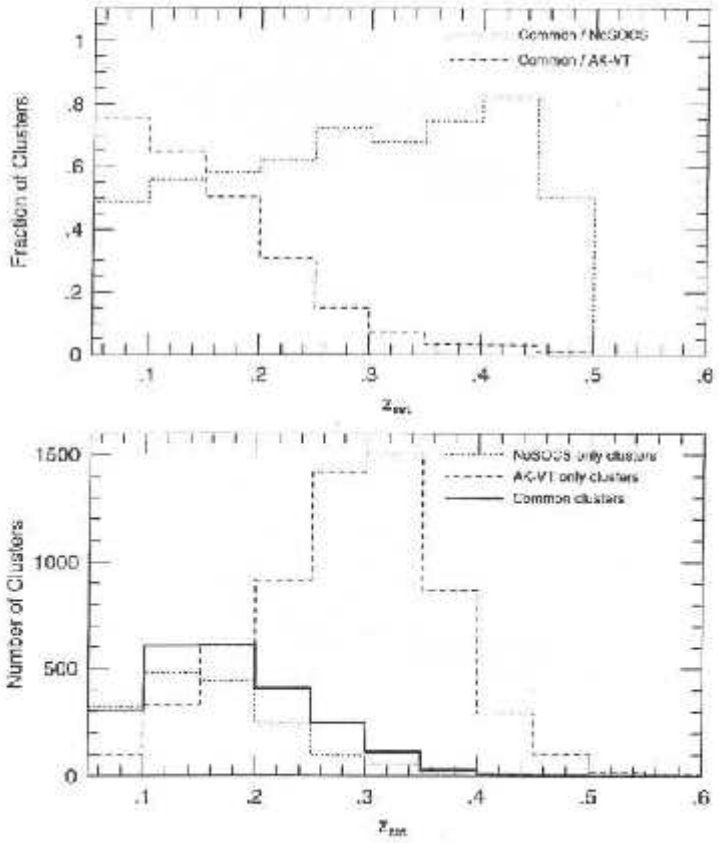


Figure 4.33: Estimated redshift distributions (bottom panel) for clusters detected only in Papers II & III (dotted line), only by this paper (dashed line) and by both surveys (heavy solid line). The top panel shows the ratio of common clusters to the NoSOCS catalog (dotted line) and to the supplemental catalog presented here (dashed line).

It is important to stress that poor clusters at $z < 0.2$ detected in Papers II & III might not be easily recovered in the current paper. The AK catalog is obviously biased against these systems due to the bright magnitude cut ($m_r \sim 19$), while the VT code might have problems recovering these nearby, low contrast structures. The binning scheme adopted to enhance the cluster contrast might not be powerful enough to enable recovering of these systems. Note also that the matching radius employed when comparing the two catalogs is crucial to the analysis. Poor systems might have large differences in the centroid determination, especially

if the centroid is given by galaxies in different magnitude ranges. Goto *et al.* (2002) adopted a 6′ radius to compare their catalog to other SDSS cluster catalogs. We then adopted this matching radius (6′) to compare the catalogs in the regime where they are expected to have high completeness level. At $0.1 < z < 0.2$ the cluster catalog from Papers II and III has 38 rich clusters ($N_{gals} \geq 65$). The VT catalog finds 33 of those, while the AK catalog finds 30. The overlap is $\sim 80\%$. For the few missing clusters we pointed the richness estimator code to the coordinates given in Papers II & III, also adopting the redshift estimate given there. We found these systems to be poor clusters in the deeper catalog used here. In other words, the contrast of these clusters decreases when seen in a wider magnitude range.

In Figure 4.34 we plot the NoSOCS photometric redshift estimates *vs.* the ones from this paper (left panel). If we assume the estimates from Papers II & III are more correct than those presented here, then we have a slight indication of overestimation of the redshifts presented here. On the right panel we compare the richness estimates for nearby clusters ($z < 0.2$). As stated in section 4.5.2 there are many differences between the richness estimates presented here and in the previous DPOSS papers. First the galaxy catalogs are deeper and have a statistical correction applied to them, which could affect the cluster contrast. Second, we now adopt a local background correction, while the background counts were previously taken from a plate scale (Papers II & III). In the previous papers the γ correction factor considered all galaxies to be elliptical, while here we consider only 60% to be ellipticals.

Finally we compare our catalog to the SDSS cluster catalogs presented at Kim (2001) and Goto *et al.* (2002). Both catalogs cover an equatorial strip of $\sim 350\square^{\circ}$. However, most of the common clusters (between SDSS and DPOSS) are found in the RA $\leq 44^{\circ}$ region, as we have only two plates in the North Galactic Pole Cap region, which are also in this equatorial strip. The precise coordinate limits used to trim both SDSS catalogs, as well as ours, are $-2.1^{\circ} < \alpha \leq 43.75^{\circ}, -1.27^{\circ} \leq \delta \leq +1.27^{\circ}$. There are 863 clusters from Kim (2001), 1288 from Goto *et al.* (2002) and 463 AK-VT candidates in this region.

In Figure 4.35 we show the estimated redshift distributions of clusters detected only by DPOSS (dashed line), common clusters (heavy solid line) and SDSS only clusters (dotted lines). In the top panel the DPOSS cluster catalog is compared to the SDSS *Cut & Enhance* (CE) catalog (Goto *et al.* 2002), while in the bottom panel the comparison is made to the catalog presented by Kim (2001) (a combination of two catalogs, detected by a *Hybrid*

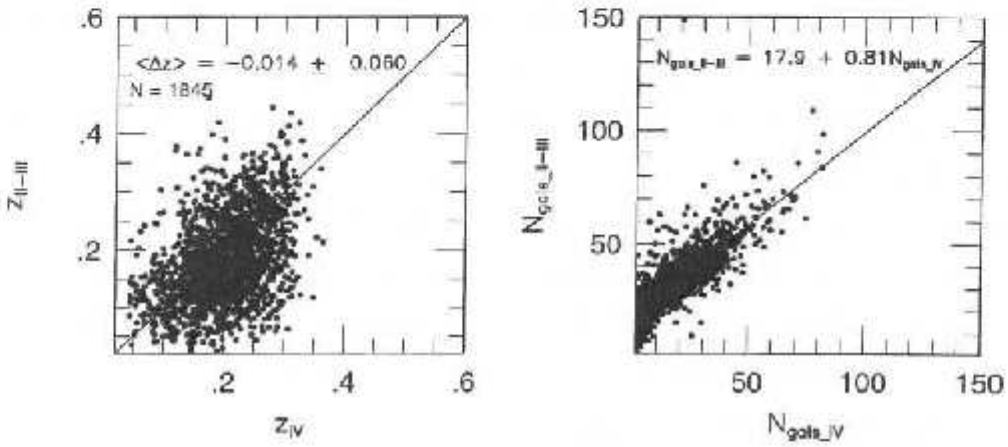


Figure 4.34: Comparison between our catalog and those from Papers II and III. On both panels the abscissa has the results for the current paper (IV), while the ordinate shows the results from Papers II and III. The left panel shows the redshift estimates for 1,845 common clusters between this paper and the previous two. On the right richness estimates for $z < 0.2$ clusters are compared.

Matched Filter and a Voronoi Tessellation techniques, hereafter called HMF-VT). In both panels the single method detections have redshift estimates coming from the own method. In the bottom panel, matched clusters have our redshift estimates, while in the top panel the common clusters have the CE estimates. The HMF-VT catalog (Kim 2001) recovers many more nearby (probably poor) clusters than DPOSS, while the CE catalog (Goto *et al.* 2002) goes deeper than DPOSS. This result also points to intrinsic differences in the clusters sampled by the two SDSS catalogs.

Our redshift measurements are compared to the ones given in both SDSS catalogs. The estimates given by Kim (2001) have an *rms* comparable to ours, while Goto *et al.* (2002) measurements are much more precise ($\Delta z < 0.02$). The upper left panel of Figure 4.36 indicates that either Kim (2001) is underestimating the redshifts or we are overestimating them. However, an overestimation from our technique should be much smaller than that suggested by this plot. Because no strong trend is evident when comparing our estimates to the ones from Goto *et al.* (2002) (upper right panel), we conclude that Kim (2001) might also underestimate the cluster redshifts.

When comparing richness measurements we re-measure N_{gals} for the common clusters, using the coordinates and redshift estimates provided in the SDSS catalogs. This procedure

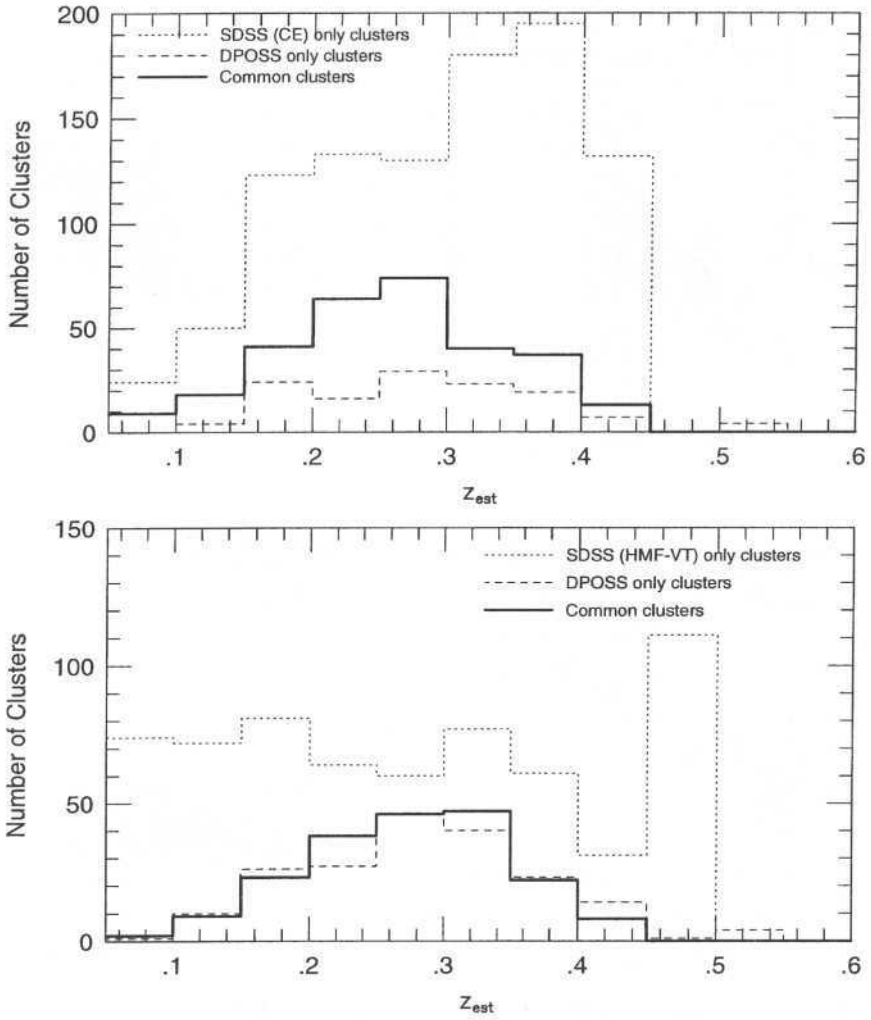


Figure 4.35: Estimated redshift distributions of DPOSS only clusters (dashed lines), common clusters (heavy solid line) and SDSS only clusters (dotted line). In the bottom panel DPOSS is compared to the catalog presented in Kim (2001), while the comparison to the CE catalog Goto *et al.* (2002) is shown in the top panel.

minimizes the effects seen in Figures 4.28 to 4.30. The richness measure given by Kim (2001) is called Λ_{cl}, which represents the total cluster luminosity within 1.0 h^{-1} Mpc in units of L^*. N_{gals} also considers a 1 h^{-1} Mpc counting radius, but we only sample galaxies at $M^* - 1 \leq M \leq M^* - 2$. Goto *et al.* (2002) count galaxies from m_3 to $m_3 + 2$ (where m_3 is the third brightest cluster galaxy) within the detection radius provided by the CE algorithm. As they state, this is similar to the Abell richness, except for the counting radius, which is

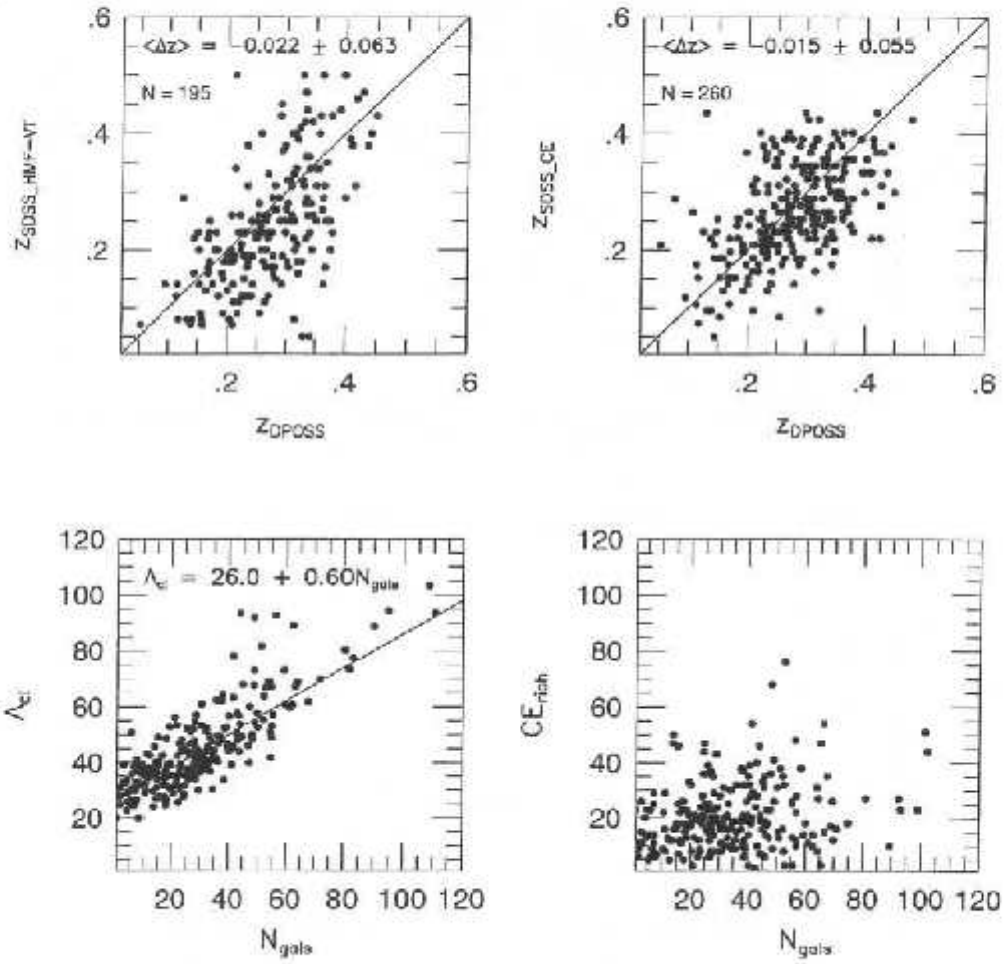


Figure 4.36: Comparison between our catalog and those from Kim (2001) & Goto et al. (2002). In the top panels we compare our redshift estimates to those from Kim (2001) (left) and Goto et al. (2002) (right). The bottom panels show the richness comparisons. More details in the text.

chosen differently for the CE method. The bottom left panel shows the relation between Λ_{ct} and N_{gals}. As we adopt the same coordinates and z_{est} provided by Kim (2001), we expect the scatter to originate mainly from the differences in the galaxy catalogs used for the richness estimation. Considering the higher quality of the SDSS data compared to the DPOSS plate data, the small scatter shown in this plot is an encouraging result. The bottom right panel shows the correlation between the CE richness and N_{gals}. The scatter is extremely large and no obvious relation is found. We believe that has nothing to do to the nature of the data,

but mainly to the choice of an apparent aperture to count galaxies for the CE richness. We believe that even a comparison between Λ_{cl} and the CE richness is difficult.

The number density of the clusters presented here is more than 2 times that found in Papers II and III. This is sensible as we sample many more distant clusters when compared to our previous papers.

The interesting comparison, however, is to the SDSS catalogs. The Gunn r-band has a similar response to the SDSS r^* filter, which renders our $m_r = 21.1$ limit similar to the one adopted by Kim (2001) and a little brighter than the choice of Goto *et al.* (2002). Thus, the catalogs' depth should not be too different. The cluster catalogs will differ more due to the quality of the data used, the algorithms and the contamination level allowed.

The number density of the HMF clusters is 7.8 clusters per square degree Kim (2001), 13.3 for the CE catalog Goto *et al.* (2002), and 3.7 for our catalog. The main reasons for the HMF-VT catalog to find more than two times clusters than our technique are the higher quality photometric data of SDSS, as well as the higher contamination rate allowed by Kim (2001). We believe it is $\sim 15\%$, while we kept our catalog at the 5% level. The CE catalog has an extremely high number density. We believe this is due to the fact that they go 0.5 magnitude deeper when looking for clusters, and their $\sim 30\%$ contamination rate in the regime of poor clusters (where most candidates are). The CE method is a powerful technique to detect clusters and minimize background/foreground contamination, but that might not justify such a high number density of real clusters.

An illustration of the higher density of the HMF-VT in comparison to the AK-VT catalog is given in Figure 4.37. We show the candidate distribution for the center region of DPOSS field 824. HMF-VT candidates are shown as solid circles, while the AK-VT are plotted with dashed circles. The circles represent a 1.5 h^{-1} Mpc radius at the estimated redshift. Two features are readily noticed. First, the HMF-VT only clusters are generally low-z systems as indicated by the large radii of the candidates missed in our survey. Second, the HMF-VT is able to recover cluster candidates at different redshifts along the same line of sight, while our method cannot distinguish between these superposed systems. In most such cases, we recover the fainter, more distant clusters instead of the closer ones.

A closer examination of our redshift distribution (Figure 4.17) when compared to Figure 25 of Goto *et al.* (2002) reveals that in the regime of rich clusters our catalog goes as deep

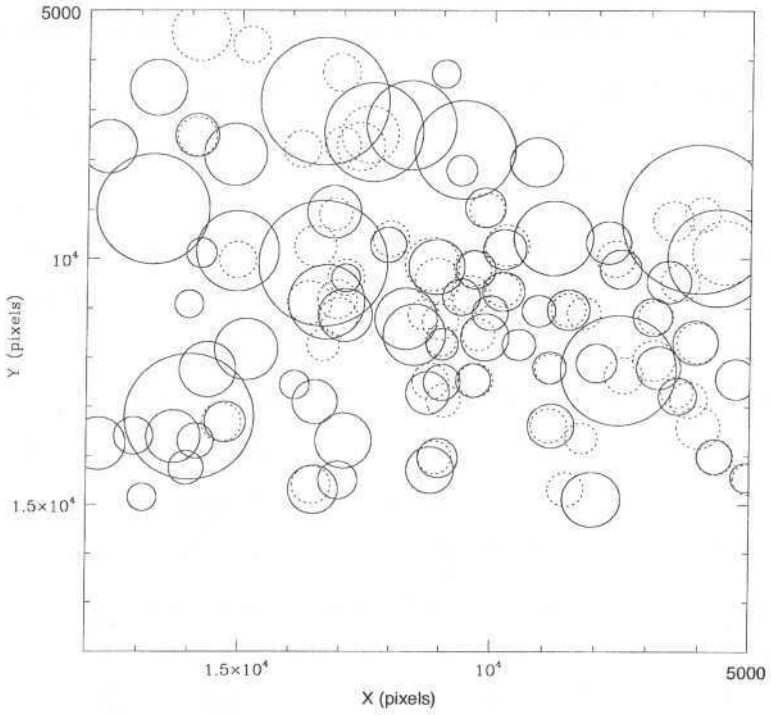


Figure 4.37: The sky distribution of candidates detected in the central region of DPOSS Field 824 ($\alpha = 5.65°$, $\delta = 0.26°$). AK-VT candidates are indicated by dashed circles and the HMF-VT clusters by solid circles. We adopt a 1.5 h^{-1} Mpc radius for this illustration. When the redshift estimator code fails we adopt $z = 0.3$ to draw the circle.

as all the SDSS catalogs. However, the HMF-VT method sample many more poor/nearby systems than us, while the CE catalog detects many more high-redshift candidates ($z_{est} > 0.25$). Nonetheless, it seems that the SFs presented in Goto *et al.* (2002) may not be in good agreement with their redshift distribution. They appear to actually go deeper than indicated by the SFs.

Finally, as done in the comparison to the catalogs presented in Papers II & III, we look for differences to the SDSS catalogs where SDSS and DPOSS predict high completeness levels. Using the relation between Λ_{cl} and N_{gals}, we convert the HMF-VT richness estimates to N_{gals}. We look for clusters with $N_{gals} > 65$ at $0.2 < z_{est} < 0.3$. There are 3 clusters in this regime in the HMF-VT catalog and 9 in our catalog. All the 3 found by Kim (2001) have a match in our catalog, while 7 of the 9 AK-VT candidates have a match in the HMF-VT

list. As we can not convert the CE richness estimate to N_{gals} we simply chose a richness cut for the CE catalog. We selected all clusters with $CE_{rich} > 40$ at $0.2 < z_{est} < 0.3$. Out of 18 CE clusters in this regime 15 have a counterpart in our catalog.

4.9 Summary

This paper presents an intermediate redshift galaxy cluster catalog ($0.1 \leq z_{est} \leq 0.5$) covering 2,708 $\square^{\circ}$ of DPOSS data. The catalog is the result of two different density estimators (AK & VT) and its combined version contains $\sim$ 10,000 cluster candidates. To facilitate the selection of subsamples for follow-up studies we provide redshift, as well as richness estimates. The redshift estimates are as accurate as the matched filter redshift output. As we keep the contamination rate at the 5% level, the probability of selecting false clusters is minimized. However, we needed to clean our catalog after detections were made, in order to discard obvious false candidates due to bright stars, nearby galaxies and groups, etc.

The main goal of this project is to provide a catalog of rich clusters to $z_{est} \sim 0.5$ covering a large area of the Northern sky. It is a valuable source for follow-up studies, such as substructure, alignment of galaxy cluster members, galaxy evolution, the Butcher-Oemler effect, the evolution of the luminosity function and color magnitude relation, etc. We plan to carry out a spectroscopic survey of a subsample of this catalog to confirm the reality of these systems; as well as a deep imaging follow-up.

We present a detailed comparison of the results obtained by the different algorithms. We show the dependence of the redshift and richness estimates on cluster centroid, as well as the richness variation associated to the redshift estimate. We also found a high level of overlap in the regime where both catalogs are expected to have high completeness.

Our catalog is also compared to the results of the most recent surveys, showing an excellent agreement with those where all catalogs are expected to have high completeness rate (rich clusters at $0.2 < z_{est} < 0.3$). However, its is evident that the cluster catalogs derived from the higher quality multi-color data of SDSS are more complete than our catalog. The HMF-VT detects more poor/nearby systems, while the CE catalog is more complete for higher-z systems. Nonetheless, this work provides the largest resource for galaxy clusters spanning this redshift range to date. It will certainly be superseded in the nearby future by

the whole sky catalogs generated from SDSS. However, it represents the first effort to search for clusters out to the faintest limits of a sky survey over such a large area. This cluster catalog constitutes an excellent resource for comparison with the final SDSS catalog, as well to deep X-ray cluster catalogs.

The processing of DPOSS and the production of the Palomar-Norris Sky Catalog (PNSC) on which this work was based was supported by generous grants from the Norris Foundation, and other private donors. Some of the software development was supported by the NASA AISRP program. We also thank the staff of Palomar Observatory for their expert assistance in the course of many observing runs. Finally, we acknowledge the efforts of the POSS-II team, and the plate scanning team at STScI. PAAL was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), under processes 145973/99-9 & 200453/00-9. Several undergraduates participated in the data acquisition and processing towards the photometric calibration of DPOSS. PAAL would like to thank Massimo Ramella and Walter Boschin for the use of the VT code, as well for helpful discussions on the use of this code. This project made use of more than 15,000 hours of CPU time in the computers of the Center for Advance Computing Research (CACR), located at Caltech. We are thankful to Roy Williams and Mark Bartelt for help on using the CACR facilities. This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Capítulo 5

Conclusões e Perspectivas Futuras

Resumo

Esta tese apresenta (capítulo 3) um dos maiores catálogos de candidatos a aglomerados de galáxias no universo local ($< z > \sim 0.15$). No capítulo 4 apresentamos um projeto para estender este catálogo até $z \sim 0.5$, sendo $< z > \sim 0.3$. Como subproduto apresentamos uma comparação entre DPOSS e SDSS, utilizada para estimarmos os limites do DPOSS (capítulo 2). Esta comparação ainda aponta para a questão de multiplicidade, que pretendemos estudar no futuro.

Como mencionado na introdução os catálogos aqui apresentados tem várias utilizações. Alguns dos estudos que planejamos desenvolver são brevemente descritos abaixo. Estes estudos fazem uso de dados de placa e/ou dados CCD (provenientes do levantamento PACOS), além de dados Raios-X provenientes da literatura. Alguns outros estudos, como a função de massa de aglomerados, dependem da aquisição de dados espectroscópicos para determinação da dispersão de velocidade dos aglomerados e por conseguinte a massa dos mesmos.

5.1 Subestrutura, Efeitos de Alinhamento e a Função de Luminosidade

O processo de formação de aglomerados de galáxias sugere que sub-aglomerações são desfeitas quando o sistema torna-se virializado. Desta forma, a presença de subestrutura em aglomerados de galáxias pode fornecer uma importante indicação do estágio evolutivo destas

estruturas (West & Bothun 1990; West *et al.* 1995; Solanes *et al.* 1999). A existência de subestrutura pode também causar uma super-estimativa da massa de aglomerados (Girardi *et al.* 1997; Pinkney *et al.* 1996; Bird 1995). Outro importante efeito que pode ser o resultado de um modelo de formação hierárquica é o alinhamento de galáxias, aglomerados e a estrutura em grande escala. Estudos recentes tem indicado que a presença de subestrutura pode estar relacionada a efeitos de alinhamento (West *et al.* 1995; Tormen 1997; Plionis & Basilakos 2002). Aglomerados seriam formados através do acúmulo de material ao longo de estruturas filamentares, com aglomerados jovens exibindo alto grau de subestrutura, alinhada com a estrutura em grande escala. O estágio evolutivo de aglomerados pode também influenciar a população de galáxias característica dos mesmos (Durret *et al.* 2002; Baldi *et al.* 2001). Aglomerados relaxados apresentam uma maior concentração de galáxias na região central, sendo que estas galáxias são resultado de um maior número de processos de fusão quando comparados com sistemas dinamicamente jovens (Bingelli *et al.* 1987; Sodré *et al.* 1989; Biviano *et al.* 1992; Stein 1997).

Recentemente, iniciamos um programa para estimar o grau de subestrutura de aglomerados a partir de sua distribuição projetada (Lopes *et al.* 2000). Para tal dispomos de dados de placa para centenas de aglomerados próximos ($0.1 < z < 0.16$) e dados CCD para aglomerados com desvio para o vermelho maior que 0.16 (cerca de 200 aglomerados do PACOS). Tanto os dados de placa como os dados CCD permitem-nos selecionar galáxias dentro de 1.0 h^{-1} Mpc e entre $M^* - 1$ e $M^* + 2$, nos respectivos intervalos de desvio para o vermelho mencionados acima. Desta forma, podemos estudar subestrutura num mesmo intervalo de magnitude absoluta e num mesmo raio para um grande número de aglomerados. Um estudo mais completo teria necessariamente de incluir mapas raios-X e dados espectroscópicos. No entanto, a análise da distribuição bi-dimensional de galáxias constitui um método simples para estimarmos a presença de subestrutura numa grande amostra de aglomerados (West & Bothun 1990; Pinkney *et al.* 1996; Kriessler & Beers 1997). Estas estimativas podem em seguida ser utilizadas como base para levantamentos espectroscópicos. O principal problema de estudos bi-dimensionais é contaminação por galáxias de fundo. Este efeito é minimizado pela seleção de galáxias com cores e magnitudes típicas de uma população de aglomerados.

Na Figura 5.1 exibimos mapas de densidade de galáxias na região de dois aglomerados de galáxias (Abell 1062 e Abell 1445). Para cada aglomerado, apresentamos 3 mapas de

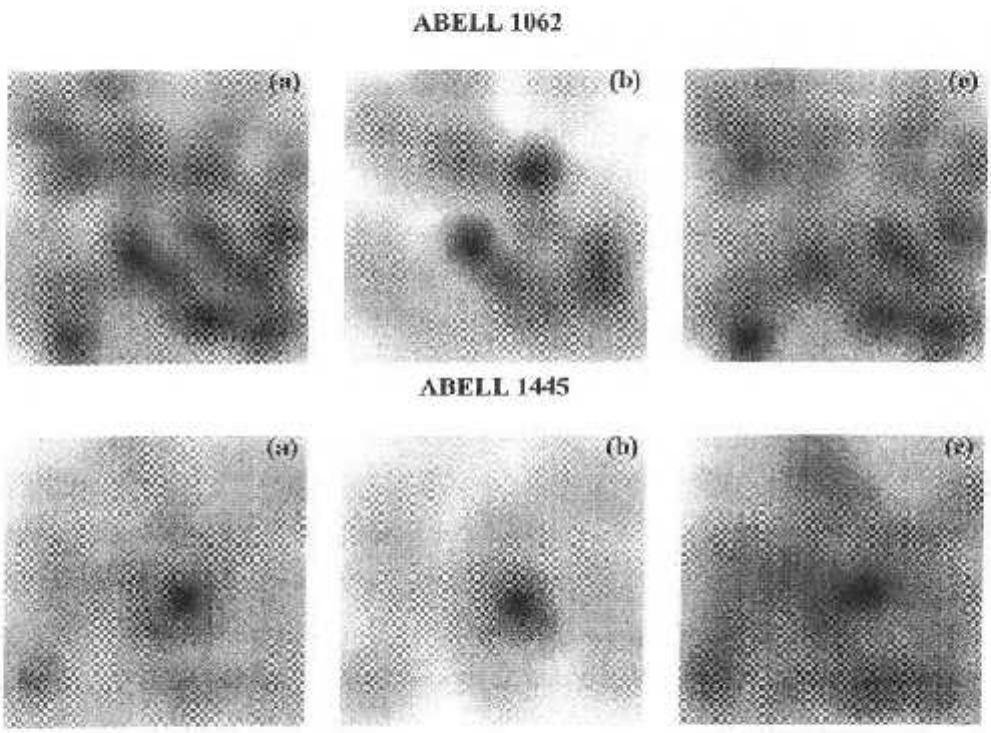


Figura 5.1: Mapas de densidade para Abell 1062 e Abell 1445. Para cada aglomerado cada mapa representa: (a) todas as galáxias na região do aglomerado; (b) galáxias com cores e magnitudes típicas de galáxias elípticas; e (c) galáxias com cores típicas de galáxias de campo. A seleção das galáxias exibidas em (b) e (c) é baseada em duas relações cor-magnitude, $(g-r)\times r$ e $(r-i)\times r$.

densidade para: (a) todas as galáxias na região do aglomerado; (b) galáxias com cores e magnitudes típicas de galáxias elípticas; e (c) galáxias com cores típicas de galáxias de campo. A seleção das galáxias exibidas em (b) e (c) é baseada em duas relações cor-magnitude, $(g-r)\times r$ e $(r-i)\times r$. Dois aspectos são facilmente distinguíveis desta figura. Primeiramente, fica claro o quão poderoso é o emprego da relação cor-magnitude para seleção de galáxias com cores típicas de galáxias de aglomerados (galáxias mais vermelhas, em geral elípticas e S0s). Podemos notar que o contraste da distribuição de galáxias vermelhas é fortemente realçado após a exclusão de galáxias de fundo (compare os painéis (a) e (b)). O segundo aspecto que notamos é a distinção na localização de galáxias vermelhas e azuis. A comparação dos painéis (b) e (c), para cada aglomerado, sugere que galáxias vermelhas estão mais próximas da região central do que galáxias azuis, mais dispersas na região do algomerado.

Após a criação dos mapas de densidade é necessário que estimemos a significância dos picos de densidade encontrados nos mapas. Em outras palavras, temos de testar se a subestrutura indicada para parte da amostra é significante. Um grande número de trabalhos apresenta diferentes testes empregados para quantificar subestrutura (Pinkney *et al.* 1996; Kriessler & Beers 1997; Salvador-Solé *et al.* 1993; Ashman *et al.* 1994; Rhee *et al.* 1991). Pretendemos testar pelo menos dois métodos, o algoritmo KMM (Ashman *et al.* 1994) e o programa DEDICA (Pisani 1996).

Em seguida ao estudo de subestrutura, pretendemos estimar a função de luminosidade e investigar efeitos de alinhamento nestes aglomerados. Paolillo *et al.* (2001) derivaram a função de luminosidade para 39 aglomerados de Abell usando dados do DPOSS. Pretendemos estender este estudo para os mesmos aglomerados para os quais estimarmos subestrutura. O objetivo final é verificar a existência de possíveis correlações entre estes resultados. A população de galáxias e por conseguinte a função de luminosidade depende do estado evolutivo de aglomerados? O aumento da fração de galáxias azuis com desvio para o vermelho (Margoniner *et al.* 2001) correlaciona-se com o grau de subestrutura; aglomerados em $z \sim 0.3$ são dinamicamente diferentes de aglomerados em $z \sim 0.1$? A relação entre subestrutura e a estrutura em grande escala detectada por Plionis & Basilakos (2002) com base em aglomerados do catálogo APM é real?

5.2 Comparação de Aglomerados de Galáxias Selecionados no Óptico e em Raios-X

Como mencionado na seção 1.2.1, aglomerados de galáxias de alta massa além de serem caracterizados por um grande número de galáxias, também possuem emissão em raios-X. Aglomerados ricos apresentam mais forte emissão, no entanto, grande número de aglomerados pobres (detectados no óptico) são também detectados em raios-X (Briel & Henry 1993; Vikhlinin *et al.* 1998; Scharf 1997; de Grandi 1999). Nesta seção apresentamos resultados preliminares de uma comparação do catálogo suplementar do NoSOCS (capítulo 4) com três catálogos de aglomerados detectados em raios-X. No futuro, pretendemos comparar propriedades ópticas de aglomerados (riqueza e dispersão de velocidade) com propriedades em raios-X (luminosidade e temperatura). Estes estudos podem ser úteis na determinação

de efeitos sistemáticos na determinação da massa de aglomerados (Yee & Ellingson 2003).

Os três catálogos com aglomerados selecionados em raios-X são o *Northern ROSAT All-Sky Galaxy Cluster Survey* (NORAS) (Böhringer *et al.* 2000), o *ROSAT Brightest Cluster Sample* (BCS) (Ebeling *et al.* 1998) e o catálogo apresentado por Vikhlinin *et al.* (1998), contendo 223 aglomerados detectados em campos do levantamento ROSAT PSPC. Os dois primeiros catálogos abrangem o universo local, contendo algomerados até $z \sim 0.3$, ao passo que o catálogo de Vikhlinin *et al.* (1998) contém aglomerados até $z \sim 0.7$. No entanto, o processo de seleção de Vikhlinin *et al.* (1998) não visa a construção de uma amostra completa.

Selecionamos todos os aglomerados de cada um destes catálogos na região de ~ 2700 graus quadrados do catálogo apresentado no capítulo 4. Ainda excluímos aglomerados raios-X que estejam dentro de uma área excluída (devido a um objeto brilhante, estrela ou galáxia) de nosso levantamento. A comparação dos 9982 candidatos do catálogo suplementar do NoSOCS é feita em relação a 70 aglomerados do NORAS, 43 aglomerados do BCS e 39 aglomerados de Vikhlinin *et al.* (1998). Na Figura 5.2 apresentamos a distribuição de separações (em Mpc) entre aglomerados raios-X e do NoSOCS. A comparação com o NORAS é apresentada como linha contínua, com o BCS como linha pontilhada e com o catálogo de Vikhlinin *et al.* (1998) como linha tracejada. É interessante notar que, em geral, a separação entre os centros óptico e raios-X é pequena (< 0.4 Mpc). Somente 2 aglomerados do NORAS (3%) não apresentam contrapartida no NoSOCS, ao passo que todos os 43 aglomerados do BCS são detectados no NoSOCS, e 13 (33%) dos objetos de Vikhlinin *et al.* (1998) não são encontrados no DPOSS.

As distribuições de desvios para o vermelho dos aglomerados detectados no óptico e em raios-X são apresentadas como linhas contínuas na Figura 5.3 (para os três casos, NORAS, BCS e Vikhlinin *et al.* 1998), enquanto que as distibuições de aglomerados não encontrados no DPOSS são exibidas como linhas pontilhadas (para os três catálogos em raios-X). Os aglomerados detectados somente por Vikhlinin *et al.* (1998) não apresentam nenhuma concentração particular em desvio para o vermelho. Na Figura 5.4 é apresentada uma comparação dos desvios para o vermelho dados nos catálogos raios-X e as estimativas do NoSOCS (capítulo 4).

Na Figura 5.5 são apresentadas as distribuições de riqueza (N_{gals}, capítulo 4) dos aglomerados raios-X encontrados no DPOSS (linha contínua), e detectados somente em raios-X (linha pontilhada). Para obter estas estimativas utilizamos as coordenadas e desvios para o

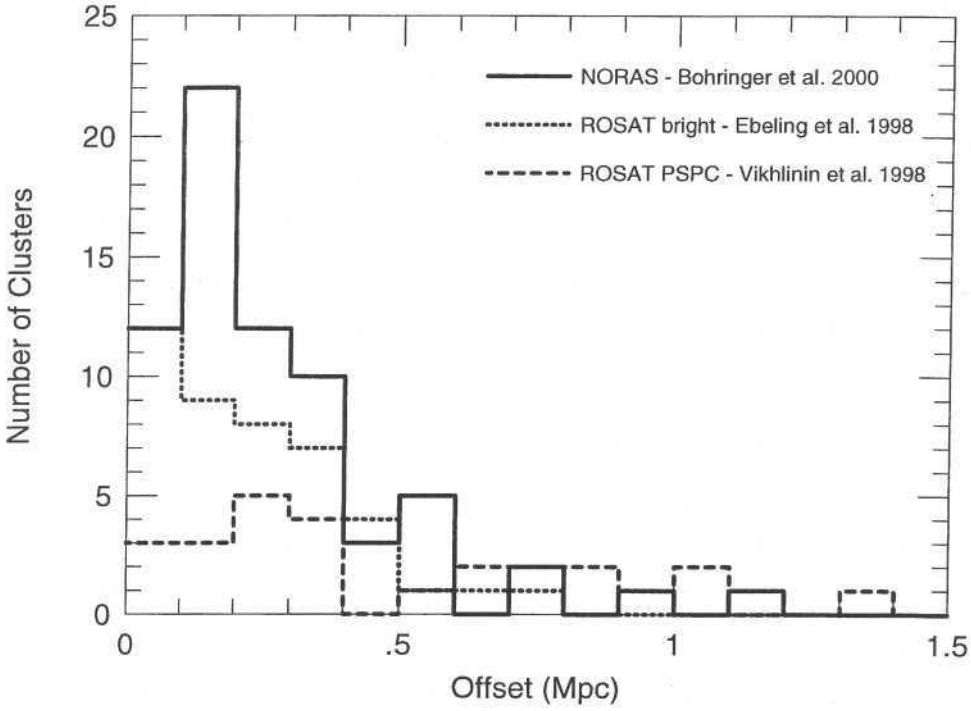


Figura 5.2: Distribuição de separações (em Mpc) entre aglomerados detectados em raios-X e pelo NoSOCS. A comparação com o NORAS é apresentada como linha contínua, com o BCS como linha pontilhada e com o catálogo de Vikhlinin *et al.* (1998) como linha tracejada.

vermelho dados em cada um dos 3 catálogos raios-X. Simplesmente apontamos nosso estimador de riqueza para estas coordenadas, contando galáxias entre $M^* - 1$ e $M^* + 2$ e em 1.0 h^{-1} Mpc no desvio para o vermelho do aglomerado. Como mencionado na seção 4.7 a estimativa de riqueza é extremamente sensível as determinações do centróide e do desvio para o vermelho do aglomerado. Tendo isto em mente, também apresentamos a título de comparação as estimativas de riqueza do NoSOCS (adotando o centróide óptico e as estimativas de z do capítulo 4). Portanto, para os aglomerados raios-X encontrados no DPOSS (linhas contínuas) as distribuições das estimativas de riqueza do NoSOCS são também exibidas como linhas pontilhada-tracejadas na Figura 5.5. As distribuições de riqueza do NoSOCS representam mais aglomerados ricos do que quando estimando riqueza com as coordenadas e z fornecidos pelos catálogos raios-X.

Finalmente, é importante mencionar que os dois aglomerados do NORAS não encontrados

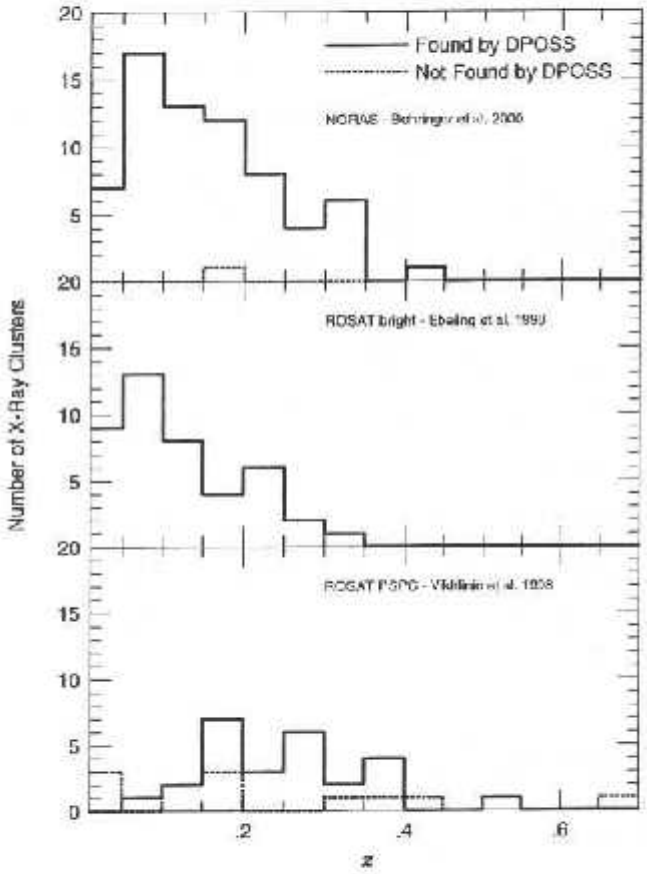


Figura 5.3: Distribuições de desvios para o vermelho de aglomerados detectados em raios-X. As distribuições de aglomerados com contrapartida no DPOSS são exibidas como linhas contínuas, ao passo que as distribuições de aglomerados não encontrados no DPOSS são exibidas como linhas pontilhadas. Resultados do NORAS são exibidos no painel superior, do BCS no painel intermediário e de Vikhlinin *et al.* (1998) no painel inferior.

no DPOSS encontram-se num regime de riqueza ($N_{gals} \sim 50$) em que não esperamos 100% de completeza em nenhum valor de desvio para o vermelho. Um dos aglomerados tem $N_{gals} = 55$ em $z \sim 0.42$, ao passo que o outro tem $N_{gals} = 50$ em $z \sim 0.18$. No que diz respeitos aos 13 aglomerados de Vikhlinin *et al.* (1998) sem contrapartida no óptico, todos estes aglomerados possuem $N_{gals} < 48$. Três aglomerados não tem estimativa de desvio para o vermelho em Vikhlinin *et al.* (1998), cinco tem $z > 0.3$ e os outros cinco tem $0.14 < z < 0.18$, com $N_{gals} < 38$.

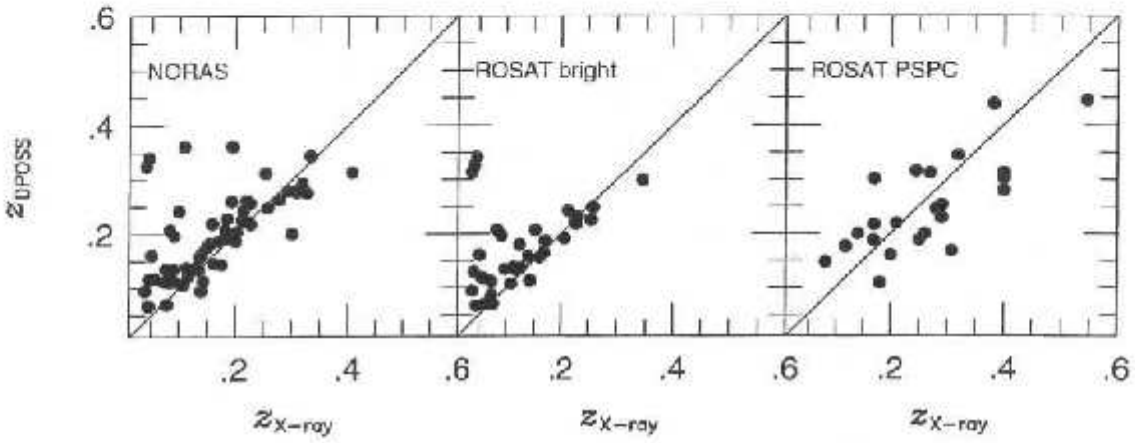


Figura 5.4: Comparação entre os desvios para o vermelho de aglomerados detectados em raios-X com as estimativas do DPOSS. A comparação com o NORAS é exibida no painel esquerdo, com o BCS no painel intermediário e com Vikhlinin *et al.* (1998) no painel direito.

5.3 Correlações entre Quasares e Aglomerados de Galáxias

Na últimas duas décadas um grande número de estudos tem sugerido que a atividade de quasares está fortemente correlacionada ao ambiente da galáxia hospedeira (Stockton 1982; Yee & Green 1987; Ellingson *et al.* 1991; Yee & Ellingson 1993; Fisher *et al.* 1996; Smith *et al.* 2000). Estes estudos sugerem que quasares e galáxias rádio estão localizados em ambientes densos, típicos de grupos e aglomerados de galáxias. No entanto, existem controvérsias na diferença do ambiente de quasares com emissão rádio significante ou não (Fisher *et al.* 1996; Croom & Shanks 1999). De qualquer forma, o estudo do ambiente típico de quasares pode contribuir siginificativamente para o entendimento de efeitos evolutivos na atividade destes objetos. Por outro lado, se esta relação for válida em mais altos desvios para o vermelho ($z > 1$) quasares podem ser usados para seleção de regiões densas, possivelmente associadas com aglomerados ou proto-aglomerados de galáxias.

Outro efeito de interesse é a determinação de possíveis correlações entre quasares distantes ($z > 1$) e galáxias e aglomerados no universo local ($z < 0.5$). Alguns autores (Arp *et al.* 1990; Burbidge *et al.* 1990) sugerem que estas associações projetadas entre quasares distantes

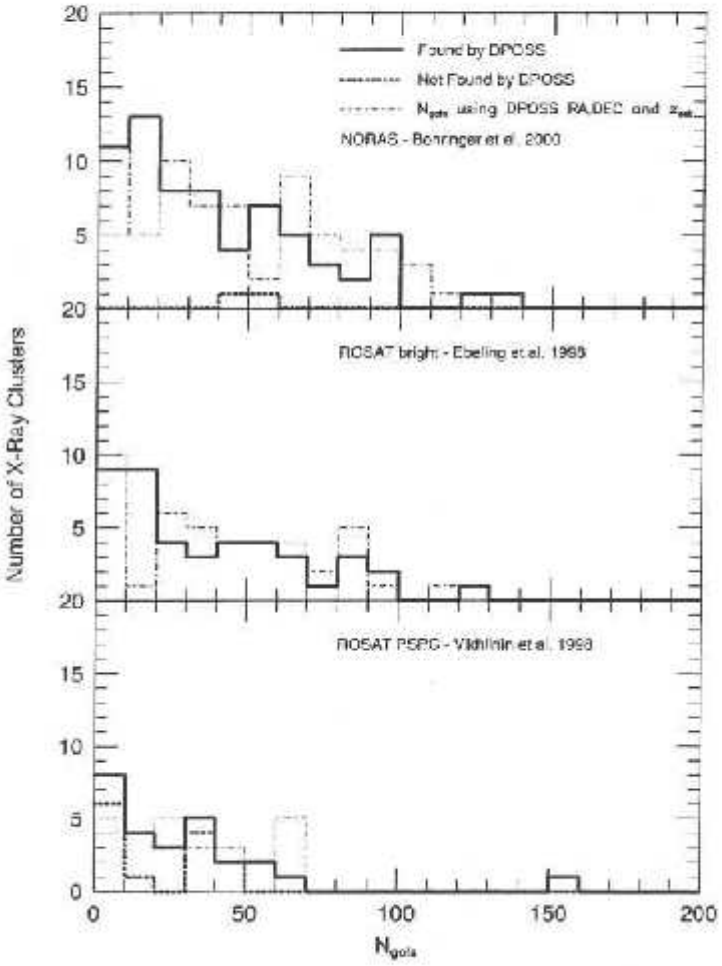


Figura 5.5: Distribuições de riqueza (N_{gals}) dos aglomerados raios-X. As distribuições de aglomerados encontrados no DPOSS são exibidas como linhas contínuas, ao passo que as distribuições de aglomerados sem contrapartida no DPOSS são exibidas como linhas pontilhadas. As distribuições das estimativas de riqueza do NoSOCS são exibidas como linhas pontilhada-tracejada. Resultados do NORAS são exibidos no painel superior, do BCS no painel intermediário e de Vikhlinin *et al.* (1998) no painel inferior.

e galáxias próximas seriam explicadas por desvios para o vermelho não-cosmológicos dos quasares. Uma alternativa mais simples sugere que as galáxias com pequeno desvio para o vermelho podem amplificar gravitacionalmente o fluxo de objetos de fundo (quasares) próximos na linha de visada (Webster & Hewett 1990; Schneider 1992; Norman & Impey 1999). No entanto, este efeito de magnificação é pequeno para justificar a amplitude da

correlação entre quasares e galáxias próximas. Além disso, a amplitude destas associações também depende das amostras utilizadas (Benítez *et al.* 2001). Em resumo, as contradições encontradas nos estudos recentes indicam que ainda estamos longe de um bom entendimento da natureza destas associações.

Considerando que os dados do DPOSS cobrem todo o hemisfério norte, podemos abordar as duas questões mencionadas acima (o ambiente de quasares e associações entre quasares distantes e galáxias no universo local) utilizando uma grande amostra de quasares. Num estudo preliminar, utilizamos a lista de quasares compilada por Véron-Cetty & Véron (2001). Esta lista contém 23760 quasares distribuídos em todo o céu. Selecionamos 2665 quasares na região de $\sim$ 2700 graus quadrados utilizada para detecção de aglomerados no capítulo 4. Destes 2665 quasares, 300 tem desvio para o vermelho $z < 0.5$. Numa tentativa de estimar se os quasares são encontrados em aglomerados de galáxias, comparamos a posição destes 300 quasares com os 9982 candidatos a aglomerado do catálogo suplementar do NoSOCS (capítulo 4). No painel superior da Figura 5.6 exibimos com a linha sólida a distribuição de separações de pares de quasares e aglomerados (77 do total de 300 possíveis). Nosso raio de busca é de 2 h^{-1} Mpc, adotando o desvio para o vermelho dos quasares (Véron-Cetty & Véron 2001). Obviamente, estes pares podem ser o resultado de efeitos de superposição. O histograma sombreado representa a distribuição do subconjunto de pares com diferenças no desvio para o vermelho ($\Delta z = z_{quasar} - z_{est-aglom}$) ≤ 0.1 (adotamos este valor levando em consideração a incerteza na estimativa de z para os aglomerados, como visto no capítulo 4). Se estes pares representarem associações reais, isto implica que os quasares associados a aglomerados, são preferencialmente encontrados na periferia de aglomerados. A fração de pares é pequena, o que não é uma surpresa se considerarmos que quasares, em geral, estão associados a pequenos excessos de densidade (grupos ou aglomerados pobres). O catálogo suplementar do NoSOCS (como quase todos os catálogos de aglomerados de galáxias) é exteremamente incompleto no regime de aglomerados pobres. Portanto, a probabilidade de encontrarmos quasares associados com candidatos a aglomerados pobres (do nosso catálogo) é pequena.

No painel inferior da Figura 5.6 exibimos (linha pontilhada) a distribuição de riqueza (N_{gals}) dos 77 aglomerados que são possíveis pares aos quasares de Véron-Cetty & Véron (2001). A distribuição é aproximadamente constante, apresentando um excesso de aglom-

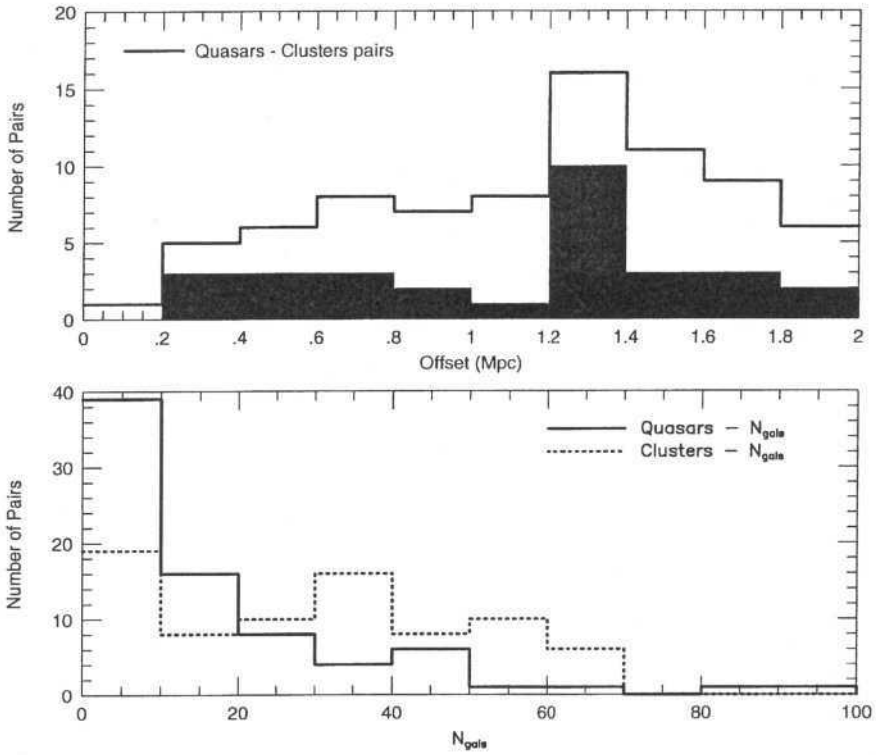


Figura 5.6: Distribuições de separações (em Mpc) entre quasares e aglomerados (painel superior). A linha sólida representa a distribuição de 77 pares, ao passo que o histograma sombreado representa os 30 pares para os quais $\Delta z \leq 0.1$. No painel inferior exibimos a distribuição de riqueza (N_{gals}) dos 77 aglomerados possivelmente associados a quasares (linha pontilhada). A distribuição das estimativas de riqueza do ambiente destes 77 quasares é exibida como linha sólida (para estas estimativas adotamos a posição e desvios para o vermelho dos quasares).

erados pobres ($N_{gals} < 10$) e um corte em $N_{gals} = 70$. Uma outra forma de estudar o ambiente de quasares é através do cálculo da função de correlação cruzada entre quasares e galáxias ou aglomerados (Romani & Maoz 1992; Wold *et al.* 2000). Também podemos estimar o excesso de galáxias na região do quasar, ou em outras palavras, estimar a riqueza típica destas regiões. No painel inferior da Figura 5.6 a linha sólida representa a distribuição da estimativa de N_{gals} adotando a posição e desvios para o vermelho dos quasares. É interessante notar que o ambiente de quasares é em geral pobre ($N_{gals} < 20$, o que não seria surpresa se

os quasares encontrarem-se na periferia de aglomerados). Estes resultados estão em acordo com estimativas anteriores (Yee & Green 1987; Smith *et al.* 2000; Wold *et al.* 2000). Alguns estudos ainda sugerem que quasares mais distantes são encontrados em ambientes mais ricos do que quasares mais próximos (Yee & Green 1987; Yates *et al.* 1989; Ellingson *et al.* 1991; Wold *et al.* 2000). Quando dividimos a amostra com 77 quasares exibida na Figura 5.6, temos que dos 22 quasares com $N_{gals} > 20$, 19 estão em $z > 0.25$.

Pretendemos investigar esta questão em mais detalhe no futuro. Dividiremos a amostra entre aglomerados com pequeno ou grande fluxo na faixa rádio, para verificar se o ambiente de quasares com grande emissão rádio é diferente dos com pouca emissão. Investigaremos também as associações projetadas entre quasares distantes $z > 1$ e galáxias próximas ($z < 0.5$).

5.4 As Funções de Correlação e Massa de Aglomerados

Nas seções 1.2.3 e 1.2.4 discutimos a importância das estimativas da função de massa e da função de correlação de aglomerados para o entendimento da formação e da distribuição de matéria em grandes escalas. A função de massa já tem sido estudada com base nos aglomerados do NoSOCS (Gal & de Carvalho 2002). Um bom conhecimento dos efeitos de seleção do catálogo, em termos de contaminação e completeza (como função da riqueza e desvio para o vermelho) é fundamental para estudos deste tipo. Obviamente, a disponibilidade de estimativas de desvio para o vermelho e riqueza também são imprescindíveis. Estas estimativas, a taxa de contaminação e a função de seleção já são conhecidas para o NoSOCS. No entanto, a aquisição de mais dados espectroscópicos é necessária para uma melhor determinação de massa de uma grande amostra dos aglomerados.

No que diz respeito a função de correlação, dividiremos nossa amostra em diferentes classes de riqueza para verificar a dependência da amplitude da correlação com a riqueza. Também pretendemos estudar a função de correlação angular de galáxias e investigar sua dependência com o limite de fluxo da amostra (Maddox *et al.* 1996). Na Figura 5.7 exibimos, como exemplo, a função de correlação angular para galáxias na região do campo 444 do DPOSS. Os resultados são exibidos em diferentes intervalos de magnitude. Apesar de preliminares, estes resultados apontam para a possibilidade de estimarmos a função de cor-

relação angular numa região significativa do céu, em duas bandas fotométricas e com melhor controle dos efeitos sistemáticos do que o APM.

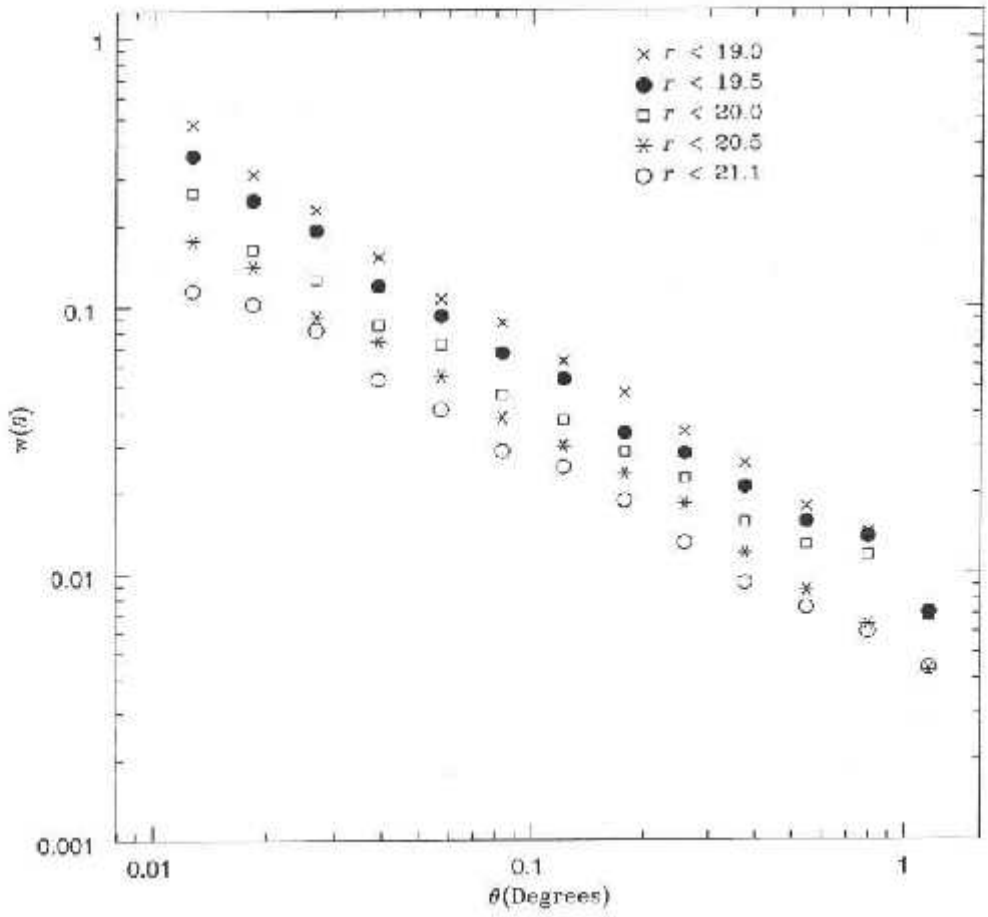


Figura 5.7: Função de correlação angular de galáxias na região do campo 444 do DPOSS. Os resultados são exibidos em diferentes intervalos de magnitude.

Apêndice A

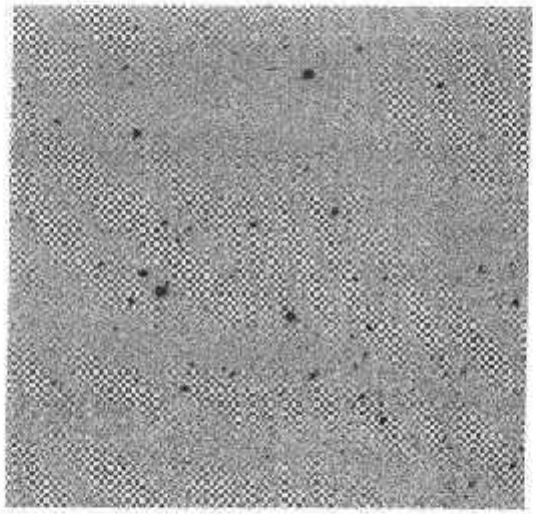
Imagens de Candidatos a Aglomerados de Galáxias

Neste apêndice apresentamos imagens[1] de 32 candidatos a aglomerados de galáxias detectados no DPOSS (capítulo 4). Devido a dificuldade de visualização destes candidatos nas placas do DPOSS, extraímos imagens do *Sloan Digital Sky Survey*, representando observações nas bandas g^*, r^* e i^*. Uma imagem composta destas três bandas é apresentada para cada candidato. Selecionamos sistemas com $N_{gals} > 50$ e $z_{est} > 0.25$, limitados por $-1.5° < \delta < 1.5°$. As imagens provenientes do SDSS[2] estão no formato *jpeg* e cobrem uma região de $300'' \times 300''$ para cada candidato a aglomerado.

[1]Estas imagens estão disponíveis em `http://www.astro.caltech.edu/~paal/thesis/colorimages/`

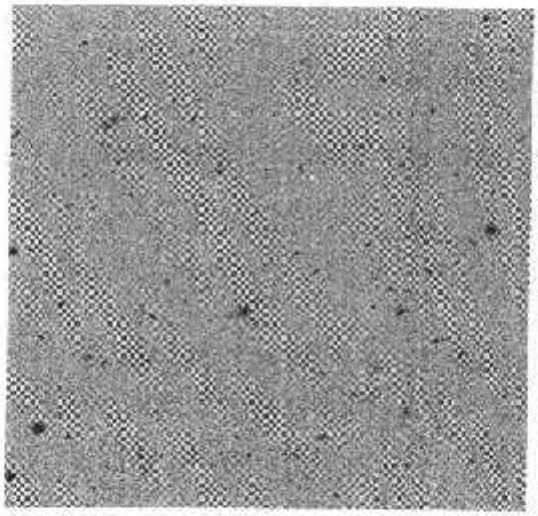
[2]Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.



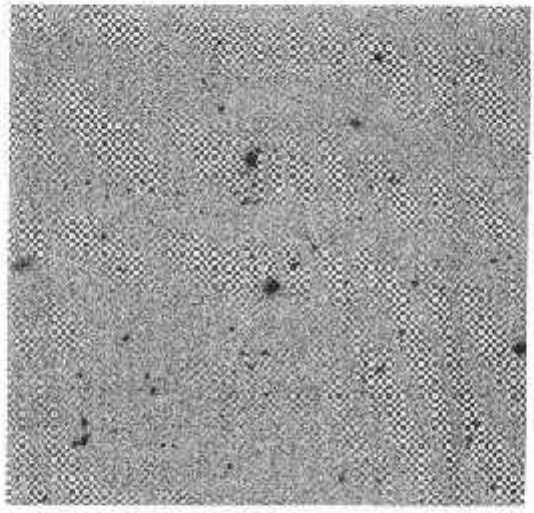
NSCS J002411-002711

$z_{est} = 0.34$ $N_{gals} = 65.0$



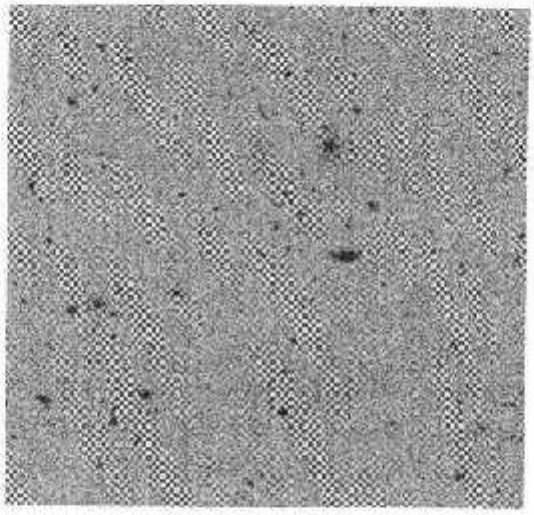
NSCS J002433+000026

$z_{est} = 0.34$ $N_{gals} = 83.5$



NSCS J002824-000220

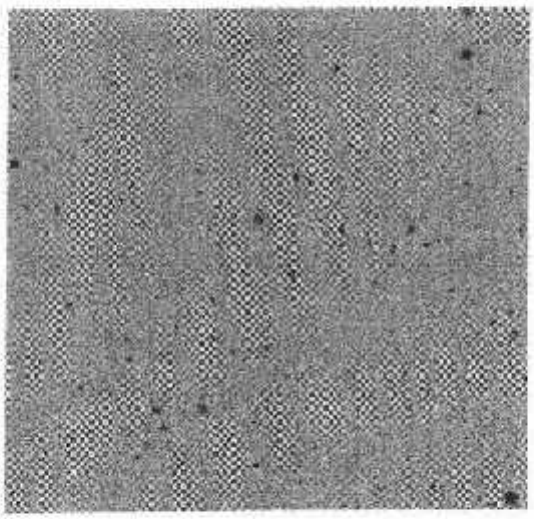
$z_{est} = 0.29$ $N_{gals} = 51.5$



NSCS J002051+000951

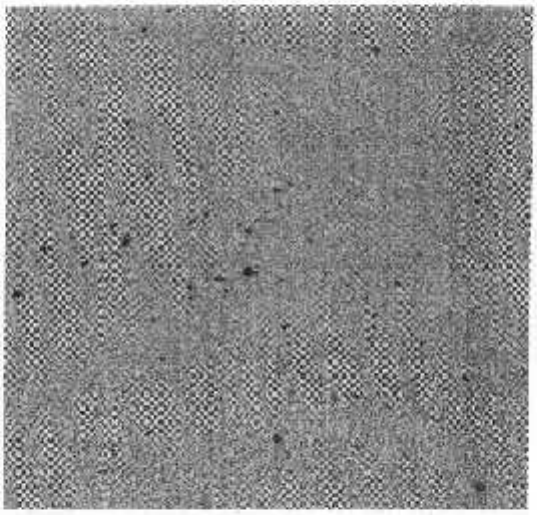
$z_{est} = 0.28$ $N_{gals} = 73.9$

Figura A.1: Imagens de candidatos a aglomerados de galáxias. Cada imagem cobre $300'' \times 300''$, sendo o resultado da composição de observações feitas nas bandas g^*, r^* e i^*.



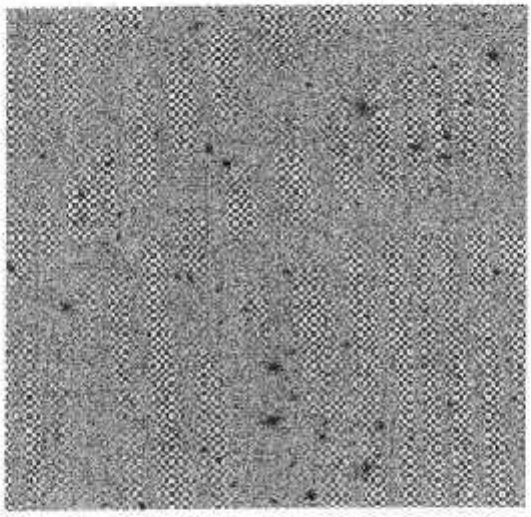
NSCS J002358+003137

$z_{est} = 0.34$ $N_{gals} = 61.9$



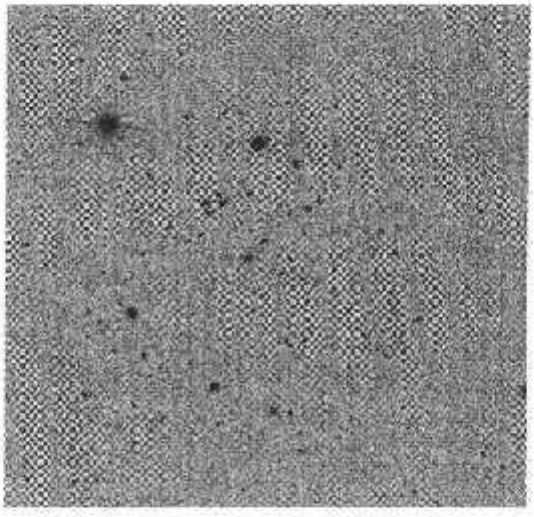
NSCS J002958+010527

$z_{est} = 0.39$ $N_{gals} = 67.8$



NSCS J003744+000349

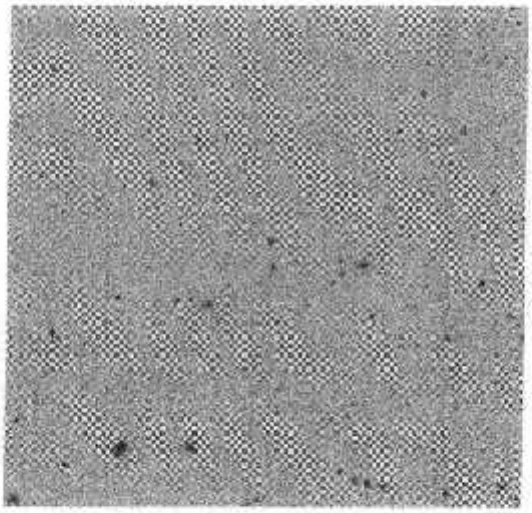
$z_{est} = 0.27$ $N_{gals} = 58.4$



NSCS J004249+001257

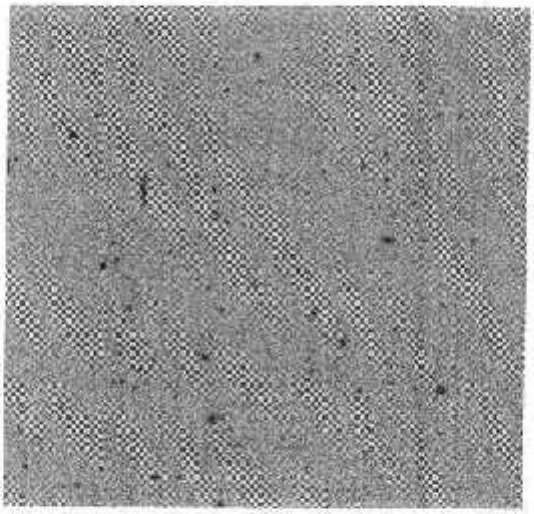
$z_{est} = 0.28$ $N_{gals} = 53.4$

Figura A.1: continuação.



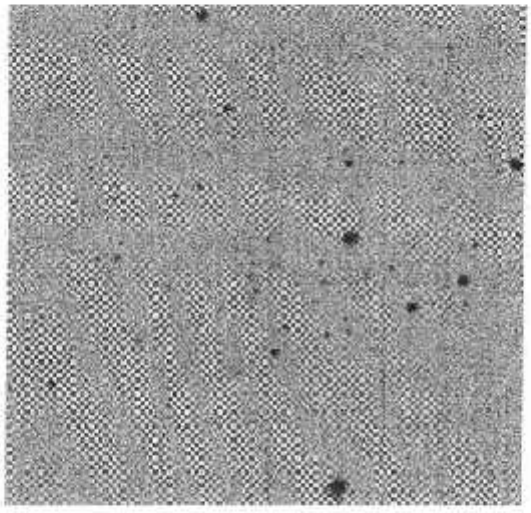
NSCS J010240-005003

$z_{est} = 0.27$ $N_{gals} = 66.2$



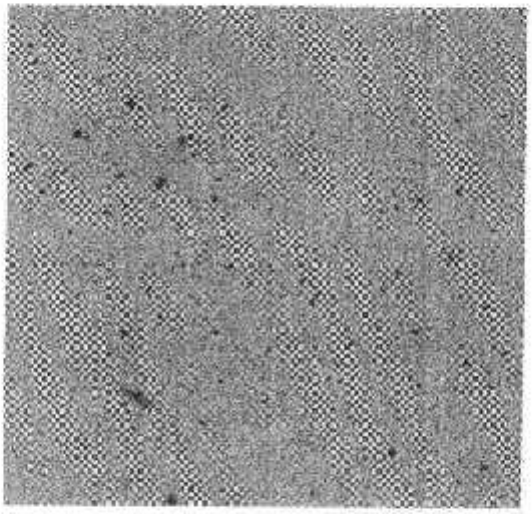
NSCS J010550-001927

$z_{est} = 0.28$ $N_{gals} = 58.0$



NSCS J010852-002827

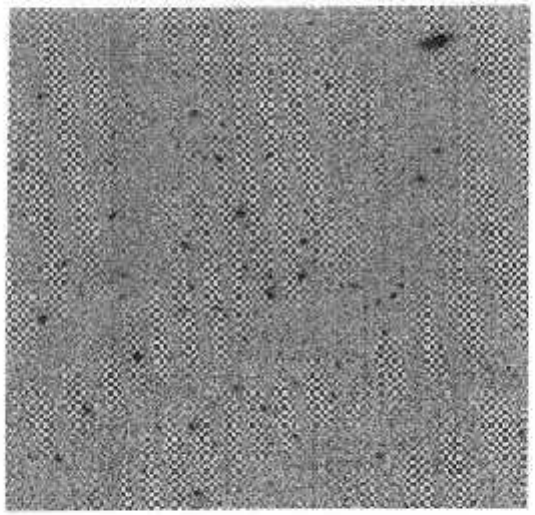
$z_{est} = 0.44$ $N_{gals} = 64.1$



NSCS J010451+000252

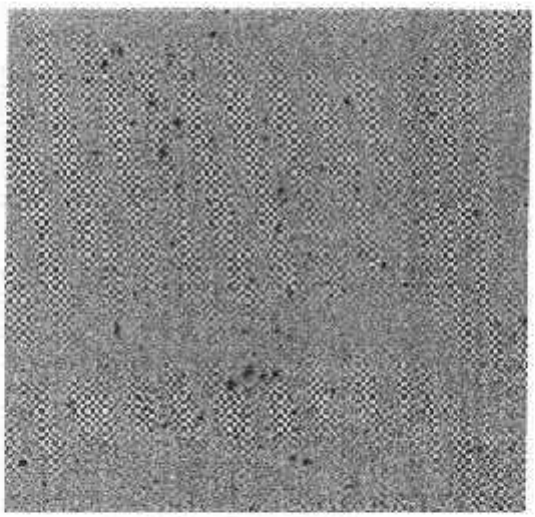
$z_{est} = 0.28$ $N_{gals} = 101.5$

Figura A.1: continuação.



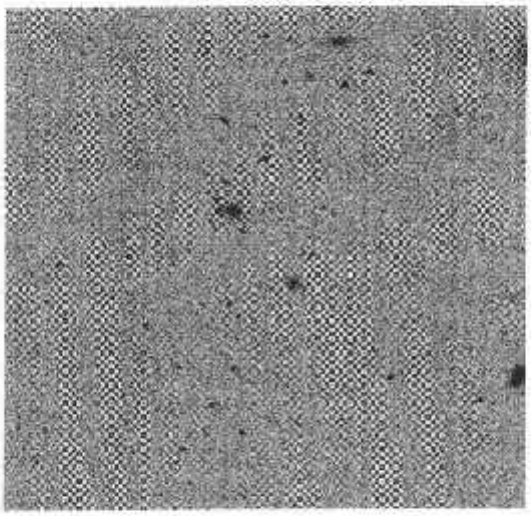
NSCS J010649+010344

$z_{est} = 0.31$ $N_{gals} = 71.5$



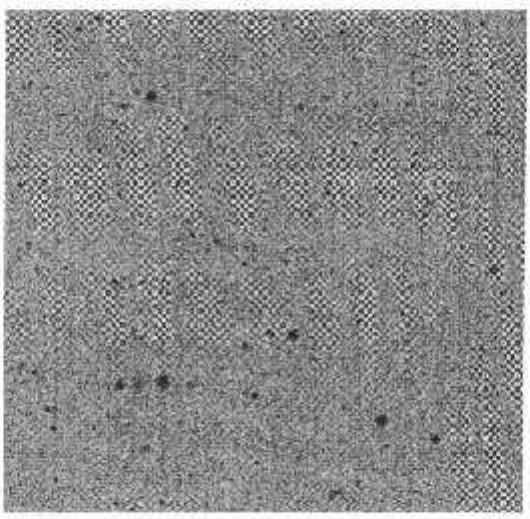
NSCS J010606+005016

$z_{est} = 0.29$ $N_{gals} = 75.3$



NSCS J011623-000828

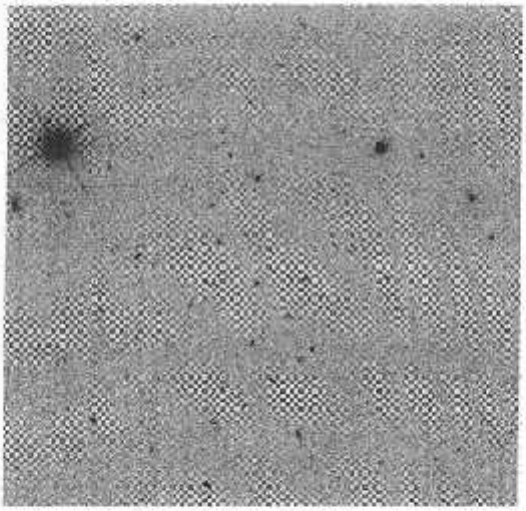
$z_{est} = 0.28$ $N_{gals} = 53.3$



NSCS J012712+002155

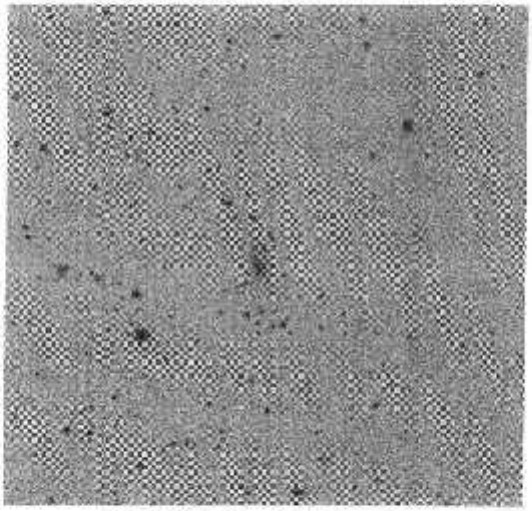
$z_{est} = 0.33$ $N_{gals} = 57.8$

Figura A.1: continuação.



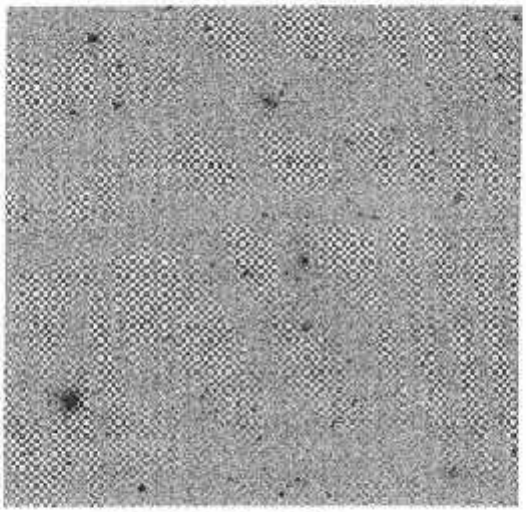
NSCS J020724-004825

$z_{est} = 0.37$ $N_{gals} = 62.6$



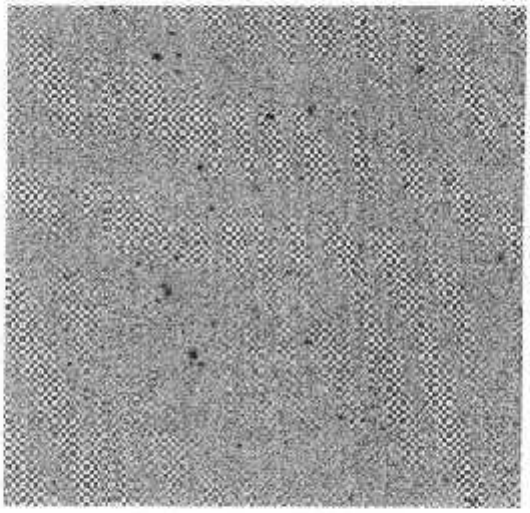
NSCS J015242+010032

$z_{est} = 0.26$ $N_{gals} = 106.0$



NSCS J021058+011049

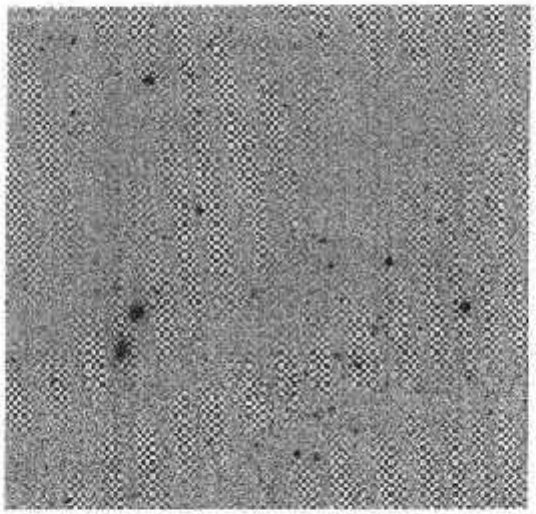
$z_{est} = 0.34$ $N_{gals} = 59.9$



NSCS J021838-002112

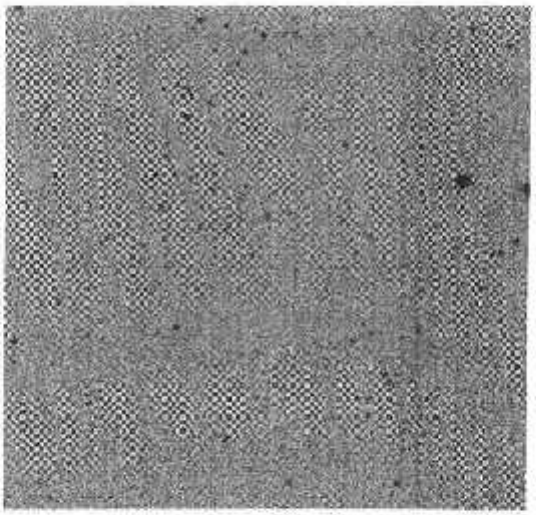
$z_{est} = 0.41$ $N_{gals} = 57.2$

Figura A.1: continuação.



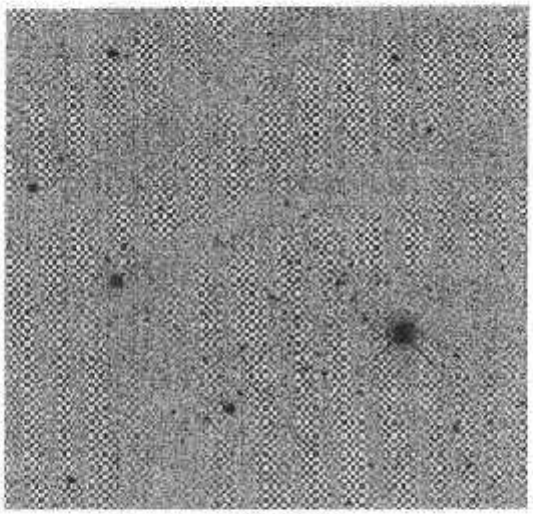
NSCS J023858-010539

$z_{est} = 0.31$ $N_{gals} = 70.1$



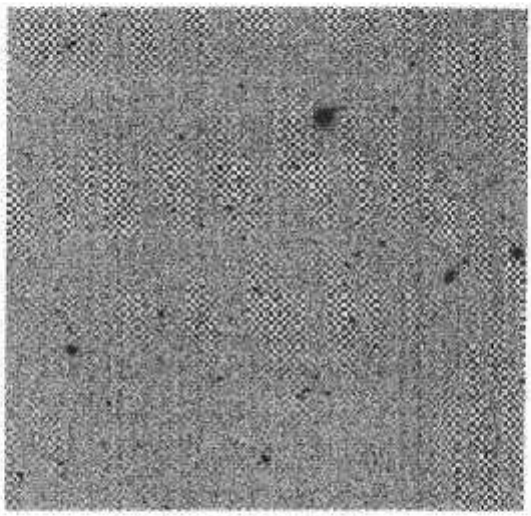
NSCS J024750-003044

$z_{est} = 0.30$ $N_{gals} = 52.9$



NSCS J122640+001608

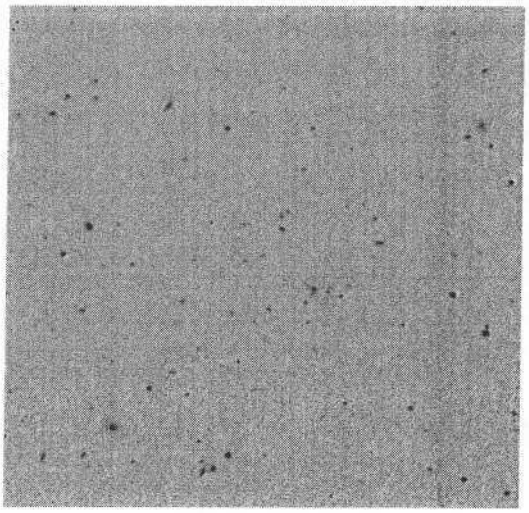
$z_{est} = 0.42$ $N_{gals} = 76.1$



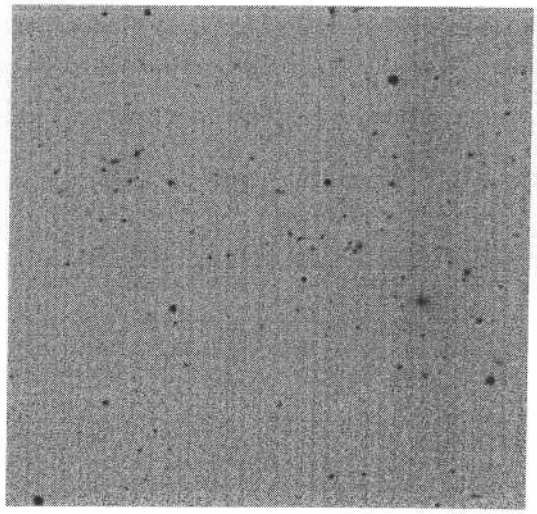
NSCS J231953-011214

$z_{est} = 0.32$ $N_{gals} = 65.6$

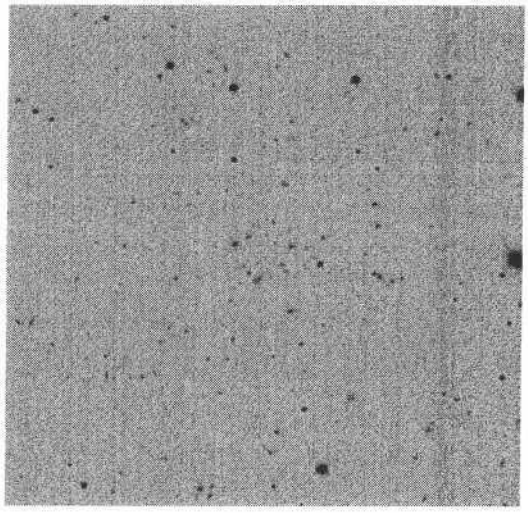
Figura A.1: continuação.



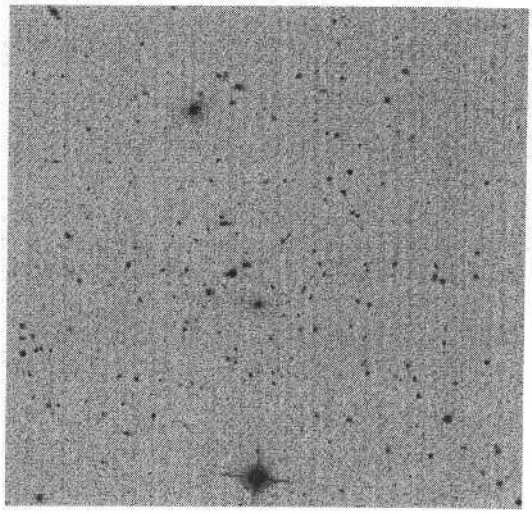
NSCS J231453+002826
$z_{est} = 0.29$ $N_{gals} = 51.6$



NSCS J232738+005700
$z_{est} = 0.32$ $N_{gals} = 76.1$

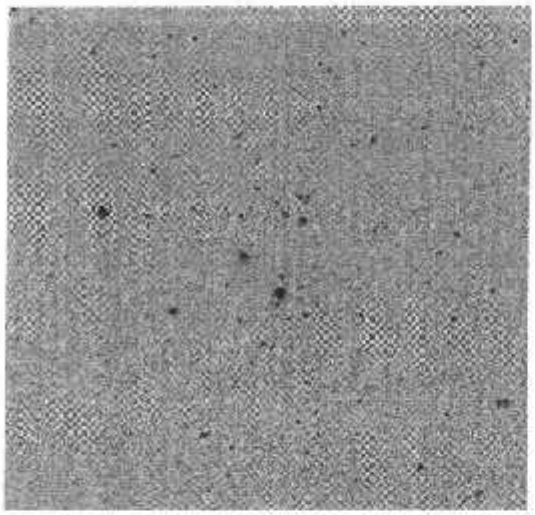


NSCS J231544+005306
$z_{est} = 0.33$ $N_{gals} = 87.8$

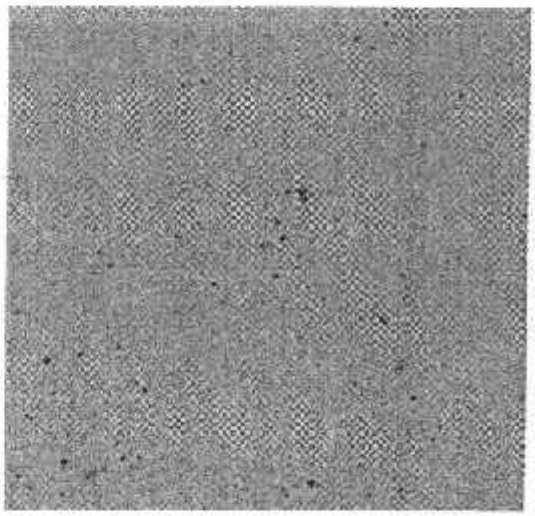


NSCS J233739+001647
$z_{est} = 0.26$ $N_{gals} = 93.0$

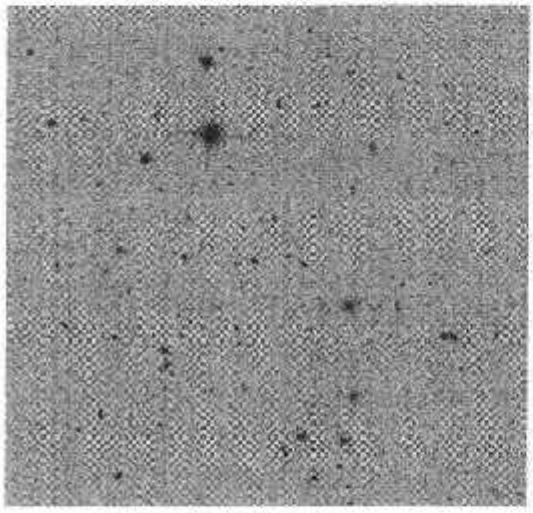
Figura A.1: continuação.



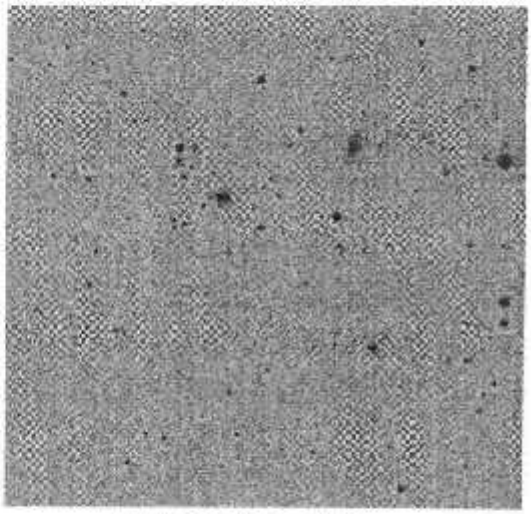
NSCS J235120+003659
$z_{est} = 0.35$ $N_{gals} = 58.1$



NSCS J234816+004627
$z_{est} = 0.38$ $N_{gals} = 50.0$



NSCS J234627+004529
$z_{est} = 0.25$ $N_{gals} = 59.4$



NSCS J234105+001759
$z_{est} = 0.36$ $N_{gals} = 63.1$

Figura A.1: continuação.

Bibliografia

Abell, G. O. 1957, Tese de Doutorado, Caltech

Abell, G. O. 1958, ApJS, 3, 211

Abell, G. O., Corwin, H. G. & Olowin, R. P. 1989, ApJS, 70, 1

Alpher, R. & Hermann, R. 1948, *Nature*, 162, 774

Annis, J. et al. 1999, *Bulletin of the American Astronomical Society* 195, 12.02

Aragón-Salamanca, A., Ellis, R., Couch, W., carter, D. 1993, MNRAS, 262, 764

Arp, H. C. *et al.* 1990, *Nature*, 346, 807

Ashman, K., Bird, C., Zepf, S. 1994, AJ, 108, 2348

Bahcall, N. A. & Soneira, R. M. 1983, ApJ, 270, 20

Bahcall, N. A. 1988, ARA&A, 26, 631

Bahcall, N. A. & Cen, R. 1992, 398, 81

Bahcall, N. A. & West, M. J. 1992, ApJ, 392, 419

Bahcall, N. A., Lubin, L. M., Dorman, V. 1995, ApJ, 447, 81

Bahcall, N. A., Fan, X., Cen, R. 1997, ApJ, 485, L53

Bahcall, N. A. & Fan, X. 1998, ApJ, 504, 1

Bahcall, N. A. 1999, *'Clusters and superclusters of galaxies'*, em *Formation of Structure in the Universe, Cambridge University Press*, p. 135

Bahcall, N. A., Cen, R., Dav, R., Ostriker, J. P., Yu, Q. 2000, ApJ, 541, 1

Bahcall, N. A. *et al.* 2003a, ApJ, 585, 182

Bahcall, N. A. *et al.* 2003b, *em preparação*

Baldi, A., Bardelli, S., Zucca, E. 2001, MNRAS, 324, 509

Benítez, N. *et al.* 2001, MNRAS, 320, 241

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393

Binggeli, B., Tammann, G. A., Sandage, A. 1987, AJ, 94, 251

Bird, C. 1993, PASP, 105, 1495

Bird, C. 1995, ApJ, 445, L81

Biviano, A. *et al.* 1992, ApJ, 396, 35

Blanchard, A. & Bartlett, J. G. 1998, A&A, 332, 49

Böhringer, H. *et al.* 2000, ApJS, 129, 435

Bramel, D. A., Nichol, R. C. & Pope, A. C. 2000, ApJ, 533, 601

Briel, U. G. & Henry, J. 1993, A&A, 278, 379

Brodbeck, D. *et al.* 1998, ApJ, 495, 1

Burbidge, G. *et al.* 1990, ApJS, 74, 675

Butcher, H. & Oemler, A. 1978, ApJ, 226, 559

Butcher, H. & Oemler, A. 1984, ApJ, 285, 426

Carlberg, R., Morris, S., Yee, H., Ellingson, E. 1997, ApJ, 479, L19

Colberg, J. M. *et al.* 2000, MNRAS, 319, 209

Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393

Colless, M. 1999, *'First results from the 2dF Galaxy Redshift Survey', Royal Society of London Philosophical Transactions Series A*, 357, 105

Collins, C. A.*et al.* 2000, MNRAS, 319, 939

Croft, R. & Efstathiou, G. 1994, MNRAS, 267, 390

Croft, R. *et al.* 1997, MNRAS, 291, 305

Croom S. & Shanks T. 1999, MNRAS, 307, L17

Dalton, G. *et al.* 1992, ApJ, 390, L1

Dalton, G. *et al.* 1994, MNRAS, 271, L47

Dalton, G. B., Maddox, S. J., Sutherland, W. J. & Efstathiou, G. 1997, MNRAS, 289, 263

David, L. P., Jones, C., Forman, W. 1995, ApJ, 445, 578

Davis, M., Summers, F., Schlegel, D. 1992, *Nature*, 359, 393

de Grandi *et al.* 1999, ApJ, 514, 148

de Sitter, W. 1917, *'On the relativity of inertia: remarks concerning Einstein's latest hypothesis' Proc. Koninkl. Akad. Wetensch. Amsterdam 19*, 1217

Djorgovski, S. G., Gal, R. R., de Carvalho, R. R., Odewahn, S. C., Mahabal, A., Brunner, R. J. & Lopes, P. A. A. 2003, AJ, *em preparação*

Donahue, M. 1998, ApJ, 502, 550

Doroshkevich, A., Gottlober, S., Madsen, S. 1997, A&AS, 123, 495

Dressler, A., Oemler, A., Couch, W., Smail, I., et al. 1997, ApJ, 490, 577

Durret, F., Adami, C., Lobo, C. 2002 A&A, 393, 439

Ebeling, H., Wiedenmann, G. 1993, Phys. Rev. E, 47, 704

Ebeling, H., *et al.* 1998, MNRAS, 301, 881

Ebeling, H., Edge, A. C., Allen, S. W., Crawford, C. S., Fabian, A. C., & Huchra, J. P. 2000, MNRAS, 318, 333

Efstathiou, G., Dalton, G. B., Sutherland, W. J. & Maddox, S. J. 1992, MNRAS, 257, 125

Einstein, A. 1915, *'Zur allgemeinen Relativitätstheorie', Preuß. Akad. Wiss. Berlin, Sitzber.*, 778

Einstein, A. 1917, *'Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie', Preuß. Akad. Wiss. Berlin, Sitzber.*, 142

Eke, V. *et al.* 1998, MNRAS, 298, 1145

El-Ad, H., Piran, T. da Costa, L. 1996, ApJ, 462, 13

Ellingson, E. *et al.* 1991, ApJS, 76, 455

Ellingson, E., Lin, H., Yee, H. K. C., Carlberg, R. G. 2001, ApJ, 547, 609

Ettori, S., Fabian, A. C. 1999, MNRAS, 305, 834

Evrard, A. 1989, ApJ, 341, L71

Fan, X., Bahcall, N., Cen, R. 1997, ApJ, 490, 123

Fisher K. *et al.* 1996, ApJ, 468, 469

Friedmann, A. 1922 & 1924, *'Über die Krummung des Resumes' Z. Phys. 10, 377 & Z. Phys. 21, 326*

Gal, R. R., de Carvalho, R. R., Odewahn, S. C., Djorgovski, S. G. & Margoniner, V. E. 2000a, AJ, 119, 12

Gal, R. R., de Carvalho, R. R., Brunner, R., Odewahn, S. C., Djorgovski, S. G. 2000b, AJ, 120, 540

Gal, R. R. 2001, Tese de Doutorado, Caltech

Gal, R. R. & de Carvalho, R. R. 2002, *Bulletin of the American Astronomical Society* 201, 42.09

Gal, R. R., de Carvalho, R. R., Lopes, P. A. A., Djorgovski, S. G., Brunner, R. J., Mahabal, A., Odewahn, S. C. 2003a, AJ, 125, 2064

Gal, R. R., de Carvalho, Lopes, P. A. A., Djorgovski, S. G., Brunner, R. J., Mahabal, A., Odewahn, S. C. 2003b, AJ, *em preparação*

Gal, R. R., de Carvalho, R. R., Odewahn, S. C., Djorgovski, S. G., Mahabal, A., Brunner, R. J., Lopes, P. A. A. 2003c, AJ, *accito para publicação*

Gamow, G. 1946, *Physical Review*, 70, 572

Gilbank, D. 2001, Tese de Doutorado, *University of Durham*

Girardi, M., Escalera, E., Fadda, D., Giuricin, G., Mardirossian, F., Mezzetti, M. 1997, 482, 41

Girardi, M., Manzato, P., Mezzetti, M., Giuricin, G., Limboz, F. 2002, ApJ, 569, 720

Gioia, I. M. & Luppino, G. A. 1994, APJS, 94, 583

Gladders, M. D. & Yee, H. K. C. 2000, AJ, 120, 2148

Gonzalez, A. H., Zaritsky, D., Dalcanton, J. J., & Nelson, A. 2001, ApJS, 137, 117

Gonzalez, A. H., Zaritsky, D., Wechsler, R. H. 2002, ApJ, 571, 129

Goto, T. et al. 2002, AJ, 123, 1807

Gunn, J. E., Hoessel, J. G. & Oke, J. B. 1986, ApJ, 306, 30

Guth, A. 1981, Physical Review D, 23, 347

Guzzo, L., Collins, C., Nichol, R., Lumsden, S. 1992, ApJ, 393, 5

Heydon-Dumbleton, N. H., Collins, C. A. & MacGillivray, H. T. 1989, MNRAS, 238, 379

Hoekstra, H., Franx, M., Kuijken, K., van Dokkum, P. 2002, MNRAS, 333, 911

Hφg, E., Fabricius, C., Makarov, V., Urban, S., Corbin, T., Wycoff, G., Bastian, U., Schwekendiek, P., Wicenec, A. 2000, A&A, 355, 27

Holden, B., Nichol, R., Romer, A., Metevier, A., Postman, M., Ulmer, M., Lubin, L. 1999, AJ, 118, 2002

Holder, G. 2000, ApJ, 544, 629

Hradecky, V., Jones, C., Donnelly, R. H., Djorgovski, S. G., Gal, R. R., Odewahn, S. C. 2000, ApJ, 543, 521

Hubble, E. 1926, ApJ, 64, 321

Hubble, E. 1929, PNAS, 15, 168

Ikeuchi, S. & Turner, E. 1991, MNRAS, 250, 519

Iovino, A., de Carvalho, R.R., Gal, R., Odewahn, S.C., Lopes, P.A.A., Mahabal, A., & Djorgovski, S.G. 2003, AJ, 125, 1660

Jarvis, J. F. & Tyson, J. A. 1981, AJ, 86, 476

Jeans, J. 1902, *Phil. Trans. Roy. Soc. 199A*, 49

Jenkins, A. *et al.* 2001, MNRAS, 321, 372

Kant, I. 1798, *Allgemeine Naturgeschichte und Theorie des Himmels*,

Katgert, P. et al. 1996, A&A, 310, 8

Katz, L. & Mulders, G. F. W. 1942, ApJ, 95, 565

Kells, W., Dressler, A., Sivaramakrishnan, A., Carr, D., Koch, E., Epps, H., Hilyard, D., & Pardeilhan, G. 1998, PASP, 110, 1487

Kawasaki, W., Shimasaku, K., Doi, M., Okamura, S. 1998, A&AS, 130, 567

Kepner, J., Fan, X., Bahcall, N., Gunn, J., Lupton, R. & Xu, G. 1999, ApJ, 517, 78

Kiang, T. 1996, *Zeitschrift für Astrophysik*, 64, 433

Kim, R. S. J. 2001, Tese de Doutorado, *Princeton University*

Kim, R. S. J. et al. 2002, AJ, 123, 20

Kitayama, T., Suto, Y. 1997, ApJ, 490, 557

Klypin, A., Nolthenius, R. Primack, J. 1997, ApJ, 474, 533

Kriessler, J. R. & Beers, T. C. 1997, AJ, 113, 80

Lange, A. *et al.* 2001, PhRvD, 63, 2001

Lasker, B. M., Doggett, J., McLean, B., Sturch, C., Djorgovski, S., de Carvalho, R. R. & Reid, I. N. 1996, ASP Conf. Ser. 101: Astronomical Data Analysis Software and Systems V, 5, 88

Lewis, A. D., Ellingson, E., Morris, S. L., Carlberg, R. G. 1999, ApJ, 517, 587

Lobo, C., Iovino, A., Lazzati, D. & Chincarini, G. 2000, A&A, 360, 896

Lopes, P. A. A., Gal, R., Djorgovski, S., de Carvalho, R., Odewahn, S. 2000, *Bulletin of the American Astronomical Society* 197, 106.11

Lopes, P. A. A., de Carvalho, R. R., Gal, R. R., Djorgovski, S. G., Odewahn, S. C., Mahabal, A. A. & Brunner, R. J. 2003a, AJ, *em preparação*

Lopes, P. A. A., Djorgovski, S. G., Mahabal, A. A., Gal, R. R., Kohl Moreira, J. L., de Carvalho, R. R., Brunner, R. J. Odewahn, S. C. 2003b, PASP, *em preparação*

Lubin, L. Bahcall, N. A. 1993, ApJ, 415, 17

Lubin, L. M., Oke, J. B., Postman, M. 2002, AJ, 124, 1905

Lucey, J. R. 1983, MNRAS, 204, 33

Lumsden, S. L., Nichol, R. C., Collins, C. A. & Guzzo, L. 1992, MNRAS, 258, 1

Maddox, S. J., Efstathiou, G. & Sutherland, W. J. 1990, MNRAS, 246, 433

Maddox, S. J., Efstathiou, G. & Sutherland, W. J. 1996, MNRAS, 283, 1227

Mandelbrot, B. 1975, *Academie des Sciences Paris Comptes Rendus Serie Sciences Mathematiques*, 280, 1551

Margoniner, V. E., de Carvalho, R. R., Gal, R. R., Djorgovski, S. G. 2001, ApJ, 548, L143

Mather, J. *et al.* 1990, ApJ, 354, 37

Mathiesen, B., Evrard, A. E. 1998, MNRAS, 295, 769

Miller, C. J., Batuski, D. J., Slinglend, K. A. & Hill, J. M. 1999, ApJ, 523, 492

Norman, D. J.; Impey, C. D. 1999, AJ, 118, 613

Odewahn, S. C. & Aldering, G. 1995, AJ, 110, 2009

Odewahn, S. C., de Carvalho, R. R., Gal, R. R., Djorgovski, S. G., Mahabal, A., Brunner, R. J., Lopes, P. A. A., Kohl Moreira, J. L., Stalder, B. 2003, AJ, *submetido*

Olivier, S., Primack, J. R., Blumenthal, G. R., Dekel, A. & Stanhill, D. 1990, ApJ, 356, 1

Olsen, L. F. et al. 1999, A&A, 345, 363

Padmanabhan, T., Sethi, S. K. 2001, ApJ, 555, 125

Paolillo, M., Andreon, S., Longo, G., Puddu, E., Gal, R. R., Scaramella, R., Djorgovski, S. G., & de Carvalho, R. 2001, A&A, 367, 59

Peebles, P. J. E. & Groth, E. J. 1975, ApJ, 196, 1

Pennington, R. L., Humphreys, R. M., Odewahn, S. C., Zumach, W. & Thurmes, P. M. 1993, PASP, 105, 521

Perlmutter, S. *et al.* 1999, ApJ, 517, 565

Picard, A. 1991, Tese de Doutorado, Caltech

Pierpaoli, E. *et al.* 2001, MNRAS, 325, 77

Pinkney, J. *et al.* 1996, ApJS, 104, 1

Pisani, A. 1996, MNRAS, 278, 697

Plionis, M. & Basilakos, S. 2002, MNRAS, 329, L47

Postman, M., Geller, M. & Huchra, J. 1986, AJ, 91, 1267

Postman, M., Huchra, J., & Geller, M. 1992, ApJ, 384, 404

Postman, M., Lubin, L. M., Gunn, J. E., Oke, J. B., Hoessel, J. G., Schneider, D. P. & Christensen, J. A. 1996, AJ, 111, 615

Postman, M., Lauer, T., Szapudi, I., Oegerle, W. 1998, ApJ, 506, 33

Postman, M. 1998, *'Clusters as Tracers of Large Scale Structure'*, em *Evolution of Large-Scale Structure: From Recombination to Garching*

Postman, M., Lauer, T. R., Oegerle, W., & Donahue, M. 2002, ApJ, 579, 93

Press, W. H.; Schechter, P. 1974, ApJ, 187, 425

Puddu, E., Andreon, S., Longo, G., Strazzullo, V., Paolillo, M., & Gal, R. R. 2001, A&A, 379, 426

Puddu, E., De Filippis, E., Longo, G., Andreon, S. & Gal, R. R. 2003, A&A, 403, 73

Ramella, M., Boschin, W., Fadda, D., Nonino, M. 2001, A&A, 368, 776

Reid, I. N. *et al.* 1991, PASP, 103, 661

Reichart, D. E. *et al.* 1999, ApJ, 518, 521

Rhee, G. *et al.* 1991, A&A, 246, 301

Ribeiro, A. L. B., Wuensche, C. A., Letelier, P. S. 2000, ApJ, 539, 1

Romani, R. & Maoz, D. 1992, ApJ, 386, 36

Romer, A. K. *et al.* 2000, ApJS, 126, 209

Sadat, R. *et al.* 1998, A&A, 329, 21

Sadat, R., Blanchard, A. 2001, A&A, 371, 19

Salvador-Solé *et al.* 1993, ApJ, 402, 398

Schaeffer, R. & Silk, J. 1988, ApJ, 332, 1

Scharf, C. A., Jones, L. R., Ebeling, H., Perlman, E., Malkan, M. & Wegner, G. 1997, ApJ, 477, 79

Schneider, P. 1992, A&A, 254, 14

Schuecker, P., Böhringer, H. 1998, A&A, 339, 315

Shane, C. D. & Wirtanen, C. A. 1954, AJ, 59, 285

Sheetman, S. A. 1985, ApJS, 57, 77

Shewchuk, J. 1996, em *First Workshop on Applied Comptutational Geometry*, ACM

Silverman, B. W. 1986, *Monographs on Statistics and Applied Probability*, London: Chapman and Hall

Slipher, V. M. 1917, *'Spectrographic observations of nebulae and star clusters'*, *Popular Astronomy*, 25, 36

Smith, R. J., Boyle, B. J., Maddox, S. J. 2000, MNRAS, 313, 252

Smoot, G. *et al.* 1992, ApJ, 396, L1

Sodré, L., Capelato, H., Steiner, J., Mazure, A. 1989, AJ, 97, 1279

Solanes, J., Salvador-Solé, E., González-Casado, G. 1999, 343, 733

Squires, G., Kaiser, N., Babul, A., Fahlman, G., Woods, D., Neumann, D. M., & Böhringer, H. 1996, ApJ, 461, 572

Stein, P. 1997, A&A, 317, 670

Stephani, H. 1990, *General Relativity, An Introduction to the Theory of Gravitational Field, ISBN 0521370663, Cambridge University Press*

Stockton, A. 1982, ApJ, 257, 33

Struble, M. F. & Rood, H. J. 1999, ApJS, 125, 35

Sunyaev, R. A. & Zeldovich, Y. B. 1980, ARA&A, 18, 537

Sutherland, W. & Efstathiou, G. 1991, MNRAS, 248, 159

Szalay, A. S. & Schramm, D. N. 1985, *Nature*, 314, 718

Thuan, T. X. & Gunn, J. E. 1976, PASP, 88, 543

Tormen, G. 1997, MNRAS, 290, 411

Tyson, J. A. & Fischer, P. 1995, ApJ, 446, L55

van Haarlem, M. P., Frenk, C. S. & White, S. D. M. 1997, MNRAS, 287, 817

Véron-Cetty, M. P & Véron, P. 2001, A&A, 374, 92

Viana, P. T. P., Liddle, A. R. 1996, MNRAS, 281, 323

Viana, P. T. P., Liddle, A. R. 1999, MNRAS, 303, 535

Viana, P. T. P., Nichol, R. C., Liddle, A. R. 2002, ApJ, 569, 75

Vikhlinin, A., McNamara, B, Forman, W., Jones, C., Quintana, H., Hornstrup, A. 1998, ApJ, 502, 558

Walker, T. *et al.* 1991, ApJ, 376, 51

Walter, C. & Klypin, A. 1996, ApJ, 462, 13

Webster, R. & Hewett, P. 1990, *'Quasar-galaxy associations'*, em *Gravitational lensing, Proceedings of the Workshop, Toulouse, France*, p. 73

Weir, N., Fayyad, U. M. & Djorgovski, S. 1995a, AJ, 109, 2401

Weir, N., Djorgovski, S. & Fayyad, U. M. 1995b, AJ, 110, 1

Weir, N., Fayyad, U. M., Djorgovski, S. G. & Roden, J. 1995c, PASP, 107, 1243

West, M. & Bothun, G. 1990, ApJ, 350, 36

West, M. *et al.* 1995, ApJ, 451, L5

White, S. D. M., Efstathiou, G., Frenk, C. S. 1993, MNRAS, 262, 1023

Willick, J. A. 2000, ApJ, 530, 80

Wittman, D., Tyson, J. A., Margoniner, V. E., Cohen, J. G., Dell'Antonio, I. P. 2001, ApJ, 557, 89

Wold, M. *et al.* 2000, MNRAS, 316, 267

Yates, M. *et al.* 1989, MNRAS, 240, 129

Yee, H. K. C. & Green, R. 1987, ApJ, 319, 28

Yee, H. K. C., Ellingson, E. 1993, ApJ, 411, 43

Yee, H. K. C. & López-Cruz, O. 1999, AJ, 117, 1985

Yee, H. K. C., Ellingson, E. 2003, ApJ, 585, 215

York, D. G. et al. 2000, AJ, 120, 1579

Zaninetti, L. 1995, A&AS, 109, 71

Zaritsky, D., Nelson, A. E., Dalcanton, J. J. & Gonzalez, A. H. 1997, ApJ, 480, L91

Zwicky, F. 1938, PASP, 50,

Zwicky, I. F. 1942, ApJ, 95, 555

Zwicky, F., Herzog, E. & Wild, P. 1968, *Catalogue of Galaxies and of Clusters of Galaxies, Pasadena: California Institute of Technology* (CIT), 1961-1968