Universidade de São Paulo
Instituto de Astronomia, Geofísica e Ciências Atmosféricas
Departamento de Astronomia

Marcelle Soares-Santos

Cosmography with galaxy clusters: shedding light on dark energy

São Paulo

2010

Marcelle Soares-Santos

Cosmography with galaxy clusters: shedding light on dark energy

Thesis presented in partial fulfillment of the requirements for the degree of Doctor in Sciences at the University of São Paulo, Institute of Astronomy, Geophysics and Atmospheric Sciences, Department of Astronomy.

Research field: Astronomy
Advisor: Prof. Dr. Jorge E. Horvath

São Paulo
2010

Acknowledgments

I would like to express my gratitude to my formal and informal advisors in São Paulo, Jorge Horvath and Reinaldo de Carvalho, whose expertise, understanding and enthusiasm have made this thesis possible. Without their encouragement and persistence, this enterprise would have been abandoned four years ago.

I also thank James Annis and Scott Dodelson for accepting me into their group at the Fermi National Accelerator Laboratory (Fermilab). Their advice and support added considerably to my graduate experience.

A very special thanks goes to Jeff Kubo, Ribamar Reis, Hee-Jong Seo, Jiangang Hao, Emily Drabek, Huan Lin, Nikolay Kuropatkin, Tom Diehl, Paulo Lopes and Francesco La Barbera for the work, knowledge and experience shared.

These have been great productive times and I am thankful for the support from the Brazilian Nacional Research Council (CNPq) and from Fermilab.

Good friends, however, have made these years into even better times. Maiko Takahashi, Krisztian Peters, Antoinette Dickens, Christian Schwanenberger, Oliver PinkPank, Rob Forest, Marco Verzocchi and Stefan Soldner-Rembold are among those who have made my time in Chicago an amazing experience. Carolina Ghidetti, Daniel Chaia, Fabio Candotti, Maitê Parente, Gustavo Picanso, Gustavo Guerrero, Denise Gonçalves, Ignacio de La Rosa and my brother, Eduardo, have made me wish to return to São Paulo. Márcia Helena Paiva, my parents, Elena and Antônio, and my sister, Míriam, make Vitória my eternal home. A special thanks goes to each of them, for their support and friendship. And to the one who has made the many overnight hours spent in my office seem worthwhile after all, Björn.

Thesis typeset using LaTeX style iagtese v2.1.

Abstract

Observations in the last ten years established dark energy as a component of unknown nature likely to be responsible for the accelerated expansion of the Universe. In its simplest formulation, dark energy features an equation of state $p = w\rho$ with $w < -1/3$. Several ways of generating such an equation of state are known and it is important to establish the w parameter and its time-variation experimentally. A powerful approach to this problem is the study of optically selected galaxy clusters as cosmological probes. For such a study, the Sloan Digital Sky Survey (SDSS) is the state of the art data set to be surpassed only by the next generation of experiments. In this thesis, we report the current status of the analysis leading to the first measurement of w using a cluster abundance test based on optical data. Our sample is based on the SDSS Coadd, a 250 square degrees area where ~ 30 exposures were stacked to produce deep images reaching magnitude $r = 23.5$, corresponding to redshift $z \sim 1$. In general, measuring w requires galaxy photometric redshifts, cluster finding algorithms, cluster mass calibration, cosmological parameter estimation and a data set of sufficient scope. For the Coadd, photometric redshifts with mean error $\sigma_z/(1+z) = 0.038$ were obtained with a neural network algorithm. A cluster catalog from this sample of 13M galaxies is being constructed using the Voronoi Tessellation (VT) cluster finder. The cluster selection function is computed using mock galaxy catalogs, showing that the VT cluster catalogs have completeness and purity $> 80\%$ for clusters above $10^{13.5}$ solar masses up to $z \sim 1$. Calibration of a cluster mass proxy is done by fitting a mass model to the weak lensing shear profile of a cluster sample split into mass proxy bins. We obtain weak lensing mass measurements for individual nearby clusters ($z < 0.1$) using this technique. On the Coadd, this technique will be applied to stacked clusters detected with the VT. A likelihood method using the mean abundance and spatial distribution of clusters is

implemented to obtain cosmological constraints expected from galaxy cluster surveys. We find that w can be measured with 40% uncertainty using the VT cluster catalog based on the SDSS Coadd data.

Resumo

Observações nos últimos dez anos estabeleceram a energia escura como uma componente de natureza desconhecida provavelmente responsável pela expansão acelerada do Universo. Em sua formulação mais simples, a energia escura caracteriza-se por uma equação de estado $p = w\rho$ com $w < -1/3$. Diversas maneiras de gerar essa equação de estado são conhecidas e é importante estabelecer o parâmetro w e sua variação temporal experimentalmente. Uma poderosa abordagem para este problema é o estudo de aglomerados de galáxias opticamente selecionados como provas cosmológicas. Para tal estudo, o Sloan Digital Sky Survey (SDSS) é o estado da arte em dados a serem superados somente pela próxima geração de experimentos. Nesta tese, reportamos o atual estágio da análise levando à primeira medida de w usando um teste de abundância de aglomerados com base em dados ópticos. Nossa amostra é baseada no SDSS Coadd, uma área de 250 graus quadrados onde ~ 30 exposições foram combinadas para produzir imagens profundas atingindo magnitude $r = 23.5$, correspondendo a redshift $z \sim 1$. Em geral, medir w requer redshifts fotométricos, algoritmos de detecção de aglomerados, calibração de massa de aglomerados, estimativa de parâmetros cosmológicos e um conjunto de dados de suficiente escopo. Para o Coadd, redshifts fotométricos com erro médio $\sigma_z/(1+z) = 0.038$ foram obtidos com um algoritmo de redes neurais. Um catálogo de aglomerados a partir dessa amostra de 13M de galáxias está sendo construido usando o algoritmo de deteção de aglomerados Voronoi Tessellation (VT). A função de seleção de aglomerados é computada usando catálogos de galáxias simulados, mostrando que os catálogos do VT têm completeza e pureza $> 80\%$ para clusters acima de $10^{13.5}$ massas solares encontrados até $z \sim 1$. Calibração de um indicador de massa é feita através do ajuste de um modelo de distribuição de massa ao perfil de shear produzido por efeito de lentes gravitacionais fracas de uma

amostra de aglomeardos dividida em intervalos do indicador de massa. Obtemos medidas de massa para aglomerados individuais próximos ($z < 0.1$) usando essa técnica. No Coadd, essa técnica será aplicada a aglomerados combinados, detectados com o VT. Um método de verossimilhança usando a abundância média e distribuição espacial de aglomerados é implementado para obter vínculos cosmológicos esperados a partir de levantamentos de aglomerados de galáxias. Mostramos que w pode ser medido com incerteza de 40% usando o catálogo de aglomerados do VT baseado nos dados do SDSS Coadd.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Observations of the last ten years established dark energy as a component of unknown nature responsible for the accelerated expansion of the Universe. Reflecting deep aspects of the fundamental theory of the underlying space, but yet apparently accessible only through astronomical observation, dark energy is a unique link between Astronomy and High-Energy Physics. Large sections of the two communities have converged in support of astronomical projects motivated by the quest for understanding the physics of dark energy. This gave rise to an exciting and fast developing field and has an extensive impact on our scientific culture. In this thesis we focus on optically selected galaxy clusters as cosmological probes. We analyze data from the Sloan Digital Sky Survey, a state of the art data set to be surpassed only by the next generation of experiments, and perform preparatory work for the upcoming Dark Energy Survey. In addition to a powerful data set, such a study requires reliable photometric redshift estimation for galaxies, cluster finding algorithms, cluster mass estimation methods and cosmological parameter estimation. Our research program, detailed and discussed here, encompasses each of these items.

1.1 The dark energy problem

An outstanding problem of present-day physics – the gravitational effect of the zero-point energy of particles and fields – has emerged from empirical evidence that the total energy density of the Universe is dominated by Einstein's cosmological constant, Λ, or an alternative slowly varying fluid, the dark energy. The most egregious problem is the disagreement between the meV energy scale observationally inferred for dark energy and that theoretically predicted, which is nearly 30 orders of magnitude larger (see Weinberg,

1989; Carroll, 2001; Peebles & Ratra, 2003). Despite this fact, theorists have long neglected the problem, postulating that for some fundamental and unknown reason the vacuum energy should be zero. But the current data puts this assumption to rest. This discrepancy supports the (still speculative) hypothesis that the nature of the cosmological constant, and dark energy in general, is related to fundamental physics and may be the key to properly connect quantum mechanics (microphysics) and gravitation (macrophysics).

The dynamics of the Universe is described by the Einstein equations, which given the set of symmetries resulting from the assumption of a homogeneous and isotropic Universe filled with a perfect fluid, can be solved into an analytic form

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{K}{a^2}\,, \tag{1.1}$$

$$\dot{H} = -4\pi G(p+\rho) + \frac{K}{a^2}\,, \tag{1.2}$$

where H is the Hubble parameter[1], ρ and p denote the total energy density and pressure of all the species present in the Universe at a given epoch and a is the scale factor. The constant K describes the geometry of the space-time as closed, flat or open, corresponding to $K = +1, 0, -1$ respectively. G is the gravitational constant. Eliminating K/a^2 from Eqs. (1.1) and (1.2), we obtain

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\,(\rho + 3p) \tag{1.3}$$

and find that the accelerated expansion occurs for $\rho + 3p < 0$. The equation of state parameter of a perfect fluid is

$$w = p/\rho. \tag{1.4}$$

One can rewrite Eq. (1.1) in the form:

$$\Omega(t) = 1 + \frac{K}{(aH)^2}\,, \tag{1.5}$$

where $\Omega(t) \equiv \rho(t)/\rho_c(t)$ is the dimensionless density parameter and $\rho_c(t) = 3H^2(t)/8\pi G$ is the critical density. This can be splitted into the various energy components of our Universe as

$$\Omega(z) = \Omega_m(z) + \Omega_r(z) + \Omega_k(z) + \Omega_{de}(z) \tag{1.6}$$

[1] The value of the Hubble parameter in the present Universe is referred to as H_0 or h, where $h \equiv H_0/100$ km s^{-1}Mpc^{-1}.

where $\Omega_i(z) = 8\pi G\rho_i/3H^2$ represents the i-th component – matter (m), radiation (r), curvature (k) and dark energy (de). Redshift and time are related by

$$\frac{1}{1+z} = a(t) = (t - t_0)^{\frac{2}{3(1+w)}} \tag{1.7}$$

and the present value, $\Omega_i(z = 0)$, is denoted by Ω_i.

Observational evidence for a flat, $\Omega_k \simeq 0$, low mass, $\Omega_m \sim 0.3$, Universe dominated by dark energy include:

1. The lower limits on the age of the Universe obtained from the age of the oldest stellar populations (Sandage, 1961).

2. Baryonic mass fraction obtained from X-ray galaxy clusters (Briel et al., 1992; David et al., 1995; Lin et al., 2003; LaRoque et al., 2006).

3. The position of the first acoustic peak of the cosmic microwave background spectrum (Spergel et al., 2003; Page et al., 2003; Spergel et al., 2007).

4. Luminosity-redshift relation obtained from observations of type Ia Supernovae (SNe) (Riess et al., 1998; Perlmutter et al., 1999; Knop et al., 2003; Riess et al., 2004; Astier et al., 2006; Wood-Vasey et al., 2007).

5. Baryon acoustic oscillation (BAO) peak observed in the SDSS (Eisenstein et al., 2005) and 2dFGRS (Cole et al., 2005) data.

6. Matter power spectrum probed through Ly-α forest observations (Viel et al., 2004; Seljak et al., 2005).

7. Angular cross-correlation between cosmic microwave background (CMB) and the large scale structure, due to the integrated Sachs-Wolfe effect (Fosalba et al., 2003; Rassat et al., 2007).

8. Angular diameter test using radio galaxies as standard rulers (Guerra et al., 2000; Daly et al., 2007).

The best fit to current observations of these astrophysical phenomena places the cosmological constant in the role of dark energy (the Λ-CDM model), but the constraints on

w are not definitive. For instance, analysis of CMB in combination with SNe data results in $w = -0.967^{+0.073}_{-0.072}$ when a flat Universe is assumed ($\Omega_k = 0$) (Spergel et al., 2007). When Ω_k is considered as a free parameter, the results are $w = -1.08 \pm 0.12$. However, Blanchard et al. (2003) have shown that, if the assumption of a single power law to the primordial fluctuations is relaxed and a small fraction of massive relic neutrinos is included, an Einstein-de Sitter $\Omega_m = 1$ model can accommodate most of the data from the large scale structure of the Universe (WMAP, 2dFGRS and Ly-alpha forest), implying that only the Hubble diagram from SNe data remains as an evidence for a non-zero Λ.

However, the incompatibility of nearly 30 orders of magnitude between the energy scale inferred from observations and the theoretical predictions have led to a plethora of alternative models (see Copeland et al., 2006, for a broad review). Among the most important ones are:

1. *X-fluid* (Turner & White, 1997), with a constant equation of state $p = w\rho$, where $w < 0$. This model is called *phantom fluid* when $w < -1$ (Caldwell, 2002). The *cosmological constant*, Λ, is the particular case where $w = -1$.

2. *Quintessence* (Ratra & Peebles, 1988; Carroll, 1998), characterized by a scalar field that takes on a dynamical but slow-varying equation of state with several possible parametrization schemes (the X-fluid may be interpreted as a particular case of this model).

3. *Quartessence*, a model whose most famous example is the Chaplygin gas (Kamenshchik et al., 2001; Bento et al., 2002) ($p \propto \rho^\alpha$), also based on a scalar field that provides a single-fluid interpolation between the dark matter and energy dominated eras.

4. *K-essence* (Chiba et al., 2000; Armendariz-Picon et al., 2000), which relies on the kinetic energy of the scalar field (not the potential energy, as in the case of quintessence) through a non-canonical kinetic term in the Lagrangian.

An interesting approach for the study of dark energy is to consider that its equation of state can be approximated by a Taylor expansion around $z = 0$

$$w(z) = w(z = 0) + w'(z = 0)z + \cdots \tag{1.8}$$

and attempt to constraint the coefficients to the highest possible order (e.g. Huterer & Turner, 2001).

When attempting to achieve the first order term, w', the current state of the art measurement, obtained from CMB analysis combined with SNe and BAO, impose very weak constraints: $w = -0.93 \pm 0.12$ and $w' = -0.38 \pm 0.65$ assuming a flat prior (Komatsu et al., 2010). These are to be significantly improved by the next generation of experiments (Albrecht et al., 2006), which aim to constraint w and w' to less then 5% and 30% uncertainty, respectively. Ongoing experiments have focused on constraining the zero-th order term, w, improving from the current $\sim 10\%$ to $\sim 1\%$ uncertainty level.

1.2 Galaxy clusters shedding light on dark energy

The discussion presented above shows that a better theoretical understanding of dark energy demands additional (independent) evidence and definitive improvements in the present constraints on w. An alternative to fulfil these demands is the cluster abundance test. The abundance of clusters as a function of mass is sensitive to cosmological parameters such as the normalization of the power spectrum, σ_8, and the matter density parameter, Ω_m. Observations at low-redshifts allowed to establish $\sigma_8 \Omega_m^{0.5} \simeq 0.5$ (e.g. Evrard, 1989; Peebles et al., 1989; Frenk et al., 1990; Bahcall & Cen, 1992; White et al., 1993). This degeneracy between σ_8 and Ω_m was broken with measurements of the evolution of the cluster abundance with redshift. The results thus obtained constrained σ_8 and Ω_m to be ~ 0.9 and $\lesssim 0.6$, repectively (Oukbir & Blanchard, 1992; Eke et al., 1996; Viana & Liddle, 1996; Oukbir & Blanchard, 1997; Bahcall et al., 1997; Fan et al., 1997; Eke et al., 1998; Donahue & Voit, 1999). The effect of w in the cluster abundance becomes measurable if a clean and well-understood selection of a large sample of clusters within a large area coverage and depth up to $z \sim 1$ is obtained (Wang & Steinhardt, 1998; Haiman et al., 2001a; Levine et al., 2002a). Recent and ongoing experiments using X-ray (e.g., Borgani et al., 2001; Reiprich & Böhringer, 2002; Seljak, 2002; Allen et al., 2003; Pierpaoli et al., 2003; del Popolo et al., 2005) and optically selected clusters (e.g., Bahcall et al., 2003b; Gladders et al., 2007; Rozo et al., 2007a) have improved the constraints above (see also Bahcall, 2000; Borgani & Guzzo, 2001; Mohr, 2005; Borgani, 2006), but their results do not allow to firmly establish the w parameter because they are either not deep or not wide

enough in angular coverage. Gladders et al. (2007), for instance, obtained $\Omega_m = 0.3^{+0.12}_{-0.11}$ and $\sigma_8 = 0.70^{+0.27}_{-0.15}$ from the deepest RCS images ($0.35 < z < 0.95$, 72.07 deg^2), while Rozo et al. (2007b) obtained $\sigma_8 = 0.92 \pm 0.10$ and $\Omega_m h^2 = 0.128 \pm 0.01$ from the SSDS data ($0.1 < z < 0.3$, 7500 deg^2, Koester et al. (2007a)). Only the next generation of surveys will be able to perform the cluster abundance test with the required precision (Albrecht & Bernstein, 2007).

Wide and deep cluster surveys planned to the near future include: the South Pole Telescope (SPT, Ruhl et al., 2004), which is a Sunyaev-Zel'dovich effect survey; the Dark Universe Observatory (DUO, Griffiths et al., 2004), which is an X-ray dedicated mission; and the Dark Energy Survey (DES, Abbott et al., 2005), which is a deep optical survey. Each of these projects will cover thousands of square degrees with depth beyond $z = 1$. Data from the recently completed Sloan Digital Sky Survey (SDSS, Adelman-McCarthy et al., 2007; Annis et al., 2010) however, allows us to perform a cluster abundance test in an area of ~ 250 square degrees up to redshift $z \sim 0.8$. The latter is the data set in which this thesis is focused.

1.2.1 Signatures of dark energy in galaxy cluster counts

The presence of dark energy results in an accelerated expansion of the Universe and influences the structure formation process. No gravitationally bound structures can be formed in a static and completely symmetric (homogeneous and isotropic) Universe, but any small overdensities present in the cosmic fluid will cause the surrounding matter to fall into these potential wells and the initial small fluctuations will grow exponentially. The expansion of the Universe acts as a damping factor, forcing the fluctuations growth to obey a power law, and this damping is enhanced if the expansion is accelerated (see Mukhanov et al., 1992, for a detailed review). The Universe on scales $\gtrsim 10h^{-1}$Mpc is dominated by linear or weakly non-linear gravitational dynamics, preserving memory of the initial conditions and not being strongly affected by complex astrophysical processes related to galaxy formation and evolution. These processes become dominant on smaller scales, where the non-linear regime is fully established. For this reason, galaxy clusters are especially interesting to probe cosmology, since they are the largest gravitationally bound structures, arising from initial perturbations with a typical comoving scale of $\sim 10h^{-1}$Mpc. However,

they are on the threshold of these two regimes and to use galaxy clusters as cosmological tools it is not enough to have a well-established theory of structure formation. We need to take into account the astrophysical effects that could compromise the accuracy of the cosmological test.

Concerning the cluster abundance test, we refer to theoretical formulations for the comoving number density of clusters per mass interval, dn/dM (the mass function), first obtained by Press & Schechter (1974) and improved through numerical simulations (e.g. Lacey & Cole, 1994; Sheth & Tormen, 1999; Sheth & Lemson, 1999) whose precision has increased rapidly (Jenkins et al., 2001; Hu & Kravtsov, 2003a; Tinker et al., 2008). The cluster abundance is the number of clusters per redshift interval and unit solid angle, $dN/dzd\Omega$, which is the product of the volume probed, $dV/dzd\Omega$, and the number density $n(z)$. The volume is proportional to the angular diameter distance $d_A(z)^2$ times $(1+z)^2/H(z)$, while $n(z)$ is the integral of the mass function above a certain mass threshold. The sensitivity of the redshift distribution of clusters to the dark energy parameter w has been shown to increase with the redshift and mass threshold by Munshi et al. (2004), for example. They used the Press & Schechter (1974) formula to demonstrate that varying w from -1.9 to $-1/3$ results in a difference of one order of magnitude in the cumulative number density, $dN(>M)/dVdz$, at $z = 1.5$; this difference is 50% lower at $z = 1$.

Figure 1.1 shows the result of a similar calculation using Tinker et al. (2008) fitting formula, with an X-fluid-like equation of state ($p = w\rho$, $w < 0$) for the dark energy (Soares-Santos et al., 2010). But instead of the cumulative number density as a function of z, Fig. 1.1 displays the actual number of clusters $N(>M)$ per square degree in redshift bins of 0.25 width in the range $0.5 \leq z \leq 1.5$. This approach is comparable to that of Evrard et al. (2002a) for simulations using the ΛCDM and τCDM ($\Omega_m = 1$, $\Omega_\Lambda = 0$) models, where he used the Jenkins et al. (2001) formula. Figure 1.1 clearly demonstrates that different dark energy models are better distinguished using more massive clusters at higher redshifts. A cosmological test that would take advantage of this fact is easy to conceive: we have to obtain a sample of clusters spanning a reasonable range in mass and redshift and compare the counts in both spatial cells and mass bins with the prediction from the various theoretical models. Note that the cluster abundance decreases with both quantities and therefore large angular areas and deep magnitude limits are necessary to

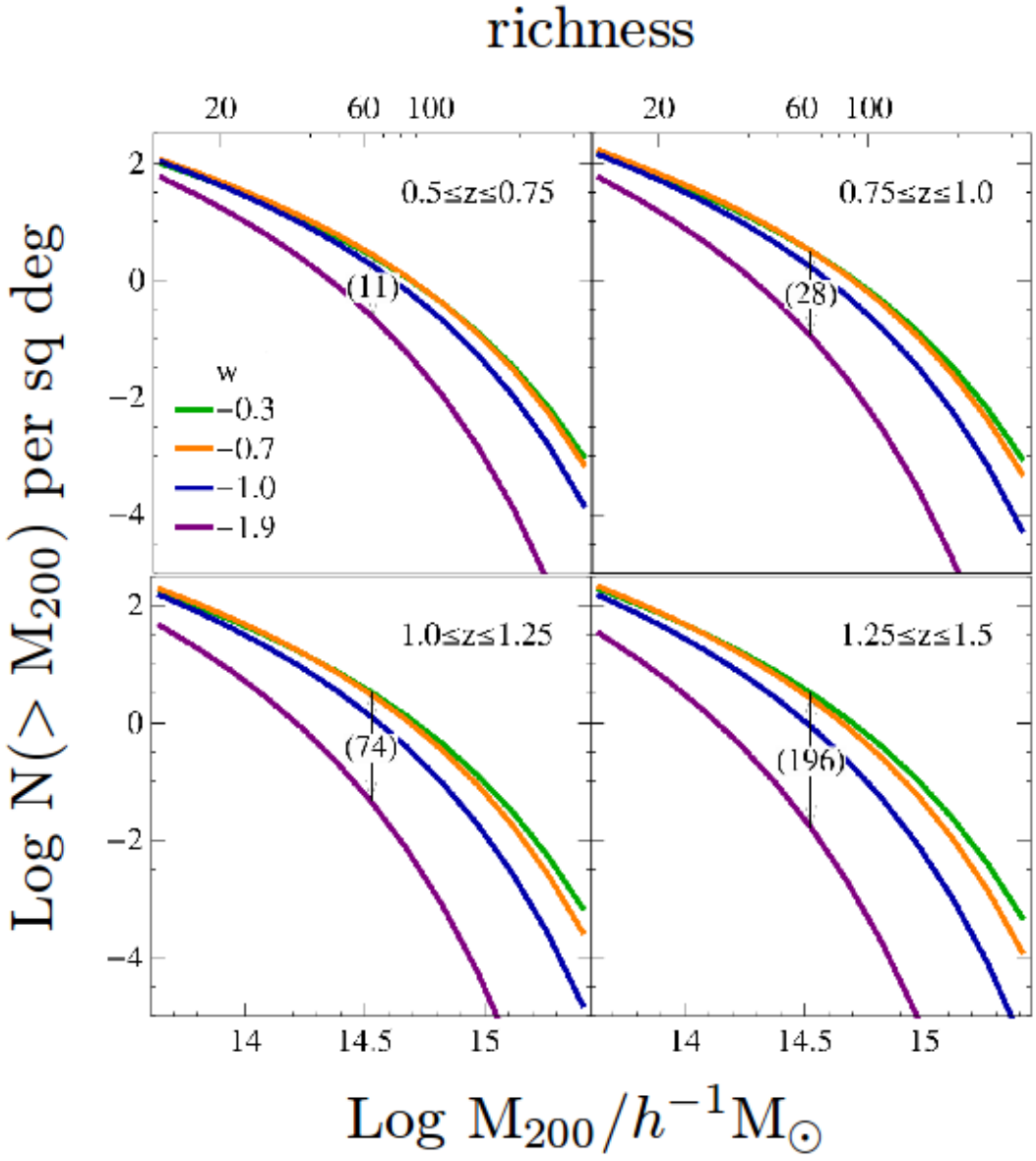

Figure 1.1: Cumulative mass function in four dark energy models, shown as the number, $N(>M)$, of clusters expected above mass M per square degree in each redshift bin of width $\Delta z = 0.25$ in the range $0.5 \leq z \leq 1.5$. The cosmological models considered (with $\Omega_m = 0.3$, $\Omega_{de} = 0.7$ and $h = 0.7$) differ only by the dark energy equation of state parameter w. From top to bottom, the solid curves in each panel correspond to $w = -0.3$, -0.7, -1.0 (ΛCDM), and -1.9, respectively, as shown in the upper-right panel. The number counts are also shown as a function of richness (upper scale), using a mass-richness relation derived from a weak lensing analysis by Johnston et al. (2007). We use the mass function fitting formula by Tinker et al. (2008).

reliably perform this test.

It is important to consider the effect of these different mass functions on number counts in comparison to the effect of w because, if that effect is large, the accuracy of the test proposed here will be compromised. Press & Schechter (1974) mass function has been shown to overestimate the number density for low-mass clusters and underestimate it for massive ones when compared to simulations (Sheth & Tormen, 1999; Jenkins et al., 2001; Evrard et al., 2002a). A generalization of Press & Schechter (1974) was formulated by Sheth & Tormen (1999), considering the effect of non-spherical collapse to obtain a fitting function whose parameters are determined through large N-body simulations. An alternative fitting expression was proposed by Jenkins et al. (2001), and both fit well the simulations. Recently, Tinker et al. (2008) has studied the redshift evolution of the mass function and has proposed a more precise fitting formula.

The mass function can be written as a function of the "mass" variable $\sigma(M)$, which is the rms fluctuation in spheres containing, on average, mass M. The mass functions expressed in terms of this quantity are easily compared since they are almost independent of the cosmological parameters, with the dependence on cosmology being contained within the function $\sigma(M)$ (Linder & Jenkins, 2003). The formula given by Jenkins et al. (2001) shows fractional accuracy better than 20% in the range $-1.2 \leq \ln \sigma^{-1} \leq 1$ ($10^{11} h^{-1} \lesssim M/M_\odot \lesssim 10^{16} h^{-1}$) and the expression of Sheth & Tormen (1999) achieves the same accuracy for $-1.2 \leq \ln \sigma^{-1} \leq 0.9$. The PS formula, in turn, agrees with these fitting formulas for $\ln \sigma^{-1} \sim 0.3$, but the maximum overestimate is $\sim 60\%$ at $\ln \sigma^{-1} = -0.52$ and the underestimates are even higher in the high-mass tail.

However, when the number counts are calculated, one finds that the discrepancy between the various fitting functions is only a small perturbation compared to the effect of the w parameter. For masses around $10^{14.5} M_\odot$, in the ΛCDM model, the difference in $\log N(> M)/\mathrm{sq\,deg}$ is less than 20% at all redshifts and is negligible at $z > 0.5$ (Evrard et al., 2002a), while the differences due to w are more than an order of magnitude at $1.25 < z < 1.5$ as seen in Fig. 1.1. Thus, dark energy tests via cluster number counts are not strongly affected by details of the cluster mass function fitting formula.

It is important to stress that when an accuracy of 5% or less is to be achieved on the mass function, it cannot be described by a single formula at all redshifts and a more precise

formulation is required (Tinker et al., 2008). This will become specially important when a high-precision measurement, aiming at percent level uncertainty in w and w', is to be done based on data from future surveys.

1.2.2 *Extracting cosmological information from a cluster sample*

Galaxy cluster surveys have the potential to provide precise constraints on the dark energy equation of state due to the exponential sensitivity of the abundance of massive dark matter halos to the growth of structure. However, the wealth of information contained in the cluster counts is difficult to extract. Building a cluster catalog covering the appropriate range of mass and redshifts is the first challenge to be faced but, it is equally complex to reliably measure the cluster mass. The mass calibration of a cluster sample is a fundamental problem because the masses of the clusters are not directly observed.

Techniques to relate cluster observables (such as Sunyaev-Zel'dovich (SZ) flux, X-ray temperature and surface brightness, optical richness and weak lensing shear) to cluster masses in the process of extracting cosmological information are known in the literature as *self-calibration* methods (Levine et al., 2002b; Hu, 2003; Majumdar & Mohr, 2004).

Extracting the full information from cluster number counts requires a joint analysis of the mean abundance and spatial distribution. It also requires that the clustering of clusters be considered as the source of sampling errors and that the cluster power spectum estimator accounts for non-Gaussian error distributions (Hu & Kravtsov, 2003b). A likelihood method fullfiling this requirement has been recently developed (Hu & Cohn, 2006) but was not yet applied to any particular data – in part due to the lack of a sample of suficient scope.

In this method, the data is the number of clusters $N_{i\mu}$ in bins of some observable proxy for mass ($\mu = 1, \ldots, n_{\rm m}$) and 3-dimensional spatial cells defined by a solid angle and a redshift interval ($i = 1, \ldots, n_{\rm p}$). The likelihood is

$$\mathcal{L}(\mathbf{N}|\bar{\mathbf{m}}, \mathbf{b}, \mathbf{S}) = \left[\prod_{i=1}^{n_{\rm p}} \int_{-b_{\rm max}^{-1}}^{\infty} d\delta_i \left(\prod_{\mu=1}^{n_{\rm m}} \frac{m_{i\mu}^{N_{i\mu}}}{N_{i\mu}!} e^{-m_{i\mu}}\right)\right] \times \frac{1}{\sqrt{(2\pi)^{n_{\rm p}} \det \mathbf{S}}} e^{-\frac{1}{2}\boldsymbol{\delta}^T \mathbf{S}^{-1} \boldsymbol{\delta}} \quad (1.9)$$

where

$$\bar{m}_{i\mu} = V_i \int d\ln M \, p_{i\mu}(M) \frac{dn}{d\ln M} \quad (1.10)$$

and

$$b_{i\mu} = \frac{V_i}{\bar{m}_{i\mu}} \int d\ln M\, p_{i\mu}(M) b(M) \frac{dn}{d\ln M} \tag{1.11}$$

are respectively the expected mean number and bias of clusters in each bin. $\mathbf{S}$ is the covariance matrix computed as the integral of the power spectrum convolved with a top-hat window function. $dn/d\ln M$ is the mass function and $p_{i\mu}(M)$ is the selection function which we must determine using simulations.

In the case of coarse pixels the Poisson term in Eq. 1.9 becomes negligible and the likelihood is much simplified. To estimate the constraints expected from a given survey using the Fisher Matrix technique as described in Lima & Hu (2004, 2005), we use the simplified likelihood. The results, assuming a survey of 1500 square degrees in the range $z < 1.5$ are shown in Fig. 1.2. We set constraints on the $\Omega_m - w$ space and assume that all relevant cosmological parameters are fixed: flat cosmology, $\Omega_m = 0.3$, $\Omega_{de} = 0.7$, $\sigma_8 = 1.0$ and $h = 0.7$. These are very optimistic assumptions, but this exercise is usefull to picture

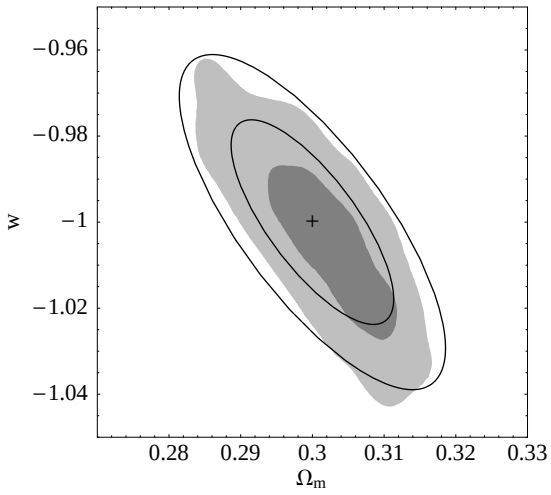

Figure 1.2: 1σ and 2σ contours around the fiducial ΛCDM model ($\Omega_m = 0.3$, $\Omega_{de} = 0.7$, $w = -1$, $\sigma_8 = 1.0$ and $h = 0.7$), considering that 100% of the clusters above $10^{14.5} M_\odot$ in an area of 1500 square degrees up to $z = 1.5$ were detected. Ellipses correspond to the Fisher Matrix result. Shaded regions result from a Monte Carlo Markov Chain analysis as a cross-check of our method.

the strength of the program we have been pursuing.

1.3 Galaxy clusters for dark energy in the SDSS Coadd (the challenges)

In short, the baseline model for dark energy assumes a constant equation of state depending on a single parameter, w. Galaxy clusters have the potential to constrain w down to the percent uncertainty level because their formation and growth are strongly affected by the expansion history of the Universe. The challenges arise mainly because the abundance of such objects is as low as a few per square degree at the high-mass end of the distribution – which implies that the survey must have a large area coverage – and because the cluster mass is not a direct observable.

A measurement of w using galaxy clusters requires:

1. A well-understood galaxy catalog with controlled completeness and contamination, with depth $r \sim 24$ or fainter, covering at least a few thousands square degrees.

2. Accurate photometric redshift measurements.

3. A cluster catalog, with completeness and purity well measured as a function of redshift and mass.

4. Calibration of a proxy-mass relation for clusters, with an estimate of the intrinsic scatter in addition to the mean relation.

5. A robust likelihood method to extract the cosmological information encoded in the cluster catalog after mass calibration.

In pursuing a research program that fulfills these requirements, a powerful data set currently in use is the SDSS Coadd (Annis et al., 2010; Abazajian et al., 2009). The Coadd is a ∼250 square degrees survey where stacking of 20-30 exposures allowed to achieve $r = 23.5$, 2 magnitudes fainter than the SDSS depth over the north galactic cap ($\sim 10,000$ square degrees). The Coadd area consists of a 2.5 degrees stripe (the Stripe 82) around the equator within $-50 \leq \mathrm{RA} \leq 60$ degrees.

Cosmological constraints from the SDSS Coadd cluster catalog can be obtained by counting clusters as a function of mass and redshift. These counts are predicted as a function of multiple cosmological parameters. A likelihood analysis is used to find the best fit model. The likelihood expressed in Eq.(1.9) incorporates the purity and completeness

of the cluster catalog, the measurement errors in the mass and redshift and the intrinsic scatter in the mass-observable relation due to underlying astrophysical properties of each cluster (Lima & Hu, 2004, 2005; Hu & Cohn, 2006).

Mass estimates for galaxy clusters are obtained via calibration of an observable proxy for mass (e.g. richness). This can be done, for instance, by applying a Virial analysis using the redshift of the member galaxies (e.g. Becker et al., 2007; Lopes et al., 2010). This method, however, requires spectroscopic redshifts and is not feasible for the mass calibration of a sample from the Coadd, specially at high-z. An alternative is the weak lensing method. By averaging all the clusters of a given richness and measuring the weak lensing shear due to their average mass, one can establish the desired mass-richness calibration. The calibration has been performed for the maxBCG cluster catalog (Koester et al., 2007b; Johnston et al., 2007; Sheldon et al., 2009), which is based on the SDSS DR5 data and covers the redshift range $0.1 < z < 0.3$.

A valuable cross-check for the weak lensing mass-calibration is to compare it with the results from a velocity dispersion analysis such as in Lopes et al. (2010), in which 8083 clusters from the Berlind et al. (2010) catalog are used. The catalog is based on SDSS DR7 and the Friends-of-Friends method was used to find the clusters (Berlind et al., 2006). This is doable only at low redshift ($z \sim 0.1$) but it is an interesting comparison that can improve the mass calibration and help to understand the systematics..

For the weak lensing mass calibration of clusters in the Coadd region, the shear is measured from the galaxy shapes, which are computed by the photometric software of the SDSS team. To that end, one method is to extract a catalog of shape parameters from the SDSS skyserver and apply a correction to undo the smearing of the shapes caused by the instrumental and atmospheric point spread function (Hirata & Seljak, 2003). Comparison with other methods is an important cross-check here as well. Partial comparison with an independent software (2DPHOT, see Chapter 2.2) shows a reasonable agreement with the results from the SDSS pipeline.

The shear-shear correlation function and its Fourier transform, the cosmic shear power spectrum, have been measured in a complementary study for the shear-cluster correlation function, which is the basis of the cluster mass calibration. It is more difficult to measure the cosmic shear than to perform the mass calibration because systematics are more im-

portant to the shear-shear than to the shear-cluster correlation function. But this study serves as a basis for understanding the systematics relevant to the mass calibration. A shear signal was measured in the SDSS Coadd at 1σ confidence level at 1^o scales (Seo et al., 2010). It is also a fact that the cosmic shear is an independent probe for cosmology (Fu et al., 2008). But the Coadd data does not distinguish between different dark energy model signatures in the cosmic shear power spectrum.

In addition to mass calibration of stacked clusters, individual mass measurements are also possible. The Sloan Nearby Cluster Weak Lensing Survey (Kubo et al., 2009b) has measured individual masses for nearly 1000 clusters with data from the SDSS DR7 and it is the largest weak lensing survey of individual galaxy clusters to date[2].

Photometric redshifts for the SDSS Coadd galaxies are crucial for this work, as well as related projects. A neural network algorithm which was used in the SDSS DR6 (see Oyaizu et al., 2008, for details) was applied to the Coadd galaxies (Reis et al., 2010). Obtaining accurate photometric redshifts for the Coadd or any other high-redshift galaxy catalog is very difficult because the spectroscopic samples needed for the training of the neural network are very sparse. Taking all the spectroscopic redshift catalogs available in the area, a training set of $\sim$ 124k galaxies was gathered (see Table 1.1).

Special care must be taken because the DEEP2 sample, which constitutes the majority of the spectra beyond $z = 0.7$, was pre-selected in the color-color space, which can introduce strong biases. This issue was minimized by Reis et al. (2010), who applyed the DEEP2 color cuts to the photometric sample. The sub-sample which passes the cuts is processed with the DEEP2 spectra included in the training set. For the remaining sub-sample, the DEEP2 is excluded.

With 13M galaxies, the Coadd galaxy catalog with photometric redshift information added (Reis et al., 2010) forms the base catalog for the cluster finding work and the weak lensing project. The photometric redshifts are well measured up to $z \sim 0.8$ (Fig. 1.3) and its mean error is $\sigma_z/(1+z) = 0.038$.

[2] At Fermilab, a user friendly framework for this survey, the Mathematica *notebook*, was derived from the work of Jeffrey Kubo by Fermilab scientist Albert Stebbins and, in a collaboration of summer student Dylan Nelson and myself, updated to handle the demands of production running and to handle stacked clusters. The notebook formed the basis of several years of mentored student investigations by high-school students of the Illinois Mathematics and Science Academy.

Table 1.1 - Spectroscopic samples in the SDSS coadd.

# of spectra	Survey	Reference	url
72,239	SDSS DR7	Abazajian et al. (2009)	sdss.org
17,677	CNOC2	Yee et al. (2000)	astro.utoronto.ca/~cnoc
8,656	WiggleZ	Drinkwater et al. (2010)	wigglez.swin.edu.au
7,766	SDSS-III/BOSS	Ross et al. (2010)	sdss3.org
6,975	DEEP2	Weiner et al. (2005)	deep.berkeley.edu
5,537	VVDS	Garilli et al. (2008)	oamp.fr/virmos/vvds.htm
2,614	2SLAQ	Cannon (2006)	2slaq.info

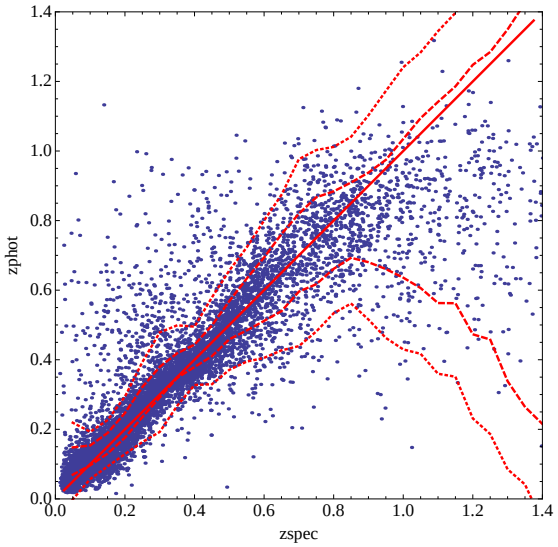

Figure 1.3: Photoz validation plot for the Coadd data. SDSS DR7, 2SLAQ, DEEP2, CNOC2, VVDS, SDSS III and WiggleZ were used as training sets. The solid line indicates $z_{spec} = z_{phot}$. Dashed lines are 68% and 95% confidence regions. Up to $z_{spec} = 0.8$ the photometric redshifts are very accurate. The mean error σ_z in the photometric redshift is $\sigma_z/(1+z) = 0.038$ (Reis et al., 2010).

With a galaxy catalog such as the Coadd, covering 250 square degrees, with depth up to redshift 0.8, we may ask how good are the prospects for constraints on w. Figure 1.4 displays the Fisher matrix contours (solid line) obtained assuming a survey with depth and area of the Coadd, where all clusters above 10^{14} solar masses have been detected and the mass calibration is error-free. The forecasts for the uncertainty in w increase by a factor of $\sim$3 in comparison to the case of a hypothetical survey covering 1500 square degrees with

depth $z = 1.0$, indicating that large areas and deep redshifts will be an important feature in upcoming surveys.

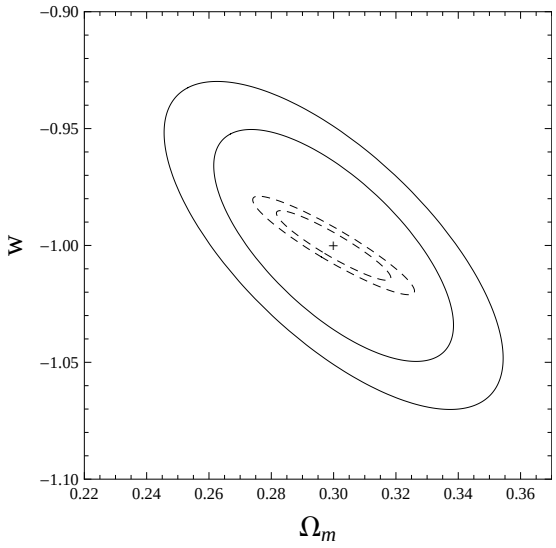

Figure 1.4: 1σ and 2σ Fisher matrix contours for the Coadd (solid) and a hypothetical survey covering 1500 square degrees with depth $z = 1.0$ (dashed). The cross marks the the fiducial ΛCDM model ($\Omega_m = 0.3$,$\Omega_\Lambda = 0.7$,$h = 0.7$,$\sigma_8 = 0.9$). We consider that 100% of the clusters above $10^{14} M_\odot$ were detected. The Coadd area and depth are 250 square degrees and $z = 0.8$, respectively.

These forecasts are very optimistic. If the uncertainty in the mass calibration and a more realistic selection function are taken into acount the figure of merit degrades (Lima & Hu, 2004; Levine et al., 2002b) and the expected uncertainty is $\sim 40\%$. In a combined analysis including CMB and SNe, a measurement consistent with the current constraints on w can be obtained. This will be the first measurement of w from an optically selected sample of clusters. Vikhlinin et al. (2009) have accomplished this for X-ray clusters over an area of 400 square degrees. Improved constraints, reaching the percent uncertainty level on w and 10% in w' will require the next generation of experiments.

The efficiency of the cluster finder depends on the completeness and contamination of the input galaxy catalog. 2DPHOT (La Barbera et al., 2008b) is a computational package to build galaxy catalogs from wide field images and compute its completeness and contamination levels. 2DPHOT has been used in several projects since then (e.g., Soares-Santos et al., 2008). Application of 2DPHOT on 24 sub-fields of the Coadd shows (Figure 1.5) that its magnitude limit is $r \sim 23.5$ and that the completeness curve falls as

$$C_g(r) = \frac{f_0}{1 + \exp((r - \mu)/\sigma)} \tag{1.12}$$

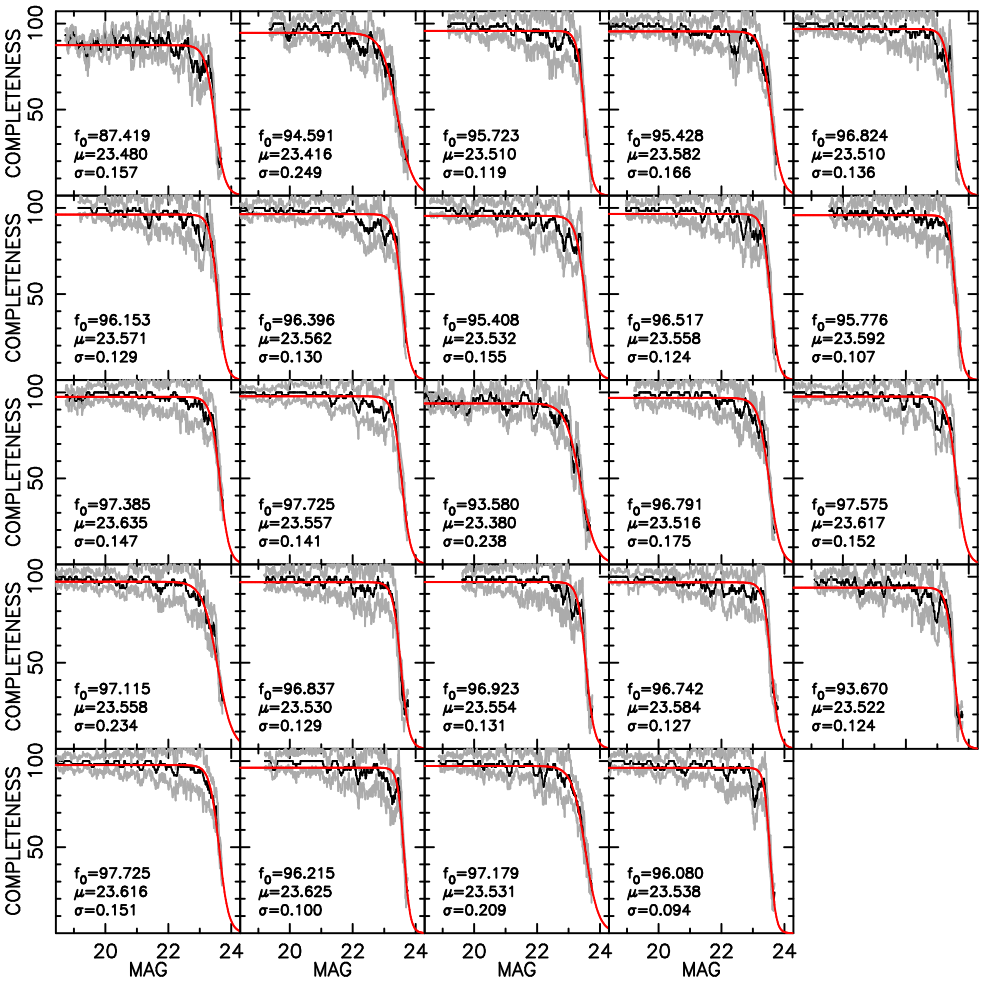

Figure 1.5: Completeness as function of r-band magnitudes, for a set of Coadd sub-images. Note that the magnitudes do not change much across the fields, indicating that the catalog is quite homogeneous.

where μ is the magnitude limit of the catalog, f_0 is a normalization constant and the parameter σ controls how fast the completeness falls when the magnitude limit is reached. For the Coadd, $\mu = 23.5$ and $f_0 = 0.95$ and $\sigma = 0.2$. These values do not change much within the fields, which indicates a homogeneous photometry.

As we proceed into the quest for a better understanding of dark energy, an important issue is to establish a framework appropriate for handling large and complex data sets as never seen before. The Brazilian Virtual Observatory (BRAVO, de Carvalho et al., 2010) is a response to the astronomical community demands for such a framework. 2DPHOT and VT are two software components being included in the project.

1.4 What is this thesis all about?

The goal of our research program is to establish dark energy constraints from galaxy cluster surveys. In accomplishing such a program, a number of challenges have emerged, encompassing all of the fundamental steps from the images registered in our survey camera to the measurement of w. We have met each of those challenges at different levels and we discuss them in the following chapters.

The work featured in Chapter 2, addresses the issue of assuring a highly complete and pure galaxy catalog. We describe 2DPHOT, a general purpose program for source detection and analysis in deep wide-field images. 2DPHOT is an automated tool to obtain both integrated and surface photometry of galaxies in an image, to perform reliable star-galaxy separation with accurate estimates of contamination at faint flux levels, and to estimate completeness of the galaxy catalog. We describe the analysis strategy on which 2DPHOT is based, and provide a detailed description of the different algorithms implemented in the package. This new environment is intended as a dedicated tool to process the wealth of data from wide-field imaging surveys. To this end, the package is complemented by 2DGUI, a graphical interface that allows multiple processing of data using a range of computing architectures.

Chapter 3 is dedicated to the discussion of the Virtual Observatory (VO) concept, a response to the astronomical community demands for improved and homogenized access to the otherwise overwhelming amount of data, combined with the tools to manipulate and exploit them. As we invest in larger projects investigating topics such as dark energy,

Astronomy will soon accumulate an unprecedented amount of data, on the order of 100 PB, while adding 2-4 PB/year – an astonishing five orders of magnitude greater than in the year 2000. This will impose the need for a complex enterprise with a decentralized, webcentric nature, implying that astronomers need to rethink the old ways of conducting their scientific programs. Chapter 3 describes our approach to creating a roadmap for the VO in Brazil and some technical developments on which we have already embarked, such as: database development and basic infrastructure, data grid and processing grid implementation, and data mining.

In Chapter 4 we focus on the issue of mass measurements for a galaxy cluster sample. We describe and present initial results of a weak lensing survey of nearby ($z \lesssim 0.1$) galaxy clusters in the SDSS single pass data. In this first study, galaxy clusters are selected from the SDSS spectroscopic galaxy cluster catalogs of Miller et al. (2005) and Berlind et al. (2006). We report a total of seven individual low redshift cluster weak lensing measurements which include: A2048, A1767, A2244, A1066, A2199, and two clusters specifically identified with the C4 algorithm. Our program of weak lensing of nearby galaxy clusters in the SDSS has evolved fast and is currently the largest weak lensing survey of individual galaxy clusters to date. This work is deeply connected to the issue of the richness-mass calibration of the galaxy cluster sample for dark energy studies. Having the opportunity to develop a methodology by working on data is an invaluable advantage.

At Chapter 5 we report the discovery of seven new, very bright gravitational lens systems from our ongoing gravitational lens search the Sloan Bright Arcs Survey (SBAS). Two of the systems are confirmed to have high source redshifts $z = 2.2$ and $z = 2.94$. Three other systems lie at intermediate redshift with $z = 1.33, 1.82, 1.92$ and two systems are at low redshift $z = 0.66, 0.86$. The lensed background galaxy in all of these systems is bright, each having an i band magnitude ranging from 19.73 to 22.06. We present the spectrum of each of the source galaxies in these clusters along with estimates of the Einstein radius for each system. In total the SBAS has now discovered eighteen strong lens systems in the SDSS imaging data, seven of which are among the highest surface brightness lensed $z \simeq 2-3$ galaxies known. Strong lensing systems are rare and their analysis are unlikely to add much to the problem of mass calibration of a cluster sample. However, those are mass measurements depending on a totally different systematics and we may learn something

from a cross-check with other methods. In addition, bright arc systems represent a unique oportunity to study high redshift galaxies $z \sim 2$ or above and learn about galaxy evolution, a related problem.

One of the most important issues – the cluster finding algorithms – was extensively explored and produced exciting results shown in Chapter 6. There we present a detailed description of the Voronoi Tessellations (VT) cluster finder algorithm in 2+1 dimensions, which improves on past implementations of this technique. The need for cluster finder algorithms able to produce reliable cluster catalogs up to redshift 1 or beyond and down to $10^{13.5}$ solar masses is paramount specially in light of upcoming surveys aiming at cosmological constraints from galaxy cluster number counts. We build the VT in photometric redshift shells and use the two-point correlation function of the galaxies in the field to both determine the density threshold for detection of cluster candidates and to establish their significance. This allows us to detect clusters in a self consistent way not relying in any assumption about their astrophysical properties. We apply the VT on mock catalogs which extends to redshift 1.4, reproducing the ΛCDM cosmology and the clustering properties observed in the SDSS data. An objective estimate of the cluster selection function in terms of the completeness and purity as a function of mass and redshift is as important as having a reliable cluster finder. We measure these quantities by matching the VT cluster catalog with the mock truth table. We show that the VT can produce a cluster catalog with completeness and purity $> 80\%$ for the redshift range up to ~ 1 and mass range down to $\sim 10^{13.5}$ solar masses.

We finally summarize and discuss future work on this and other closely related programs in Chapter 7.

Chapter 2

2DPHOT: a multi-purpose environment for the two-dimensional analysis of wide-field images[1]

2.1 Introduction

In the past decade, wide-field surveys have provided the scientific community with a huge amount of spectroscopic and photometric data, allowing significant progress in our understanding of the Universe. Perhaps the most widely known example is the Sloan Digital Sky Survey (SDSS), whose seventh data release has recently provided photometry in five bands for more than $3 \cdot 10^8$ astronomical objects, as well as spectra of about 1.6 million sources (see Abazajian et al. 2009) over more than $11,000$ square degrees on the sky. One key to the success of the SDSS has been its capability to effectively store, process, and analyze, in a fully automated fashion, the vast amount of data gathered during survey operations. This goal was achieved by using dedicated and well-designed software pipelines, updated during survey operations with reprocessing for the delivery of new data releases. In the coming years, many general purpose astronomical surveys are slated to begin taking data. These wide-field imaging projects will gather deeper and deeper multi-waveband data over large sky areas, producing ever greater data flows. The scientific community must manage and analyze the huge wealth of information contained in these enormous datasets.

In this environment, we have undertaken the development of a new image analysis tool called 2DPHOT, designed to derive two-dimensional information by analyzing both the surface brightness distributions of individual astronomical sources and the spatial distribu-

[1] Chapter published as: La Barbera, F. et al. 2008, PASP 120, 681.

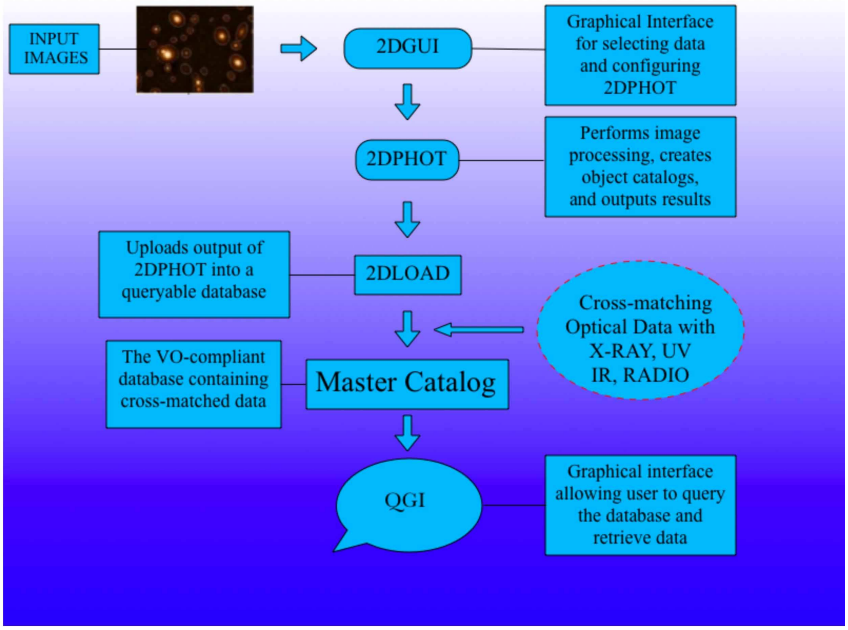

Figure 2.1: Schematic representation of the 2DPHOT environment.

tion of these sources in the image. The package includes several tasks, such as star/galaxy classification, measurement of both integrated and surface photometry of galaxies, PSF modeling, and estimation of catalog completeness and classification accuracy. The package is complemented by a graphical interface named 2DGUI. A schematic view of the 2DPHOT environment is shown in Fig. 2.1. Briefly, the environment is conceived as follows. To start processing, the input images are uploaded to a computer system (e.g. a local cluster or a grid computer) via the 2DGUI interface. 2DGUI also allows the user to configure the 2DPHOT input parameters. A scheduler is also included, allowing timed and sequential execution of several 2DPHOT runs to be performed on the same computer. The actual image analysis is done by the 2DPHOT package, which is the core of the whole environment. During execution, several output tables and plots are produced, showing the different steps of the image analysis and providing means of quality control. These data, which can also be directly downloaded from the user through the 2DGUI interface, are all uploaded into a database system (by the 2DLOAD application, see Fig. 2.1). This system produces a master catalog, by cross-matching 2DPHOT output results with information provided from other VO-compliant web services, and allows the user to perform data queries on this master catalog.

There are several survey projects for which the 2DPHOT environment has already been partly implemented or will be implemented. We have automatically processed the g- and r-band images from the Palomar Abell Cluster Survey (Gal et al., 2000), with the main goal of measuring structural parameters, i.e. the effective radius, the corresponding mean surface brightness, and the Sérsic index n of galaxies in clusters with different richnesses, in the redshift range of 0.05 to 0.2. The structural parameters have been used to estimate the environmental dependencies of internal color gradients in early-type galaxies (see La Barbera et al. 2005). Some examples of general purpose imaging surveys to be analyzed by 2DPHOT are those carried out with the VLT Survey Telescope (VST), a 2.6m diameter imaging telescope equipped with a large format (16k x16k pixels) CCD camera yielding a 1 square degree field of view. The VST, which will be located at the ESO Cerro Paranal Observatory (Chile), has been designed and constructed under a joint venture of ESO and the Capodimonte Astronomical Observatory (OAC). Several survey projects will be carried out with Capodimonte's VST guaranteed time[2].

One of the most interesting science cases for the development of 2DPHOT is the Kilo-Degree Survey with VST (KIDS, see Arnaboldi et al. 2007), a public survey project which will image 1500 square degrees of the southern sky in the *ugri* bands. As shown in Chapter 1, applying 2DPHOT to a moderately deep survey such as VST-KIDS allows detection and measurement of massive galaxy clusters up to redshift $z \sim 1.2$ with high completeness. This cluster abundance measurement can be used to set strong constraints on the dark energy equation of state, which is one of the most crucial issues of modern observational and theoretical cosmology. Reliable star-galaxy separation, with accurate estimates of contamination at very faint flux levels, as well as an accurate cluster detection algorithm are among the 2DPHOT features of paramount importance for such a dark energy project.

This Chapter presents the 2DPHOT package[3], describing the image analysis strategy on which it is based, as well as all the algorithms which are implemented for the different tasks the package performs. We also describe briefly the web-based graphical interface. This Chapter is intended as a reference for all current and forthcoming scientific applications of 2DPHOT. The layout of the Chapter closely follows the order of execution of the 2DPHOT

[2] See http://vstportal.oacn.inaf.it/

[3] The source code of the package is available on request to the authors in its standard form, namely without the VO structure and 2DGUI interface, which will be made available in the near future.

tasks. In Sec. 2.2, we give general, short descriptions of these tasks, and how they are linked during image analysis. Section 2.3 describes the initial components of the first analysis step, i.e. how 2DPHOT produces the image catalog and identifies those objects which are classified as sure stars in the input image. The analysis of each source in the catalog is performed by extracting a stamp image from the input frame and constructing a corresponding mask file (Sec. 2.4). The package performs PSF modeling and derives rough structural parameters for all sources in the image as described in Secs. 2.5 and 2.6, respectively. Section 2.7 deals with the star/galaxy separation, while Sec. 2.8 describes the final fitting of galaxy stamps with seeing-convolved Sérsic models. The isophotal analysis of galaxy stamps is then described in Sec. 2.9, while the determination of the seeing-corrected galaxy aperture magnitudes is outlined in Sec. 2.10. Sections 2.11 and 2.12 describe how 2DPHOT estimates the completeness of the galaxy catalog and the uncertainty in the star/galaxy separation. Section 13 shows how 2DPHOT performs in estimating contamination and completeness at faint magnitudes. Finally, Sec. 2.14 presents the graphical interface (2DGUI). A summary is given in Sec. 2.15. The input parameters and output quantities measured by 2DPHOT are provided in Appendices A.1 and A.2, respectively.

2.2 The 2DPHOT package: tasks and analysis strategy

2DPHOT is designed to have a simple structure consisting of a shell script running a suite of C and Fortran77 programs developed using freely available software libraries. 2DPHOT works on both single-chip and wide-field (up to 16000x16000 pixels) images with a set of input parameters provided either at the invocation of the shell script through a command line syntax or a corresponding graphical interface (see Sec. 2.14). Thus, the package can be used either as a standalone application or via a dedicated web-based interface. The list of input parameters along with short descriptions is provided in Appendix A.1.

The main tasks of 2DPHOT are:

1. Producing a cleaned catalog of the image.

2. Performing reliable star/galaxy classification.

3. Estimating the completeness of the galaxy catalog and the contamination due to star/galaxy misclassification.

4. Constructing an accurate model of the Point Spread Function (PSF) of the input image, taking into account possible spatial variations of the PSF as well as deviations of stellar isophotes from circularity.

5. Deriving structural parameters of galaxies by fitting galaxy images with two-dimensional PSF-convolved Sérsic models.

6. Measuring galaxy isophotes by fitting them with Fourier-expanded ellipses, and derivation of one-dimensional surface brightness profiles of galaxies.

7. Measuring the growth curve of seeing corrected aperture magnitudes of galaxies.

All of these tasks are part of an image analysis flow and are strictly linked with each other such that the output from one task is used as input to the subsequent tasks. Figure 2.2 provides a graphical representation of this flow, where the boxes represent different steps in the image analysis and the arrows follow the image processing timeline. The package starts by running S-Extractor (Bertin & Arnouts, 1996) on the input image through an iterative procedure, allowing simultaneous measurement of the seeing FWHM and removal of spurious object detections. Stamp and mask images are then extracted for each object in the cleaned catalog, and are used to model the PSF across the field and to obtain a coarse estimate of the Sérsic parameters of the detected sources. Using both the S-Extractor stellarity index and the coarse effective radius estimates, 2DPHOT performs star/galaxy classification. The selected galaxies are then analyzed using a two-dimensional fitting procedure as well as a full isophotal analysis. Seeing corrected aperture magnitudes are also estimated. At this point, simulations are performed to estimate completeness and contamination of the final catalog.

The following sections describe all of the image analysis steps, following the diagram in Fig. 2.2. The output quantities measured by 2DPHOT are summarized in Appendix A.2.

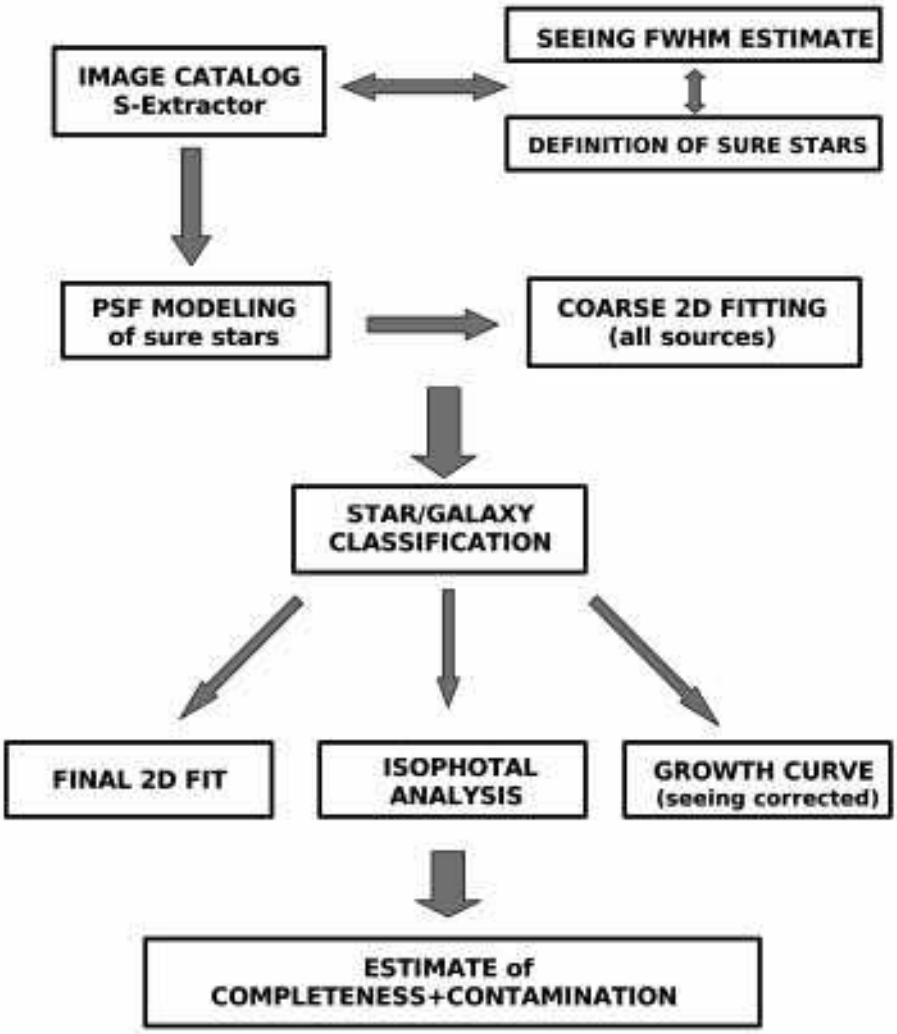

Figure 2.2: Image analysis flow of 2DPHOT.

2.3 The image catalog and the definition of 'sure stars'

2DPHOT produces the source catalog from the input image using the S-Extractor package (Bertin & Arnouts, 1996). Star/galaxy separation is performed on the basis of the S-Extractor stellarity index SI and the effective radius parameter r_e (see Sec. 2.7.1). In order to obtain a reliable estimate of the stellarity index, the seeing FWHM of the input image has to be provided to S-Extractor via the $SEEING\_FWHM$ input parameter (see the S-Extractor documentation[4]). To measure this, 2DPHOT produces a catalog from the input image via the following two-step procedure. S-Extractor is first run for the sole purpose of detecting sources in the input image and calculating their Kron magnitudes and FWHM and ELLIPTICITY parameters. By applying a 3σ clipping procedure to the FWHM and ELLIPTICITY distributions of all the bright ($S/N > 100$) unsaturated objects, 2DPHOT generates a preliminary list of candidate stars. The peak value f and the width σ of the FWHM distribution of these objects is derived using the bi-weight estimator (Beers et al. 1990). The values of f and σ define what we call the *sure star locus*, with the *sure stars* being the objects that lie within $\pm 2\sigma$ of f. Given the values of f and σ, S-Extractor is run a second time by setting the $SEEING\_FWHM$ parameter to the value of f. As an example of this procedure, Fig. 2.3 shows the FWHM versus S/N ratio diagram for all the detected sources in two CCD images of the galaxy cluster Abell 2495, which has been observed twice, under different seeing conditions, as part of the Palomar Abell Cluster Survey (Gal et al. 2000). Similar figures showing the sure star locus and the sure stars are automatically produced during each run of 2DPHOT.

After sure stars are defined, the catalog is cleaned of spurious detections by excluding all sources 3σ below the sure star locus. Objects whose distance from the image edges, in units of their FWHM value, is smaller than *REDGE*, where *REDGE* is one of the input 2DPHOT parameters (see Appendix A), are also excluded from the analysis since their photometry can be incomplete and/or corrupted.

[4] http://terapix.iap.fr/rubrique.php?id_rubrique=91/index.html

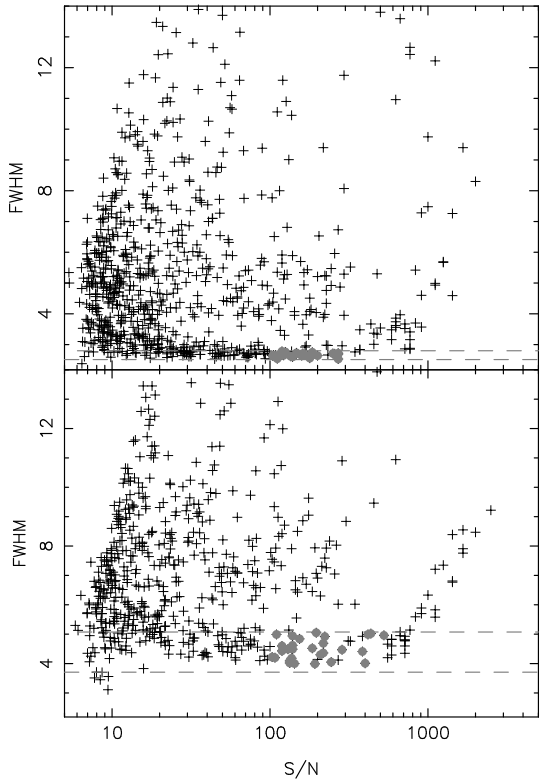

Figure 2.3: Locus of sure stars for two g-band images from the Palomar Abell Cluster Survey (see Gal et al. 2000). The images were obtained with a SITe 2048×2048, AR-coated CCD, at the Palomar 1.5m telescope, and cover an area of $12.56' \times 12.56'$ around the cluster of galaxies Abell 2495 (at $z \sim 0.09$), with a pixel scale of 0.368 arcsec/pixel. Both panels plot the $FWHM$ vs. S/N diagram for all the sources in the same field around the cluster center, with the lower panel for the image taken in worse seeing conditions ($FWHM \sim 1.7''$). The $FWHM$ is given in pixel units. The S/N ratio is computed as the inverse of the uncertainty on the S-Extractor Kron magnitude (neglecting readout noise). The locus of sure stars is defined by the two horizontal dashed gray lines that mark the $\pm 2\sigma$ region around the peak value of the FWHM distribution of star candidates. Sure stars are defined as the bright ($S/N > 100$) non-saturated star candidates which lie inside the sure star locus, and are plotted as grey circles (see text).

2.4 Extraction of stamps and mask images

For each detected source, 2DPHOT extracts an image section (stamp) centered on the source. The area of the stamp is proportional to the $ISOAREA$ output by S-Extractor such that a wide sky region around the central object is also included in the stamp. This allows a reliable estimate of the local background to be obtained from the two-dimensional fitting program (see Sec. 2.8). For each stamp, a mask image is also produced by flagging all the pixels that belong to all the other sources in the input image whose isophotal areas overlap the given stamp. The isophotal areas are defined through the $ISOAREA$, $ELLIPTICITY$, and position angle (PA) parameters from S-Extractor, by multiplying the $ISOAREA$ value by an expansion factor $EXPND$ (with a default value of 1.5), which is an input parameter of 2DPHOT. This expansion factor allows us to mask also the faintest diffuse external regions of each object. Sources whose isophotal areas overlap the central source by more than 50% are not masked out and are analyzed simultaneously with the central object (see Secs. 2.6 and 2.8). For each stamp, the number of sources treated simultaneously is written into the $NOBJ$ keyword of the corresponding mask file header. The local background value and its standard deviation are also estimated by applying biweight statistics to all the pixels which do not belong to the isophotal area of the central source and are not flagged in the mask file. These values are stored in the keywords M_BK and S_BK of the mask file header, respectively. Fig. 2.4 shows some examples of the stamp and mask images automatically produced by 2DPHOT.

2.5 PSF modeling

The Point Spread Function (PSF) is modeled by fitting the images of sure stars (Sec. 2.3) with a sum of two-dimensional Moffat functions. In order to account for PSF asymmetries, the isophotes of each Moffat function are described by ellipses, whose shape is modulated with a sin/cos angular expansion, similar to that adopted for describing deviations of the isophotal shape of early-type galaxies from pure ellipses (see e.g. Bender & Moellenhoff 1987). The number of fitted stars is given by the lesser of the number of available sure stars and the 2DPHOT input parameter $NSMAX$. The value of $NSMAX$ is chosen as a compromise between the computation time for the fitting algorithm and

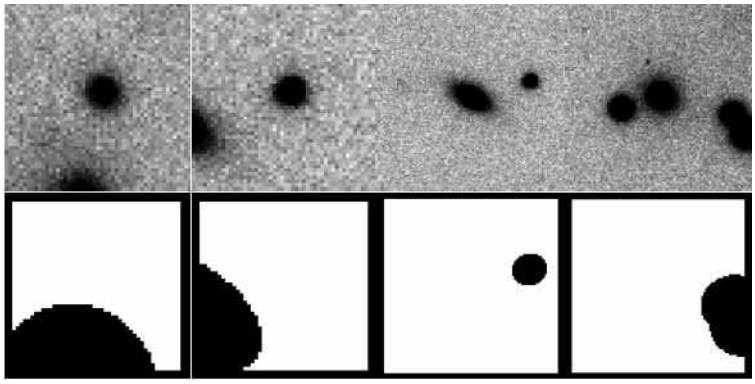

Figure 2.4: Examples of stamp and mask images produced by 2DPHOT. Stamps are shown in the upper panels, while lower panels plot the corresponding mask images. The pixels which are flagged in the mask files are plotted in black. Notice that pixels very close to the stamp edges are also flagged in each mask image. This is done to reduce computational overhead in the 2D fitting algorithm.

the accuracy of the PSF model. Increasing $NSMAX$ yields more accurate PSF models at the cost of larger computational times. Usually, values of $NSMAX$ in the range of 3 to 5 give reliable results[5]. To account for possible spatial variations of the PSF across the chip, 2DPHOT provides two PSF modeling options. In the first case, a global PSF model is obtained by simultaneously fitting $NSMAX$ stars randomly extracted from the entire list of sure stars. As a second option, 2DPHOT can construct a two-dimensional grid on the input image and derive a PSF model independently for each cell, by randomly selecting up to $NSMAX$ stars among the available sure stars. PSF models are only derived for cells including at least two sure stars. The cell size has to be provided through the 2DPHOT input parameter $NSIZE$. 2DPHOT associates to the PSF model of each cell the median values of the x and y coordinates of the corresponding fitted sure stars and the two-dimensional modeling of each galaxy (Sec. 2.8) is performed by using its closest PSF model. In order to avoid a discretely varying PSF across the chip, the user can also choose to adopt a locally interpolated PSF model. For each galaxy, 2DPHOT selects the PSF models of the cells around the galaxy itself, and performs a bi-linear interpolation of the selected models at the galaxy position. Since there is a strong correlation between the fitting parameters of each PSF model, 2DPHOT does not derive the local PSF model by interpolating each single fitting parameter. Instead, the interpolation is performed inde-

[5] Processing several images, we found that increasing the value of $NSMAX$ to more than 5 stars does not significantly change the output of 2DPHOT.

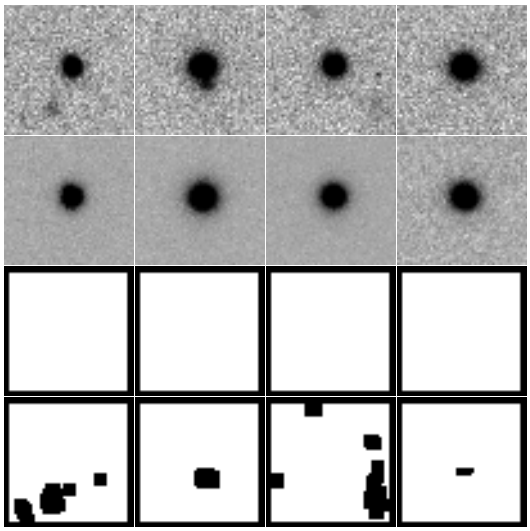

Figure 2.5: Examples of the clipping procedure applied to four stellar images. Each column corresponds to a different star. From top to bottom, the star stamps, the corresponding median images, the original mask images and the updated mask images are shown. The four stars have been selected because S-Extractor fails to detect the faint sources around them, and therefore the corresponding mask images are blank. The 2DPHOT clipping procedure detects the missed faint sources, and masks them in the updated mask images. The star stamps and median images use the same intensity scale.

pendently for each pixel of the PSF models, by interpolating the corresponding intensity values.

Prior to fitting the PSF, 2DPHOT applies a clipping procedure to remove stars that might be contaminated by nearby objects. For each star, all of the other sure stars are co-registered to the same center coordinates and median stacked. An rms image is constructed by estimating, at each position, the standard deviation of the stacked pixels. The mask images of the sure stars are then updated by flagging all the pixels which deviate by more than 5σ from the corresponding median images. If the fraction of flagged pixels is larger than 20% of the total mask image area, the sure star is considered to be strongly contaminated and it is excluded from the PSF fitting. This procedure allows faint sources which may not have been detected by S-Extractor to be masked, and to exclude objects which are misclassified or blended with nearby sources. Some examples of the clipping and mask update algorithms are shown in Fig. 2.5, while Fig. 2.6 plots an example of the PSF modeling results. The latter figure is automatically produced by 2DPHOT.

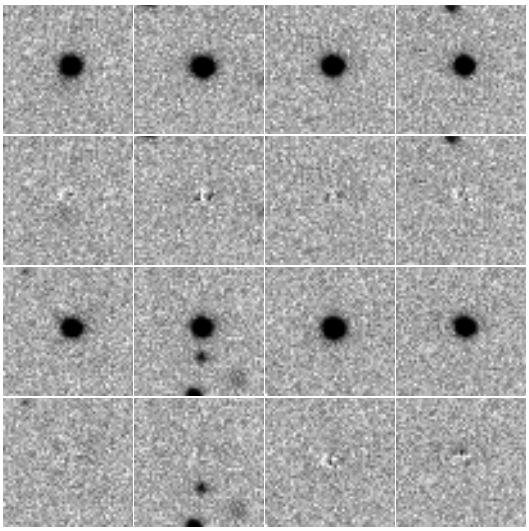

Figure 2.6: Examples of PSF modeling. Eight star images have been fitted simultaneously using a sum of three two-dimensional Moffat functions (see the text). For each star, two panels are shown, with the upper panel plotting the stamp image and the lower panel showing the fitting residuals.

2.6 Coarse 2D fitting

2DPHOT produces an initial estimate of structural parameters for all objects in the input image using a discrete, coarse two-dimensional fitting algorithm (INI2DF). For each object, a set of PSF-convolved Sérsic models is constructed by varying the effective radius $r_{\rm e}$, the total magnitude $m_{\rm T}$, and the Sérsic index n. 'Geometric' parameters, such as the center coordinates, the axis ratio and the position angle of the models are estimated by fitting the object image with a single 2D Moffat function, and are kept fixed during the coarse fitting. The Moffat fit is performed by excluding the inner part of the object, which is strongly affected by seeing. The local background value is also not changed in the fit and is obtained from the keyword M_BK in the mask image header (see Sec. 2.4). INI2DF changes the effective radius of the Sérsic model using an adaptive grid of 10 values computed on the basis of both the pixel scale and the seeing FWHM of the image. Four different values are considered for the Sérsic index parameter, $n = \{1, 3, 5, 7\}$, while the total magnitude can take the values $m_{\rm T} = m_{\rm K}, m_{\rm K} - 0.2, m_{\rm K} - 0.4$, where $m_{\rm K}$ is the Kron magnitude of the source (S-Extractor $MAGAUTO$). We point out that the grids of $r_{\rm e}$, $m_{\rm T}$ and n values have been empirically chosen by analyzing several images with a wide range of characteristics (e.g. optical and near-infrared data as well as ground-based and HST

images). We found that further increasing the grid size does not change significantly the 2DPHOT results. With these sizes for the r_e, m_T and n grids, INI2DF produces a total of 120 discrete models, each of which is compared to the object image by computing the corresponding χ^2 value. The coarse structural parameters are given by the parameters of the model with lowest χ^2.

In the case that, for a given stamp, several objects have to be treated simultaneously (see Sec. 2.4), the above procedure is modified as follows. A simultaneous fit is performed by using a single two-dimensional Moffat function for each object. Then, for each overlapping object, the others are subtracted using the fitted Moffat models. A corresponding updated mask image is also produced, by flagging all pixels for which the sum of the subtracted Moffat models exceeds the local background standard deviation (S_BK in the mask image header, see Sec. 2.4) by $> 50\%$. Coarse structural parameters are then obtained by fitting each object in the stamp as a single source, applying the same procedure outlined above. Some examples of this procedure are shown in Fig. 2.7. We see that there are some cases where the single Moffat models do not result in accurate subtraction of overlapping sources. Nevertheless, we found that the above approach allows reliable estimation of structural parameters, with the great advantage of significantly reduced computational times compared to an approach where overlapping galaxy models are fitted simultaneously (see also Sec. 2.8).

2.7 *Identification of stars in the 2DPHOT package*

The classification of stars and galaxies is one of the most challenging issue in the analysis of astronomical images, and there is no method that works in all scenarios as the optimum classifier. In the current version, 2DPHOT adopts a simple method of star/galaxy (hereafter S/G) separation, which is based on both the S-Extractor and the coarse structural parameters estimated by the INI2DF procedure (Sec. 2.6). The parameters which are used for S/G classification have been chosen on the basis of Monte-Carlo simulations as detailed in Sec. 2.7.1, while the S/G classification algorithm is described in Sec. 2.7.2. In the future, we plan to implement more complex classification techniques (such as wavelet approaches), and provide a detailed comparison of their performance. Since there is no method that correctly classifies all sources in a given image, particularly at the faintest flux levels, it

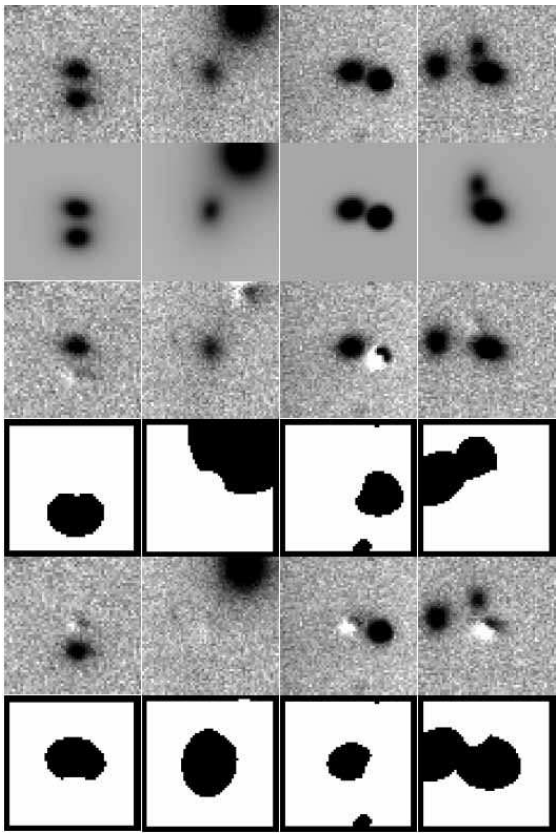

Figure 2.7: Coarse fitting of double objects. Panels in each column of the figure correspond to a different stamp image. From the top row down, the first and second rows show the stamp image and the corresponding Moffat model, obtained by simultaneously fitting two single Moffat functions. The third panel shows the subtraction of one Moffat function from the stamp. The corresponding updated mask image is shown in the fourth panel. The lowest two panels show the Moffat subtracted image and the corresponding updated mask image for the second object in the stamp.

is crucial that every classification framework provide an estimate of contamination due to misclassified sources as a function of the S/N ratio. As described in Sec. 2.7.2, 2DPHOT accurately estimates such contamination using simulated stars and galaxies added to the input processed frame.

2.7.1 Reliable parameters to identify stellar sources

We adopt a two-step procedure to establish useful parameters for star/galaxy separation. First, we look for reliable classifiers of point-like sources, i.e. 2DPHOT output parameters whose values for stellar sources lie in a narrow region of parameter space over wide ranges of the S/N ratio, seeing, and sampling characteristics of the images. Then, we analyze the ability of such classifiers to separate stars and galaxies by examining the distribution of values they assume for both kinds of objects. To address the first point, we created simulated CCD images, each with a random spatial distribution of stars. The simulations were generated using the same pixel scale ($0.369''/pix$), image size (2048×2048 pixels), and the noise properties as the r-band images of the Palomar Abell Cluster Survey (hereafter PACS, see Gal et al. 2000). The PACS data have been extensively processed from the authors through the 2DPHOT package in order to investigate the effects of environment on internal color gradients of early-type galaxies (see La Barbera et al. 2005).

Stellar images were simulated using both the Gaussian profile and the Moffat law:

$$P(r) = \left[1 + \left(\frac{r}{r_{\rm c}}\right)^2\right]^{-\beta}, \tag{2.1}$$

where $P(r)$ is the surface brightness of the star as a function of the distance r to its center, β is the shape parameter of the profile, and $r_{\rm c}$ is the Moffat scale radius, which is related to the FWHM by $FWHM = 2r_{\rm c}(2^{1/\beta} - 1)^{0.5}$. For the Moffat fits, we set $\beta = 3$, which is the mean value for stellar images in the PACS, and we varied the $FWHM$ from one star to another within each simulated image according to a normal deviate with central value $< FWHM >$ and width σ_{FWHM}. Four simulated fields were created, labeled F1, F2, F3, and F4. The main simulation parameters are summarized in Table 2.1. For each field, we randomly created $N = 500$ stars, and we set the parameters $< FWHM >$ and σ_{FWHM} as follows. For field F1, both the $< FWHM >$ and σ_{FWHM} are set to the median values measured from the r-band PACS images. Fields F2 and F3 simulate observations

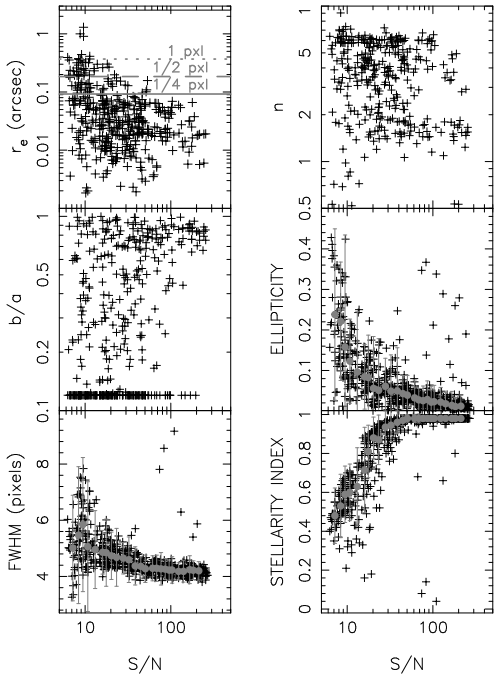

Figure 2.8: Measured parameters for sources in field F1 as a function of their S/N ratio. The quantities r_e, b/a, and n are the effective radius, axis ratio, and Sérsic index obtained from 2DPHOT through the two-dimensional fitting procedure (see Sec. 2.8). The other parameters (ELLIPTICITY, FWHM, and STELLARITY INDEX) are those measured by S-Extractor. As shown in the plot, the solid, dashed and short-dashed grey lines in the upper-left panel mark the values of r_e corresponding to different pixel fractions. Grey circles and corresponding error bars in the ELLIPTICITY, FWHM, and STELLARITY INDEX panels have been obtained by binning the data with respect to the S/N ratio, and correspond to the mean and 1σ interval in each bin.

with worse seeing conditions. F2 has the same $< FWHM >$ as F1 while σ_{FWHM} is doubled, mimicking the case of large scatter in the seeing $FWHM$ across the image. Field F3 has the same σ_{FWHM} as field F1 but higher $< FWHM >$, corresponding to either observations taken in worse mean seeing or data with better PSF sampling. Finally, for Field F4, we used the same $< FWHM >$ and σ_{FWHM} values as F1, but stellar images were created with Gaussian profiles. We note that the above simulated images span all the possible cases that have been found when processing the PACS images, and because of their wide range of seeing parameters, they also reproduce the seeing properties of a variety of ground-based images.

Catalogs of the simulated stellar fields were generated as described in Sec. 2.3. For each field, all detected sources were fit with PSF convolved Sérsic models, following the

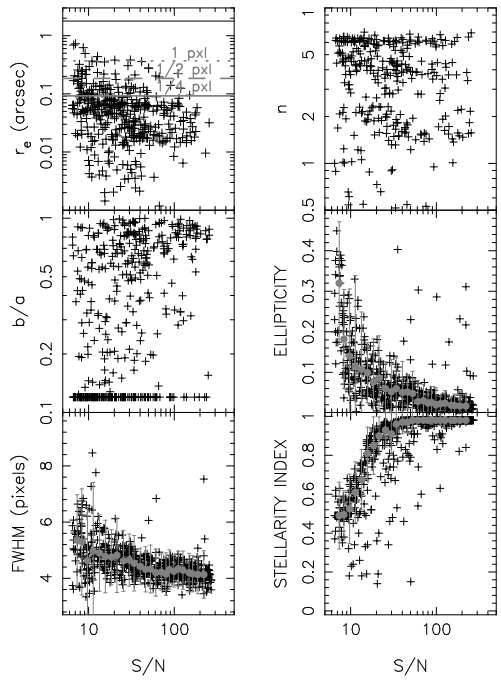

Figure 2.9: Same as Fig. 2.8 but for Field F2.

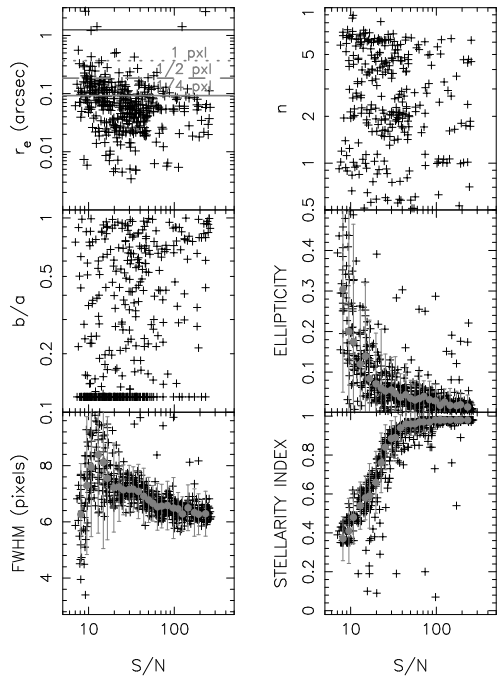

Figure 2.10: Same as Fig. 2.8 but for Field F3.

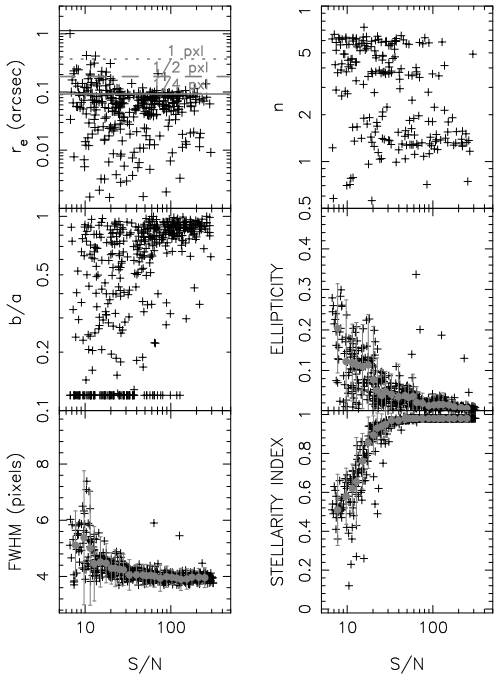

Figure 2.11: Same as Fig. 2.8 but for Field F4.

procedure described in Secs. 3–5 and running the final two-dimensional fitting program (see Sec. 2.8). Figs. 2.8, 2.9, 2.10 and 2.11 plot the Sérsic parameters, i.e. the effective radius r_e, the Sérsic index n, and the axis ratio b/a, as well as the the ELLIPTICITY, FWHM and stellarity index (hereafter SI) parameters from S-Extractor as a function of the S/N ratio of sources in fields F1, F2, F3, and F4, respectively. The S/N ratio was computed as the inverse of the uncertainty on the S-Extractor Kron magnitude. From Figs. 2.8, 2.9, 2.10 and 2.11, we draw the following conclusions:

i) The Sérsic index and the b/a parameters are not reliable classifiers. The scatter in these quantities is large compared to the range of values they can assume.

ii) The effective radius is a reliable classifier, in the sense that its values are always well limited to a given region of the corresponding parameter space. Whatever the seeing conditions are, the effective radius of stars is always smaller than ~ 1 pixel, and for $S/N > 30$, the values of r_e are always smaller than ~ 0.5 pixel.

iii) As one would expect (Bertin & Arnouts, 1996), the SI parameter of S-Extractor is a reliable classifier. Its values can range from 0 to 1, but for the simulated stars with

Table 2.1 - Parameters of simulated stellar fields. Cols. 2, 3 and 4 give the $FWHM$, $\sigma_{FWHM}/FWHM$ and β parameters (see text). The gain, zero-point and read-out noise are in Cols. 5, 6 and 7, respectively. In the case of field 4, stellar images have Gaussian profiles.

$Field\#$	$FWHM$ (pxls)	$\frac{\sigma_{\rm FWHM}}{FWHM}$ %	β	$gain$ e^-/ADU	$zpoint$	$rnoise$ e^-
1	4	3	3	1.62	30.75	6.3
2	4	3	3	1.62	30.75	6.3
3	6	6	3	1.62	30.75	6.3
4	4	3	Gaussian	1.62	30.75	6.3

$S/N > 20$, the values of SI are always larger than ~ 0.7.

iv) The FWHM and ELLIPTICITY parameters are good potential classifiers as well, although the values of FWHM are obviously strongly dependent on the seeing characteristics of the analyzed image. Generally, we find that using the FWHM and ELLIPTICITY parameters does not lead to any significant improvement in star/galaxy separation, and thus we elected not to use these parameters.

The reliability of the above star/galaxy separation scheme, based on the SI–r_e diagram, will be further addressed in Section 13, where we will assess contamination and completeness as measured by 2DPHOT.

2.7.2 Final rules for identifying stellar sources

After INI2DF parameters have been derived for all the sources in a given field, 2PHOT performs S/G separation. Simulated images are constructed by adding a spatially random distribution of stars to the input image. The surface density of stars is chosen so that 50 stars are added to an input image area of 2000×2000 pixels, while the number of simulations is such that we have a total of 2000 artificial stars. Stars are created from the actual PSF model, with photon noise added based on the GAIN provided in the *default.sex* file of S-Extractor. The magnitude of each artificial star is extracted according to a uniform random S/N ratio distribution, with an upper cutoff of $S/N = 200$. For each simulation, a new catalog is generated and the SI parameter is computed for all of the artificial stars detected by S-Extractor. 2DPHOT defines star candidates on the basis of the distribution of these

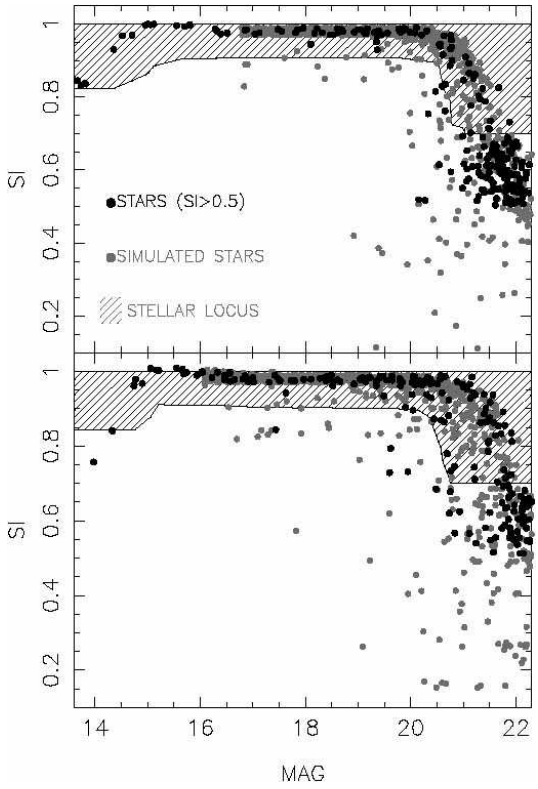

Figure 2.12: Definition of the stellar locus in the 2DPHOT package (see text). The plot shows the stellar index versus Kron magnitude diagram for two g-band images from PACS. Grey points are the simulated stars added to each field by 2DPHOT, while black circles show the objects with stellar index larger than 0.5. The hatched area marks the region used to select star candidates.

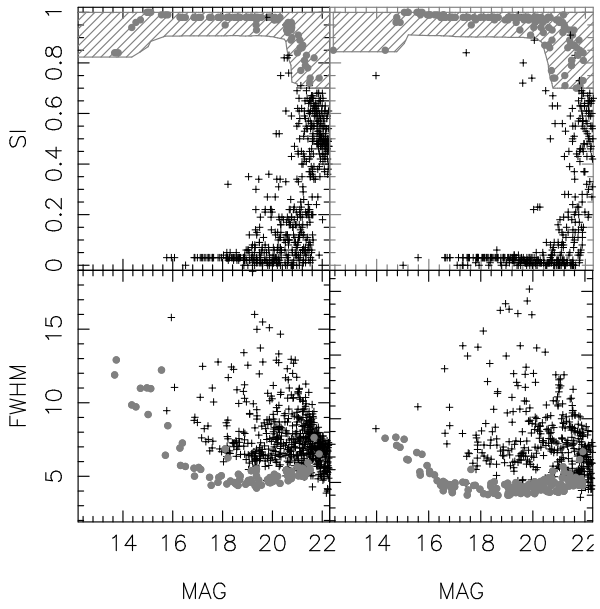

Figure 2.13: Selection of stars in 2DPHOT for two g-band images of the PACS. The upper panels show the stellar index versus Kron magnitude diagrams, with the hatched region showing the stellar locus (see text). Grey circles are the selected stars, while black crosses plot all the remaining sources. Lower panels show the corresponding FWHM versus Kron magnitude plots for the same PACS images.

artificial stars in the SI versus magnitude diagram. First, artificial stars are ordered by ascending magnitude, and the 50 stars with magnitudes closest to that of each artificial star are selected. Then, for each artifial star, the 10% percentile of the distribution of SI values (SI_{10}) of the 50 selected artificial stars is computed. In order to minimize the number of galaxies that are misclassified at the faintest magnitudes of the catalog, a minimum cutoff of $SI_{10} = 0.7$ is imposed (see Sec. 2.12 for details). As a second option, that turns out to be more suitable in the case of deep images (see Sec. 13), 2DPHOT can define the star locus by applying the same procedure outlined above but replacing the 10% percentile of the SI distribution of simulated stars with the quantity $\theta - p\sigma$, where θ and σ are the location and width of the SI distribution, while p is a parameter input to 2DPHOT. With a suitable choice of p this second definition allows a narrower stellar locus to be defined in the SI versus magnitude diagram (see Sec. 2.12), and thus it can be more suitable at faint magnitudes where galaxies with small size are more likely to be misclassified as stars. Hereafter, unless stated explicitly, we will consider only the first definition of the star locus.

Since saturated stars have lower SI than bright unsaturated stars, at magnitudes brighter than those of artificial stars the value of SI_{10} is set to the minimum SI for

observed sources with $SI > SI_{\rm min} = 0.5$. This procedure allows even saturated stars to be correctly classified by 2DPHOT. The value of $SI_{\rm min}$ was chosen empirically based on several processed images where we found that saturated stars always have $SI > 0.5$, while the brightest galaxies all have lower SI. As shown in Fig. 2.12, star candidates are selected using the region between $SI = SI_{10}$ and $SI = 1$ in the SI versus magnitude diagram. Plots like those in Fig. 2.12 are automatically produced by each run of 2DPHOT. If required, the value of $SI_{\rm min}$ can be changed by the user after inspection of the stellar locus plot. As final rules for S/G separation, we define an object as a candidate star if it belongs to the star candidate locus and its INI2DF effective radius is smaller than 1 pixel, with the latter criterion from the results of Monte-Carlo simulations discussed in Sec. 2.7.1. The locus of star candidates should include most of the point-like sources in the input image, with the percentage of misclassified objects increasing as the magnitude increases. The selection of stars through this procedure is shown in Fig. 2.13, where we consider two CCD images from the PACS. The plots in Fig. 2.13 are automatically produced by 2DPHOT. As one would expect, at faint magnitudes the star/galaxy classification becomes progressively more uncertain. The distribution of artificial stars with respect to the locus of star candidates provides a quantitative way to estimate the magnitude (and/or) the S/N limit above which the S/G classification is reliable (see Sec. 2.12).

2.8 *Final 2D fitting*

Objects identified as galaxies through the 2DPHOT S/G classification scheme are then fit with PSF convolved Sérsic models. This 'final' fitting differs from that of Sec. 2.6 since a full χ^2 minimization algorithm is adopted, without using any discrete (coarse) grid of reference convolved models (as for INI2DF), providing a precise estimate of structural parameters at the cost of longer computation times[6]. The χ^2 minimization is performed through the Levenberg-Marquardt algorithm, assigning zero weight to all the flagged pixels in the mask image. The 2D fitting routine adopted in the 2DPHOT package is also described in La Barbera et al. (2002), where several tests of its accuracy have been performed. The initial conditions for the optimization routine are set to the output values of INI2DF, which are on average quite close to the best fitting final parameters. This largely

[6] The CPU time required for the final two-dimensional fitting is 4-5 times longer than for the coarse fit.

reduces the well known issue of spurious convergence that can characterize strongly non-linear optimization problems. The case of overlapping objects is treated with an analogous approach to that described in Sec. 2.6 for the coarse fit. Instead of using the multiple single Moffat fits described in Sec. 2.6, 2DPHOT takes advantage of the INI2DF best fitting models to reduce the final fitting of overlapping objects to that of separate single sources. For each blended galaxy, the overlapping objects are subtracted using the INI2DF models and the mask image is correspondingly updated as described in Sec. 2.6. Although a suitable treatment of overlapping galaxies would require a simultaneous fit to be performed (see e.g. van Dokkum & Franx 1996), reducing the problem to that of fitting single sources greatly decreases computation times. Comparing both approaches, we verified that the 2DPHOT procedure does not produce any significant change in the final structural parameters of *multiple* objects. Some examples of two-dimensional fitting are shown in Fig. 2.14. The plots in this figure are automatically generated by 2DPHOT.

2.9 Isophotal analysis

To analyze the isophotal properties of galaxies, 2DPHOT performs an elliptical fit of galaxy isophotes and measures the deviations of such isophotes from purely elliptical shapes. Details on how the package performs these tasks are given in Sec. 2.9.1. The isophotal fit allows the radial surface brightness profile of galaxies and stars to be extracted. As described in Sec. 2.9.2, the package uses these brightness profiles to obtain a further estimate of the galaxy structural parameters, hence providing an independent estimate of these parameters than that obtained with the full two-dimensional fitting approach (Sec. 2.8). The isophotal analysis is also used to extract a growth curve for each galaxy's aperture magnitude. The aperture magnitudes are computed and corrected for seeing effects as described in Sec. 2.10.

2.9.1 Isophotal fitting

For the measurement of galaxy isophotes, the package first defines the corresponding isophotal intensity values. For each stamp, a rough estimate of the object center coordinates are obtained as the intensity-weighted means of the x and y pixel coordinates. The mean values are computed in a section of 5x5 pixels around the intensity peak of the

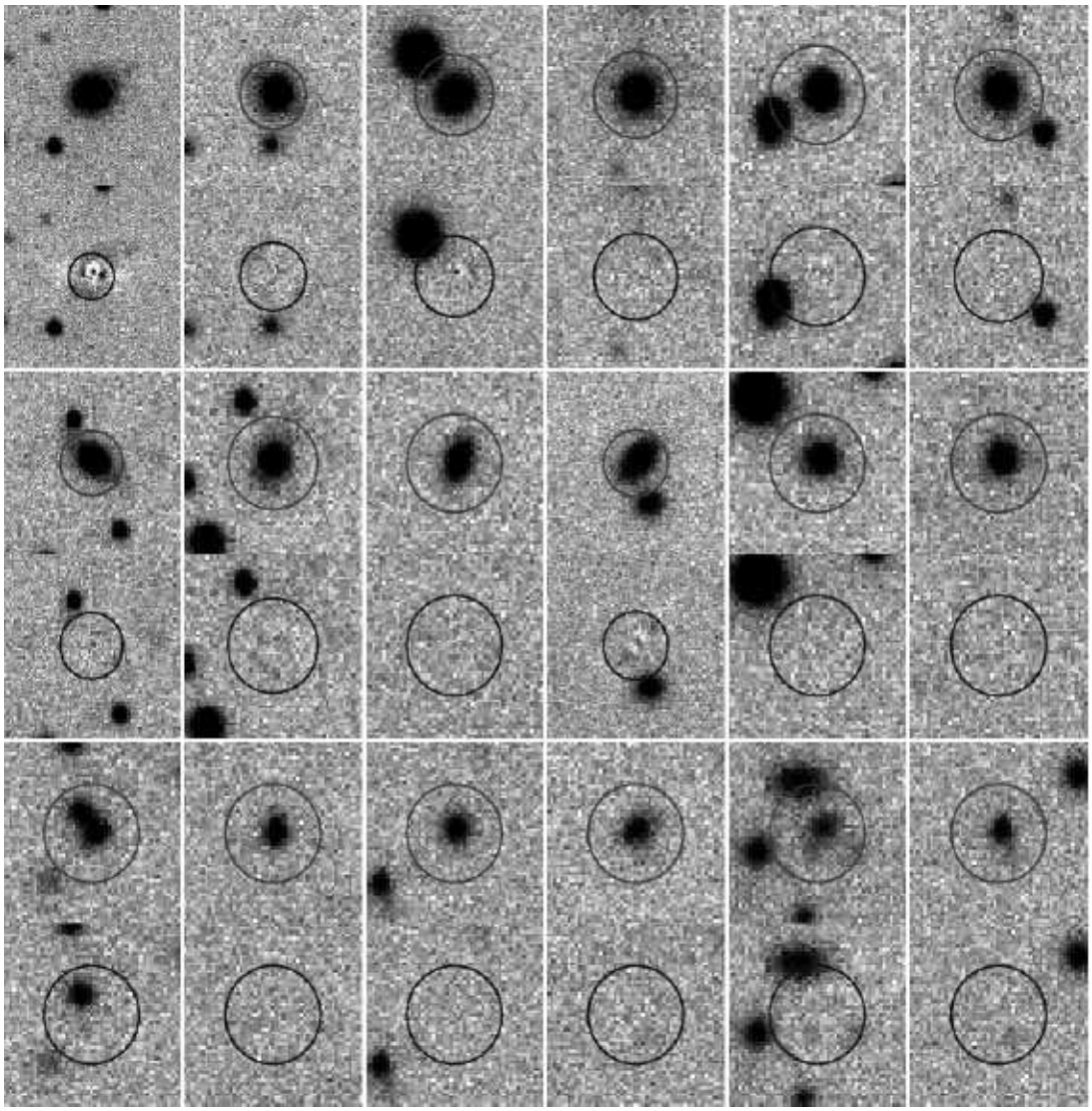

Figure 2.14: Two-dimensional fitting of galaxy stamps with seeing-convolved Sérsic models. Subpanels show the galaxy stamps and the corresponding residual images, obtained by subtracting the model fit from the galaxy stamp. For each galaxy stamp, the fitted galaxy is marked by a grey circle of radius $2''$, and the lower subpanel shows the residual image, where the same circle is plotted in black. From top to bottom and left to right, galaxies are shown in order of decreasing magnitude, from $r \sim 16.5$ for the upper-right panel to $r \sim 20$ for the lower-right panel. Images are drawn from the PACS image of the cluster Abell 574 at $z \sim 0.185$.

object. Using these center coordinates, a set of concentric circles is constructed, with radii equally spaced by 0.5 pixel. For each circle, the mean value of 90 intensity samples equally spaced in polar angle is computed via cubic interpolation of the stamp intensity values at the corresponding radial and polar coordinates. The mean intensity values provide an initial estimate of the object surface brightness profile, and are used to derive the isophotal contours of the object. This procedure allows us to construct isophotal contours whose equivalent radii are approximately equally spaced by 0.5 pixels. For a given isophotal intensity level, I, the isophote is defined by a set of x and y pairs on the stamp. These isophotal samples are defined as follows. For a given polar angle θ, different intensity values I_j are computed at several radii r_j from the galaxy center. The algorithm selects the smallest radius at which the intensity brackets the value of I (i.e. $I_j \leq I \leq I_{j+1}$ or $I_{j+1} \leq I \leq I_j$). The radius r corresponding to this intensity is then computed by linear interpolation of the r_j values with respect to I_j. The isophotal samples are directly computed from r and θ by varying θ such that the number of samples is proportional to the isophote length and by excluding those points flagged in the mask file. To exclude low signal-to-noise regions, the isophotal computation is stopped when the background-subtracted value of I falls below four times the background standard deviation within the stamp. As a default, to exclude galaxies whose isophotal contours are overly affected by seeing, 2DPHOT performs the isophotal analysis only for galaxies whose S-Extractor isophotal radius[7] is larger than four times the seeing FWHM. Figure 2.15 plots some example of isophotal analysis for galaxies from one PACS r-band image. The panels shown in the plot are automatically produced by 2DPHOT.

Galaxy isophotes are modeled as described in Bender & Moellenhoff (1987), by fitting each isophote with an elliptical contour modulated by the following sin/cos angular expansion:

$$\sum a_n \cdot cos(n\theta) + b_n \cdot sin(n\theta), \tag{2.2}$$

where θ is the polar angle, and the sum is done with respect to the index n. For $n \geq 3$, the coefficients a_n and b_n describe the deviations of the isophotes from the elliptical shape. In particular, the a_4 term is used to describe the boxy ($a_4 < 0$) and disky ($a_4 > 0$) isophotal shapes of early-type galaxies. Each ellipse is characterized by five fitting parameters, which

[7] This is defined as $\sqrt{ISOAREA/\pi}$, where $ISOAREA$ is the ISOAREA parameter of S-Extractor.

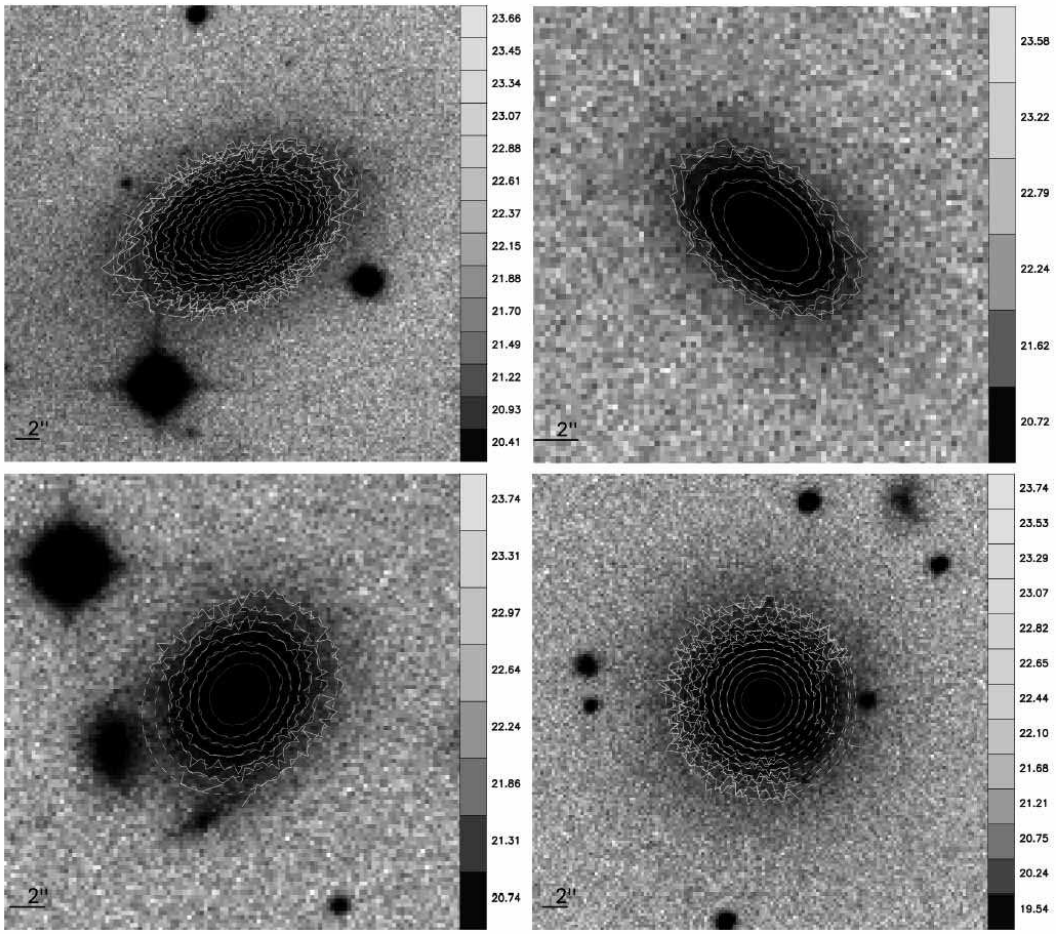

Figure 2.15: Fitting of galaxy isophotes with elliptical contours modulated by a series of sin/cos angular functions. The four panels plot different galaxy stamps, extracted from one PACS r-band image. Solid lines are the isophotal contours, derived as described in Sec. 2.9.1, while the fitted ellipses are plotted as dashed curves. The fits were performed by including only the a_4 term in the sin/cos expansion. Isophotes are plotted with different gray levels, with the grayscale proportional to the corresponding surface brightness value. The relations between gray intensity and surface brightness are shown on the grey scales at the right of each panel. Each surface brightness value on these gray scales corresponds to a different isophote. Surface brightness values are given in units of $mag/arcsec^2$, and become brighter as the isophotal color changes from white to black. The spatial scale is shown in the lower-left corner of each panel. The plots are produced automatically from the 2DPHOT package.

are its center coordinates, equivalent radius, ellipticity, and position angle of the major axis. The sin/cos terms which have to be included in Eq. 2.2 are defined as input parameters of 2DPHOT. The isophotal parameters are derived by a χ^2 minimization procedure, through a Levenberg-Marquardt algorithm. Examples of isophotal fits are shown in Fig. 2.15, while Fig. 2.16 shows the radial profiles of isophotal parameters measured by the fitting procedure. All of these plots are automatically produced by 2DPHOT. Global values of a_n and b_n are computed as follows. Following Bender & Moellenhoff (1987) and Bender et al. (1989, hereafter B89), only the range of a_n and b_n profiles between a minimum radius R_{min} and a maximum radius R_{max} is selected. R_{min} is set to four times the seeing FWHM of the image, while R_{max} is set to twice the galaxy effective radius. The global a_n and b_n values are then defined as the average of their profiles within the selected radial range. Fig. 2.17 compares the a_4 values of B89 with those derived by running 2DPHOT on the r-band images of 42 galaxies from B89 with available photometry from the Sloan Digital Sky Survey Data Release 5 (SDSS DR5). Ten out of the 42 galaxies have been observed multiple times in the SDSS, and we used these repeated observations to check the reliability of the a_4 values. Looking at the figure, we see that there is good agreement between the two sets of measurements. Moreover, there is excellent agreement among repeated a_4 measurements. We note that 2DPHOT measures global a_n and b_n values somewhat differently than Bender et al. (1989), where either the peak values or the values of a_n and b_n at one effective radius were considered. Using the mean values has the advantage of producing more robust estimates, reducing the effects of possible spurious peaks in the a_n and b_n profiles that can arise from noise fluctuations. Furthermore, as shown above, the two methods give, on average, fully consistent results.

2.9.2 Measuring surface brightness radial profiles

For all galaxies with final 2D fitting parameters, 2DPHOT extracts a one dimensional surface brightness profile. Four galaxy isophotes, corresponding to intensity values of 4, 6, 8, and 10 background standard deviations over the background level are computed, and are fitted by elliptical contours, as described in Sec. 2.9.1. The values of center coordinates, axis ratio, and position angle of the fitted ellipses are averaged, and are used to construct several concentric ellipses on the galaxy stamp, with their equivalent radii equally spaced

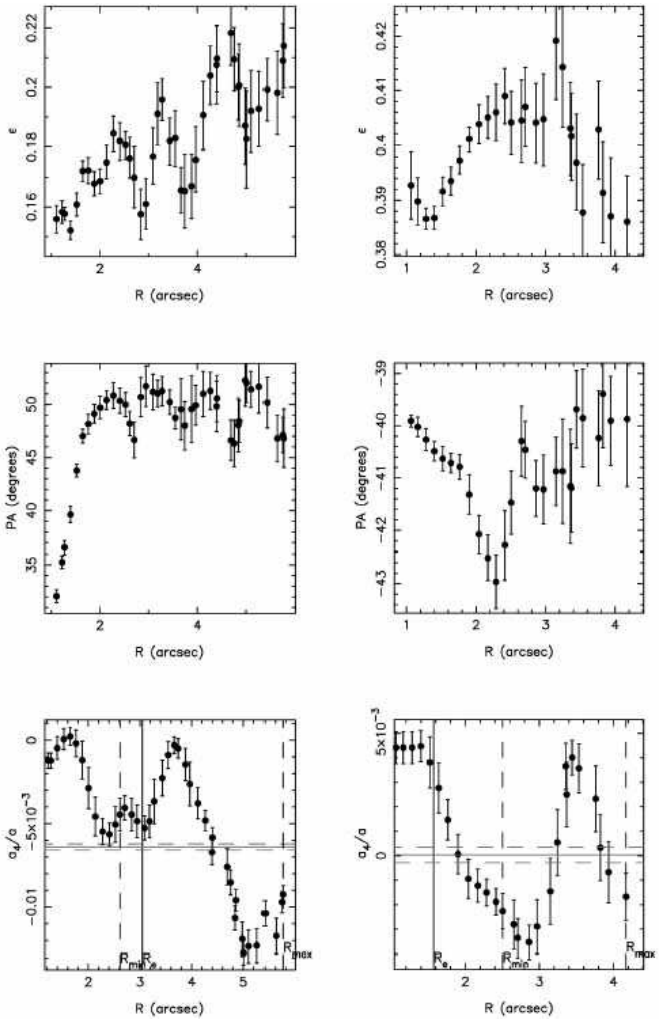

Figure 2.16: Radial profiles of isophotal parameters, as derived by the isophotal fitting algorithm (Sec. 2.9.1). From top to bottom, the panels show the profiles of ellipticity, position angle of the ellipse's major axis, and a_4 coefficient as a function of the equivalent radius of the fitted isophotes. Left panels show the profiles of the galaxy in the lower-left panel of Fig. 2.15, while right panels correspond to the galaxy shown in the upper-right panel of Fig. 2.15. Error bars mark one sigma standard uncertainties. In the bottom panels, the minimum and maximum radii to define the global a_4 value are shown as vertical dashed lines. The effective radius of the galaxy is marked by a vertical solid line. The solid horizontal gray line denotes the mean value of a_4 in the selected radial range, while the dashed gray lines mark the corresponding 1σ interval.

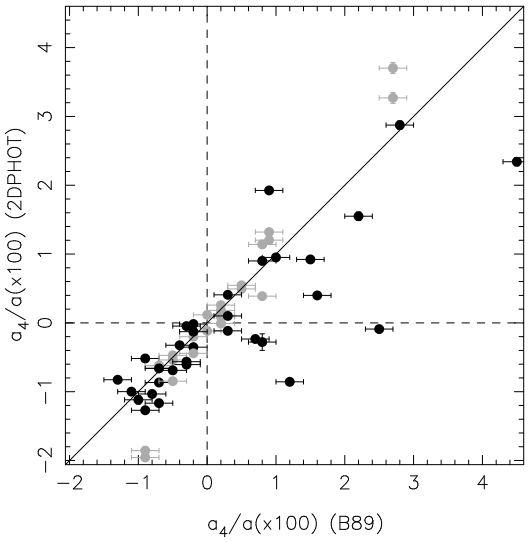

Figure 2.17: Comparison of a_4 values as estimated from Bender et al. (1989) (horizontal axis) and from the 2DPHOT package (vertical axis). The new a_4 values were obtained by running the 2DPHOT package on r-band images of the 42 galaxies of Bender et al. (1989) with available photometry from the SDSS DR5. The horizontal error bars mark the typical uncertainty on the a_4 values of Bender et al. (1989) (see their sec.2). The vertical bars denote one sigma standard uncertainties as estimated from 2DPHOT. Most of these error bars are smaller than the symbol size in the plot. From left to right, the following galaxies are plotted: NGC4261, NGC4365, NGC4387, NGC5322, NGC3605, NGC5127, NGC4478, NGC5532, NGC3894, NGC4551, NGC4406, NGC5576, NGC4649, NGC4374, NGC4472, NGC3842, NGC6411, NGC4636, NGC4489, NGC3608, NGC3640, NGC4486, NGC5638, NGC3379, NGC3193, NGC4494, NGC5490, NGC5831, NGC3613, NGC4382, NGC4168, NGC5845, NGC4125, NGC4473, NGC2693, NGC3377, NGC4621, NGC4550, NGC4564, NGC3610, NGC4660, NGC4251, NGC4570. In several cases, a galaxy has repeated SDSS observations. Such cases have been processed independently by 2DPHOT, and the corresponding values are plotted as gray symbols in the figure.

by 0.5 pixel. The one dimensional surface brightness profile is then obtained as described in Sec. 2.9.1, by computing the mean intensity value in each ellipse as a function of the ellipse equivalent radius. The brightness profile is sky-subtracted by applying a similar procedure to that described in Jorgensen et al. (1995). The outermost part of the surface brightness profile intensities is fit with a power law, $\alpha \cdot r^{-\beta} + bg$, where the α and β parameters as well as the local background value, bg, are estimated by a χ^2 minimization procedure. The outermost part of the profile is defined as that radial range where the mean isophotal intensity minus an approximated median background falls below twice the background standard deviation. Some examples of surface brightness profiles for the same galaxies as in Fig. 2.15 are shown in Fig. 2.18. The profiles are used to obtain a further estimate of galaxy structural parameters independent of the 2D fitting approach. 2DPHOT follows the procedure described by Bendinelli et al. (1982, hereafter BPZ82). In this approach, one assumes the surface brightness distributions of both the galaxy and the PSF to have circular symmetry. With this assumption, it can be shown that the 2D seeing convolution is reduced to a one dimensional integral, with the integrand given by the product of the surface brightness radial profile of the galaxy model with that of the PSF surface brightness profile, modulated by a zero-order modified Bessel function (see BPZ82 for details). Drawbacks and advantages of the one and two-dimensional methods have been discussed in many papers (see Kelson et al. 2000, La Barbera et al. 2002 and references therein). To summarize, the one dimensional approach allows one to significantly reduce the computation time of galaxy structural parameters. However, the uncertainties in the 1D parameters are larger, due to the circular symmetry approximation as well as to the interpolation of intensity values which is required to derive the galaxy and PSF one-dimensional profiles. On the other hand, the 2D approach is more time consuming, but allows more accurate estimates of structural parameters by taking advantage of all the information contained in the galaxy image. The one dimensional fitting procedure is included in the 2DPHOT package for completeness, particularly for cases where galaxy isophotes are strongly distorted and this distortion changes as a function of galaxy radius. In these situations the 2D approach can provide a poorly constrained fitting model, while useful parameters can still be obtained by the 1D approach. In order to apply the BPZ82 method, for each cell of the two-dimensional grid over which the two dimensional PSF

modeling is done (see Sec, 2.5), a one dimensional PSF model is computed. To this end, the circular surface brightness profiles of all the sure stars in a given cell are derived (see Sec. 2.9.1) and averaged together after sky subtraction and flux scaling. The 1D combined profiles are fit with a sum of Moffat or Gaussian functions applying a procedure similar to that described in Sec. 2.5. The one dimensional structural parameters are then derived by the BPZ82 method, convolving one dimensional Sérsic models with the derived 1D PSF models. The best fitting 1D parameters, i.e. the central surface brightness, the effective radius and the Sérsic index are then derived using χ^2 minimization with the Levenberg-Marquardt algorithm. Examples of one dimensional fitting results are shown in Fig. 2.18 for the same galaxies as in Fig. 2.15.

2.10 Growth curves

The aperture magnitude growth curve of each galaxy is derived by direct integration of the corresponding one dimensional surface brightness profile. The integration is performed for each of the concentric ellipses used to extract the 1D profile. As described in Sec. 2.9.2, all of the ellipses are defined by the average ellipticity and position angle of the galaxy. In order to correct the growth curve for seeing effects, the following procedure is adopted. The 2D seeing convolved Sérsic model obtained from the final 2D fitting analysis is used to extract a growth curve following the same procedure as for the galaxy image. The growth curve is also computed for the deconvolved Sérsic model, which is defined by the output parameters of the final 2D fit. The model is integrated over concentric ellipses using an adaptive 2D integration algorithm, and aperture magnitudes are extracted within the same apertures that define the galaxy growth curve. The difference between the growth curves of the seeing convolved and the seeing deconvolved models are used to correct the galaxy aperture magnitudes. We note that since the Sérsic model appears in the difference between the convolved and the deconvolved curves, the correction is expected to be largely independent of the choice of galaxy model, especially for the outermost parts of the galaxy where seeing corrections are small. The seeing corrected growth curve is used to estimate the half-light radius of the galaxy, and the corresponding mean surface brightness within that radius. We note that the seeing corrected growth curve and surface brightness profile allows the so-called eta function to be computed, which is defined by the ratio of

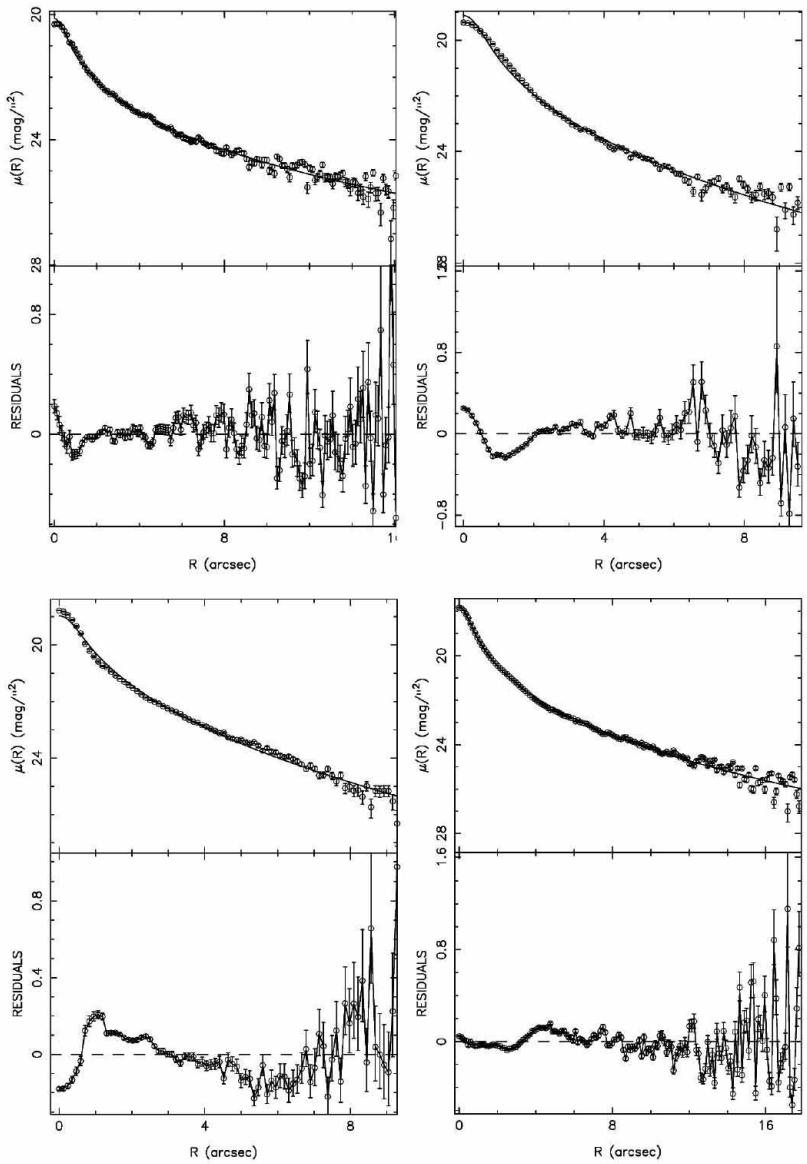

Figure 2.18: Surface brightness profiles of the same galaxies as in Fig. 2.15. In the upper plot of each panel, the surface brightnesses computed over different elliptical contours are plotted as a function of the ellipses' equivalent radii. The surface brightness values have been sky subtracted as described in the text. The error bars denote 1σ uncertainties, computed by adding in quadrature the standard deviation of the intensity values in each ellipse with the uncertainty in the background estimate. The solid line is the best-fitting one dimensional Sérsic model. The lower plot of each panel shows the residuals, in units of $mag/arcsec^2$, obtained after subtracting the model from the data. The four panels correspond to the same galaxies as in Fig. 2.15.

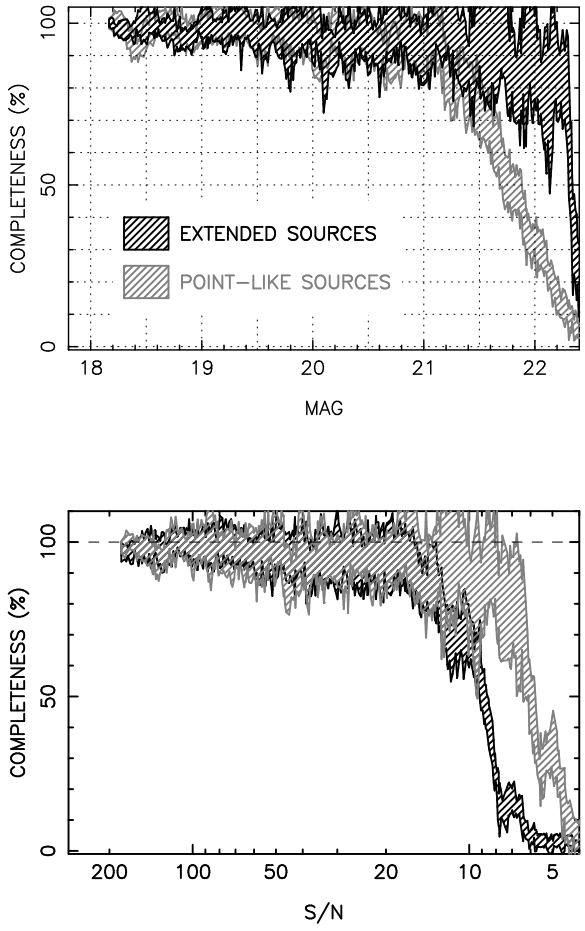

Figure 2.19: Completeness as a function of magnitude (upper panel) and S/N ratio (lower panel) for sources in the PACS image of the cluster Abell 1081 at $z \sim 0.16$. Hatched regions mark 1σ confidence intervals of the completeness. Stars and galaxies are plotted as grey and black regions, respectively, as shown in the upper panel. The dashed line in the lower plots marks the 100% completeness level.

the surface brightness value at a given radius to its mean value within the same radius (see Sandage & Perelmuter 1990). This function can be used to compute Petrosian metric radii and corresponding mean surface brightnesses. This feature will be implemented in the 2DPHOT package.

2.11 Completeness

In order to estimate the completeness of the galaxy catalog, we follow a procedure similar to that described in Sec. 2.7.2. 2DPHOT creates a set of simulated images by adding to the input image a random spatial distribution of artificial galaxies. The surface density of artificial galaxies in each simulated image and the total number of simulations are chosen using the same criteria outlined in Sec. 2.7.2. Artificial galaxies are created with seeing-convolved Sérsic models. The parameters of each model are chosen to match the distribution of galaxy structural parameters as a function of galaxy magnitude obtained from the input image. The coordinates of each artificial galaxy are chosen randomly within the input image, while its total magnitude $m_{\rm g}$ is extracted from a uniform random distribution spanning the same range as the observed galaxies[8]. 2DPHOT randomly selects one of the fifty objects[9] in the catalog with magnitudes closest to $m_{\rm g}$ and with corresponding S/N ratio larger than a cutoff value, $S/N_{\rm min}$. The INI2DF parameters of this object and the PSF model that correspond to the extracted center coordinates are then used to create the artificial galaxy. The S/N cutoff $S/N_{\rm min}$ is introduced because at very low S/N ratios the catalog is highly incomplete, biasing the distribution of galaxy structural parameters toward objects with a higher detection probability, such as galaxies with smaller effective radii and/or higher central concentrations (i.e. higher Sérsic index). Since the distribution of galaxy parameters at magnitudes below the completeness limit is not known, we adopt the working assumption that this distribution is similar to that of galaxies which are 'close' to the completeness limit of the catalog. In other words $S/N_{\rm min}$ is chosen as the lowest value of the S/N ratio for which the catalog is still nearly 100% complete. 2DPHOT

[8] i.e. the same magnitude range as objects classified as extended sources by 2DPHOT

[9] This number is chosen to sample the full range of galaxy structural parameters at a given galaxy magnitude. For several kinds of images, we verified that varying this number from twenty to eighty does not affect significantly the completeness estimates.

adopts a default value of $S/N_{\rm min} = 25$. However, processing several images, we found that changing $S/N_{\rm min}$ from 25 to 50 does not significantly change the completeness function.

For each simulated image, a catalog is generated using S-Extractor with the same settings as for the observed data. The galaxy completeness function is then derived by binning the artificial galaxies in magnitude and measuring the fraction of detected objects in each bin. The uncertainties on the completeness function are estimated by shifting magnitudes of artificial galaxies according to their corresponding uncertainties and recomputing the fraction of detected sources in each given bin. The same procedure is applied to the simulated stars created by 2DPHOT to define the locus of star candidates (Sec. 2.7.2). In this way, the completeness functions of both extended and point-like sources are estimated. Fig. 2.19 shows the results of processing one r-band image from the PACS. The figure has been automatically produced by 2DPHOT. The completeness of the catalog is shown as a function of both magnitude and S/N ratio. For the latter, artificial data are binned by S/N ratio and the fraction of detected sources is measured in S/N bins. We see that both the galaxy and the star catalogs are almost 100% complete down to $S/N \sim 20$ ($r \sim 21^m$).

2.12 Contamination

Using simulated stars and galaxies described in Sec. 2.7.2 and Sec. 2.11, 2DPHOT estimates the fractions of galaxies and stars which are misclassified as a function of their magnitudes. We examine the distribution of the artificial galaxies and stars added to the input image in the SI versus magnitude diagram, using the definition of the star locus (Sec. 2.7.2) to perform S/G classification. This procedure is illustrated in Fig. 2.20, where the results obtained for one of the r-band PACS images are displayed. These figures are automatically produced by 2DPHOT. The upper panel shows the star locus as well as the distribution of both artificial stars and artificial galaxies in the SI - Mag diagram. We note that almost all of the artificial galaxies have $SI \leq 0.7$, which holds true for all images we processed with 2DPHOT. This implies that adopting a lower cutoff of $SI = 0.7$ for the definition of the star locus minimizes the fraction of misclassified galaxies at low S/N ratios, as noted in Sec. 2.7.2. The lower panel of Fig. 2.20 plots the fraction, $\phi_{\rm s}$, of misclassified stars, i.e. the fraction of artificial stars that lie outside the locus of star candidates, and the fraction, $\phi_{\rm g}$, of misclassified galaxies, i.e. the fraction of artificial

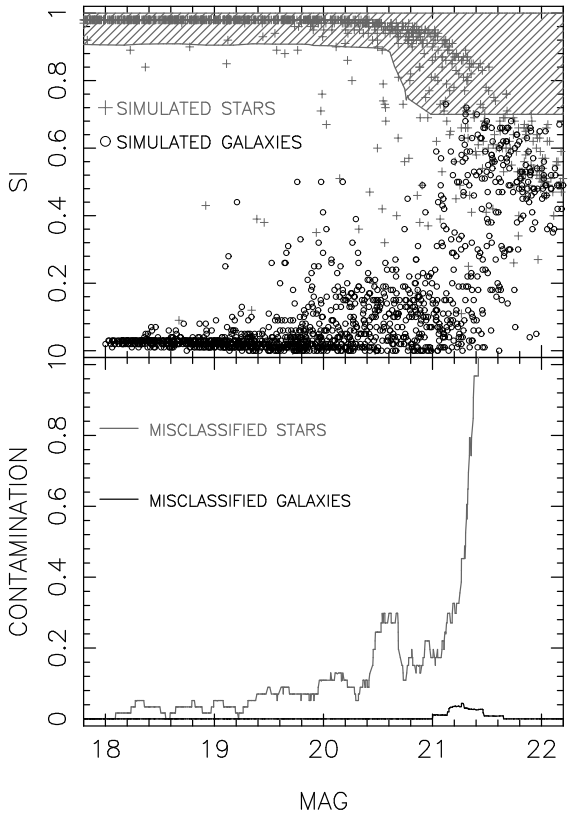

Figure 2.20: The upper panel plots the stellarity index from S-Extractor as a function of the Kron magnitude of simulated stars (grey crosses) and simulated galaxies (black circles), respectively. The hatched region corresponds to the stellar locus, which is defined as described in Sec. 2.7.2. The fractions of stars and galaxies that are erroneously classified on the basis of the stellar locus are plotted as grey and black curves, respectively. The plots have been obtained by processing the PACS image of the cluster Abell 1081, as for Fig. 2.19.

galaxies that are classified as stars, as a function of their magnitude. We see that ϕ_g is always smaller than a few percent, while the fraction of misclassified stars increases rapidly at faint magnitudes ($MAG > 21$). For bright magnitudes, at $MAG < 20$, where one would expect that stars and galaxies are always properly identified, we find that the value of ϕ_s does not reach zero, but is typically $\sim 5\%$, varying from $\sim 3\%$ to $\sim 10\%$ between $MAG = 18$ and $MAG = 20$. In order to understand why there is such a small fraction of misclassified stars, we considered the PACS frame whose 2DPHOT contamination plots are shown in Fig. 2.20 and selected those misclassified stars for which $MAG \leq 20$, yielding 25 out of 726 total stars. Fig. 2.21 shows the regions where each of these 25 simulated stars are randomly added to the PACS image. The S-Extractor stellarity index and FLAG values are also reported in the plot. Looking at the figure, we can clearly see that the small misclassification fraction at the bright magnitudes is caused by blending. In fact, the figure shows that bright misclassified stars can be identified as follows: (i) they lie just on top of some other object in the field, (ii) strongly blended with bright galaxies, and (iii) embedded within the extended halo of a bright saturated star. We also find that for $\sim 70\%$ (18 out of 25) of the simulated stars the FLAG value estimated by S-Extractor is 3, corresponding to the case of blended sources (see Bertin & Arnouts 1996). We notice that the blending issue does in principle affect any star/galaxy separation algorithm, and can be more or less important depending on how crowded is the image being processed. On the other hand, adding simulated stars and galaxies to a given image as done by 2DPHOT (Sec. 2.7.2) one can estimate the star and galaxy contamination fractions by taking into account also misclassification due to blended sources.

2.13 *Testing the star/galaxy separation at faint magnitudes*

So far, we have tested the star/galaxy separation obtained with 2DPHOT using images from the Palomar Abell Cluster Survey. As shown in Sec. 2.12, with PACS data we achieve reliable star/galaxy separation down to $r \sim 21$. On the other hand, many scientific programs are expected to reach significantly deeper limits, where the small size of galaxies and blending issues can make the star/galaxy separation far more troublesome. In order to discuss how the star/galaxy separation in 2DPHOT performs at faint magnitudes, we use two deep i-band image pointings taken with the Large Format Camera (LFC) at the

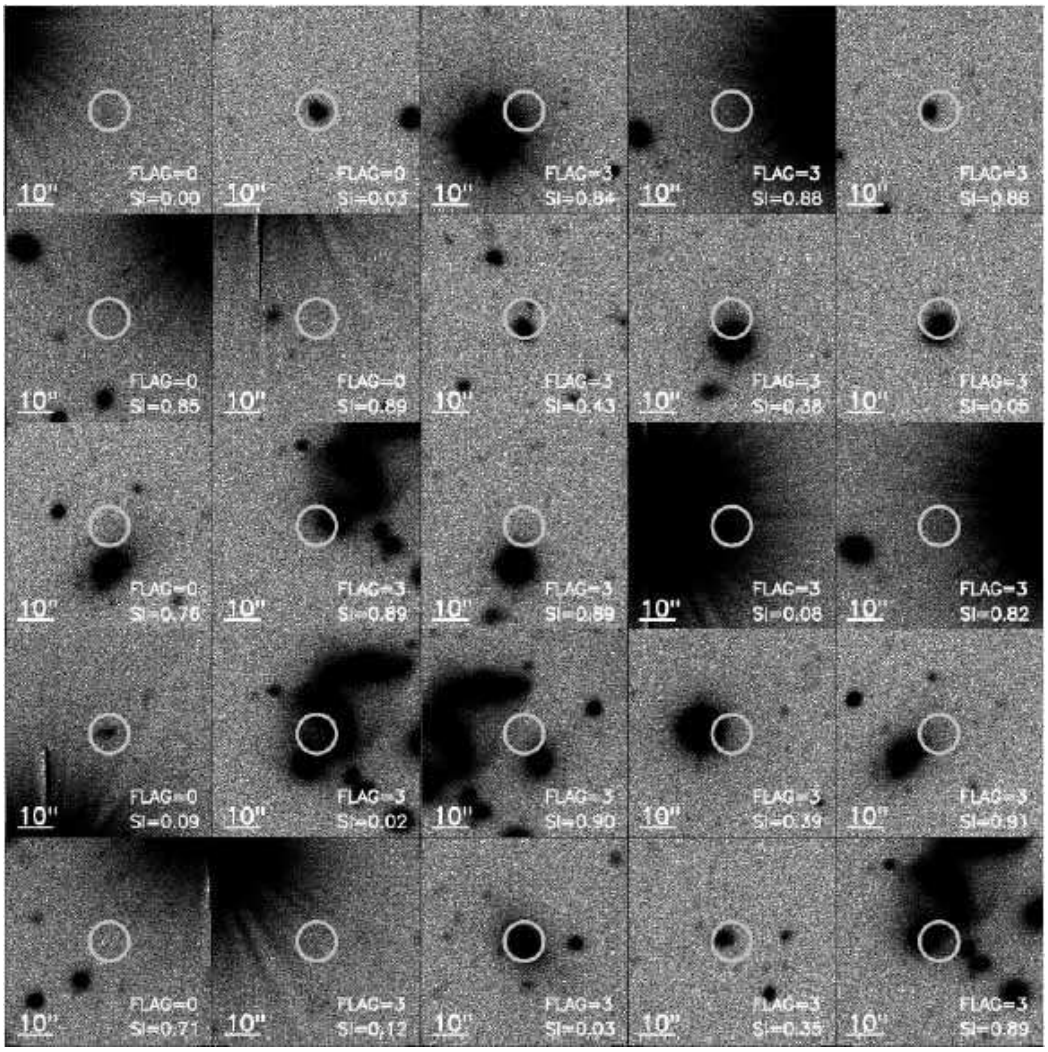

Figure 2.21: Regions of the same PACS image analyzed in Fig. 2.20 showing the position of bright simulated stars erroneously classified as galaxies by 2DPHOT (see the text). Each panel corresponds to a different simulated star. For clarity, we do not show the simulated star images added to the real one, but their positions on each panel are marked by a grey circle with a radius equal to three times the average FWHM value (1.8") of the PACS image. The stellarity index and the FLAG parameters estimated by S-Extractor are reported in the lower-right corner of each panel, while the spatial scale is shown in the lower-left.

Palomar 200" telescope. Each LFC pointing covers a circular area of $24'$ in diameter, with a pixel scale of $0.182''/pixel$. Data for the same sky area were also taken with the Advanced Camera for Surveys (ACS) onboard of HST and consist of 15 pointings taken with the F814W filter. The ACS data were drizzled to a pixel scale of $0.03''/pixel$, covering a total area of $\sim 13\, arcmin^2$. For more details on the data quality and main characteristics of the images, we refer the reader to Gal et al. (2005).

2.13.1 Comparing the HST and ground-based classification

We ran 2DPHOT on each of the ACS and LFC images and obtained the corresponding catalogs of stars and galaxies. All the ACS and LFC catalogs were matched, resulting in a final list of 3825 sources in common. In order to define the star locus, we adopted here the second option provided by 2DPHOT (see Sec. 2.7.2), where the locus is defined by setting the minimum value of the stellarity index SI equal to $\theta - 2.5\sigma$, where θ and σ are the location and width values of the SI distribution of artificial stars. As shown in Fig. 2.22, where the star locus is plotted for one of the LFC and one of the ACS images, the above definition establishes a narrower stellar region reducing the number of small faint galaxies which can be potentially misclassified as galaxies.

As shown in Fig. 2.22, the HST data go about two magnitudes fainter than the LFC imaging. Moreover, HST allows a sharp separation of stars and galaxies down to $i_{AB} \sim 25$, while for LFC the two classes begin to overlap by $i_{AB} \sim 22$. Assuming that the HST data provide the 'true' classification, we can estimate the fraction of HST stars and galaxies which are not properly classified from LFC and compare these fractions with those estimated by 2DPHOT. Fig. 2.22 compares the 'true' misclassified fractions with those predicted by 2DPHOT, as computed by averaging those obtained for the two LFC fields. The figure shows that the 2DPHOT results are in good agreement with the 'true' contamination estimates. The fraction of misclassified galaxies is always very close to zero, reaching $\sim 10\%$ at $i_{\rm AB} \sim 24$ for both the 'true' and 2DPHOT estimates. For stars, the fraction of both 'true' and 2DPHOT misclassified stars increases smoothly with magnitude, becoming larger than 50% at $i_{\rm AB} \sim 24$. We notice that the 'true' fraction is slightly larger than that estimated by 2DPHOT in the magnitude range of $i_{\rm AB} \sim 21$ to $i_{\rm AB} \sim 23$. However, considering the uncertainty on the 'true' fraction of misclassified stars

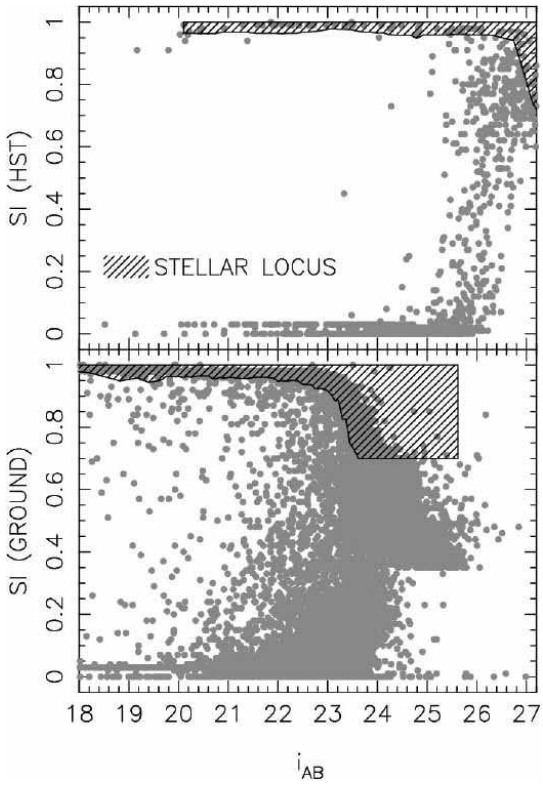

Figure 2.22: Definition of star locus obtained from 2DPHOT for one of the HST (upper panel) and one of the LFC (lower panel) images. Each panel plots the stellarity index versus the i_{AB} magnitude. Grey circles mark all the sources in a given image, with the corresponding stellar locus being represented by the hatched region.

the above difference is only marginally significant. Hence, we conclude that also at the fainter magnitudes sampled by the LFC photometry, 2DPHOT is able to provide realiable estimates of the contamination in the star and galaxy catalogs.

2.14 2DGUI: An interface for 2DPHOT

The 2DPHOT package requires installation of supporting software packages[10] and its performance varies depending on the compilers used. After installation, the user has to run the package by configuring both the input files for S-Extractor as well as some additional parameters specific to the package itself, which control the different steps of the image analysis (see Sec. 2.2). To simplify deployment and provide a uniform interface, we have developed a front-end called 2DGUI. To allow the timely execution of potentially time-

[10] such as S-Extractor and the cfitsio and pgplot libraries.

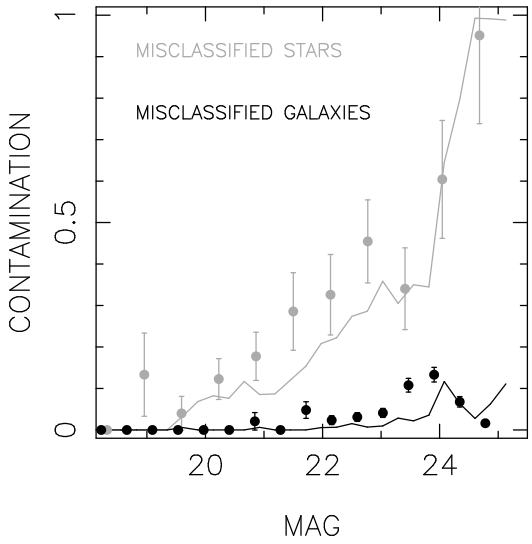

Figure 2.23: Comparison of the fractions of misclassified stars and galaxies provided by 2DPHOT with the 'true' values obtained by comparing the HST and LFC data (see text). The fractions of stars and galaxies are plotted in grey and black, respectively. The 'true' fractions are plotted as filled circles, with the error bars marking 1σ standard uncertainties, estimated by accounting for poissonian errors on counts in each magnitude bin. The fraction of misclassified sources provided by 2DPHOT are shown as continuous curves.

consuming processing jobs and manage parallelization, we have also included a simple scheduler system.

The 2DGUI package consists of three basic components. First, we provide an interface where the user can execute several 2DPHOT runs (jobs) through a local 2DPHOT installation. Second, a small, local (i.e. server-independent) database provides user access to the output files of 2DPHOT. Finally, the simple scheduling system allows timely execution of several jobs to be performed without server overloading. All of these components were either developed or adapted from well-known, portable, royalty-free software. Since the database is included in the 2DGUI package, no additional software is required. Since the scheduler is based on the *cron* utility and on a bash– or csh–system shell, the server must use a Unix-like operating system.

The first step includes user identification and job creation through the form shown in panel 1 of Fig. 2.24. Currently, only data available on the server can be processed. After creating a job, the user must configure parameters for both S-Extractor and 2DPHOT, using the two forms shown in panels 2 and 3 of Fig. 2.24, respectively. Both forms show

the command-line equivalent parameter names, their default values, and short comments. At this point, the job is created and scheduled, and information on the job execution is provided in the 2DGUI interface, as shown in panel 4 of Fig. 2.24.

2DGUI then creates a user directory (if necessary), along with a subdirectory for each job defined by the user identification, the 2DPHOT and S-Extractor parameters, and the filename of the input image. Four files are stored in this directory: 1) the original FITS image uploaded by the user; 2) a shell script (runme.sh), that includes the command-line syntax[11] for running 2DPHOT; 3 and 4) The S-Extractor configuration and parameter files of S-Extractor, named *default.sex* and *default.param*.

During execution, the 2DPHOT main script dumps textual information on each step of the image analysis in a log file. 2DGUI reads this file and informs the user of the job processing status by automatically updating the form shown in panel 4 of Fig. 2.24. When processing is finished, all the files generated by 2DPHOT are listed in the 2DGUI interface, and the user can select and download 2DPHOT output results. The execution of several *runme.sh* files on a given server is done by a scheduling program which runs only a predefined maximum number of *runme.sh* scripts, in such a way as to avoid overloading the server.

Currently, 2DGUI is still under development and its components are actively being improved. The major planned improvements include the following features: 1) the ability for a given user to submit several jobs, using the same parameters to process several images simultaneously; 2) use of the local database to store users' preferred parameters; 3) offline processing alerts (e.g. sending an e-mail to the user when a task is completed); and 4) a remote dispatcher, that allows tasks to be executed remotely, e.g. on a local cluster or a grid computer.

2.15 Summary

We have presented 2DPHOT, a new computational tool for astronomical image processing, designed to analyze the output data of wide-field imaging surveys in a completely automated fashion. The package includes several tasks, such as star/galaxy classification, measurement of both integrated and surface photometry of galaxies, PSF modeling,

[11] An example of this syntax is shown in panel 4 of Fig. 2.24.

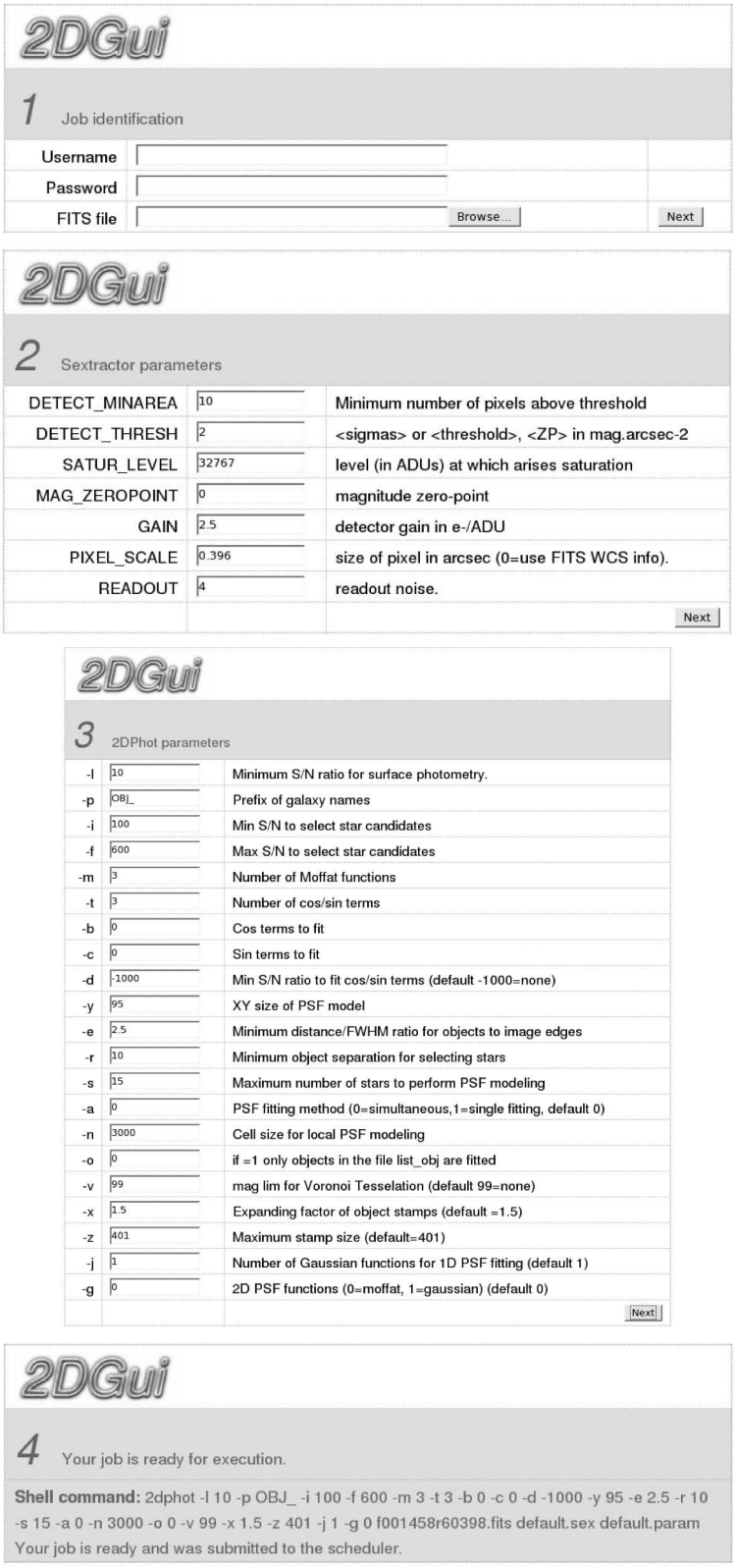

Figure 2.24: Web forms used by the 2DGUI package. The upper form, panel 1, is used for the user login and image upload. The middle forms, panels 2 and 3, are used to set the S-Extractor and 2DPHOT parameters. The bottom form, panel 4, provides information to the user about the image processing.

estimation of catalog completeness and classification accuracy. 2DPHOT incorporates a variety of quality control plots, which have historically been left to a separate step in image analysis, and is complemented by a graphical interface named 2DGUI. In addition, to accommodate the extensive output of 2DPHOT, both in terms of object catalogs and quality control figures, we are developing a database architecture which will comply with the standards proposed by the IVOA (International Virtual Observatory Alliance). All these components make 2DPHOT a powerful environment to analyze, handle, and store the output data coming from large area surveys. Some examples of surveys where we plan to apply this new analysis environment are those that will be carried out with the VLT-Survey Telescope (VST). We emphasize that 2DPHOT is conceived as a general purpose package, whose possible applications can span different research topics.

Chapter 3

The Brazilian Virtual Observatory - A New Paradigm for Astronomy in Brazil[1]

3.1 Introduction

Astronomy is now an enormously data rich science, and currently produces terabytes of raw data per day, with a few petabytes already in various archives. Both the data volume and data rate are increasing exponentially, with a doubling time of ~ 1.5 years. Even more important is the growth of data complexity (expressed, e.g., as the dimensionality of the parameter space spanned by the measurements of the detected sources) and heterogeneity. These data are now being federated in a global data grid under the umbrella of the Virtual Observatory (VO). A complete and effective scientific exploitation and exploration of these large and complex data spaces is a highly non-trivial task, requiring a new generation of software (databases, scalable data mining tools, interfaces), hardware (computing power, storage, network infrastructure), and expertise. The absence of these resources is a key bottleneck in data-rich astronomy: the data are there, but the means of extracting knowledge from them are not.

Figure 3.1 demonstrates the severity of these problems. We see the rapid increase in data volume from only a decade ago, where the Digitized Second Palomar Observatory Sky Survey provided single-epoch observations of half the sky in just 3 bands, to current projects like Pan-STaRRS1, which provides imaging of three-quarters of the sky in 5 filters but now at hundreds of epochs. Including the time domain not only increases the storage and computational requirements, but challenges the community with the need for new

[1] Chapter published as: de Carvalho, R. et al. 2010, PACIS, 1(3), 187.

algorithms and tools. Incredibly, we see that astronomy is generating data at the same pace as experiments in particle physics. This is extraordinary, considering that the number of researchers and the worldwide financial investment is much less in astronomy. Figure 1 clearly exhibits the necessity of efficient data storage, data processing and data mining, which are specific areas addressed by this project.

Typical research paths taken in the scientific exploitation of large sky surveys are either construction of statistical samples of objects or populations of interest (e.g., normal galaxies, quasars, etc.) and their study (e.g., to probe their evolution, large-scale structure, etc.), or selection of interesting targets (e.g., peculiar galaxies, distant quasars, brown dwarfs, supernovae, etc.) for follow-up observations. The scientific potential of such studies is greatly enhanced by federating data sets (e.g., combining optical, infrared, and radio sky surveys), which often reveal important features and populations of objects not easily distinguishable in any of the data sets taken separately. For example, a typical VO data enabled project would be a complete clustering and correlation analysis of combined source catalogs, using a federation of multi-wavelength data from several major astronomical surveys, ranging from radio, through infrared, optical, UV, to X-ray, or even γ-ray. Data federation of the source catalogs from these surveys generally results in a parameter space of $\sim 10^8 - 10^9$ data vectors in $\sim 10^2 - 10^3$ dimensions. The existing tools and algorithms do not scale well to such hyper-dimensional data sets, so we must assemble, test, improve, and deploy the necessary data mining, statistical, and visualization tools for this exploration. Concurrently, we must develop the necessary computational and network infrastructure and human expertise to develop, implement, and utilize these tools. Examples of specific challenges will be presented later in this Chapter.

The main objective of the BRAVO@INPE project is to address these strategic issues. More specifically, this project intends to generate investment in information technology, with particular emphasis on Computational Infrastructure, Data Grid, Data Processing, and Data Mining. We present not only a brief history of what has been done in the recent past but also elucidate the specific needs for the near future. This effort aims to prepare the Brazilian astronomical community for the avalanche of data and massive data processing needs that are a reality now, and which will increase rapidly in the coming years with the advent of the large telescopes and surveys currently under development (GMT, TMT;

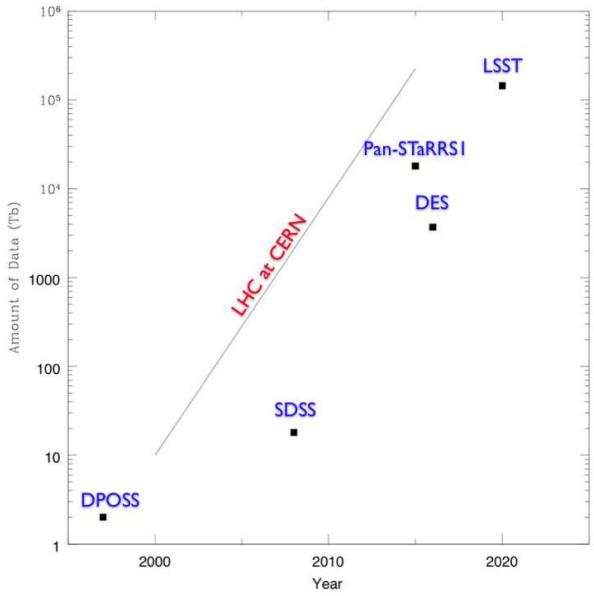

Figure 3.1: The rapid increase of astronomical data, considering only the most important optical surveys carried out in the past 20 years. For comparison we show the data rate from the Large Hadron Collider experiment at CERN (for details see http://lhc.web.cern.ch/lhc/)..

LSST, Pan-STaRRS, VISTA, VST).

This Chapter is organized as follows. Section 3.2 outlines the general concept of the Virtual Observatory, establishing a context for the more specific components described later. Section 3.3 describes the initial stages taken in generating a roadmap for the VO in Brazil, while Sec. 3.4 introduces two of the basic elements of any VO: computational infrastructure and databases. The fundamental concepts of data grids and processing grids are presented in Sec. 3.5. Section 3.6 describes an image processing pipeline developed by our group, 2DPHOT, and the main characteristics of the Astro-Wise environment. Section 3.7 provides an overview of the ongoing developments within Brazil in terms of astrophysical applications. Section 3.8 reviews data mining and describes specific projects we are undertaking in this field. Section 3.9 focuses on the four main areas in which we plan to invest resources, while section 3.10 summarizes the BRAVO@INPE project.

3.2 The VO Concept

For more than two decades, the international astronomical community has witnessed an exponentially growing capacity for accumulating astronomical data. Today, informa-

tion is gathered in large surveys from the ground and from space, covering virtually the entire electromagnetic spectrum, from X-rays through the ultraviolet, optical, infrared, and beyond. Individual projects yield complementary data through specific, targeted scientific programs. Much of these data are made available to the community through public servers, usually in several different formats, and distributed at many institutions. The data quality, metadata, interfaces, and accessibility are heterogeneous, since each project typically curates its own data, presents it in a custom database, and even data formats in astronomy are instrument dependent with little effort made to unify them.

An underlying concept of the VO is that by providing improved and homogenized data access combined with the tools to manipulate and explore the data, the need for new observations will be reduced even as the scientific output is increased. All gathered data, can be accessed via the VO, enriching the international community. Large surveys would take precedence over individual, targeted observations, providing added coherence to the VO structure. Therefore, the VO is not an enterprise driven by a single institute or even one country. It is rather a community endeavor aimed at the democratization of information that will certainly expand to other scientific areas like meteorology, geophysics, and space science, allowing new interactions and the exchange of methods and technology. Thus, the VO today represents to the astronomical community what the Internet was for the academic world in the 1980s. It is clear today that science, especially in developing countries, would be shockingly different without the Internet in the same way that we envisage in the future saying that astronomy would not be the same without the VO.

The Virtual Observatory (VO) concept is the astronomical community's response to the scientific and technological challenges posed by massive and complex data sets. At its heart, the VO is a set of standards - metadata (data describing data), interoperability, and other minimum requirements to be a VO-compliant database (Ball et al., 2006; Hanisch & Quinn, 2010). Such standards allow the development of VO tools that can then be deployed to operate on any VO compliant data set. Disparate surveys and individual observers' programs can thus be federated, queried, and manipulated by a single tool. Achieving these basic goals is not simple. To exemplify the obstacles to dealing with a modestly large amount of data, its complexity, and the challenge of processing it over a reasonable timescale, we examined the re-processing of reduced galaxy images from the

seventh and final data release from the Sloan Digital Sky Survey (SDSS DR7). These data cover ∼8400 square degrees on the sky and provide images in five bands. We consider only the re-measurement of photometric parameters using a custom pipeline (described later) - not the reprocessing of raw data to calibrated images, and ignore the spectroscopic data entirely. Even for this modest task, if we want to process the imaging data in one week, we would need 6577 processors, which is a factor of 15 more computing power than everything currently available to Brazilian astronomers. We use a timescale of one week as an upper limit for what a user would accept to retrieve important information from such a large data set - and this data set is almost trivial compared to upcoming surveys.

Computational hardware requirements are just one small part of the issues that arise when dealing with such vast data sets. Processing takes a lot of time, so once completed, it is of paramount importance that querying and retrieving data be done quickly. This requires investment not only in database software, but the astronomical and computational expertise to design and implement efficient and scientifically useful data models. This information, once structured in such a database, needs to be retrieved efficiently, demanding high-speed internet connections to which most research centers in Brazil do not have access. For these reasons, our top priorities include implementing grid computing to enable the processing of massive data sets; creating a dedicated network for astronomy to enable access to the resulting data, and training astronomers and computer scientists to develop these tools to produce cutting-edge science. Figure 3.2 summarizes this critical situation.

Nevertheless, many VO compliant databases and tools have already been developed. DS9, a commonly used astronomical image viewer, can communicate with databases using VO protocols. The European Southern Observatory has developed VirGO, a visual tool that allows both amateur and professional astronomers to search for available data in any part of the sky using planetarium-style software, and then retrieve data based on user supplied constraints. The US Virtual Observatory provides tools to convert ASCII tables to VOTable format, search multiple databases around a given position with one query, cross-match objects across databases, and more. One thing is clear - while the VO today provides standards and tools to find and extract data, few VO compliant tools exist to analyze the data. As we describe, the Brazilian contribution to the VO will focus on deployment of such resources.

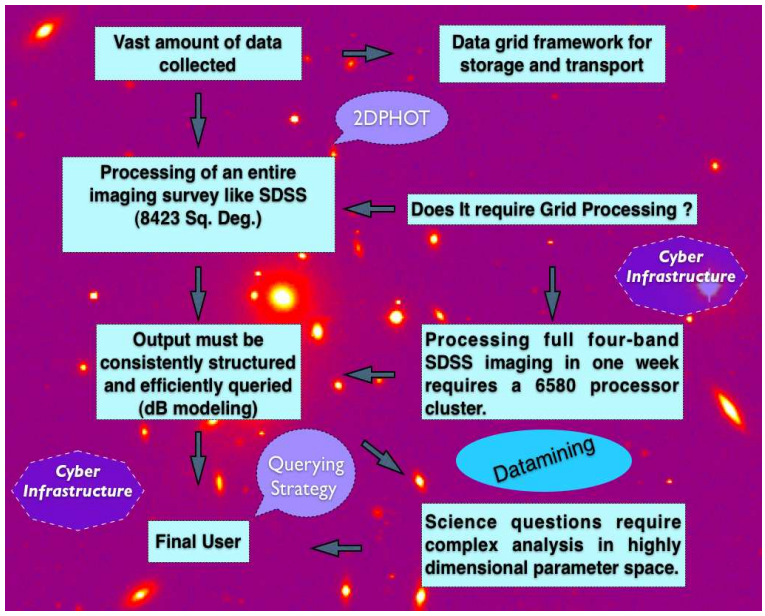

Figure 3.2: The complexity of implementing a VO structure considering all the infrastructural elements and addressing the demands of the astronomical community. The needs expand rapidly and will grow beyond the computational resources currently available, especially in Brazil.

3.3 The Status of Information and Communication Technologies (ICT) in Brazil

The new era of large data sets and the co-requisite data processing needs led to the recognition two years ago that we must modernize the tools for astrophysics in Brazil. In addition to the large photometric and spectroscopic surveys being carried out in both hemispheres, Brazil has committed significant resources to new facilities (including SOAR and Gemini). As a result, we have access to extraordinary amounts of data in all portions of the electromagnetic spectrum, but without standard techniques for storage, retrieval, distribution, processing or analysis. Thus, the underlying concept of BRAVO@INPE is to federate these resources, using a common framework, standard interfaces, computational infrastructure, and analysis tools. BRAVO@INPE is a branch of BRAVO, which is part of INCT. BRAVO comprises a committee in charge of planning VO activity in Brazil, with branches located at various universities and institutes (e.g. BRAVO@IAG at University of São Paulo, BRAVO@ON at Observatório Nacional). BRAVO orchestrates and coordinates the specific developments at different branches.

All of these developments are embraced within the concept of Information and Communication Technologies, encompassing all means for processing and communicating information. ICT is often used to describe digital technologies including methods for communication, transmission techniques, communications equipment, and techniques for storing and processing information. The term has gained popularity partially due to the convergence of information technology (IT) and telecom technology.

Before embarking on a major enterprise to develop BRAVO@INPE, we must understand our current hardware/software/personnel resources and their ability to meet our needs both today and in the future. The partners in this project are the thirty-one institutes comprising the INCT-Astronomy (National Institute for Science & Technology) recently created by the MCT (Ministry of Science & Technology). The central repository of knowledge about computational hardware, software and personnel in BRAVO@INPE will be the BNPGA (Brazilian Network for Processing Grid in Astronomy). A list of the institutes that will compose the BNPGA and their representatives is available in (de Carvalho et al., 2010, Appendix A). This new program will allow us to trace the roadmap of what is really needed for the future.

We have performed an initial census of the capabilities of the INCT-Astronomy member institutes. Here we provide a brief synopsis of the results; the complete list of questions and results can be found in (de Carvalho et al., 2010, Appendix B). We find that most users of our community have access to at least a desktop computer with moderate computational capacity. This conclusion must be seen with caution, as many members of our community use data of low complexity and in small volumes. This situation is changing dramatically with the next generation of large surveys and telescopes. In this context, the current computational facilities may be adequate today, but it is clear that the current cyberinfrastructure will be obsolete when dealing with the extremely large amount of data coming from both stellar and extragalactic projects.

These new programs will often require large computing clusters. We examined access to modern servers with more than 8 processors each (Class A) and to beowulf types, composed of mono-processed nodes and internal networks of 100 Mbps (Class B). Only 12 out of the 20 institutes that responded have access to a cluster and only 7 out of these 12 have access to a Class A cluster. It is important to note that in some cases the clusters are shared

with researchers from different disciplines like Physics since the small groups of researchers developing Astronomy in Brazil are contained within large Physics departments.

Adding up all the available processors in the different clusters gives, in principle, the total number of processors available for grid processing (see de Carvalho et al. (2010, Appendix B, Table 4)). This total, 419, is only 6% of the required number to processing the entire SDSS DR7 in one band, in one week, for example. This is only a crude estimate considering that all the processors are different, some better than others - fifty are old types of processors that would add little to the total processing capacity. It is to be noted that modern surveys like Pan-STARRS use a single 512-node cluster with the latest 3+GHz processors to analyze their data - more powerful than all of the Brazilian astronomical community's computers combined.

Beyond processing power, the total disk storage available to our clusters is approximately 45 TB. While this satisfies the needs of individual groups, it is clearly incompatible with the needs of the coming decade where telescopes will produce data at a rate of 2 PB/year. Moving any fraction of such data quantities also requires high-speed network connections, which many of our institutions still do not have.

The results of this census demonstrate the extreme deficiency of the current hardware, software and network infrastructure in Brazil. An often overlooked (and underfunded) aspect of any computational project is the need for personnel with expertise in all aspects of the program. In BRAVO@INPE, we cannot expect a computer scientist with experience in commercial database applications to understand and implement astronomical databases without new training. Similarly, we would not expect an astronomer to develop efficient computational algorithms for, say, clustering analysis, without learning about recent advances in such applications.

To address some of these issues, we organized two workshops in 2007 at INPE where formal presentations were given by a number of researchers from around the world engaged in e-science. At the second workshop we had specific presentations by researchers from Brazil engaged in VO-related projects, showing the tremendous potential that we have to actively participate in this international effort (see www.ivoa.net). For more information and access to the presentations of the lectures see www.lac.inpe.br/projetos/bravo/. A major component of this project will be the training of technical staff, which is of paramount

importance for us. We have already begun a program of visits by Brazilian astronomers and computer scientists to foreign institutions with extensive astronomy database involvement, including Caltech, Johns Hopkins University, and the Institute for Astronomy in Hawaii. The recent inclusion of BRAVO within the IVOA will help increase the interaction with other foreign groups.

3.4 Database Development and Basic Infrastructure

Many scientists view databases as simply a form of data storage. Perhaps you could do simple computations on columns on your personal computer and output the results. Typical tables might have tens to thousands of entries. Even a large food market has only about 50,000 different items available - and we are tempted to imagine that their warehouse database must be large and complex. We would be very wrong.

Astronomical data sets have far surpassed the largest commercial databases in size and complexity. Almost twenty years ago the Digitized Second Palomar Observatory Sky Survey (Djorgovski et al., 1998) database contained over 100 million objects, measured in three bands, with a total of 100 properties per object. Information about the survey (calibration, plate metadata, related CCD imaging, classification schemes) was spread over ∼50 different tables. Information from these tables often had to be joined (such as calibration and raw fluxes) to yield scientifically meaningful data. This database never became easily accessible to the public, which would have required the creation of an added layer of interfaces and query tools.

The current *gold standard* of databases in astronomy is the Sloan Digital Sky Survey. The imaging catalog has almost half a billion objects, in five filters, with nearly 500 columns of data on each object. While the volume of this single table (many TB) is itself daunting, the SDSS database has nearly 100 unique tables, with an additional 50 views offering easy access to scientifically useful subsets of specific tables. The complexity of this database required years of consideration to design a workable schema, decide on which columns to generate indices to speed up queries, understand how to load and update tables with new data, and how to provide public access. Just writing a portion of the table documentation was a full time job for a postdoctoral researcher for almost two years. Beyond the nearly 20TB of catalog data, SDSS also allows users to access a comparable volume of images.

While the SDSS is quite complicated for an astronomical database, it pales in comparison to those from next-generation projects. Upcoming surveys such as Pan-STaRRS and LSST will yield a comparable amount of data - every time they survey the sky. These projects will create a new SDSS every few months. Not only do they produce multi-filter imaging, which must be processed, cataloged, stored, and distributed, they will also produce time series data. Every object detected in one image must be matched to its corresponding detection in all earlier images of that same area. Optimal methods for differencing images must be developed to look for astronomical sources that vary or move. An entire pipeline is necessary to take moving objects, find them at different locations in images taken at different times, associate them, and generate orbits. Light curves for both stationary and moving objects must be created. All of this must be done almost instantaneously, because rare, one-time events such as supernovae must be found and notifications for follow-up observations disseminated before they fade. This means processing one gigapixel image every minute. The resulting database is correspondingly more difficult to model and populate. A static sky database must be created with everything detected, and updated as repeated observations allow for the creation of ever deeper images. Variable and moving objects must have all of their detections stored so that light curves and orbits can be derived.

The evolution of these surveys vividly demonstrates that we must contend with a new paradigm in astronomy. We must have the resources to store, disseminate and access large databases. We must have the knowledge of how such databases are structured, and how we can develop our own tools to create novel science. We must have our own databases for Brazilian programs, and enable interoperability with VO tools to maximize their scientific potential. We must also remember that so far we have only discussed large optical surveys. Multi-wavelength and multi-epoch studies demand new tools to cross-identify sources observed across the electromagnetic spectrum, with different spatial and time resolutions. This fundamental problem too has been approached but is far from solved. In all of these arenas, Brazil has much to learn, but also much to contribute.

3.5 Data Grid & Processing Grid

We are in the midst of a revolution in data gathering that encompasses all realms of science. In particular, the volume of data in astronomy, both real and simulated, is growing exponentially. The need for tools to analyze these data is naturally creating a new branch of scientific investigation - data science. There are many challenges in this emerging enterprise. We must have methods for extracting knowledge from large amounts of data, which by itself is non-trivial. What is important and what is noise? Which correlations are fundamental and which are secondary? In addition, we must urgently develop the skills and tools for processing these data. These requirements are already being addressed by two areas of computer science: data mining and high performance computing (HPC). We discuss the former in Section 3.8; here we will focus on the latter.

There are many approaches to HPC. The first one used parallel machines - vector machines, multi-processing machines with shared memory, multi-processing machines with distributed memory, and more recently multi-core processing chips. These solutions aimed to improve the processing capacity of a single, central machine. By the late 1990s, a form of distributed computing was created, using internet connections among geographically distributed processors to spread the computational labor. This is the underlying concept of grid computing, where processors across a city, country, or the whole world can be utilized by a single program. This type of grid is a new environment for science in the current century. We should note that a computing grid is not properly a HPC implementation. This concept (HPC) is typically reserved for enhancing the performance of a single system. However, grid computing is one of the strategies for addressing the need for intensive computation.

There are many types of grids and generally they can be classified according to:

- the nature of the processing: data grid or processing grid;
- the focus of the processing: open vs. closed or general vs dedicated;
- the hardware components: homogeneous or heterogeneous

Since the accumulation rate of data in astronomy is already reaching an unprecedented level of 10 PB/year, it is becoming difficult, technically and financially, to cen-

trally store all of the data, and impossible to replicate data for personal use. A data grid provides an environment for distributing, sharing, and modifying large amounts of data. We find applications for such a grid in different fields, such as meteorology. Examples include the Earth System Grid (http://www.earthsystemgrid.org/) and SegGrid (http://seghidro.lsd.ufcg.edu.br/). In astronomy, the Montage software platform (http://montage.ipac.caltech.edu) has been prepared to run in a grid environment. Many other astronomical tools have been or are being ported to grid applications (Benacchio & Pasian, 2007; Taffoni et al., 2009)

Similarly, increasingly large and complex processing tasks are required to process and analyze these data sets, or to generate large simulations. A processing grid addresses these issues. Collaborative processing is a type of application that exemplifies this new technology. The SETI@home project (Search for Extra-Terrestrial Intelligence, http://setiathome.ssl.berkeley.edu/) was the first popular distributed computing project, and now has over 3 million users, hosted by the Space Sciences Laboratory (University of California, USA). It is only one of 50 such projects using the BOINC volunteer and grid computing platform (http://boinc.berkeley.edu/). The Large Hadron Collider (LHC) has developed an enormous computing grid joining 170 computing centers in 24 countries (http://lcg.web.cern.ch/lcg/). Processing grids in other fields include protein folding (http://folding.stanford.edu/), climate change (http://www.climateprediction.net/), and seasonal mesoscale climate prediction GBRAMS (http://www.cptec.inpe.br/brams/gbrams.shtml) and RECLIRS (http://yule.lacesm.ufsm.br/nucleus332/). Numerous online directories of grid computing projects (http://www.gridcomputing.com, for instance) provide lists of dozens to hundreds of grid projects (commercial and scientific), environments, and applications.

Thus, it is imperative that we take advantage of the computing resources available at different computer centers linked by fast network connections. This could take the form of our own, internally developed grid implementation, or the installation and deployment of existing tools such as BOINC. Today, Brazilian users contribute almost 10 Teraflops of computing power to BOINC projects, the highest in South America. One of the main goals of BRAVO@INPE is to create a processing grid to harness academic computing along with this private processing power, with initial focus on two specific astrophysical applications:

image processing with 2DPHOT (described in Section 3.6.1); and analysis of cosmological simulations with hundreds of millions of particles using the FoF algorithm (described in Section 3.7.2).

Our team is currently strongly committed to the use of grid technologies and web services within the VO context. Specifically, our focus is data modeling, within the scope of BRAVO@INPE, to develop a framework for the metadata describing both observed and simulated data. We examine the logical relationships between these metadata, with the intent of establishing a general architecture for retrieving, processing, and interpreting data from different branches of spatial science and in particular from astronomy. This is an important step for constructing protocols that will guide VO applications.

3.6 Data Processing

As described above, critical issues in VO development include the large amount of data and how it is to be processed - transformed from raw images to reduced data suitable for further analysis. Here, we describe two concrete steps to address these problems undertaken in Brazil: the installation and operation of the first Astro-Wise (AW) node in South America and the insertion of our photometry environment (2DPHOT) into AW.

3.6.1 2DPHOT

2DPHOT is an automated tool to derive both integrated and surface photometry of galaxies in an image, to perform reliable star/galaxy separation with accurate estimates of contamination at faint fluxes, and to estimate the completeness of the resulting catalog. A 2DPHOT graphical user interface (2DGUI) is also under development, allowing the user to easily set 2DPHOT input options and detection parameters. More details can be found in La Barbera et al. (2008b). We show a schematic representation of the 2DPHOT environment in Figure 2.1.

The main tasks of 2DPHOT are:

- Producing a cleaned catalog of the image.
- Performing reliable star/galaxy classification.

- Estimating the completeness of the galaxy catalog and the contamination due to star/galaxy misclassification.
- Constructing an accurate model of the Point Spread Function (PSF) of the input image, taking into account possible spatial variations of the PSF as well as deviations of stellar isophotes from circularity.
- Deriving structural parameters of galaxies by fitting galaxy images with two-dimensional PSF-convolved Sérsic models.
- Measuring galaxy isophotes by fitting them with Fourier-expanded ellipses, and deriving one-dimensional surface brightness profiles of galaxies.
- Measuring the growth curve of seeing corrected aperture magnitudes for galaxies.

The image analysis flow of 2DPHOT is presented in Figure 2.2. 2DPHOT is being utilized by several projects conducted by researchers involved in this project and by others. In a spectroscopic and photometric study of a rich cluster at intermediate redshift, it is used to measure global properties of cluster galaxies (Mercurio, 2008); and in a fundamental plane study based on SDSS and UKIDSS data (La Barbera et al., 2008a) The analysis of internal color gradients in early-type systems has been recently published in La Barbera et al. (2005)We also used 2DPHOT in a recent study of Fossil Groups (La Barbera et al., 2009). We have also begun a large-scale study (SPIDER; Spheroids Panchromatic Investigation in Different Environment Regime of the general properties of early-type galaxies (ETGs) combining SDSS and UKIDSS data). This project makes extensive use of 2DPHOT to properly measure the seeing corrected structural parameters for nearly 40,000 ETGs.

3.6.2 ASTRO-WISE

2DPHOT is only a starting point in preparation for the avalanche of data in the next few decades. 2DPHOT requires as input an already processed image. Thus, we must be able to process raw images either on an individual basis or in a pipeline. To do so, we are taking advantage of the Astro-Wise (AW) system, developed by a consortium of European astronomy research institutes. The AW environment consists of hardware and software federated over five institutes in Europe, designed to scientifically exploit the

increasing amount of data produced by experiments in different fields. AW is a general information system which was initially geared towards astronomy, but is now also used in other branches of science. This is an essential trait of AW in the context of a unified environment for data processing at INPE. It allows a user to archive raw data, calibrate data, and perform post-calibration scientific analysis. All results are stored in a single environment that links together all of the discrete steps performed when analyzing a data set. This complete linkage, including the input, output, and software code used to derive one from the other, for arbitrary data volumes, has only been feasible thanks to a novel paradigm devised by the creators of AW. The algorithms included in the software have been developed to include arbitrary optical wide field imagers. This aspect is of major importance for BRAVO@INPE, since we will be developing software that enables us to ingest data from instruments available at SOAR and in the future from LSST.

AW was designed and implemented as a fully scalable and distributed information system to properly handle the huge amount of data that will be produced by large area surveys in the near future. By allowing the end-user to trace the data products, following all dependencies from the final catalog back to the raw data, it becomes possible to re-derive the result with better calibration and/or improved analysis tools. This represents perhaps the first time that astronomers could truly reproduce each others results.

To achieve these goals, AW includes structural functions that allow for storing data models along with data, in distributed databases (Valentijn et al., 2007). It contains a file server that can access these databases along with the image data, and a processing grid that can utilize parallel clusters while retrieving and storing input and output in the databases. AW is fully scalable so that it can accommodate large and small projects, can work with data from any imaging camera. Adding new analysis code is simple, allowing users to deploy their own tools on the compute clusters accessible to AW. These properties mean that AW overcomes the limitations of traditional analysis tools, which typically reside on a user's own computer or cluster. Reduction processes would usually be run by a single user, and saving sufficient metadata to reproduce every step is up to that individual. These behaviors are simply not sufficient for the new data volumes, collaborations and complexity in modern astronomy. Hundreds of terabytes of data will start entering the system when SOAR starts operating with the complete suite of planned instruments.

3.7 Data Analysis

The processing of raw data from a telescope into images, spectra or other products suitable for further analysis is only the first computationally intensive step on the path from photons to science. The processed data must be analyzed to detect, classify and characterize individual objects and groups of objects, and obtain physically meaningful measurements.

Within BRAVO@INPE, we are focusing on a few distinct data analysis projects:

- Implementation of a decision tree for star/galaxy separation in the faint magnitude regime for wide field images.
- Development of a parallelized Friends-of-Friends (FoF) algorithm.
- Automatic morphological analysis of images using both traditional tools for structural parameter estimation (e.g. concentration/ asymmetry, (Hernandez-Toledo, 2008)) and advanced methods for image analysis such as the Euler characteristic and gradient spectral analysis (Rosa et al., 2007).
- Development of a cluster finding algorithm using Voronoi Tessellation, but considering a more realistic background distribution instead of the usual Poissonian assumption (Soares-Santos et al., 2010, see Chapter 6).
- Virial analysis of galaxy clusters, allowing us to measure the most important dynamical quantities including total mass, based on the gapper technique described below.

3.7.1 Decision Tree (DT)

A decision tree is a computational method for splitting data into distinct classes, either based on pre-existing knowledge of the subgroups (supervised) or on inherent characteristics (unsupervised). Let a data set be described by a collection of attributes for each object in the data set. Each attribute is a measurement of some characteristic of an object (such as magnitude or size). These objects could belong to different classes or clusters (such as stars and galaxies). Imagine a data set for training, where the class of each object

is already known. Our task is to develop a classification rule to determine the class of an object based on its various attributes. If two objects have the same attributes, but they belong to different classes, then it is impossible to separate these objects based on this set of attributes. In this case, the data set with these attributes is not appropriate for a training set for the induction task. Thus, we must also determine the appropriate attributes to separate the objects into the desired classes.

Unlike other techniques for clustering analysis, the DT does not rely on distance metrics but instead makes a series of branching decisions based solely on numerical values of the attributes. A DT is a simple structure, where the final leaves define to which cluster an object with a specific set of attributes belongs. The nodes represent tests on a given attribute, with a branch for each possible output. For classifying an object, the starting point is the root of the tree; a test is applied to one attribute and the appropriate output branch is determined. The process is repeated using other attributes until the last leaf. Therefore, the object will belong to the cluster represented by that leaf.

There are many induction algorithms for decision trees. The ID3 algorithm, developed by (Quinlan, 1986), is the most popular. The algorithm was improved, allowing continuous parameters (Quinlan, 1993). A package has been developed, called WEKA (Waikato Environment for Knowledge Analysis) where several standard machine learning techniques were incorporated into a "workbench". Several decision trees were designed for classifying objects detected in the SDSS (Sloan Digital Sky Survey) data for 5 passbands (u, g, r, i, z), employing WEKA (see Ruiz et al., 2008; Suchkov, 2005). Our main goal is to provide a VO service to deal generally with the problem of star-galaxy separation - for any training set provided by the user, allow the generation of an appropriate DT using different methods and cross-validate the final obtained trees.

3.7.2 Parallel Friend-of-Friends Algorithms

The friend-of-friends (FoF) algorithm is commonly used to join galaxies within a linking volume around each galaxy. This method has several attractive features, like being independent of the particular geometry of the galaxy distribution. For a given linking volume a unique group catalog is defined. One of the main problems in using this algorithm is the time it takes to process large numbers of objects, scaling with N2 log N. It is necessary

to weaken this dependence on the total number of objects and thus be able to treat the hundreds of millions of particles found in current large cosmological simulations.

First experiments on reducing the dependence on N have shown that after a domain decomposition (subdividing the data in redshift shells) combined with a post-processing step we have already reduced the scaling to N log N2, a considerable improvement. A simple domain decomposition can be implemented in a purely parallel manner, but it is insufficient because some objects artificially separated by sub-domain boundaries could in reality belong to the same group. Therefore, a post-processing procedure is applied to examine objects close to a boundary but with a valid friend in an adjacent sub-domain. Our parallel version has fully reproduced previous results (Caretta et al., 2008) for computing the potential gravitational energy spectrum for galaxies and clusters of galaxies at many redshifts. A VO service will be made available allowing the user to run the FoF algorithm over the most important cosmological simulations available to date and those inputed by the user.

3.7.3 Advanced Tools for Morphological Analysis

As spatial information becomes ever more accessible through high resolution digital images, the need for robust techniques for complex pattern characterization is obvious. An obvious example is the mathematical description of galaxy images. Considerable attention has been paid to morphological classification of E/SO/Sa/Sab/Sm/Irr galaxy morphologies using Sloan Digital Sky Survey imaging. The data to be analyzed usually are (1) sky-subtracted, cleaned and log scaled g-band images; (2) filtered-enhanced versions of the g-band images; (3) the corresponding RGB composite images; and (4) a set of measured parameters, including surface brightness, position angle, ellipticity and spectral coefficients. In this sense, some useful mathematical and statistical approaches have been proposed (e.g. Lotz et al., 2004) to estimate the CAS (concentration, asymmetry and clumpiness) structural parameters. Motivated by the data analysis challenges in the context of BRAVO@INPE, we have developed an alternative and complementary approach for characterization of inhomogeneity and radial asymmetry in galaxy images. Inhomogeneity is calculated using the Euler characteristic from the Minkowski functional. Radial asymmetry is obtained by applying gradient pattern analysis to 2D wavelet multi-resolution

samples of the image. The combination of both structural characteristics is proposed as an effective measurement for galaxy morphology. The main objective here is to implement a VO service to deal with morphological analysis in general and in particular to analyze the entire SDSS (DR7) and explore the relationships between morphology and stellar population parameters, for instance.

3.7.4 A Modified Voronoi Tesselation Code to Search for Clusters of Galaxies

We are currently developing a cluster finder algorithm in 2+1 dimensions based on Voronoi tesselation (VT). The method is non-parametric and does not smooth the data, making the detection independent of the cluster shape. It uses all of the available galaxies, going as far down in the luminosity function as the input catalog permits. It does not rely on the existence of features such as a unique brightest cluster galaxy or a tight ridgeline in color-magnitude space. It works in shells of redshift, treating each shell as an independent 2-dimensional distribution of galaxies. The core of the VT algorithm is the background above which an overdensity must rise to be identified as a cluster. In contrast to earlier implementations of the VT algorithm, we do not assume a Poissonian background. We use a more realistic assumption that the angular two-point correlation function of the background distribution has a power-law shape, similar to what is actually observed. In a given redshift shell, we build a Voronoi diagram and compare the distribution of cell areas with the distribution expected from a background-dominated field. We set as a threshold the cell size below which the distribution starts to increase faster than its background counterpart. The clumps of contiguous cells found with density significantly above their respective cells are flagged as potential clusters.

The Voronoi diagram of a 2-dimensional distribution of points is a unique, non-arbitrary and non-parametric fragmentation of the area into polygons. A simple pseudo-algorithm to perform such fragmentation is the following: starting from any point P1 , we label its nearest neighbor P2 and follow the perpendicular bisector between those points. We stop when we reach the first point Q1 on this bisector that is equidistant from P1 , P2 and a third data point P3 . We now walk along the perpendicular bisector between P1 and P3 until we reach the point Q2 and identify the next point P4 by the same criterion. Successive repetition of this process will eventually bring us back to Q1 after a finite number of steps,

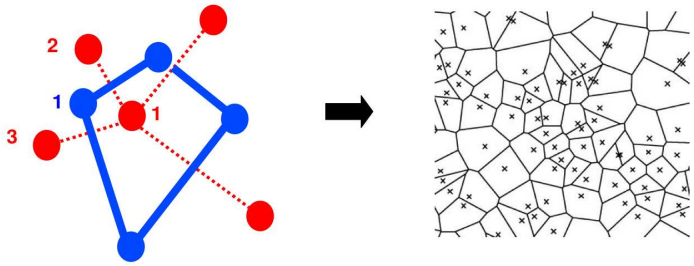

Figure 3.3: Diagram showing how the cells are built and a galaxy distribution on the sky with its corresponding tesselation.

creating a polygonal shape, the Voronoi cell, enclosing P1 and having vertices at the points Qi. By repeating this process for each point Pi (every galaxy in the redshift shell) we will have built the VT corresponding to this field. An example is shown in Fig. 5, which plots the distribution of galaxies on the sky (black dots) along with the Voronoi cells surrounding each galaxy. There are several robust and efficient computational algorithms to build a Voronoi diagram from a given distribution. In our code we use the so-called divide-and-conquer algorithm implemented in the Triangle library (Shewchuk, 1996).

There are no arbitrary parameters in constructing the VT for a given data set. Cells will be smaller in the high-density regions and since each cell contains one and only one point, the inverse of the cell area gives the local density. The VT cluster finder takes advantage of this fact in the process of detection. We plan to implement a VO service where the user can input a galaxy catalog over a given area of the sky and receive a cluster catalog as output.

3.7.5 The Virial Analysis Tool for Understanding Cluster Dynamics

Removal of interlopers and proper selection of galaxy cluster members is an essential step in the dynamical modeling of clusters and investigations of environmental effects affecting bound galaxies. There are several different approaches for interloper removal available in the literature. A recent comparison of the performance of many different methods applied to N-body cosmological simulations is given by (Wojtak et al., 2007). In particular, they found that differences in mass estimates may be explained by the number of interlopers a given method selects or rejects. These could also explain the discrepant

estimates from other methods of mass estimation (e.g., based on X-ray observations or lensing analysis). The shifting gapper method has two main advantages: (i) it is based on combined information for both position and velocity; (ii) it is independent of any hypotheses regarding the dynamical state of the cluster. The procedure we consider is similar to the approach adopted by (Fada et al., 1996). The input data consists of the radial and velocity offsets of each galaxy from the cluster center, being visualized as a phase-space diagram. It works through the application of the gap technique (Katgert et al., 1996; Olsen et al., 2005) in radial bins from the cluster center. This technique is used to identify gaps in the redshift distribution, resulting in the identification of groups in z-space. The bin size we consider for the shifting gapper is 0.60 Mpc or larger to force the selection of at least 15 galaxies (Fada et al., 1996, consistent with). Galaxies not associated with the main body of the cluster are eliminated. This procedure is repeated until the number of cluster members is stable (no more galaxies are rejected as interlopers). After a final list of members is reached, they can be used to measure the cluster velocity dispersion, from which we can estimate the cluster mass through virial analysis. While other procedures are based on physical assumptions about the cluster mass profile, the shifting gapper makes no physical hypotheses about the clusters dynamical state. Further details of this method can be found in (Lopes et al., 2009). This technique will be integrated in the VO service described in the previous subsection and will allow the user to carry out a dynamical analysis for the clusters detected with VT and having sufficient redshift measurements.

All of the applications described here already exist or are close to completion. However, they have not yet been deployed as VO-compliant tools. To accomplish this goal we must first implement the foundational concepts of the VO - network infrastructure, grid processing, and more important interoperability which is still lacking in the Brazilian astronomical panorama.

3.8 Data Mining

After the completion of image processing and derivation of meaningful quantities through data analysis, we are now confronted with an enormous collection of numerical quantities describing our data. An extant example is SDSS imaging, with 500 million objects, each with nearly 500 measured attributes. Which of these parameters are connected to funda-

mental physical properties? How do different types of objects cluster in high-dimensional parameter spaces? How do we find rare classes of objects, especially in the presence of errors or catastrophic mismeasurements? The 21st century will be a period of data-driven science, with the development of techniques to uncover the hidden knowledge in these kinds of massive data sets. This is the primary concern of a long standing branch of computer science - data mining (DM) - that has been applied extensively to astronomical data. It embraces a set of techniques for dealing with classification (neural networks (Bishop, 1994; Haykin, 1994), decision trees (Quinlan, 1986, 1993), clustering analysis (Hartigan, 2009; Kuntsche et al., 2003) visualization (Rieber, 1995; Hege & Polthier, 2002; Wright, 2006; Telea, 2007), pattern recognition (Jain et al., 2000; Bishop, 2006) and statistical analysis of massive data sets with extremely high dimensionality (Dryden, 2005; Francois, 2008). However, astronomers have yet to implement many of these techniques in easily accessible, cross-database tools. The incredible dimensionality and complexity of astronomical data also challenges conventional implementations of these algorithms (Hey et al., 2009).

In the space science domain, although there are extensive archival data resources available over the web, the ability of scientists to access and analyze this content is becoming more and more limited. The large data volumes cannot be moved to a personal workstation to be processed by an individual's own software, while the software cannot be easily placed on the data host. Thus, DM in the context of this project, refers to specific computational methodologies, working in a logical system, to extract information and find hidden patterns embedded in the large amounts of data from space science surveys. Generally speaking, any computational methodological tool performed to transform data into information is called a Data Mining System (DMS). It is notable that in space science (astronomy, astrophysics, cosmology and solar system studies) many existing data archives are unsuitable for DM because key pieces of metadata are missing. Hence, our goal in the first part of this project is to outline the major components of such a DMS, logically connected to the data processing and data grid requirements described previously.

3.9 A New Era for BRAVO

In the past two years we have gained important experience and knowledge of VO development, and our project was realigned to be economically feasible. New collaborations

were established in all facets of our planned investment. Within the context of the INCT-Astronomy, the priority is to devise a roadmap for the near future to coherently invest in hardware and software that can meet our researchers needs. BRAVO@INPE aims to create this synergy and contribute in strategic areas of the global VO.

Below we list the main strategic points of this enterprise. We emphasize that these are the overarching items defining this project and can be seen as pillars of a consistent investment in VO:

3.9.1 Network Infrastructure

From the results of our IT census we see the level of insufficiency of the hardware used by the astronomical community in Brazil, especially for astronomers located in smaller and more isolated institutes. High speed and secure network connections are of paramount importance not only for simple tasks in our daily work but also for establishing a national grid processing facility, such as the one we are developing with the BNPGA. The Brazilian National Research and Education Network (RNP) has been enhancing the network infrastructure throughout Brazil and has recently started the development of a plan to improve network access within the astronomical community in Brazil (see http://www.rnp.br/en/backbone/index.php). This is one of the main points of this project - to coordinate a study of the current situation and establish a schedule for implementing a modern network infrastructure for all institutes of astronomy in the country.

3.9.2 Creating the BNPGA

The census we did within the INCT-Astronomy community indicated that we not only need to upgrade our network infrastructure but also invest in creating a Grid Processing facility that can meet the growing demands of the astronomy community, not only because of the increase in the amount of data but also due to its increasing complexity. The BNPGA is the response to community's need for processing a large amount of data and reliably publishing the results in an environment meeting VO standards. BNPGA will commence as an exercise of processing the entire SDSS in one band; by doing so we will be able to implement the environment before upgrading to more powerful clusters with thousands of modern processors.

3.9.3 Astro-Wise as a national environment for data reduction and analysis

Several pipelines were developed in recent years to address the demand of large area surveys like SDSS. In these cases, users do not have to worry about data reduction. However, more and more sophisticated algorithms for object detection, star-galaxy separation, photometric redshift estimates, morphological analysis and other tasks are flourishing and there is an obvious need for reprocessing data in some of these cases. As discussed above, we are implementing AW (developed to be VO compliant) as the environment for large amounts of data processing. Ours will be the first AW node in South America, as it is currently extant only in Europe in compliance with IVOA standards.

3.9.4 The Virtual Lab for Advanced Data Analysis (VLADA)

The Virtual Laboratory for Advanced Data Analysis is a project initiated at the Lab for Computing and Applied Mathematics-INPE which aims to provide a new virtual environment for scientific analysis tools to extract statistical and physical information from time series, images and hypercube data. Its preliminary version consists of a PHP user interface through which the user can input the data and receive specific measures characterizing the data (statistical moments, power spectra, generalized dimensions, Euler characteristics, asymmetry coefficients, etc). In the context of BRAVO@INPE, VLADA can be seen as a virtual tool box for data analysis in general. VLADA will be made available as a VO service, which will require high speed connectivity and a high performance server. The project is being developed under the same umbrella as BRAVO@INPE since the technical operational issues are similar.

In order to tackle all of these issues within BRAVO@INPE we are planning to invest in personnel in three main categories. First, the project includes a collaboration with computer scientists from the Laboratory for Computing and Applied Mathematics at INPE (LAC) who are mainly engaged in projects dedicated to data mining and database development. They will be the main personnel involved in BRAVO@INPE making use of the large data sets acquired in the upcoming decades and pursuing solutions to problems astronomers will be facing in the Petabyte era. Second, LAC is recruiting staff specializing in database management, web design, and operating system management. They will be responsible for keeping the VO services operational. Third, we will support graduate

students and postdocs involved in the project. They will be developing their theses on subjects related to topics described in Section 3.7.

3.10 Summary

The Virtual Observatory is rapidly becoming a reality. The combination of growing data volumes and data complexity, coupled with computational and algorithmic advances, has made the VO a necessity. We have described some of the ongoing projects to implement databases, general-purpose computational algorithms, grid networks, and other VO-enabling technologies in Brazil. A common theme among all of these developments is the dire need for computational resources (CPUs, storage and network), software, and the expertise to design, install, and bring to life these complex systems. The international nature of astronomy implies that everyone can benefit and everyone should contribute to this enterprise. We have described the specific contributions that the Brazilian astronomical and computer science communities have made and will be making to this effort. Our growing partnerships in large telescopes and unfettered access to large public data sets demands that we develop our own tools and expertise to leverage these investments and strengthen our scientific output. Finally, we have described the necessary next steps in terms of hardware, software and personnel to advance BRAVO@INPE from an incipient program to a fully functioning project.

Chapter 4

The Sloan Nearby Cluster Weak Lensing Survey[1]

4.1 Introduction

The number density of galaxy clusters as function of mass and redshift is a well known probe of cosmological parameters (Haiman et al., 2001b). Several methods now exist to calibrate the cluster mass (optical dynamics, galaxy infall, X-ray, Sunyaev Zel'dovich, and weak or strong gravitational lensing) each of which have their own advantages and drawbacks. Of particular interest to future optical imaging surveys is weak lensing which uses the shape distortion of background galaxies induced by the gravity of a foreground cluster to measure the cluster mass. This lensing distortion is especially useful since it is independent of the dynamical state of the cluster and allows the cluster halo to be probed out to very large radii. Much progress has been made in cluster lensing studies since its initial detection (Tyson et al., 1990), however surprisingly a recent compilation estimates that only ~ 150 individual clusters have been studied with weak lensing (Dahle, 2007). Typically these clusters are imaged deeply to obtain a sufficient lensing signal but usually only cover a limited area around each cluster, probing the cluster halo in the inner few Mpc's. At low redshift in particular weak lensing has not been well studied since this requires imaging an area of $\sim 1^\circ$ or more around each cluster (Joffre et al., 2000). Studying a large sample of clusters with weak lensing at low redshift allows for a direct comparison to other methods of mass calibration which have been well studied in the low redshift regime.

To date the largest area imaging survey is the Sloan Digital Sky Survey (SDSS) (York et al., 2000). Previous studies in the SDSS have measured weak lensing by stacking clusters

[1] Chapter published as: Kubo, J. et al. 2009, ApJ, 702, L110

(Sheldon et al., 2009) in the redshift range $0.1 < z < 0.3$. In spite of the shallow imaging, weak lensing of *individual* clusters in the SDSS is also possible (Stebbins et al., 1996; Gould & Villumsen, 1994) provided the clusters lie at low redshift. In the low redshift limit ($z_l << z_s$) the lensing shear signal is

$$\gamma_t \propto \frac{D_l D_{ls}}{D_s} \propto D_l \tag{4.1}$$

where D_l and D_s are the angular diameter distances of the lens and source, and D_{ls} is the angular diameter distance between the lens and source. The lensing noise is

$$\delta\gamma_t \sim \sqrt{\text{area}} \sim D_l, \tag{4.2}$$

so the lensing signal-to-noise ratio is equal to a constant. Therefore by taking advantage of the large imaging area provided by the SDSS and "going wide" (Stebbins et al., 1996) weak lensing can be used to probe the mass of low redshift galaxy clusters.

Using the SDSS we recently reported a weak lensing measurement of the Coma Cluster (Kubo et al., 2007) which is the lowest redshift cluster ($z = 0.0236$) ever measured with weak lensing. Since this study, we have begun a program to measure other nearby, low redshift clusters in the SDSS. In this Chapter we present the first results from our campaign using clusters selected from two publicly available spectroscopic galaxy cluster catalogs in the SDSS: the C4 catalog of Miller et al. (2005) and the Berlind et al. (2006) catalog.

4.2 Data

4.2.1 Imaging and Spectroscopy

For our study we use data from the SDSS, an 8000 deg^2 imaging and spectroscopic survey using a dedicated 2.5m telescope (Gunn et al., 2006) at Apache Point Observatory in New Mexico. Imaging is obtained in a time-delay-and-integrate (or drift scan) mode in five filters *ugriz* (Fukugita et al., 1996) using the SDSS imaging camera (Gunn et al., 1998). The astrometric calibration of the SDSS is described in detail in Pier et al. (2003) and the photometric calibration pipeline is described in Tucker et al. (2006) and Hogg et al. (2001). Targets for the SDSS spectroscopic survey are selected using automated algorithms described in Strauss et al. (2002) and spectra are obtained using two fiber-fed

double spectrographs. The main spectroscopic galaxy sample is complete to a magnitude of $r < 17.77$.

4.2.2 Source Galaxies

In our lensing analysis we use galaxies drawn from the SDSS Data Release Six (Adelman-McCarthy et al., 2008). Shape measurement is performed using the PHOTO pipeline (Lupton et al., 2001) which measures the shapes of objects using adaptive moments (Bernstein & Jarvis, 2002). We correct for the effects of the point spread function (PSF) using the linear PSF correction algorithm described in Hirata & Seljak (2003). Source galaxies used in our lensing analysis are required to be detected in each of the $ugriz$ bands, classified by PHOTO as galaxies (type=3), and have extinction corrected model magnitudes (Stoughton et al., 2002) in the range $18 < r < 21.5$. Other lensing studies in the SDSS have used more sophisticated star-galaxy classifiers (Sheldon et al., 2009), however similar to Mandelbaum et al. (2005) we use the PHOTO classification and restrict the sample to larger galaxies with a resolution factor R > 0.33 (or $1.5 \times$ PSF). Here R is given by

$$\mathrm{R} = 1 - \frac{\mathrm{M}_{\mathrm{rrcc}}^{\mathrm{PSF}}}{\mathrm{M}_{\mathrm{rrcc}}} \tag{4.3}$$

where M_{rrcc} and $M_{\mathrm{rrcc}}^{\mathrm{PSF}}$ are the sum of the second order moments (in the CCD row and column directions) of the object and PSF respectively. Our catalog is further restricted in that we only use shape measurements from the r band which is the filter that typically has the best seeing (Adelman-McCarthy et al., 2008). Photometric redshifts are drawn from the SDSS photoz2 table (Oyaizu et al., 2008) and we restrict our sample to galaxies between $0.2 < z_{\mathrm{phot}} < 0.8$ with error $z_{\mathrm{err}} < 0.4$.

4.2.3 Cluster Sample

We have selected low redshift clusters in the SDSS from two publicly available spectroscopic catalogs: (1) the C4 cluster catalog of Miller et al. (2005) and (2) the Berlind et al. (2006) catalog. The public C4 catalog is based on the SDSS Data Release Two (Abazajian et al., 2004) and contains 748 clusters in the redshift range $0.02 < z < 0.17$ over a 2600 deg^2 region. The C4 algorithm searches for clusters in a seven dimensional parameter space which includes position (R.A., Decl.), redshift, and four colors ($u-g$,$g-r$,$r-i$,$i-z$). The

use of galaxy colors in this algorithm is found to minimize cluster projection effects (Miller et al., 2005). The public C4 catalog contains three different cluster centers which include the peak in the C4 density field, a luminosity-weighted mean centroid, and the position of the BCG (brightest cluster galaxy).

The Berlind catalog uses the friends-of-friends algorithm of Huchra & Geller (1982) to search for groups and clusters in the SDSS NYU Value Added Galaxy Catalog (NYU-VAGC) (Blanton et al., 2005) which is equivalent to the SDSS Data Release Three (Abazajian et al., 2005). The Berlind sample consists of three volume limited catalogs and we use the Mr20 (absolute r magnitude brighter than -20) catalog to search for clusters since it extends over the broadest redshift range $0.015 < z < 0.1$. For each system the Berlind catalog measures several cluster parameters including an unweighted group centroid, a mean redshift, and a richness estimate. For this study we select five of the richest clusters from the C4 catalog and two rich clusters from the Berlind catalog, described further in §4.3.2.

4.3 Weak Lensing Analysis

4.3.1 Mass Model

The lensing shear due to a cluster is given by

$$\gamma_t = \frac{\bar{\Sigma}(\leq r) - \Sigma(r)}{\Sigma_{\rm crit}} \tag{4.4}$$

where $\bar{\Sigma}(\leq r)$ is the average projected mass density within a radius r, $\Sigma(r)$ is the projected mass density at r, and $\Sigma_{\rm crit}$ is the critical surface mass density (Miralda-Escude, 1991). We fitted the data for each cluster to a Navarro, Frenk, & White profile (NFW) (Navarro et al., 1996) which has been found in numerical simulations to provide an excellent description of dark matter halos, ranging from galaxy to cluster sized haloes. From this mass profile a virial mass (M_{200}) can be determined from

$$M_{200} = \frac{800\pi}{3}\rho_c r_{200}^3, \tag{4.5}$$

where the virial radius (r_{200}) is defined as the radius where the density reaches a value of 200 times ρ_c (the critical density of the universe at the redshift of the cluster). The virial radius is related to the halo concentration (c_{200}) by $r_{200} = c_{200}/r_s$ where r_s is the scale radius, the radius where the density profile changes shape (Navarro et al., 1996). The

expression for the tangential shear due to an NFW halo is given in Wright & Brainerd (2000).

To measure the shear in each cluster we use an unweighted shear estimator given by

$$\gamma_t = \frac{1}{2\mathcal{R}} \frac{\sum e_t}{N}, \tag{4.6}$$

where e_t is the tangential ellipticty and N is the number of galaxies measured in logarithmically spaced annuli relative to the center of each cluster. Our use of an unweighted estimator is valid since our source galaxies are in the magnitude and size range where the shape measurement error in each galaxy is small compared to the intrinsic shape noise. We scaled the tangential ellipticity to a tangential shear in each bin using the shear responsivity $\mathcal{R} = 1 - \sigma^2_{\rm SN}$, where we assume a fixed value of $\sigma_{\rm SN} = 0.37$ in all of our measurements (Hirata et al., 2004).

To fit the data to an NFW model we adopt a likelihood approach. We define a binned likelihood given by

$$\mathcal{L} = \prod_i^{N_{\rm bins}} \frac{1}{\sqrt{2\pi}} e^{\frac{-(\chi-\chi_{\rm NFW})^2}{2}} \tag{4.7}$$

where χ and $\chi_{\rm NFW}$ are

$$\chi = \frac{\bar{e}_t}{\sqrt{\frac{\sigma_{\bar{e}_t}}{N}}}, \quad \chi_{\rm NFW} = \frac{\bar{e}_t^{\rm NFW}}{\sqrt{\frac{\sigma_{\bar{e}_t}}{N}}}. \tag{4.8}$$

Here $\bar{e}_t$ is the mean tangential ellipticity and $\sigma_{\bar{e}_t}$ is the standard deviation of the mean tangential ellipticity.

4.3.2 Results

In Figure 4.1 we show lilkelihood plots in M_{200} vs. c_{200} for five C4 clusters (A2048, A1767, A2244, C4 1003, and C4 3156) and two clusters from the Berlind catalog (A1066 and A2199). The corresponding value of r_{200} is also shown on the right panels in the figure. The maximum likelihood value for each cluster is indicated as the large white dot and contours represent the 50% (blue), 75% (green), and 87.5% (yellow) confidence regions. For all of the clusters the halo concentration is not well constrained, which is in general typical of weak lensing measurements of clusters. For both sets of clusters we find that the shear is maximized when centered on the cluster BCG. For the C4 clusters we use the reported BCG center in the catalog but for the Berlind clusters we re-center on the cluster

BCG by hand. All cluster centers reported in Table 4.1 indicate the position of the cluster BCG used in the lensing analysis.

Mass model parameters for our sample of clusters are summarized in Table 4.1. The weak lensing masses of our cluster sample varies from $\log(M_{200}) \sim 14.2 - 14.8$ with the cluster redshifts ranging from $z = 0.0306 - 0.0990$. Because of the very wide area imaging available in the SDSS the shear field due to each cluster can be measured out to large separation. With this sample we are finding that the error in the mass is slightly reduced when extending the outer radius out to $\sim 15 h^{-1}$Mpc, and this is the outer cutoff radius used in our analysis. Beyond this radius the mass determination does not appear to improve, and may be subject to the effects of large scale structure (see §4.3.3). Stacked measurement of galaxy clusters have measured the mean shear field out to $\sim$ 30Mpc (Sheldon et al., 2009), but this is the first time we know of that the shear field of individual clusters have been measured out to this large of a separation. We briefly comment on two clusters in our sample below:

Abell 2199 is a known supercluster which also contains the cluster Abell 2197. Rines et al. (2002) measured the mass of this cluster using the galaxy infall method and obtained a central mass of $3.2 \times 10^{14} h^{-1} M_{\odot}$ which is within 1σ of our measurement. Our imaging data is not deep enough to separate the contributions of A2199 and A2197 therefore our mass should be interpreted as an estimate of mass of the combined system.

Abell 2244 has a relatively small separation on the sky ($\sim$ 30' and in redshift $\Delta z \sim$ 0.01 to the cluster Abell 2245. Because of the relative proximity of these two clusters we also cannot separate out the individual contributions and report only an estimate of the combined mass for the system. In our calculation of the weak lensing mass we have assumed the redshift of Abell 2244 ($z = 0.0990$).

4.3.3 Projection effects

Because weak lensing relies on the shape distortion induced on background galaxies, intervening clusters along the line of sight could cause additional error in the mass estimates of clusters (Hoekstra, 2001). In the case of the SDSS the shape distortion due to these background clusters is expected to be small since the SDSS imaging is relatively shallow. This effect could however become non-negligible for future imaging surveys such as the

Dark Energy Survey (Annis et al., 2005) and the Joint Dark Energy Mission (Aldering, 2005). Foreground structure correlated with the cluster itself could also cause additional error as well as cause the mass estimates to be biased upward (Metzler et al., 1999) (Cen, 1997). Our sample of low-redshift clusters could in principle be used to study this effect, however a method to correct for this effect has not been developed so we leave this to future work.

4.4 Summary

Here we summarize the main points of this Chapter:

1. In addition to the Coma Cluster (Kubo et al., 2007), weak lensing measurements of other *individual* nearby ($z < 0.1$) clusters in the SDSS are possible.

2. We present the first results from our survey which includes seven low redshift clusters selected from existing SDSS spectroscopic cluster catalogs. All of these clusters have no previous weak lensing measurement.

3. With the very wide imaging area provided by the SDSS we are able to probe the shear field of individual clusters out to $\sim 15 h^{-1}$Mpc, further than other previous measurements.

The final spectroscopic cluster catalogs using these two cluster finding algorithms on the SDSS Data Release 7 (Abazajian et al., 2009) should identify a large number of other low redshift clusters for which individual weak lensing measurements are possible. We are also studying weak lensing on non-spectroscopic cluster catalogs including the Abell catalog (Abell et al., 1989), the Giriadi optical cluster catalog (Girardi et al., 1998) and the maxBCG catalog, which is based on SDSS DR5 data (Koester et al., 2007a). We have already processed a total of $\sim$ 3000 clusters from these catalogs, of which $\sim$ 300 appear to be reliable measurements. Analysis of these results, including the understanding of systematics are in progress. A publication of the final results of the survey with a detailed discussion of lensing systematics is planned. This should include weak lensing mass estimates for $>$ 200 clusters over the entire SDSS, making it the largest study of individual clusters with weak lensing. This will allow for a detailed comparison of weak lensing derived cluster masses with masses using a variety of other probes at low redshift.

Table 4.1 - Cluster data and lensing Virial masses

Name	ID[a]	R.A. (J2000)	Decl. (J2000)	z	$\log(M_{200})$
Abell 1767	3011	204.0347	59.2064	0.0704	$14.34^{+0.28}_{-0.54}$
Abell 2048	8129	228.8088	4.3862	0.0949	$14.78^{+0.22}_{-0.32}$
Abell 2244	3004	255.6771	34.0600	0.0990	$14.46^{+0.30}_{-0.56}$
C4 1003	1003	184.4213	3.6558	0.0771	$14.20^{+0.36}_{-1.17}$
C4 3156	3156	258.8017	64.3191	0.0950	$14.34^{+0.34}_{-0.80}$
Abell 1066	12289	159.7776	5.2098	0.0680	$14.78^{+0.20}_{-0.30}$
Abell 2199	16089	247.1593	39.5512	0.0306	$14.66^{+0.22}_{-0.32}$

[a]Cluster ID given in either the Berlind or C4 catalog.

Note. — The first five clusters in the table are from the C4 catalog, the last two are from the Berlind catalog. Errors bars on $\log(M_{200})$ are 1σ errors, where M_{200} is in units of $h^{-1}M_{\odot}$.

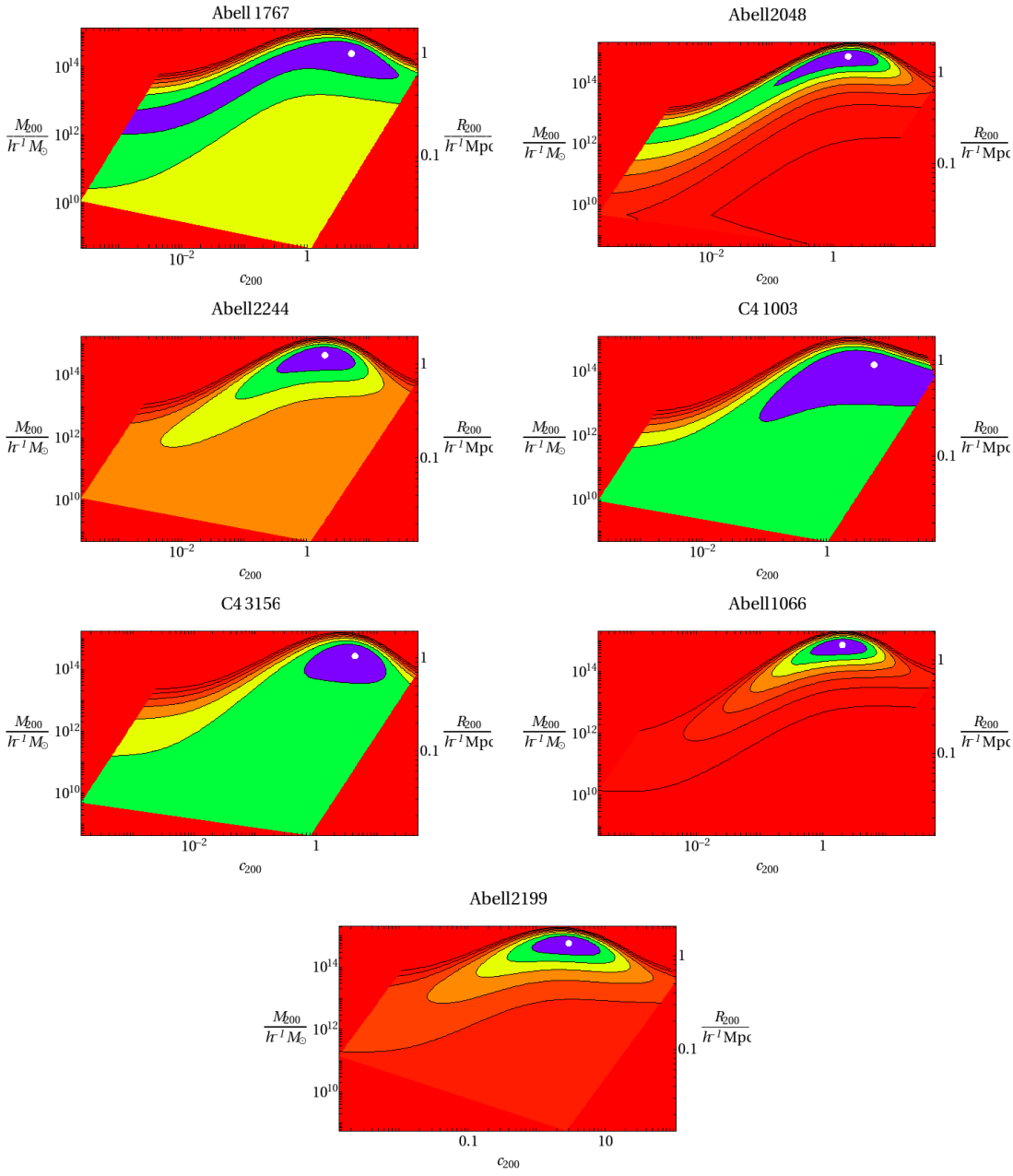

Figure 4.1: Weak Lensing virial mass (M_{200}) and halo concentration (c_{200}) contours for each cluster in our sample. On the right hand side of each panel we also indicate the corresponding virial radius r_{200}. The solid white dot in each panel is the maximum likelihood value and contours represent the 50% (blue), 75% (green), and 87.5% (yellow) confidence regions.

Chapter 5

The Sloan Bright Arcs Survey : Seven New Strongly Lensed Galaxies Between z = 0.66 – 2.94[1]

5.1 Introduction

Strong gravitational lenses allow for the detailed study of distant background galaxies through the magnification provided by lensing. Models of these systems also provide interesting constraints on the underlying foreground lens mass distribution which includes the distribution of dark matter. Beginning with the serendipitous discovery of the 8'oclock arc (Allam et al., 2007) we have initiated the Sloan Bright Arc Survey, a systematic search of the SDSS imaging database for candidate strong lensed systems. The first system from our survey was dubbed "the Clone" (Lin et al., 2009). It was followed by the discovery of six other systems (Kubo et al., 2009a). More recently we reported the discovery of four high redshift $z = 2$ systems (Diehl et al., 2009). To date we have reported the discovery of a dozen strong lens systems by mining the existing SDSS imaging data.

In this Chapter we report on a set of seven new systems in the SDSS imaging data which we have confirmed to be bona-fide strong lens systems. These systems were confirmed via spectroscopy using the Apache Point Observatory (APO) 3.5m telescope in New Mexico. Two of these systems we have confirmed to be at $z = 2$, three are confirmed to be $z \simeq 1-2$ systems, and two are at $z \simeq 1.0$. This Chapter is organized as follows: candidate selection follow-up spectroscopy are described in §5.2. In §5.3 we report the details of each system including preliminary mass models. In §5.4 we summarize our results.

[1] Chapter to be published as Allam et al. 2010, ApJL, in prep.

5.2 Data

5.2.1 Lens Search

We search for candidate strong lens systems in the 8,000 square degrees of SDSS imaging data. Specifically we performed two systematic searches of the SDSS Data Release Five (Adelman-McCarthy et al., 2007) and Data Release Six (Adelman-McCarthy et al., 2008). Our first candidate list is generated using the SDSS Catalog Archive Server (CAS) database. Two separate queries were performed in which we searched for blue objects around of catalog of (1) Luminous Red Galaxies (LRG's) (Eisenstein et al., 2001) and (2) a catalog of Brightest Cluster Galaxies (BCG's) generated using the maxbcg optical cluster finder algorithm (Hansen et al., 2005). Systems were then separated into groups depending on the number of of blue objects around each LRG or BCG. Each group of candidates were then visually examined by four different inspectors and flagged if they exhibited an arc-like morphology. The most promising systems were then chosen for follow-up. One confirmed system discovered in this search in presented in §5.3.

Our second list is generated by performing a search against a catalog of merging galaxies generated using the method outlined in Allam et al. (2004). A merging galaxy pair is defined here as two galaxies in the range $16 < g < 21.0$ which are separated less than the sum of their respective Petrosian radii (Stoughton et al., 2002). We ran this algorithm on imaging data from the SDSS DR6 and visually examined and classified the resulting catalog. A total of 5739 candidates were flagged by the algorithm and upon visual inspection a list of 2761 objects were flagged as possible strong lens candidates. In §5.3 we report on six systems that have been followed-up from this list which we have confirmed to be strong lensed galaxies.

5.2.2 APO Spectroscopy

Candidate systems were followed up with the Dual Imaging Spectrograph (DIS III) on the the Apache Point Observatory 3.5m telescope. DIS is medium dispersion double spectrograph that has separate red and blue channels. The standard medium red/low blue grating setup was used which covers a spectral range of $3600 - 9600\mathring{A}$. The resolution is $2.43\mathring{A}$ per pixel in the blue and $2.26\mathring{A}$ per pixel in the red. We targeted each of our seven

arcs with an exposure time of 3×900s. Typically we used a 1.5" or 2.0" slit oriented along the brightest segment of the arc in order to maximize the signal-to-noise ratio. The data were reduced used standard IRAF tasks which we described previously in Lin et al. (2009).

5.3 *Lens Sample*

Based on the search methods described in §5.2.1 we have confirmed a sample of seven new strong lens systems shown in Fig. 1. Each system has a dominant LRG and bluish arc located within an angular separation of 4.4"-9.4" of the central LRG. To characterize the environment of each system, we matched the lens coordinates against the GMBCG cluster catalog (Hao et al., 2010), a catalog of clusters in the SDSS DR7 generated using a red sequence based cluster finding algorithm described in Hao et al. (2009). Six out of seven systems are matched to clusters in the GMBCG catalog, and each of the LRG's are identified by the GMBCG algorithm as the brightest cluster galaxy in the cluster. One of our lens systems (SDSS J0952 + 3434) is associated with a moderately rich cluster with scaled richness $N_{gals} = 44$. The remaining four systems are associated with lower 'group-scale' objects with in the richness range $7 < N_{gals} < 11$.

Because of the relatively poor image quality of the SDSS images ($\sim$ 1.3") detailed models of the systems are not possible. Instead we adopt a simple model and describe each system with a singular isothermal sphere (Narayan & Bartelmann, 1996). The resulting velocity dispersions fall in the range $501 - 698$ kms^{-1} which correspond to enclosed masses between $2.5 - 11.3 \times 10^{12} M_{\odot}$. Magnitudes of the arc in each system are given in Table 1 and are given in SDSS model magnitudes unless otherwise stated. In some cases the arc is split into multiple knots by the SDSS image de-blender and in these cases the arc magnitude is the total sum of the individual knots. One system, SDSS J1537 + 6556, lies near the edge of the SDSS footprint and catalog data is not available in the SDSS DR7. For this case we independently measure and report the isophotal magnitude with Sextractor (Bertin & Arnouts, 1996). Below we briefly describe each system and present DIS spectroscopy for each of the arcs. A summary of all arc parameters is given in Table 5.1.

5.3.1 SDSS J0952 + 3434

SDSS J0952 + 3434 appeared in our search around LRG's and is part of a moderately rich galaxy cluster in GMBCG cluster catalog with $N_{gals} = 44$. The lens system itself consists of two LRG's with a blue arc located directly to the South. For the Eastern-most LRG ($r = 21.64$) we measure four absorption lines due to G, Hb, Mg, Na. From this we compute a spectroscopic redshift of the LRG to be $z = 0.35$. We did not obtain a spectroscopic redshift for the western LRG ($r = 19.51$) but it is a member of the cluster as identified by the GMBCG algorithm. The blue arc is split by the SDSS imaging deblender into two knots; the Eastern knot has a magnitude $r = 21.64$ and the western knot has $r = 22.04$. We measure a spectroscopic redshift of the brightest knot to be $z = 2.2$. With a measured Einstein radius of 6.9", this corresponds to an enclosed mass of $(5.6 \pm 0.5) \times 10^{12} M_{\odot}$.

5.3.2 SDSS J0957 + 0509

SDSS J0957 + 0509 appeared in our merging galaxy catalog. The system is a part of a small group of galaxies $N_{gals} = 7$. The BCG in this group ($r = 18.85$) has a spectroscopic redshift determined from DIS III to be $z = 0.5$. A blue arc to the southwest is split into three knots the brightest of which is $r = 20.45$. Along the brightest knot we measure a redshift of $z = 1.82$ (Figure 5.2). With a measured Einstein radius of 8.0" this gives an enclosed mass of $(11.3 \pm 0.9) \times 10^{12} M_{\odot}$.

5.3.3 SDSS J1207 + 5254

SDSS J1207 + 5254 was flagged as a candidate in our merging galaxy catalog. The system is a group of galaxies with $N_{gals} = 11$, with 2-3 LRG's interior to the arc. The brightest LRG ($r = 17.53$) has a spectroscopic redshift determined with DIS III of $z = 0.27$ based on the absorption lines of Ca, K, H, and G. This brightest LRG was detected in the SDSS as one object, but based on the imaging may actually be two distinct LRG's. A third LRG interior to the arc is directed to the northeast with $r = 18.64$. An extended arc is located to the Northeast of the central LRG and is also broken into several knots by the SDSS imaging deblender. The brightest knot has $r = 20.87$ and we measure strong absorption line due to CIV and $L\alpha$ placing the arc at $z = 1.92$ (Figure 5.3). With a

measured Einstein radius of 9.4" this corresponds to an enclosed mass of $(8.3 \pm 0.5) \times 10^{12} M_{\odot}$.

5.3.4 SDSS J1318 + 3942

SDSS J1318 + 3942 was selected from our merging galaxy catalog. The system is a group of galaxies ($N_{gals} = 10$) with a central LRG ($r = 18.82$) that has a spectroscopic redshift from the SDSS database of $z = 0.4751 \pm 0.0002$. To the South-East is a blue arc with the brightest knot reported in the SDSS database having a magnitude of $r = 20.59$. With DIS we identify emission lines in the arc (Figure 5.3) which places the arc at a redshift of $z = 2.94$. We measure the Einstein radius of the system to be $\theta_E = 8.5$" which along with the known redshift of the lens and source gives an enclosed mass of $(10.5 \pm 0.7) \times 10^{12} M_{\odot}$.

5.3.5 SDSS J1450 + 3908

SDSS J1450 + 3908 was also selected from our catalog of merging galaxies. The foreground lens is a group ($N_{gals} = 8$) with a BCG ($r = 17.62$) that has a spectroscopic redshift from the SDSS DR7 database of $z = 0.2893 \pm 0.0002$. To the East of the LRG is a blue arc which is detected as a single object with $r = 22.03$. With DIS III we measure four prominement emmission lines OII, OIII (5007), OIII(4959), Hb (Figure 5.4) which place the arc at a redshift $z = 0.86$. We measure and Einstein radius to be $\theta_E = 4.4$" which corresponds to and enclosed mass of $(2.5 \pm 0.3) \times 10^{12} M_{\odot}$.

5.3.6 SDSS J1537 + 6556

SDSS J1537 + 6556 was selected from our merging galaxy catalog. The system consists of a single LRG with a faint blue extended arc directed to the East. The system lies at the edge of the SDSS footprint and catalog data is not available for this system in the SDSS DR7. For the LRG DIS was used to measure absorption lines in both the red and blue part of the spectrum. In the red there are four absorption lines due to G, Hb, Mg, Na and in the blue absorption lines due to Ca, K, H, and G. From these line we compute a spectroscopic redshift for the LRG to be $z = 0.25$. In the arc we measure five prominent emission lines: OII, Hg, OIII (5007), OIII (4959), Hb shown in Figure 5.4 which places the arc at a redshift of $z = 0.66$. With a measured Einstein radius of $\theta_E = 8.1$" the enclosed

mass for this systems is $(7.6 \pm 0.6) \times 10^{12} M_{\odot}$.

5.3.7 SDSS J1723 + 6556

SDSS J1723 + 3411 was the final system selected from our merging galaxy catalog and is identified by GMBCG as a small group with $N_{gals} = 8$. The lens system consists of a single LRG with a blue arc directed to the southeast. We do not have a spectroscopic redshift for the LRG but the SDSS Photometric Reshift catalog (Oyaizu et al., 2008) places the LRG at $z = 0.41 \pm 0.02$. For the arcDIS emission lines place the arc at a $z = 1.33$. The measured Einstein radius of the system is $\theta_E = 4.7$" which yields and enclosed mass of $(3.5 \pm 0.5) \times 10^{12} M_{\odot}$.

5.4 Summary

We have presented seven new strong lens system from the Sloan Bright Arcs Survey, our search for strong galaxy-galaxy lenses in the SDSS imaging data. Two of these systems are of particular interest since they are lensing galaxies at high redshift, with $z = 2.94$ and $z = 2.2$. Because of magnification provided by lensing these are among the brightest systems known in this redshift range. The five remaining systems are lensing galaxies at intermediate and lower redshifts $z = 1.92, 1.82, 1.33, 0.86, 0.66$. To date the SBAS has reported a total of 18 strong lens systems, seven of which are lensing galaxies at high redshift $z \sim 2 - 3$. These bright, magnified galaxies are providing important windows into the star formation history and galaxy formation at high redshift. In upcoming papers we will present detailed models of these systems using high resolution HST imaging data.

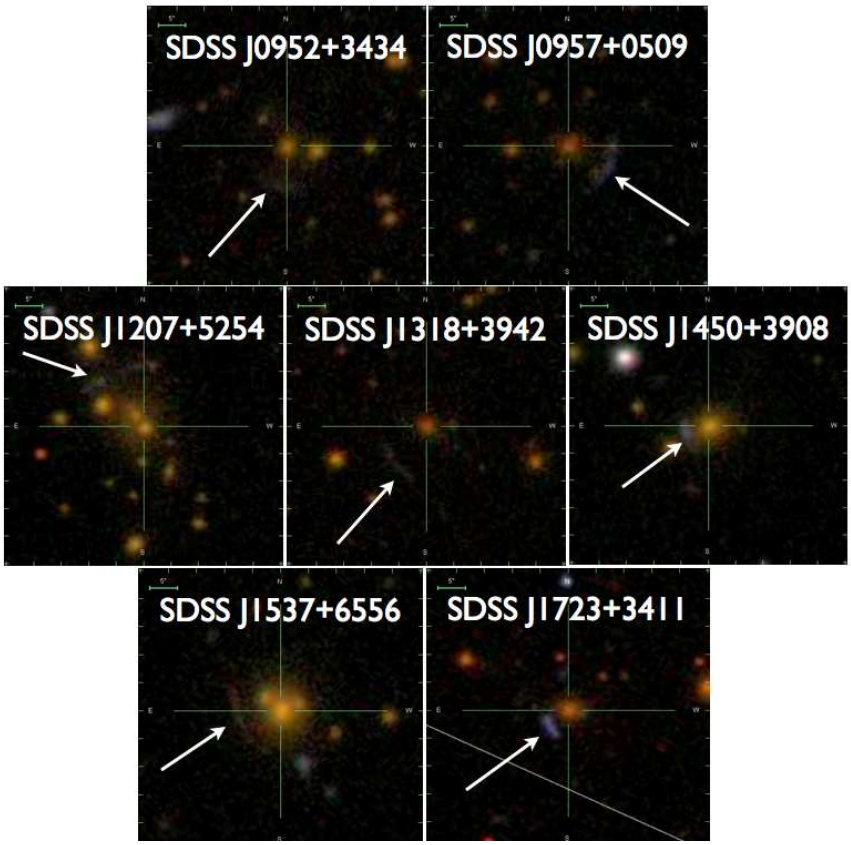

Figure 5.1: Mosaic of seven new strong lens systems discovered in the SDSS. Images are taken from the SDSS database. Systems are centered on the brighest LRG in each system. In each thumbnail image North is up, East is to the left.

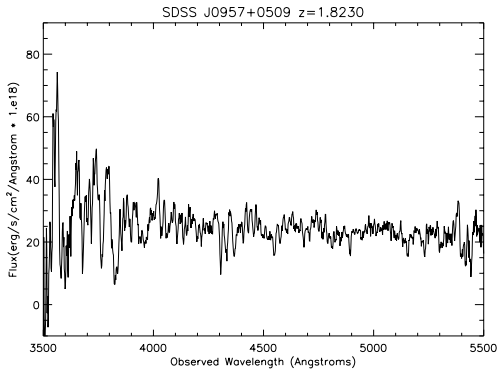

Figure 5.2: DIS Spectra of the arc in SDSS J0957 + 0509 $z = 2.2$.

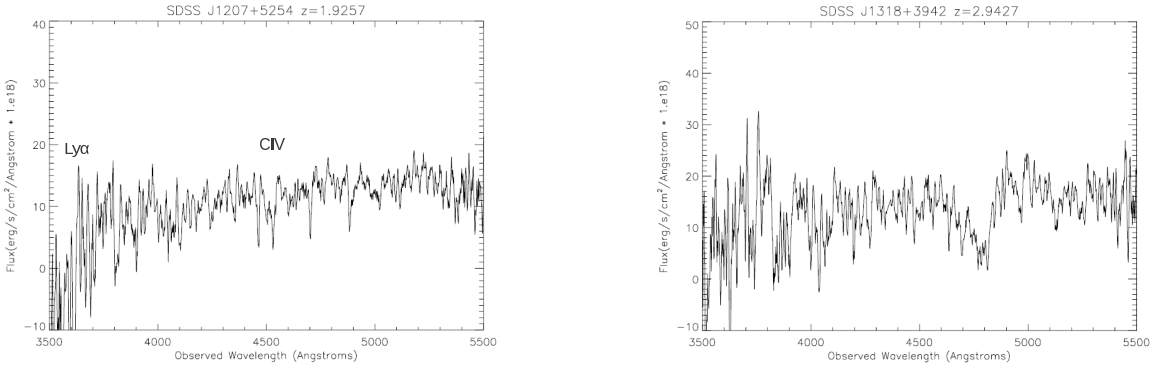

Figure 5.3: DIS Spectra of the arc in (Left) SDSS J1207 + 5254 indicating absorption lines due to CIV and $L\alpha$ placing the arc at $z = 1.92$ and in (Right) SDSS J1318 + 3942 with $z = 2.94$.

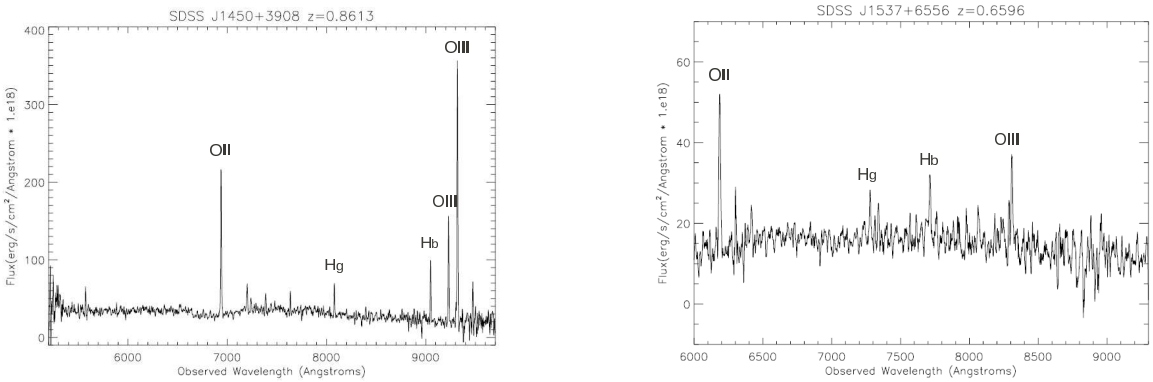

Figure 5.4: DIS Spectra of the arc in (Left) SDSS J1450 + 3908 emission lines OII, OIII (5007), OIII(4959), Hb which place the arc and $z = 0.86$ and (Right) SDSS J1537 + 6556 with emission lines OII, Hg, OIII (5007), OIII (4959) placing the arc at $z = 0.66$.

Table 5.1 - Lens parameters for each system

System	R.A. (deg)	Decl. (deg)	z_l	z_s [b]	θ_E [d] (")	σ_v (km s^{-1})	$M(<\theta_E)$ (10^{12}h^{-1}M$_\odot$)	$(g,r,i)_{\rm arc}$
SDSS J0952 + 3434	148.16760	34.5794	0.35 [b]	2.2	6.9 ± 0.3	565 ± 12	5.6 ± 0.5	(21.52,21.07,21.01) [c]
SDSS J0957 + 0509	149.41333	5.1588	0.5 [b]	1.82	8.0 ± 0.3	672 ± 13	11.3 ± 0.9	(20.55,20.04,19.73) [c]
SDSS J1207 + 5254	181.89960	52.9164	0.2717 [b]	1.9257	9.4 ± 0.3	644 ± 10	8.3 ± 0.5	(20.81,20.56,20.01) [c]
SDSS J1318 + 3942	199.54796	39.7075	0.4751 [a]	2.94	8.5 ± 0.3	642 ± 11	10.5 ± 0.7	(21.04,20.59,20.24) [c]
SDSS J1450 + 3908	222.62769	39.1386	0.2893 [a]	0.8613	4.4 ± 0.3	501 ± 17	2.5 ± 0.3	(21.50,22.03,22.06) [c]
SDSS J1537 + 6556	234.30500	65.9391	0.25 [b]	0.66	8.1 ± 0.3	698 ± 13	7.9 ± 0.6	N/A [f]
SDSS J1723 + 3411	260.90067	34.1994	0.41 [e]	1.3294	4.7 ± 0.3	519 ± 17	3.5 ± 0.5	(20.77,21.09,21.29) [c]

[a] Spectroscopic redshift from the SDSS database

[b] Spectroscopic redshift determined using DIS on the APO 3.5m

[c] Galaxy model magnitude (Stoughton et al., 2002) from the SDSS database

[d] Errors on manual fits to θ_E are estimated to be 0.3".

[e] Photometric redshift from the SDSS database

[f] Data not available in the SDSS DR7

Chapter 6

The Voronoi Tessellation cluster finder in 2+1 dimensions[1]

6.1 Introduction

Today we recognize that galaxies constitute a very small fraction of the total mass of a cluster, but they are nevertheless some of the clearest signposts for detection of these massive systems. Furthermore, the extensive evidence for differential evolution between galaxies in clusters and the field – and its sensitivity to the underlying cosmological model – means that it is imperative to quantify the galactic content of clusters. Perhaps even more importantly, optical detection of galaxy clusters is now inexpensive both financially and observationally. Large arrays of CCD detectors on moderate telescopes can be utilized to perform all-sky surveys with which we can detect clusters to $z \sim 1$, and even further with IR mosaics.

Forthcoming projects such as the Dark Energy Survey (DES, darkenergysurvey.org), Pan-STaRRS (pan-starrs.ifa.hawaii.edu) and the Large Synoptic Survey Telescope (LSST, lsst.org) will map thousands of square degrees to very faint limits (∼29th magnitude per square arcsecond) in at least five filters, allowing the detection of clusters through their weak lensing signal as well as directly through the visible galaxies. Combined with ever more efficient cluster-finding algorithms, these programs will expand optical cluster detection to redshifts greater than unity. Prospects for utilization of such data to address one of the most important scientific problems of our time by measuring the cosmological parameters with improved precision are outstanding. In fact, given the statistical power

[1] Chapter to be published as Soares-Santos, M. et al. 2010, ApJ, submitted.

of these surveys, clusters have become one of the strongest probes for dark energy (*e.g.* Haiman et al., 2001b; Holder et al., 2001; Levine et al., 2002b; Hu, 2003; Rozo et al., 2007b, 2010). Two unavoidable challenges imposed by these projects are to produce optimal cluster catalogs – with high completeness and purity – and to determine their selection function as a function of cluster mass and redshift.

To see how to proceed, we must understand the strengths and important limitations of techniques in use today, especially with respect to the characterizability of the resulting catalogs. We focus on photometric techniques rather than on cluster finding in redshift space, which also has a long story, starting with Huchra & Geller (1982), and has been succesfully applied to spectroscopic redshift survey data such as 2dFGRS (Eke et al., 2004) and DEEP2 (Gerke et al., 2005). Although the VT uses redshift information, it is a photometric technique and this motivates a discussion focused on this class of cluster finders.

The earliest surveys relied on visual inspection of vast numbers of photographic plates, usually by a single astronomer. The true pioneering work in this field did not appear until the late fifties, upon the publication of a catalog of galaxy clusters produced by Abell (1958), which remained the most cited and utilized resource for both galaxy population and cosmological studies with clusters for over forty years. Abell, Corwin, & Olowin (1989, 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 decades, even with serious concerns about their reliability. Despite the numerical criteria laid out to define clusters in the Abell and ACO catalogs, their reliance on the human eye and use of older technology and a single filter led to various biases. These old catalogs suffered as much from being black and white as they did from being eye-selected. Even more disturbingly, measures of completeness and contamination in the Abell catalog disagree by factors of a few. Unfortunately, some of these problems will plague any optically selected cluster sample, but the use of color information, objective selection criteria and a strong statistical understanding of the catalog can mitigate their effects.

Only in the past twenty 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. The hybrid tech-

nology of digitized plate surveys blossomed into a cottage industry. The first objective catalog produced was the Edinburgh/Durham Cluster Catalog (EDCC, Lumsden et al., 1992), which covered 0.5 sr ($\sim 1,600$ square degrees) around the South Galactic Pole (SGP). Later, the APM cluster catalog (Dalton et al., 1997) was created by applying Abell-like criteria to select overdensities from the galaxy catalogs. The largest, most recent, and the last of the photo-digital cluster survey is the Northern Sky Optical Survey (NoSOCS; Gal et al., 2000, 2003, 2009; Lopes et al., 2004). This survey relies on galaxy catalogs created from scans of the second generation Palomar Sky Survey plates, input to an adaptive kernel galaxy density mapping routine. The final catalog covers 11,733 square degrees, with nearly 16,000 candidate clusters, extending to $z \sim 0.3$. A supplemental catalog up to $z \sim 0.5$ was generated by Lopes et al. (2004) using Voronoi Tessellation and Adaptive Kernel maps.

With the advent of CCDs, fully digital imaging in astronomy became a reality. These detectors provided an order-of-magnitude increase in sensitivity, linear response to light, small pixel size, stability, and much easier calibration. The main drawback relative to photographic plates was (and remains) their small physical size, which permits only a small area (of order $15'$) to be imaged by a larger 4096^2 pixel detector. Realizing the vast scientific potential of such a survey, an international collaboration embarked on the Sloan Digital Sky Survey (SDSS, `sdss.org`), which included construction of a specialized 2.5 meter telescope, a camera with a mosaic of 30 CCDs, a novel observing strategy, and automated pipelines for survey operations and data processing. Main survey operations were completed in the fall of 2005, with over 8,000 square degrees of the northern sky image in five filters to a depth of $r' \sim 22.2$ with calibration accurate to $\sim 1 - 2\%$, as well as spectroscopy of nearly one million objects.

With such a rich dataset, many groups both internal and external to the SDSS collaboration have generated a variety of cluster catalogs, from both the photometric and the spectroscopic catalogs, using techniques including:

1. Voronoi Tessellation (Kim et al., 2002)

2. Overdensities in both spatial and color space (maxBCG, Annis et al., 1999; Koester et al., 2007b; Hao, 2009)

3. Subdividing by color and making density maps (Cut-and-Enhance, Goto et al., 2002)

4. The Matched Filter and its variants (Kim et al., 2002)

5. Surface brightness enhancements (Zaritsky et al., 1997, 2002; Bartelmann & White, 2002)

6. Overdensities in position and color spaces, including redshifts (C4, Miller et al., 2005)

7. Friends-of-Friends (FoF, Berlind et al., 2006)

Each method generates a different catalog, and early attempts to compare them have shown not only that they are quite distinct, but also that comparison of two photometrically-derived cluster catalogs, even from the same galaxy catalog, is not straightforward (Bahcall et al., 2003a).

In addition to the SDSS, smaller areas, but 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, Lobo et al., 2000), and many others. None of these surveys provide the angular coverage necessary for large-scale structure and precision cosmology studies, and have been specifically designed to find rich clusters at high redshift. The largest such survey to date is the Red Sequence Cluster Survey (RCS, Gladders & Yee, 2005), based on moderately deep two-band imaging using the CFH12K mosaic camera on the CFHT 3.6m telescope, covers $\sim$ 100 square degrees. This area coverage is comparable to X-ray surveys designed to detect clusters at $z \sim 1$ (Vikhlinin et al., 2009).

Any cluster survey must make many different mathematical and methodological choices. Regardless of the data set and algorithm used, a few simple rules should be followed to produce a catalog that is useful for statistical studies of galaxy populations and for cosmological tests:

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

2. The algorithm utilized should impose minimal constraints on the physical properties of the clusters, to avoid selection biases. Any remaining biases must be properly characterized.

3. The sample selection function must be well-understood, in terms of both completeness and purity, as a function of both redshift and mass. 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, including estimates of their distances and some mass proxy (richness, luminosity, overdensity) such that specific subsamples can be selected for future study.

One of the most popular and commonly used methods today is the Voronoi Tesselation (VT, Ramella et al., 2001; Kim et al., 2002; Lopes et al., 2004). Our implementation of this technique is described in detail in §6.2. Briefly, it subdivides a spatial distribution into a unique set of polygonal cells, one for each object, with the cell size inversely proportional to the local density. One then defines a galaxy cluster as a high density region, composed of small adjacent cells. Voronoi Tesselation satisfies the above criteria for generating statistical, objective, cluster samples. It requires *no* a priori assumption on galaxy colors, the presence of a red sequence, a specific cluster profile or luminosity function. Mock catalogs have been used to test the efficiency of the detection algorithm. These attractive qualities have led to its employment in numerous projects beginning almost 20 years ago (van de Weygaert & Icke, 1989; Ikeuchi & Turner, 1991; van de Weygaert, 1994; Zaninetti, 1995; El-Ad et al., 1996; Doroshkevich et al., 1997). Ebeling & Wiedenmann (1993) used VT to identify X-ray sources as overdensities in X-ray photon counts. Kim et al. (2002), Ramella et al. (2001) and Lopes et al. (2004) looked for galaxy clusters using VT. van Breukelen & Clewley (2009) included the VT as one of two methods in their 2TecX detection algorithm, an extension of their work on clusters in UKIDSS (van Breukelen et al., 2006). Barkhouse et al. (2006) detect X-ray clusters in Chandra data with the VT as well.

Here we improve on past implementations of this technique focusing on optical data. We build the VT in photometric redshift shells and use the two-point correlation function of the galaxies in the field to determine the density threshold for detection of cluster candidates and to establish their significance. This allows us to detect clusters in a self consistent way using a minimum set of free parameters and without any assumptions about the astrophysical properties of the clusters. We provide a list of member galaxies for each cluster and use the number of members as a proxy for mass. We apply the VT on mock catalogs that accurately reproduce the ΛCDM cosmology and the clustering properties

observed in the SDSS data. By comparing the VT cluster catalog with the truth table, we measure the completeness and purity of our cluster catalog as a function of mass and redshift. We show that our implementation of the VT produces a reliable cluster catalog up to redshift ~ 1 and down to $\sim 10^{13.5}$ solar masses.

The Chapter is organized as follows: §6.2 is dedicated to a detailed presentation of the algorithm; §6.3 describes the method used to compute the selection function of the cluster catalog; in §6.4 we discuss the completeness and purity results and show our ability to recover the mass function of the mock catalog at redshift close to unity; §6.5 presents a summary of this work. The work on the relation between the two-point correlation function and the VT cell areas distribution – fundamental for the development of our method – is detailed in the Appendix B.

6.2 Algorithm

We present the VT cluster finder in 2+1 dimensions. The method is non-parametric and does not smooth the data, making the detection independent of the cluster shape. It uses all galaxies available, going as far down in the luminosity function as the input catalog permits. It does not rely on the existence of features such as a unique brightest cluster galaxy or a tight ridgeline in the color-magnitude space. It works in shells of redshift, treating each shell as an independent 2-dimensional field.

Central to the VT algorithm is the background over which an overdensity must rise to be identified as a cluster. In contrast to earlier implementations of the VT algorithm (Ebeling & Wiedenmann, 1993; Ramella et al., 2001; Kim et al., 2002; Lopes et al., 2004), we do not assume a Poissonian background. We use a more realistic assumption that the angular two-point correlation function of the background galaxy distribution is represented by a power-law (e.g. Connolly et al., 2002). Another improvement over earlier works on VT-based cluster finders is the use of photometric redshifts instead of magnitudes (Ramella et al., 2001; Lopes et al., 2004) or colors (Kim et al., 2002). This eliminates the need for a percolation step and allows for a cluster finder which is not based on astrophysical properties of clusters (the luminosity function or color-magnitude relation), but on the characteristics of the large scale clustering process. This makes the VT a cluster finder subject to different systematics from color-based methods.

The fundamental inputs required for cluster detection using the VT are the coordinates RA, Dec and redshift of each galaxy and the redshift error $\sigma_z(z)$ for the full galaxy sample. The input catalog is sliced in non-overlapping 1-σ_z wide redshift shells. Note that the velocity dispersion of a typical cluster is much smaller than realistic values of σ_z. For each shell an estimate of the parameters (A,γ) of the two-point correlation function is required. This can be obtained directly from the data.

We then build a Voronoi diagram and compare the distribution of cell areas with the distribution expected from a background-dominated field. Since small cell size implies high density, this allows us to establish a size threshold below which the distribution is dominated by cluster members. The most significant clumps of contiguous cells smaller than this threshold are listed as clusters. This procedure is repeated on all redshift shells and the results are merged into a unique list of cluster candidates. The merge proceeds as follows. From the input galaxy catalog we extract 3-dimensional boxes centered at the coordinates of each candidate. We run the VT on those boxes to confirm the detection. This recursive procedure eliminates the edge effects at the interface between successive shells, reduces the number of fake detections due to projection effects and eliminates multiple detections.

In the resulting cluster catalog, we report position, redshift, redshift error, galaxy density contrast, significance of detection, richness, size and shape parameters of the clusters. We also provide a list of members with the local density of their respective cells and flags indicating the central galaxy (the galaxy found in the highest density cell).

Although it is possible to build Voronoi diagrams on a sphere, we use a rectangular coordinate system, which is easier to implement. This implies that we must process small sky areas at a time to avoid distortions due to tangential projection. We have tested different area sizes and concluded that boxes of 3×3 degrees are adequate. A buffer region is implemented to avoid edge effects and the effective area is the central 1×1 square degree box. Clusters found in the buffer regions are rejected prior to the merging of the shells' candidate lists. The size of the buffer zone corresponds to the angular scale of a large cluster at the lowest redshift (a 1 degree scale corresponds to $\sim$ 3Mpc at $z = 0.05$).

In the following, we detail each step of the cluster detection process and explain how each of the above quantities are derived, justifying the choices made in designing the

algorithm.

6.2.1 VT construction

The Voronoi diagram of a 2-dimensional distribution of points is a unique, non-arbitrary and non-parametric fragmentation of the area into polygons. A simple algorithm to perform such fragmentation is the following (see Fig. 6.1): starting from any position P_1, we label its nearest neighbor P_2 and walk along the perpendicular bisector between those points. We stop when we reach for the first time a point Q_1 equidistant from P_1, P_2 and any third point P_3. We now walk along the perpendicular bisector between P_1 and P_3 until we reach the point Q_2 and identify the next point P_4 by the same criterion. Successive repetition of this process will eventually bring us back to Q_1 after a finite number of steps. The set of points Q_i are the vertices of a polygon, the Voronoi cell, associated with P_1. If this process is repeated for each point P_i we will have built the Voronoi tessellation corresponding to this point field.

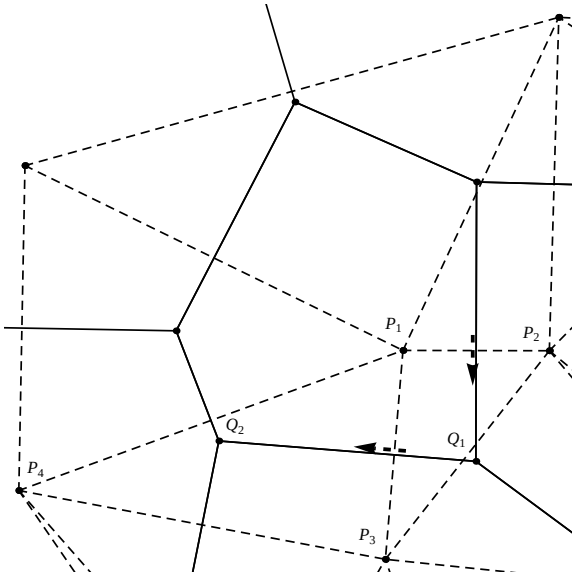

Figure 6.1: A portion of a typical Voronoi tessellation is shown together with its dual Delaunay mesh (solid and dashed lines, respectively) to illustrate the Voronoi diagram building process. For each generator set P_i, there is one and only one set of Voronoi cells given by the vertices Q_i. See text for details.

However, there are several more robust and efficient computational algorithms to build a Voronoi diagram from a given distribution. In our code we use the so-called Divide & Conquer algorithm implemented in the Triangle library (Shewchuk, 1996). The D&C is

based on recursive partition and local triangulation of the points and then on a merging stage. The total running time, for a set of n points is $O(n \log n)$.

There are no arbitrary choices in building the VT. The cell edges are segments of the perpendicular bisectors between neighbor points and each vertex is an intersection of two bisectors. This implies that the cells will be smaller in the high-density regions and since each cell contains one and only one point, the inverse of the cell area gives the local density. The VT cluster finder takes advantage of this fact in the process of detection.

6.2.2 Cluster candidate detection

Each realization of a given point process will result in a distinct unique tessellation, but the distribution of Voronoi cell areas will be the same. The case of the Poisson point process has been extensively investigated and it has been shown (Kiang, 1966) that the resulting distribution of Voronoi cell areas is well fitted by a gamma distribution

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp^{-\beta x} \tag{6.1}$$

with $\beta = \alpha = 4$ (only for the Poisson case) and x being the cell area normalized by the mean area of all cells. Here we extend Kiang's formula to a more general case.

Consider a random distribution of points in a plane with two-point correlation function given by $w(\theta) = A\theta^{1-\gamma}$, where the variable θ is the separation between point pairs and the parameters A and γ are respectively the amplitude and slope of the power-law. The Poisson distribution is the particular case where $A \to 0$. A general relation between the statistics of the point field and the VT areas distribution remains as a conjecture yet to be proved, but in the case of a point field generated from the above two-point correlation function, the gamma distribution still holds with the values of α and β modified. We have proven this fact and obtained the relation between α, β and the parameters A, γ numerically. Using the simulated annealing method described in the context of materials science (Rintoul & Torquato, 1997) we generate test fields spanning a wide range of A, γ pairs. On each test field we applied the VT algorithm and obtained the corresponding distribution of cell areas, fitting Eq. 6.1 to obtain the corresponding pair α, β. These two parameters are not independent. They are related by a simple relation: $\beta = 0.26 - \alpha$. See the Appendix for a detailed discussion of these results.

Information about the background is given to the VT code via the two input parameters A, γ. These will depend on the redshift shell and, ideally, they should be estimated directly from the data being considered. High accuracy in the parameters are not required, though. Note that no free parameters are introduced by A and γ, since they can be completely determined from the global input galaxy catalog. Clusters and groups being present in the field when the two-point correlation function is measured does not have a negative impact on the cluster finder. On the contrary, our method is based on the idea that the clustering process resulting in the power-law described by A and γ also results in the formation of clusters, which are found in the high density end of the VT cell distribution.

Taking the differential probability distribution (6.1) as a function of the normalized cell density, $\delta = 1/x$, our goal is to identify a density threshold δ^* above which the contribution of the clusters starts to dominate over the background. A schematic example is shown in Fig. 6.2. To the background distribution given by $A = 0.005$ and $\gamma = 1.7$ (upper panel, dashed line), we add a cluster contribution of 10% given by a simple Gaussian (upper panel, dotted line). As a result, the total distribution is distorted by the presence of the clusters. To perform the detection, we take the corresponding cumulative distributions. For the background, the cumulative distribution is given by,

$$P(\delta) = \frac{\Gamma(\alpha, \beta/\delta)}{\Gamma(\alpha)} \tag{6.2}$$

and depends on the input parameters A, γ through α and β. The maximum of the difference between the background (dashed) and the total (solid) distributions corresponds to the point where the total distribution increases faster than the background. This point is a natural choice for the threshold δ^* (vertical line).

In the example above an artificial cluster contribution with a particular shape was added to illustrate the principle of detection. In the actual process, we work only with the cumulative distributions. Once the threshold is computed we select all the cells with $\delta \geq \delta^*$. We then take the clumps of contiguous selected cells as cluster candidates.

Setting the threshold at the point of maximum difference between the two distributions leads to the detection only of the central regions of the most massive clusters ($M > 10^{14.5} M_\odot$). This is a consequence of the fact that the two-point correlation function of the field includes the contribution of clusters, and only the highest density peaks deviate significantly from the distribution predicted by Eq. 6.2. To improve this result, we allow

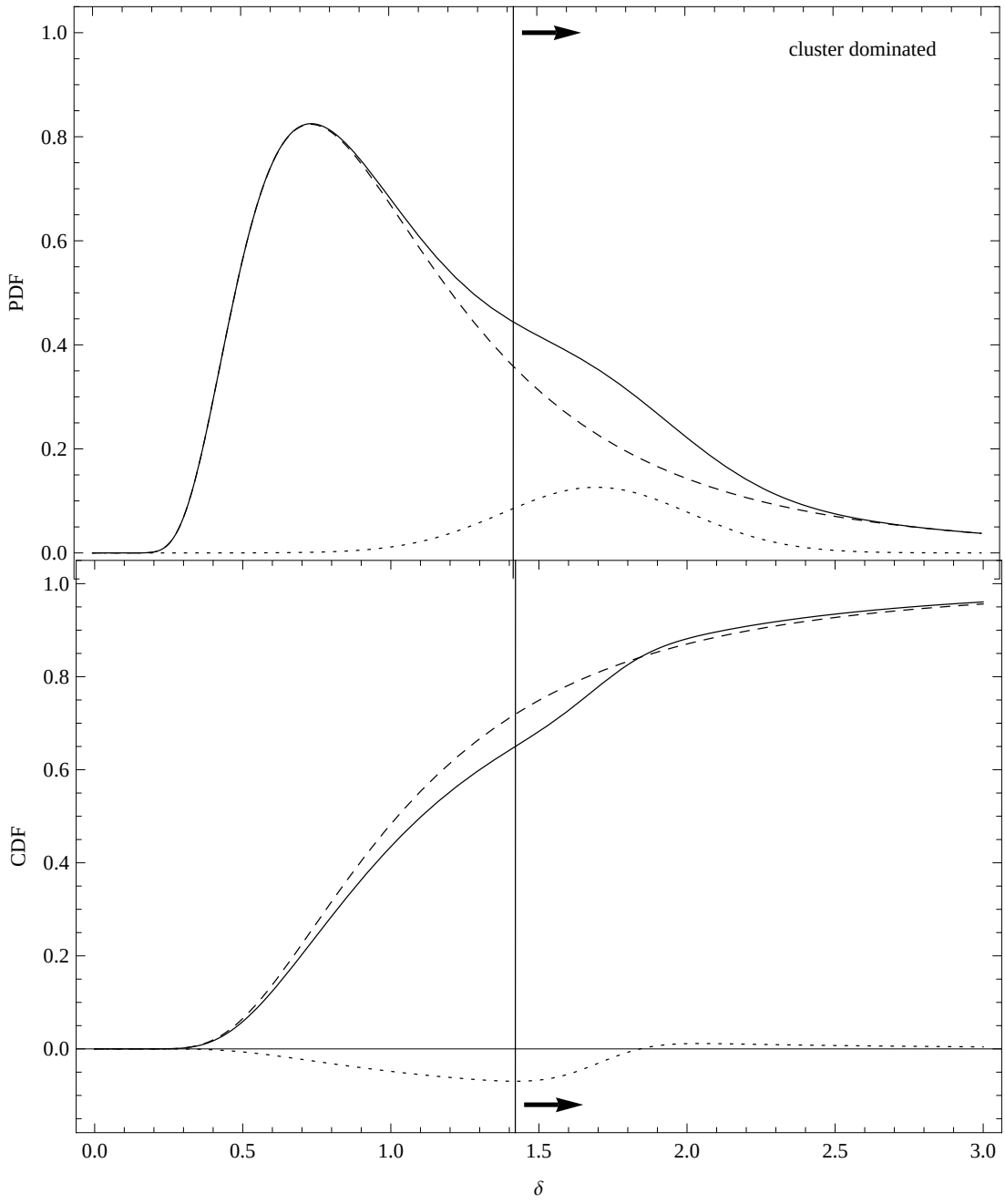

Figure 6.2: Differential and cumulative distributions of normalized cell densities illustrating the process of detection in the VT cluster finder. The dashed lines correspond to a background distribution with $A = 0.005$ and $\gamma = 1.7$. The solid lines correspond to the distributions distorted by an artificial Gaussian-shaped cluster contribution (dotted line). The vertical line is the threshold for detection δ^*. All cells above the threshold are selected as cluster member candidates.

this to be an adjustable parameter, called *scl*. By comparing the two-point correlation function of galaxies measured by Davis & Peebles (1983) in the $14.5m_B$ CfA redshift

survey with the two-point correlation function of rich ($R \geq 1$) Abell clusters measured by Bahcall & Soneira (1983), Bahcall (1986) has estimated that $\sim 25\%$ of the galaxies are associated to clusters in scales reaching ~ 30Mpc. We therefore set our threshold at the point δ^* where the cumulative distribution reaches $\sim 75\%$. As this fraction must change with redshift, magnitude limit of the galaxy catalog and lower mass limit of the cluster catalog, we determine the exact values of the cumulative distribution used to set δ^* in each redshift bin, $scl(z)$, by applying the cluster finder on simulated galaxy catalogs and maximizing the completeness and purity of the output catalog. This process does introduce a free parameter that we must tune.

6.2.3 Selection of high-significance candidates and membership assignment

For a given threshold δ^*, we assume that each cluster candidate has a probability

$$p(\delta_{min}, N_g) = 1 - \mathrm{Erf}\left(\left(\frac{\delta_{min}}{\delta^*} - 1\right)\frac{N_g}{\sqrt{2}}\right) \tag{6.3}$$

of being caused by random fluctuations of the background field. Here δ_{min} is the minimum cell density and N_g is the number of galaxies in the candidate. Note that the process of detection implies $\delta_{min} \geq \delta^*$. A confidence level of 95% is required for a candidate to be accepted. If a given candidate has $p(\delta_{min}, N_g)$ below this level, we iterate on its cells, dropping the one with lowest density and recomputing $p(\delta_{min}, N_g)$, until this candidate falls within the acceptable level or runs out of galaxies. As a result, some cluster candidates will be reduced in size and others will be eliminated. The final list of candidates is composed by clusters above the required confidence level. This cleaning process is a necessary step because δ^* is set to be permissive as the estimate by (Bahcall, 1986) that $\sim 25\%$ percent of the galaxies are associated with clusters includes their outer tails, which may extend to ~ 30Mpc. This process results in a list of cluster members, given by all the galaxies within the final VT footprint of the cluster. The galaxy belonging to the cell of highest density is taken as the central galaxy.

The accuracy of the membership assignment is limited by the errors in the redshift of the galaxies and width of the redshift shell. As discussed in section 6.2.5, the membership list is improved in the second run of the VT cluster finder, which is performed in boxes centered at the central galaxies flagged during this first run.

6.2.4 Shape measurement

To obtain the cluster shape parameters, we take the galaxies within the cluster VT footprint and compute the second moments of the galaxy distribution with respect to the coordinates (x_c, y_c) of the central galaxy, using the cell densities δ as weights. These second moments are:

$$\begin{aligned} m_{xx} &= \frac{\sum_i \delta_i (x_i - x_c)^2}{\sum_i \delta_i} \\ m_{yy} &= \frac{\sum_i \delta_i (y_i - y_c)^2}{\sum_i \delta_i} \\ m_{xy} &= \frac{\sum_i \delta_i (x_i - x_c)(y_i - y_c)}{\sum_i \delta_i} \end{aligned} \tag{6.4}$$

where the x and y directions are aligned with the RA and Dec axes, respectively. We use these quantities to compute the semi-major and semi-minor axes, a and b, respectively:

$$\begin{aligned} a &= \left[\frac{1}{2}(m_{xx} + m_{yy} + f)\right]^{1/2} \\ b &= \left[\frac{1}{2}(m_{xx} + m_{yy} - f)\right]^{1/2} \end{aligned} \tag{6.5}$$

where

$$f = (m_{xx} - m_{yy} + 4m_{xy})^{1/2}$$

The position angle is also obtained in terms of the same quantities,

$$PA = \frac{180}{\pi} \tan^{-1}\left(\frac{b^2 - m_{xx}}{m_{xy}}\right) \tag{6.6}$$

and is given in degrees.

6.2.5 Catalog construction

A global list of cluster candidates is made by merging the results of the individual shells. For each cluster in that list we extract from the full input galaxy catalog (not the z shells) a 3-dimensional box centered at its central galaxy and with the same size as in the first run: 3×3 square degrees and σ_z width. These boxes are processed with the VT algorithm, repeating the steps described in §6.2.1-6.2.4 and a new global list of cluster candidates is constructed, taking only the clusters found at the center of each box.

We perform a matching between the two global lists. In this matching scheme, candidates are considered the same cluster if they have more than 50% of shared galaxies and multiple matches are not allowed. When a matching occurs, that cluster is eliminated from the list of candidates available for matching with other candidates. The clusters found in the first run but undetected in the second run are eliminated as projection effects. The primary function of this stage, however, is to deal with photo-z slice edge effects.

Because the new boxes are allowed to cross the initial shell boundaries, edge effects in the redshift dimension are eliminated. Clusters split in several components during the initial detection will result in cluster candidates with a number of shared galaxies after the second run. For a given pair of candidates found to be the same cluster (i.e., sharing more than 50% of their galaxies), only the one with the largest number of members is added to the final cluster catalog. Otherwise, they are said to be distinct clusters with shared galaxies (which are flagged in the members list) and both are included in the cluster catalog. Setting the threshold of shared galaxies to 50% is a natural choice, between the two extremes where all candidates would be duplicated or only the clusters found with the same set of member galaxies would be accepted.

At this point the detection is completed. We have the final list of clusters containing RA, Dec, redshift and a list of member galaxies including the parameters of the corresponding VT cells. This forms the VT footprint of the cluster. The cluster redshift is estimated as the median of the redshift of the cluster members. The quantity is better estimated in the second run after a cleaner membership list is obtained, so as to avoid projection effects along the line of sight.

The output parameters of the VT cluster catalog are: ID, RA, Dec (coordinates of its central galaxy or the highest density peak), z (given by the average of all members), σ_z (rms value), δ_c (density contrast measured at the final stage of detection), σ (significance of detection), richness (number of members), size (radius of the circle enclosing all galaxies), a (semi-major axis), b (semi-minor axis), PA (position angle).

We also report a members list containing: ID, host ID (most likely host cluster), cell density, shared flag (1 if the galaxy is shared with another cluster, 0 otherwise), central flag (1 for central galaxy, 0 for regular members). Note that we do not list every possible galaxy-cluster association in the output. Galaxies not associated to any cluster are listed

with host ID, shared flag and central flag set to −1. These non-member galaxies can be used, for instance, to compute the local density of non-member galaxies around a cluster or to run afterburners to measure cluster properties such as richness and R_{200}.

Having a list of members generated by the cluster finder is highly desirable, because properties such as the optical richness and R_{200} can be estimated. The lack of membership assignment in VT implementations using magnitudes was a drawback and we improve on that matter. Also, this allows us to compute the algorithm efficiency as follows.

6.3 Algorithm efficiency

The efficiency of the algorithm is evaluated by measuring the VT catalog completeness and purity as a function of mass and redshift. These quantities are the selection function needed to understand the catalog. This can be done with mock galaxy catalogs with known relations to dark matter halos. The field can no longer be advanced by placing single clusters in the center of an image with random backgrounds.

We apply the algorithm to a mock galaxy catalog and match the resulting cluster catalog with the corresponding mock truth table of halos – the truth table. This allows to define completeness as the fraction of halos with a VT cluster counterpart and purity as the fraction of VT clusters with a matching halo. We perform this in bins of redshift and we also estimate the impact of redshift errors.

6.3.1 Mock catalogs

Our mock galaxy catalogs are created using the ADDGALS code (v1.07) (Busha & Wechsler, 2008, see also Gerdes et al. 2010, Appendix A). These galaxy catalogs reproduce the luminosity function, the magnitude dependent 2-point correlation function and the color-density-luminosity distribution measured from the SDSS data. ADDGALS takes a N-body simulation light cone and attaches galaxies to its dark matter particles. ADDGALS then builds a list of galaxies r-band luminosities drawing from a luminosity function $\phi(M_r)$. The parameters of $\phi(M_r)$ are taken from the observed SDSS r-band luminosity function at redshift ~ 0.1 assuming passive evolution of 1.3 magnitudes per unit redshift. Overdensities of dark matter are then computed using the characteristic radius R_δ, defined as the radius enclosing $1.8 \times 10^{13} h^{-1}$ solar masses of dark matter. For each galaxy, an overdensity is

chosen assuming that a galaxy of given luminosity L_r has the probability $P(R_\delta|L_r/L_*)$ to be located in an overdensity region of size R_δ. The galaxy is finally assigned to a random particle with that density. The function $P(R_\delta|L_r/L_*)$ is determined using a Monte Carlo Markov Chain analysis, imposing that the observed magnitude dependent 2-point correlation function is matched. The next step is to assign galaxy colors. The local galaxy density is computed for each galaxy in the simulation and in a training set of galaxies from the SDSS catalog. Each mock galaxy is assigned the SED of a randomly selected SDSS galaxy with similar local galaxy density and absolute magnitude M_r. The SED is then k-corrected and the appropriate filters are applied to obtain SDSS colors. At high redshift, color information is extrapolated from low redshifts: r-band magnitudes are passively evolved and k-corrected assuming that the rest-frame colors and the color-density-luminosity distribution remain unchanged. The resulting catalog reproduces the overall photometric and clustering properties of the SDSS galaxies at low redshifts ($z \sim 0.3$) and extends, using simplified assumptions, to higher redshifts ($z \sim 1.3$) and deeper magnitudes ($r \sim 24$). The brightest cluster galaxies (BCGs), however are an exception. BCGs luminosities are tightly correlated with their host halo mass and are not reproduced by this method. Therefore, a BCG luminosity is calculated for each resolved halo (of mass $\sim 5 \times 10^{13} h^{-1} M_\odot$ and above) before the usual galaxy-to-dark matter particle assignment begins. The corresponding galaxies are then removed from the initial list of galaxies and placed at the center of its host halo.

We run our cluster finder in a set of mock catalogs and compare our results with the truth table. The quantities featured in the truth table are RA, Dec, redshift and M_{200}, plus list of member galaxies of each halo. In this Chapter, we refer to the truth table as the *halo catalog*, and to the VT output as the *cluster catalog*. The quantities we use as inputs are, as always: RA, Dec, and photometric redshift. We generate photometric redshifts from the true redshifts, using a Gaussian distribution of width $\sigma_z(1+z)$. We test four different values of σ_z, namely 0.015, 0.03, 0.045 and 0.06 to assess the impact of the photometric redshift errors in our cluster finder.

The discussion so far was restricted to a perfect volume limited galaxy catalog. A real galaxy catalog, however, will have an irreducible level of contamination and incompleteness. Here we mimic the effects of these two quantities in the mocks by assuming that the

completeness of the input galaxy catalog as a function of magnitude in r band, $C_g(r)$, is well described by a fitting function that resembles a Fermi-Dirac distribution:

$$C_g(r) = \frac{f_0}{1 + \exp((r - \mu)/\sigma)} \tag{6.7}$$

where μ is the magnitude limit of the catalog, f_0 is a normalization constant and the parameter σ controls how fast the completeness falls when the magnitude limit is reached. The parameters f_0 and σ are obtained by modeling the completeness of SDSS data, as estimated by the software 2DPHOT (La Barbera et al., 2008b), with Eq. 6.7. We found that $f_0 = 0.99$ and $\sigma = 0.2$ are typical values. We degrade the mock catalogs using $\mu = 23.5$, interpreting $C_g(r)$ as the probability that a galaxy of magnitude r is detected. Similarly, from the SDSS data we infer that a small fraction of contaminants, due to misclassified stars, can be present in the input catalog. Using 2DPHOT, we find that the fraction of misclassified objects increases exponentially for magnitudes above $\mu - 1.5$, for the SDSS $r = 22$. We take this into account by generating false galaxies randomly above this limit and drawing from (6.7) the probability that this object is actually added to the catalog.

6.3.2 Membership matching

The evaluation of completeness and purity requires a well defined matching scheme between the cluster catalog and the truth table. We use a membership-based matching method. Membership matching has been used in evaluating completeness and purity of both photometric and spectroscopic catalogs (White & Kochanek, 2002; Eke et al., 2004; Gerke et al., 2005; Koester et al., 2007a). Unlike cylindrical matching, which has been largely employed in this kind of study, this method is parameter-free, unambiguous and provides the means to evaluate the efficiency of the cluster finder as a function of halo mass regardless of the observable proxy for mass. This allow us to distinguish the aspects relevant to the cluster finding problem from aspects connected to the mass-observable proxy calibration, which is a problem per se and is better addressed by a separate set of post-finding algorithms.

The inputs for the matching code are the halo catalog and the cluster catalog. The first is ranked by mass while the latter is ranked by the number of galaxies, both in descending

order and in bins of redshift. It is critical to do the ranking in bins of redshift for both the halos and the clusters. In the case of halos, the mass function is evolving, so the masses will be changing at fixed rank. In the case of the clusters, the luminosity limit will be evolving with redshift, so the ranks will be changing at fixed mass. If this is not taken into account, a massive cluster at high-z ($z \sim 1$) will get a much lower rank than a massive cluster at low-z ($z \sim 0.1$).

After ranking, the first step is to fit a rank-mass relation $R(M)$ to the halo catalog. We use the fitting formula

$$R(M) = \left(\frac{M}{M_p}\right)^{\alpha} \exp\left(\exp\left(\frac{M_0 - M}{M_e}\right) - \frac{M}{M_1}\right) \tag{6.8}$$

This relation has no motivation other than a global fitting function, valid at all redshifts provided that the ranking is performed as described above. For our mock catalogs, the best fit parameters for this fitting function are $M_p = 2.26 \times 10^{17}$, $M_e = 1.40 \times 10^{14}$, $M_0 = 1.85 \times 10^{13}$, $M_1 = 1.85 \times 10^{14}$ and $\alpha = -1.15$. We then invert the relation above to compute an "observed mass" for each cluster and proceed to the matching. If the proxy used to rank the clusters has a tight correlation with mass, the ranking will be accurate and the observed mass will show a tight correlation with the true mass for the matched pairs. It is important to notice, that the use of ranking instead of observed mass, does not require the mass-observable relation to be calibrated. Moreover, neither mass information nor the ranking is used in the matching process, which is membership-based.

A match takes place if a fraction of member galaxies is shared by a halo-cluster pair. The best match is the object sharing the largest fraction of galaxies. We require unique matching, in which a given halo/cluster is not allowed to be associated with more than one cluster/halo. As both lists are ranked by number of galaxies, uniqueness is imposed by eliminating a matched object from the list of available objects for future matches down the list. We also require two-way matching, where the best matching pair is found when the matching is performed in both directions, halos-to-clusters and clusters-to-halos.

6.3.3 Completeness & Purity

Completeness is defined as the fraction of halos having a counterpart in the cluster catalog. Purity in turn is defined as the fraction of objects in the cluster catalog that

correspond to a true halo. In both cases, only unique two-way matches are considered. Allowing for non-unique matching, where each cluster may have more than one matching halo and vice-versa, would be a more permissive approach. For instance, purity would not be affected by a halo being split in two components and completeness would not be affected by two halos appearing as a single cluster.

We count the number of matched objects in bins of mass and redshift. Therefore,

$$C(M,z) = \frac{N_{\rm matched}(M,z)}{N_{\rm halos}(M,z)} \tag{6.9}$$

$$P(M,z) = \frac{N_{\rm matched}(M,z)}{N_{\rm clusters}(M,z)} \tag{6.10}$$

Note that $C(M,z)$ can be computed using the true mass of the halos, being totally independent of the mass proxy used to rank the clusters. The true mass of the clusters, however, is available only for the matched objects. Therefore $P(M,z)$ has to be computed using the observed mass and does depend on the ranking. We fit a power law to the $M_{\rm obs} - M_{\rm true}$ relation from the matched objects and use it to transform the scale in the $P(M,z)$ plots and show both completeness and purity as a function of $M_{\rm true}$. This cannot be performed before the rank-mass relation fitting step, which is part of the matching process. This method allow us to evaluate the efficiency of any cluster finder imposing minimum requirements, namely a list of members for each cluster. The selection function can be defined in terms of completeness and purity as

$$f(M,z) = \frac{C(M,z)}{P(M,z)}. \tag{6.11}$$

This is a simplified definition. For cosmological studies with real data, $f(M,z)$ should be defined and evaluated in a likelihood analysis that includes the scatter in the mass-observable relation after calibration. Here, however, we simply want to compare the observed cluster number counts $N_{obs}(M,z)$ to the predictions from the ΛCDM cosmological model $N_{\Lambda{\rm CDM}}(M,z)$. In this case, the selection function is easily taken into account:

$$N_{obs}(M,z) = f(M,z) N_{\Lambda{\rm CDM}}(M,z). \tag{6.12}$$

This comparison allows us to develop a feel for how well we can recover the true cluster number counts using the VT catalog and our ability to perform a cosmological test using VT clusters as a probe.

However, the method described above is very simplified with respect to the procedures involved in an actual measurement of the mass function. This would require a measurement of the mass-observable relation and its scatter, rejecting outliers as a source of completeness and purity. We do not perform this because the VT cluster catalog does not provide an optimal mass proxy. We use the number of galaxies as a mass proxy, but this was not optimized to have a tight relation with mass at all redshifts and all mass bins. Optimizing a mass proxy is a necessary step if the VT is to be used in performing cosmological tests. But this problem is better addressed by a separate algorithm, specifically designed to provide a calibrated mass proxy including the mean relation and the scatter.

6.4 Results and discussion

In Fig. 6.3 we show the completeness and purity as a function of mass and redshift for different Gaussian σ_z values. The photometric redshift errors have a strong impact on both completeness and purity. For $\sigma_z = 0.015$, completeness lies above 80% for all redshift bins and masses above $\sim 10^{13.5} M_\odot$. Purity however, drops significantly at the low mass end. We attribute this to the fact that the range $10^{13.5} - 10^{14} M_\odot$ is at the lower boundary of the halo catalogs associated with the mock catalog. ADDGALS will populate some fraction of real dark matter clumps in the simulation even if they are below the threshold for detection in the halo catalog. A fraction of these halos were populated with galaxies by ADDGALS, but were not listed in the truth table. We have no means to determine the exact fraction at this point and therefore we interpret the purity curve as a lower limit.

In the high-redshift regime, completeness and purity do not change much with σ_z. The lowest redshift bin, however, shows the lowest purity and completeness in almost all cases. This may be due to the large angular size of clusters at low-z. At $z \sim 0.1$ the target area of 1 square degree corresponds to a few times the R_{200} of a typical cluster. This hypothesis requires further investigation. However, even in this case the VT catalog achieves completeness and purity above $\sim 80\%$ at all masses. Since we are most interested in a reliable catalog at high redshifts, we consider the cluster finder efficiency, as shown in Fig. 6.3, very good.

Note that the behavior of purity is qualitatively different in the last panel, $\sigma_z = 0.060$. This may be connected to low redshift clusters leaking to high redshift shells at higher

rates than the high redshift ones fall towards low redshift.

Testing the effect of changes in the cluster finder free parameters on the completeness and purity functions, we find that:

1. Changing the fraction of shared galaxies required to consider two candidates as the same cluster in the range $40-60$ percent has less than 1% impact on the results. We fix this value at 50%.

2. The selection function is very sensitive to $scl(z)$. Setting $scl(z)$ too high (> 0.97) leads to fragmentation of clusters, which affects purity at all masses, and failure to detect low contrast clusters, which affects completeness at the low mass end. Setting $scl(z)$ below 0.75 causes merging of clusters and affects completeness. An optimal value for $scl(z)$ in the range $0.75-0.97$ has to be found at each redshift bin.

3. The confidence level threshold has little effect on the detection. The final list of clusters shows less than 10% difference when this parameter varies in the range $90-99.5$ percent. But it affects the selection function by modifying the membership list.

Fig. 6.4 illustrates our ability to recover the true cluster number counts of the input catalog. We take the case $\sigma_z = 0.015$ and the redshift bin $0.9 < z < 1.1$. For a given mass bin M_i we divide the number of VT clusters detected by the selection function term $f(M_i, z)$. We then sum the corrected counts through all bins of mass $> M$ (red solid line). The curve for the truth table is done by counting all the halos above M (black dotted line). We finally plot (blue dashed line) the values expected in a ΛCDM cosmology (*e.g.* Evrard et al., 2002b) for comparison.

There is a remarkable agreement between the three curves. The tilt of the measured curve with respect to the truth table may be interpreted as low mass clusters being misplaced towards more massive bins, due to our neglect of the scatter in the mass-observable relation. As pointed out in §6.3.3, the method used here does not take into account crucial steps involved in an actual measurement of the mass function. This issue must be addressed with a full program of mass calibration and is beyond the scope of this Chapter. The result shown in Fig. 6.4 encourages the pursuit of such a program, though. Our results show that the VT is a reliable cluster finder in the redshift and mass range of interest, as

seen in the completeness and purity curves. Application of this algorithm on SDSS data is underway and will be presented in a forthcoming paper.

6.5 Summary

In this Chapter we have presented an improved implementation of the Voronoi Tessellation cluster finder. Improvements with respect to earlier works include:

1. The use of photometric redshifts instead of magnitudes.

2. A more realistic assumption that galaxy fields have two-point correlation function described by a power-law, and not a by Poisson distribution.

3. Implementation of a membership assignment scheme.

The VT cluster finder in 2+1 dimensions was tailored to fulfill the requirements of upcoming cosmological experiments aiming at using clusters as probes for dark energy. The main challenges towards this goal include the construction of reliable cluster catalogs up to high redshifts ($z \sim 1$) and down to low mass limits ($\sim 10^{13.5} M_{\odot}$) and the measurement of the selection function as a function of M and z. To achieve these goals using the VT we:

1. Adapted the VT algorithm to use photometric redshift shells and take advantage of the relation that we have discovered between the two-point correlation function of the galaxy field and its distribution of VT cell areas.

2. Defined the selection function in term of completeness and purity, establishing an objective way to measure these quantities using simulated catalogs.

3. Applied the VT to mock galaxy catalogs and computed the completeness and purity of the output cluster catalog with the truth table, showing that the VT can produce cluster catalogs with completeness and purity above 80% in the ranges of interest within the M-z parameter space.

4. Computed the cluster abundance from the VT catalog and compared it to the halo abundance in the mocks, finding a remarkable agreement at all mass bins.

These results allow us to be confident in our ability to perform a cosmological test for dark energy using the VT algorithm on a data set of sufficient scope. Analysis of the application of the VT to the SDSS data is underway and will be presented elsewhere.

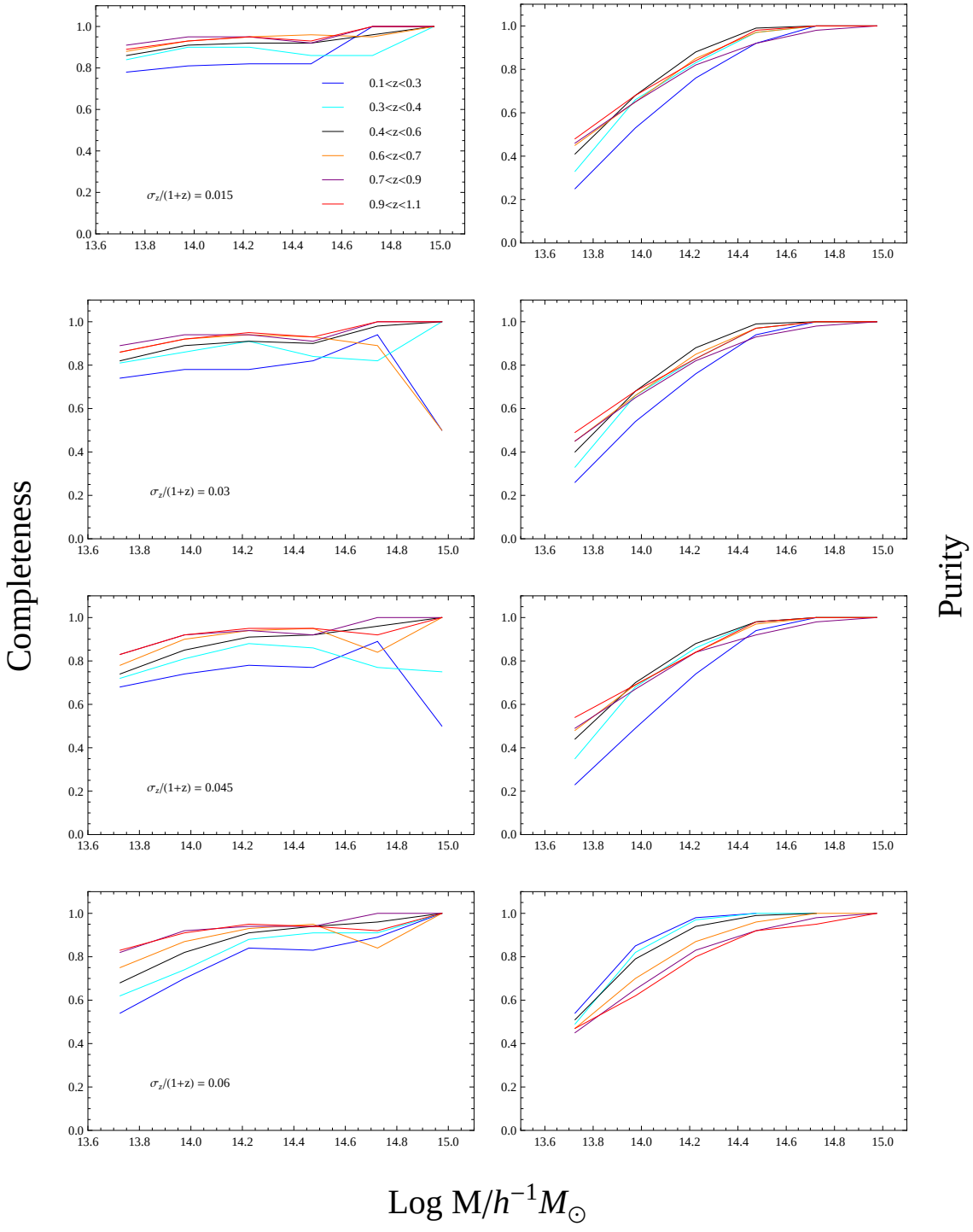

Figure 6.3: Completeness (left) and purity (right) curves as a function of mass for six redshift bins: $0.1 < z < 0.3$ (blue), $0.3 < z < 0.4$ (cyan), $0.4 < z < 0.6$ (black), $0.6 < z < 0.7$ (orange), $0.7 < z < 0.9$ (purple), $0.9 < z < 1.1$ (red). From top to bottom, the plot pairs feature different σ_z values: 0.015, 0.03, 0.045, 0.06. The photometric redshift errors have a strong impact on both completeness and purity. In the best case, completeness and purity rest above 80% for all redshift bins and masses above $\sim 10^{14.2}$. In the case of purity, this curve should be interpreted as a lower limit (see text for discussion).

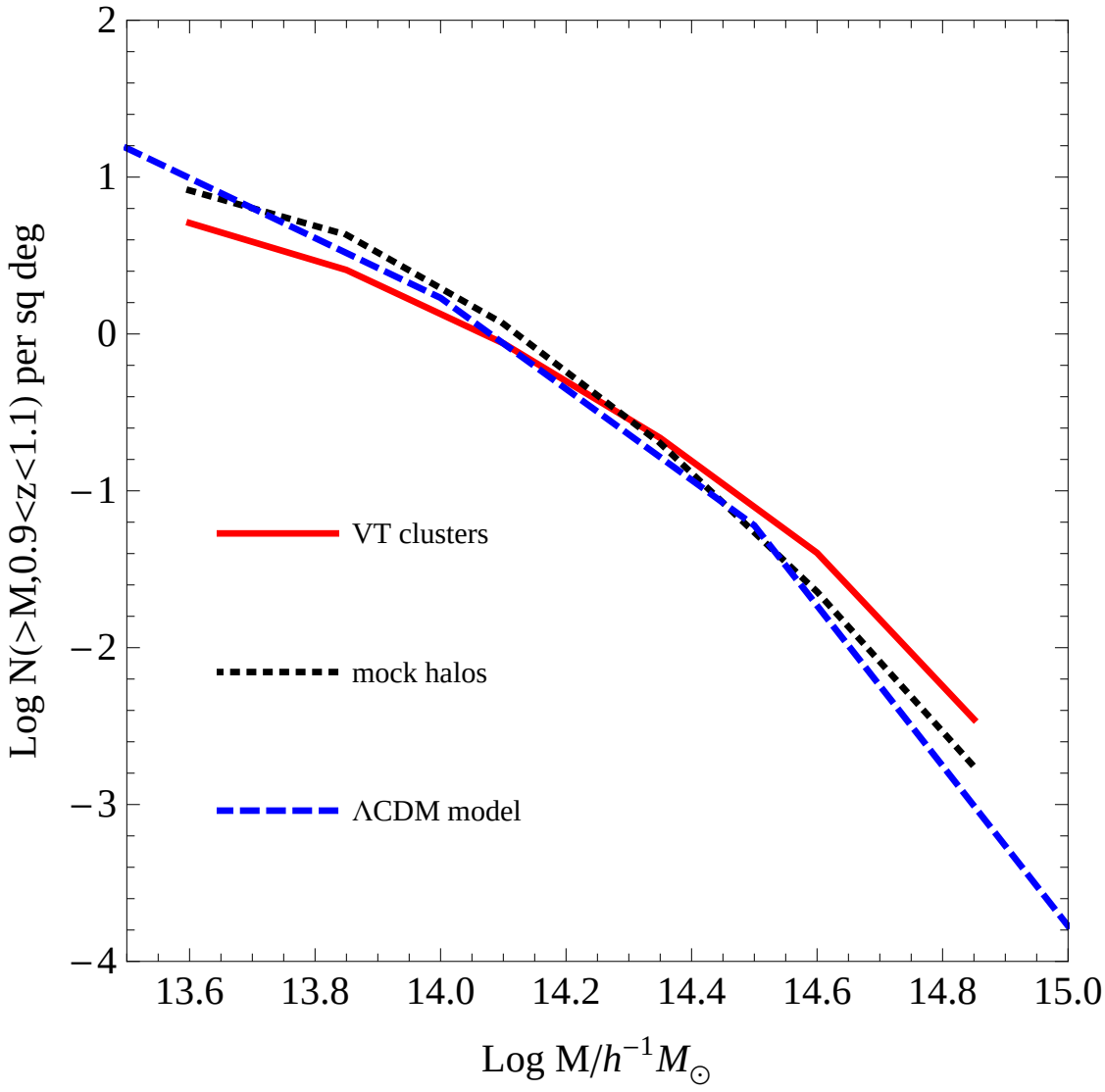

Figure 6.4: Cumulative cluster abundance as a function of mass in the redshift range $0.9 < z < 1.1$. The black (dotted) line shows the counts in the truth table; the red solid (solid) line shows the results of the VT catalog, taking $\sigma_z = 0.015$; the blue (dashed) line shows the values predicted for a ΛCDM cosmology.

Chapter 7

Conclusions & Perspectives

Results of our research program aiming at dark energy constraints from optical galaxy clusters are presented. We focus on extraction of cosmological information from the SDSS Coadd data (Annis et al., 2010), a survey covering 250 square degrees with depth up to $z \sim 1$. This is a task of great complexity that requires the following accomplishments:

1. Production of a galaxy catalog with completeness and contamination well determined and shape measurements included.

2. Photometric redshift information.

3. Construction of a cluster catalog, with high completeness and purity and well-understood selection function.

4. Calibration of a mass-proxy relation for the cluster sample.

5. Implementation of a likelihood method that takes into account the uncertainties in the selection function and the mass calibration.

The first of these requirements is fulfilled by the Coadd data, for which photometric redshifts (Reis et al., 2010), galaxy shape parameters (Seo et al., 2010) and completeness function curves have been measured. The latter was measured in selected sub-fields. Plans regarding a full run of 2DPHOT aim at extending this for the full catalog and on providing an independent set of shape measurements for comparison, which would help understand residual systematics present in the current catalogs.

2DPHOT (Chapter 2) is a general purpose analysis environment for wide-field images. An important feature of 2DPHOT is that it provides a measurement of completeness

and contamination of the galaxy catalog it creates. This information is required when evaluating the selection function of clusters. Since its release, (La Barbera et al., 2008b) 2DPHOT has been used in several other projects (e.g. Gargiulo et al., 2009; Haines et al., 2009; Merluzzi et al., 2010) and has been integrated into the Brazilian Virtual Observatory (BRAVO, Chapter 3). Therefore, 2DPHOT is an application available not only for this particular project, but also to the astronomical community. In light of the tendency of generation of vast amounts of data from various ongoing and forthcoming surveys, is in fact an important contribution.

The VT cluster finder (Chapter 6) was developed to meet the demands of building a reliable cluster catalog up to high redshift. The algorithm has been extensively tested on simulations, demonstrating that VT can produce a cluster catalog with high completeness and purity for clusters above $\sim 10^{13.5}$ solar masses for all bins of redshift up to $z \sim 1$. The VT is being used to produce a cluster catalog from the Coadd galaxy catalog.

We address the problem of cluster mass estimation of clusters via weak lensing in Chapter 4. We have reported the measurement of seven individual nearby clusters as part of the Sloan Nearby Weak Lensing Survey (Kubo et al., 2009b). Since then, we have turned our mass measurement framework into a production framework and performed measurements for nearly 1000 clusters. Mass measurement of individual clusters are invaluable information, but in the context of a dark energy project, we actually have to perform a global mass calibration of the sample. This can be done by dividing the cluster into richness bins and measuring the mean mass of each stack. Our framework allows us to follow this approach and it is currently a project being pursued using the Coadd galaxies as sources (Kubo et al., 2010) and the maxBCG clusters (at $0.1 < z < 0.3$) as lenses. The same approach can be extended for the Coadd clusters.

Strong lensing effect on the other hand, can provide an independent measurement of individual cluster masses. We report in Chapter 5 on seven systems found by our ongoing Sloan Bright Arcs Survey, which has reported a total of 18 clusters. It is not clear how we can combine this information into a mass calibration scheme. However, information from those systems are interesting for studies of galaxy evolution, as high-redshift galaxies are strongly magnified by the lens.

Dark energy and other projects involving large scale surveys require handling and processing of data of size and complexity never yet seen in Astronomy. The enterprise of establishing BRAVO (Chapter 3) is motivated by this need. We have been able to put in place a number of tools to achieve this goal. In addition to 2DPHOT, the Voronoi Tessellation cluster finder algorithm (VT, Soares-Santos et al., 2010, see Chapter 6) is being integrated to BRAVO. The development of 2DPHOT, VT and the virtual observatory itself are deeply connected to the challenge of constructing and analyzing suitable for the study of dark energy. Addressing all those challenges, we have been able to produce scientific results from the currently available data and to prepare for upcoming projects.

The problem of dark energy, established by observations in the last ten years, has been extensively studied by both theorists and experimentalists. In its simplest formulation, dark energy features an equation of state $p = w\rho$ with $w < -1/3$ and several mechanisms for generating such an equation of state are known. To push forward our understanding of dark energy, however, it is important to establish the w parameter and its time-variation, w', experimentally. Exploring the sensitivity of the cluster abundance to w, we perform a cosmological test based on the SDSS Coadd data. We have shown that the VT cluster catalogs reproduce the cluster abundance in mock galaxy catalogs (Fig. 6.4) and that w can be constrained to $\sim 40\%$ uncertainty by using this powerful approach (Fig. 1.4). This is comparable to the current constraints from other probes such as CMB, SNe and BAO. It represents the first measurement of w using cluster abundance of optically selected galaxy clusters. Future experiments such as the DES will benefit from the developments made for this thesis, being able to improve the current constraints by measuring w with $\sim 5\%$ and w' with $\sim 30\%$ uncertainty. In this thesis, we did not perform a full mass-calibration of the VT clusters, but comparing Figs. 6.4 and 1.1 we can see that the differences between the mock mass function and the VT mass function are already much smaller than the difference between ΛCDM and a $w = -1.9$ model, for instance. This is a remarkable result and encourages a pursuit of a measurement of w using the VT cluster finder applied to the SDSS Coadd data.

Bibliography

Abazajian, K. et al. 2004, AJ, 128, 502

—. 2005, AJ, 129, 1755

—. 2009, ApJS, 182, 543

Abbott, T. et al. 2005, astro-ph/0510346

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

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

Adelman-McCarthy, J. K. et al. 2007, ApJS, 172, 634

Adelman-McCarthy, J. K. et al. 2008, ApJS, 175, 297

Albrecht, A. & Bernstein, G. 2007, PRD, 75, 103003

Albrecht, A. et al. 2006, arXiv:astro-ph/0609591

Aldering, G. 2005, New Astronomy Review, 49, 346

Allam, S. S. et al. 2007, ApJ, 662, L51

—. 2004, AJ, 127, 1883

Allen, S. W. et al. 2003, MNRAS, 342, 287

Annis, J. et al. 1999, Bulletin of the Amerian Astronomical Society, 31, 1391

—. 2005, arXiv:astro-ph/0510195

Annis, J. et al. 2010, ApJ, in prep.

Armendariz-Picon, C., Mukhanov, V. & Steinhardt, P. J. 2000, PRL, 85, 4438

Arnaboldi, M. et al. 2007, The Messenger, 127, 28

Astier, P. et al. 2006, A&A, 447, 31

Bahcall, N. A. 1986, ApJ, 302, L41

—. 2000, Phys. Rep., 333, 233

Bahcall, N. A. & Cen, R. 1992, ApJ, 398, L81

Bahcall, N. A., Fan, X.—. 1997, ApJ, 485, L53+

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

Bahcall, N. A. et al. 2003a, ApJS, 148, 243

—. 2003b, ApJ, 585, 182

Ball, N. M. et al. 2006, ApJ, 650, 497

Barkhouse, W. A. et al. 2006, ApJ, 645, 955

Bartelmann, M. & White, S. D. M. 2002, A&A, 388, 732

Becker, M. R. et al. 2007, ApJ, 669, 905

Benacchio, L. & Pasian, F., eds. 2007, Grid-enabled Astrophysics

Bender, R. & Moellenhoff, C. 1987, A&A, 177, 71

Bender, R. et al. 1989, A&A, 217, 35

Bendinelli, O., Parmeggiani, G. & Zavatti, F. 1982, Ap&SS, 83, 239

Bento, M. C., Bertolami, O. & Sen, A. A. 2002, Phys. Rev. D, 66, 043507

Berlind, A. et al. 2010, in prep.

Berlind, A. A. et al. 2006, ApJS, 167, 1

Bernstein, G. M. & Jarvis, M. 2002, AJ, 123, 583

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

Bishop, C. M. 1994, Review of Scientific Instruments, 65, 1803

—. 2006, Pattern Recognition and Machine Learning

Blanchard, A. et al. 2003, A&A, 412, 35

Blanton, M. R. et al. 2005, AJ, 129, 2562

Borgani, S. 2006, arXiv:astro-ph/0605575

Borgani, S. & Guzzo, L. 2001, Nature, 409, 39

Borgani, S. et al. 2001, ApJ, 561, 13

Briel, U. G., Henry, J. P. & Boehringer, H. 1992, A&A, 259, L31

Busha, M. & Wechsler, R. 2008, 43rd Rencontres de Moriond Proceedings

Caldwell, R. R. 2002, Physics Letters B, 545, 23

Cannon, R. 2006, 372, 425

Caretta, C. A. et al. 2008, A&A, 487, 445

Carroll, S. M. 1998, Physical Review Letters, 81, 3067

—. 2001, Living Reviews in Relativity, 4, 1

Cen, R. 1997, ApJ, 485, 39

Chiba, T., Okabe, T. & Yamaguchi, M. 2000, Phys. Rev. D, 62, 023511

Cole, S. et al. 2005, MNRAS, 362, 505

Connolly, A. J.—. 2002, ApJ, 579, 42

Copeland, E. J., Sami, M. & Tsujikawa, S. 2006, International Journal of Modern Physics D, 15, 1753

Dahle, H. 2007, arXiv:astro-ph/0701598

Dalton, G. B. et al. 1997, MNRAS, 289, 263

Daly, R. A. et al. 2007, arXiv:0710.5112, 710

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

Davis, M. & Peebles, P. J. E. 1983, ApJ, 267, 465

de Carvalho, R. et al. 2010, PACIS, 187

del Popolo, A., Ercan, N. & Yeşilyurt, I. S. 2005, A&A, 432, 771

Diehl, H. T. et al. 2009, ApJ, 707, 686

Djorgovski, S. G. et al. 1998, Wide Field Surveys in Cosmology

Donahue, M. & Voit, G. M. 1999, ApJ, 523, L137

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

Drinkwater, M. J. et al. 2010, 401, 1429

Dryden, I. 2005, Annals of Statistics

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

Eisenstein, D. J. et al. 2001, AJ, 122, 2267

Eisenstein, D. J. et al. 2005, ApJ, 633, 560

Eke, V. R., Cole, S. & Frenk, C. S. 1996, MNRAS, 282, 263

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

—. 2004, MNRAS, 348, 866

El-Ad, H., Piran, T. & da Costa, L. N. 1996, ApJ, 462, L13+

Evrard, A. E. 1989, ApJ, 341, L71

Evrard, A. E. et al. 2002a, ApJ, 573, 7

—. 2002b, ApJ, 573, 7

Fada, D. et al. 1996, ApJ, 473, 670

Fan, X., Bahcall, N. A. & Cen, R. 1997, ApJ, 490, L123+

Fosalba, P., Gaztañaga, E. & Castander, F. J. 2003, ApJ, 597, L89

Francois, D. 2008, High-dimensional Data Analysis: From Optimal Metrics to Feature Selection

Frenk, C. S. et al. 1990, ApJ, 351, 10

Fu, L. et al. 2008, A&A, 479, 9

Fukugita, M. et al. 1996, AJ, 111, 1748

Gal, R. R. et al. 2000, AJ, 120, 540

—. 2003, AJ, 125, 2064

—. 2009, AJ, 137, 2981

Gal, R. R., Lubin, L. M. & Squires, G. K. 2005, AJ, 129, 1827

Gargiulo, A. et al. 2009, MNRAS, 397, 75

Garilli, B. et al. 2008, A&A, 486, 683

Gerdes, D. W. et al. 2010, ApJ, 715, 823

Gerke, B. F. et al. 2005, ApJ, 625, 6

Girardi, M. et al. 1998, ApJ, 505, 74

Gladders, M. D. & Yee, H. K. C. 2005, ApJS, 157, 1

Gladders, M. D. et al. 2007, ApJ, 655, 128

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

Gould, A. & Villumsen, J. 1994, ApJ, 428, L45

Griffiths, R. et al. 2004, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 5488, UV and Gamma-Ray Space Telescope Systems, ed. G. Hasinger & M. Turner, 209–221

Guerra, E. J., Daly, R. A. & Wan, L. 2000, ApJ, 544, 659

Gunn, J. E. et al. 1998, AJ, 116, 3040

—. 2006, AJ, 131, 2332

Haiman, Z., Mohr, J. J. & Holder, G. P. 2001a, ApJ, 553, 545

Haiman, Z., Mohr, J. J.—. 2001b, ApJ, 553, 545

Haines, C. P. et al. 2009, MNRAS, 396, 1297

Hanisch, R. & Quinn, P. 2010, http://www.ivoa.net/pub/info/TheIVOA.pdf

Hansen, S. M. et al. 2005, ApJ, 633, 122

Hao, J. 2009, PhD thesis, University of Michigan

Hao, J. et al. 2009, ApJ, 702, 745

Hao, J. et al. 2010, in prep.

Hartigan, J. 2009, Clustering Algorithms

Haykin, S. 1994

Hege, H. & Polthier, K. 2002

Hernandez-Toledo, H. 2008, A&A, 136, 2115

Hey, T., Tansley, S. & Tolle, K. 2009, The Fourth Paradigm - Data-Intensive Scientific Discovery, Microsoft Research

Hirata, C. & Seljak, U. 2003, MNRAS, 343, 459

Hirata, C. M. et al. 2004, MNRAS, 353, 529

Hoekstra, H. 2001, A&A, 370, 743

Hogg, D. W. et al. 2001, AJ, 122, 2129

Holder, G., Haiman, Z. & Mohr, J. J. 2001, ApJ, 560, L111

Hu, W. 2003, Phys. Rev. D, 67, 081304

Hu, W. & Cohn, J. D. 2006, Phys. Rev. D, 73, 067301

Hu, W. & Kravtsov, A. V. 2003a, ApJ, 584, 702

Hu, W.—. 2003b, ApJ, 584, 702

Huchra, J. P. & Geller, M. J. 1982, ApJ, 257, 423

Huterer, D. & Turner, M. S. 2001, Phys. Rev. D, 64, 123527

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

Jain, A. et al. 2000, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22, 1

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

Joffre, M. et al. 2000, ApJ, 534, L131

Johnston, D. E. et al. 2007, ArXiv e-prints

Jorgensen, I., Franx, M. & Kjaergaard, P. 1995, MNRAS, 273, 1097

Kamenshchik, A., Moschella, U. & Pasquier, V. 2001, Physics Letters B, 511, 265

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

Kelson, D. D. et al. 2000, ApJ, 531, 184

Kiang, T. 1966, Zeitschrift fur Astrophysik, 64, 433

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

Knop, R. A. et al. 2003, ApJ, 598, 102

Koester, B. P. et al. 2007a, ApJ, 660, 239

—. 2007b, ApJ, 660, 221

Komatsu, E. et al. 2010, ArXiv e-prints

Kubo, J. et al. 2010, 41, 384

Kubo, J. M. et al. 2009a, ApJ, 696, L61

—. 2007, ApJ, 671, 1466

—. 2009b, ApJ, 702, L110

Kuntsche, E. et al. 2003, Swiss Journal of Psychology/Schweizerische Zeitschrift fur Psychologie/Revue Suisse de Psychologie, 62, 202

La Barbera, F. et al. 2008a, ApJ, 689, 913

—. 2002, ApJ, 571, 790

—. 2009, AJ, 137, 3942

—. 2005, ApJ, 626, L19

—. 2008b, PASP, 120, 681

Lacey, C. & Cole, S. 1994, MNRAS, 271, 676

LaRoque, S. J. et al. 2006, ApJ, 652, 917

Levine, E. S., Schulz, A. E. & White, M. 2002a, ApJ, 577, 569

Levine, E. S., Schulz, A. E.—. 2002b, ApJ, 577, 569

Lima, M. & Hu, W. 2004, Phys. Rev. D, 70, 043504

Lima, M.—. 2005, Phys. Rev. D, 72, 043006

Lin, H. et al. 2009, ApJ, 699, 1242

Lin, Y.-T., Mohr, J. J. & Stanford, S. A. 2003, ApJ, 591, 749

Linder, E. V. & Jenkins, A. 2003, MNRAS, 346, 573

Lobo, C. et al. 2000, A&A, 360, 896

Lopes, P. et al. 2009, MNRAS, 392

Lopes, P.—. 2010, in prep.

Lopes, P. A. A. et al. 2004, AJ, 128, 1017

Lotz, J. et al. 2004, AJ, 128, 163

Lumsden, S. L. et al. 1992, MNRAS, 258, 1

Lupton, R. et al. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden Jr., F. A. Primini, & H. E. Payne, 269–+

Majumdar, S. & Mohr, J. J. 2004, ApJ, 613, 41

Mandelbaum, R. et al. 2005, MNRAS, 361, 1287

Mercurio, A. 2008, MNRAS, 387, 1374

Merluzzi, P. et al. 2010, MNRAS, 402, 753

Metzler, C. A. et al. 1999, ApJ, 520, L9

Miller, C. J. et al. 2005, AJ, 130, 968

Miralda-Escude, J. 1991, ApJ, 370, 1

Mohr, J. J. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 339, Observing Dark Energy, ed. S. C. Wolff & T. R. Lauer, 140

Mukhanov, V. F., Feldman, H. A. & Brandenberger, R. H. 1992, Phys. Rep., 215, 203

Munshi, D., Porciani, C. & Wang, Y. 2004, MNRAS, 349, 281

Narayan, R. & Bartelmann, M. 1996, ArXiv Astrophysics e-prints

Navarro, J. F., Frenk, C. S. & White, S. D. M. 1996, ApJ, 462, 563

Okabe, A. 2000, Spatial tessellations: concepts and applications of voronoi diagrams

Olsen, L. et al. 2005, å, 435, 781

Oukbir, J. & Blanchard, A. 1992, A&A, 262, L21

Oukbir, J.—. 1997, A&A, 317, 1

Oyaizu, H. et al. 2008, ApJ, 674, 768

Page, L. et al. 2003, ApJS, 148, 233

Peebles, P. J. & Ratra, B. 2003, Reviews of Modern Physics, 75, 559

Peebles, P. J. E., Daly, R. A. & Juszkiewicz, R. 1989, ApJ, 347, 563

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

Pier, J. R. et al. 2003, AJ, 125, 1559

Pierpaoli, E. et al. 2003, MNRAS, 342, 163

Postman, M. et al. 1996, AJ, 111, 615

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

Quinlan, J. R. 1986, Machine Learning, 1, 81

—. 1993, Programs for Machine Learning. San Mateo, CA: Morgan Kaufman

Ramella, M. et al. 2001, A&A, 368, 776

Rassat, A. et al. 2007, MNRAS, 377, 1085

Ratra, B. & Peebles, P. J. E. 1988, Phys. Rev. D, 37, 3406

Reiprich, T. H. & Böhringer, H. 2002, ApJ, 567, 716

Reis, R. et al. 2010, ApJ, in prep.

Rieber, L. 1995, Educational Technology Research and Development, 43, 45

Riess, A. G. et al. 1998, AJ, 116, 1009

Riess, A. G.—. 2004, ApJ, 607, 665

Rines, K. et al. 2002, AJ, 124, 1266

Rintoul, M. D. & Torquato, S. 1997, Journal of Colloid and Interface Science, 186, 467

Rosa, R. et al. 2007, Physica A, 386, 666

Ross, N. et al. 2010, Bulletin of the American Astronomical Society, 41, 517

Rozo, E. et al. 2007a, arXiv:astro-ph/0703574

—. 2007b, arXiv:astro-ph/0703571

—. 2010, ApJ, 708, 645

Ruhl, J. et al. 2004, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 5498, Millimeter and Submillimeter Detectors for Astronomy II, ed. C. M. Bradford et al., 11–29

Ruiz, R. et al. 2008, National Congress on Computing and Applied Mathematics

Sandage, A. 1961, ApJ, 133, 355

Sandage, A. & Perelmuter, J. 1990, ApJ, 350, 481

Seljak, U. 2002, MNRAS, 337, 769

Seljak, U. et al. 2005, Phys. Rev. D, 71, 103515

Seo, H. et al. 2010, in prep.

Sheldon, E. S. et al. 2009, ApJ, 703, 2217

Sheth, R. K. & Lemson, G. 1999, MNRAS, 304, 767

Sheth, R. K. & Tormen, G. 1999, MNRAS, 308, 119

Shewchuk, J. R. 1996, in Lecture Notes in Computer Science, Vol. 1148, Applied Computational Geometry: Towards Geometric Engineering, ed. M. C. Lin & D. Manocha (Springer-Verlag), 203–222, from the First ACM Workshop on Applied Computational Geometry

Soares-Santos, M. et al. 2008, 34th ICHEP Proceedings, arXiv:0810.3689

Soares-Santos, M. et al. 2010, ApJ, submitted

Spergel, D. N. et al. 2003, ApJS, 148, 175

Spergel, D. N. et al. 2007, ApJS, 170, 377

Stebbins, A., McKay, T. & Frieman, J. A. 1996, in IAU Symposium, Vol. 173, Astrophysical Applications of Gravitational Lensing, ed. C. S. Kochanek & J. N. Hewitt, 75–+

Stoughton, C. et al. 2002, AJ, 123, 485

Strauss, M. A. et al. 2002, AJ, 124, 1810

Suchkov, A. 2005, AJ, 130, 2439

Szapudi, I. et al. 2005, ApJ, 631, L1

Taffoni, G., Belikov, A. & Schaaff, A. 2009, Mem. S. A. It., 80, 493

Telea, A. 2007, Data Visualization, A K Peters

Tinker, J. et al. 2008, ApJ, 688, 709

Tucker, D. L. et al. 2006, Astronomische Nachrichten, 327, 821

Turner, M. S. & White, M. 1997, Phys. Rev. D, 56, 4439

Tyson, J. A., Wenk, R. A. & Valdes, F. 1990, ApJ, 349, L1

Valentijn, E. et al. 2007, Astronomical Data Analysis Software and Systems XVI ASP Conference Series. R. A. Shaw, F. Hill & D. J. Bell, eds, 376

van Breukelen, C. & Clewley, L. 2009, MNRAS, 395, 1845

van Breukelen, C. et al. 2006, MNRAS, 373, L26

van de Weygaert, R. 1994, A&A, 283, 361

van de Weygaert, R. & Icke, V. 1989, A&A, 213, 1

van Dokkum, P. G. & Franx, M. 1996, MNRAS, 281, 985

Viana, P. T. P. & Liddle, A. R. 1996, MNRAS, 281, 323

Viel, M., Haehnelt, M. G. & Springel, V. 2004, MNRAS, 354, 684

Vikhlinin, A. et al. 2009, ApJ, 692, 1060

Wang, L. & Steinhardt, P. J. 1998, ApJ, 508, 483

Weinberg, S. 1989, Reviews of Modern Physics, 61, 1

Weiner, B. J. et al. 2005, ApJ, 620, 595

White, M. & Kochanek, C. S. 2002, ApJ, 574, 24

White, S. D. M., Efstathiou, G. & Frenk, C. S. 1993, MNRAS, 262, 1023

Wojtak, R. et al. 2007, å, 466

Wood-Vasey, W. M.—. 2007, ApJ, 666, 694

Wright, C. O. & Brainerd, T. G. 2000, ApJ, 534, 34

Wright, H. 2006, Introduction to Scientific Visualization, Springer

Yee, H. K. C. et al. 2000, ApJS, 129, 475

York, D. G. et al. 2000, AJ, 120, 1579

Zaninetti, L. 1995, A&AS, 109, 71

Zaritsky, D. et al. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 257, AMiBA 2001: High-Z Clusters, Missing Baryons, and CMB Polarization, ed. L.-W. Chen, C.-P. Ma, K.-W. Ng, & U.-L. Pen, 133–+

Zaritsky, D. et al. 1997, ApJ, 480, L91+

Appendix

Appendix A

Input parameters and output quantities of 2DPHOT

A.1 Input parameters of 2DPHOT

Table A.1 summarizes the main input parameters of 2DPHOT. These parameters can be either set as input options for the 2DPHOT main script or passed to the package by 2DGUI (see Secs. 2.2, 2.14). In the table, we also include a short description of each parameter, as well as a reference to the sections in this paper where the 2DPHOT task influenced by that parameter is described.

A.2 Output quantities measured by 2DPHOT

Table A.2 summarizes the output quantities measured by 2DPHOT. A short description of all quantities is provided, together with a reference to sections where the corresponding 2DPHOT tasks are described. The quantities measured by running S-Extractor are not included for brevity.

Table A.1 - Summary of 2DPHOT input parameters. Column 1: Options for the 2DPHOT main script. Column 2: Description of the parameter. Column 3: The name used to denote the parameter in the paper text. Column 4: Paper sections related to the parameter.

-l	Minimum S/N ratio required to perform 2D final fitting and surface photometry.		2.8, 2.9
-x	Stamp sizes are proportional to the S-Extractor ISOAREA parameter. This parameter provides the proportionality factor.	$EXPND$	2.4
-z	Maximum size of the stamp images. This parameter can be used to prevent overly large stamp frames.		2.4
-i	Minimum S/N ratio required to define sure stars.		2.3
-f	Maximum S/N ratio required to define sure stars.		2.3
-j	Number of Moffat/Gaussian functions for 1D PSF fitting.		2.9.2
-m	Number of Moffat/Gaussian functions for 2D PSF fitting.	$NSMAX$	2.5
-g	Functions used in the 2D PSF fitting (0=Moffat, 1=Gaussian).		2.5
-t	Number of cos/sin terms used for the expansion of star isophotes in the 2D PSF fitting.		2.5
-d	Minimum S/N ratio to perform 2D fitting with expansion of the galaxy model into a cos/sin series.		2.8
-b	Label providing the cos terms used for the expansion of the galaxy model in 2D final fitting (e.g. -b 34 makes 2DPHOT calculate the a_3 and a_4 coefficients)		2.8, 2.9.1
-c	Label providing the sin terms used for the expansion of the galaxy model in 2D final fitting.		2.8, 2.9.1
-e	Minimum distance of an object to the image edges, in units of its FWHM. Objects that are closer to the edge this distance are not analyzed.	$REDGE$	2.3
-s	Maximum number of sure stars used in a cell to perform PSF modeling.	$NSIZE$	2.5, 2.9.2
-a	Flag that determines the 2D PSF fitting method. When equal to zero, this option forces all sure stars in a given cell to be fitted simultaneously. When equal to one, a single fit to each sure star is performed.		2.5
-n	Size (in pixels) of the grid cells where PSF modeling is performed.		2.5
-o	The user can choose to process only some objects in the image by providing a list of x and y coordinates on the image. This feature is enabled with -o 1.		2.3

Table A.2 - Summary of 2DPHOT output quantities. Column 1: Quantity description. Column 2: Related sections in the text.

S-Extractor quantities.	2.3, 2.4
Stellar locus quantities: stellar index vs. S/N ratio for S/G separation, mean and standard deviation values of the sure star locus.	2.3, 2.4, 2.7
PSF fitting parameters: central intensity, width, axis ratio, position angle, shape parameter (in the case of PSF fitting with Moffat functions), and cos/sin terms of each PSF fitting function; central coordinates and local background value of each sure star stamp; reduced χ^2 of PSF fitting. These quantities are obtained for both the 2D and 1D fitting methods.	2.5, 2.9.2
Coarse Sérsic parameters: center coordinates, central surface brightness, effective radius, axis ratio, position angle of the major axis, Sérsic index, total magnitude.	2.6
Final 2D fitting parameters: center coordinates, central surface brightness, effective radius, axis ratio, position angle of the major axis, Sérsic index, total magnitude magnitude, local stamp background value, reduced χ^2.	2.8
Isophotal parameters. For each isophote, the following quantities are computed: center coordinates, equivalent radius, position angle of the major axis, coefficients of the sin/cos expansion.	2.9
1D Sérsic fitting parameters: central surface brightness, effective radius, Sérsic index, total magnitude, reduced χ^2.	2.9.2
Seeing corrected parameters: aperture magnitudes and surface brightness values corresponding to elliptical and circular contours, half-light radius and the corresponding mean surface brightness is also computed, petrosian function.	2.10
Completeness function: percentage of recovered simulated stars and galaxies as a function of the S/N ratio and magnitude.	2.11
Contamination estimates: percentages of misclassified stars and galaxies as a function of the S/N ratio and magnitude.	2.12

Appendix B

The Voronoi Tessellation cell areas distribution for power-law correlated point processes

Motivated by what is known about the two-point correlation function of galaxies in the Universe, we consider a 2-dimensional point field characterized by a two-point correlation function of the form

$$w(\theta) = A\theta^{1-\gamma} \tag{B.1}$$

where θ is a distance, A is the amplitude of the correlation and γ is the slope of the power-law. $A = 0$ represents the Poisson particular case. We generate simmulated fields spanning a wide range of the parameter space (A, γ) around the measured values reported in the literature. These simulated fields are used to characterize the VT cell areas distribution.

Although aimed at application in our cluster finder algorithm, this study allows to investigate the connection between this VT property and the statistical process of the generator set of points. This topic has been extensively discussed (see Okabe (2000) for a review). For the Poisson case, simulations have been used to support the so-called Kiang's conjecture that the distribution of standardized cell sizes (size/mean size) in n-dimensional space is given by

$$p(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp^{-\beta x} \tag{B.2}$$

with $\alpha = \beta = 2n$. This has been rigorously shown for $n = 1$ and studied in simulations up to $n = 3$. Here we extend this conjecture to the case where the two-point correlation function of the field is given by a power-law. We focus on $n = 2$. Our results indicate that Eq. B.2 still holds but the parameters α and β are modified. The relation $\alpha = 0.26 + \beta$ is found to be valid within the parameters space explored.

In the following sections we describe the simulations and the modeling of the area distribution. We discuss our results in comparison to the well-studied Poisson case and provide the relevant quantities in a table.

B.0.1 Point field simulation

To generate the simulated fields with two-point correlation function given by Eq. B.1, we implement the simulated annealing method as proposed by Rintoul & Torquato (1997). This method is generally used to find the state of minimum "energy" of a given system, by sampling the different states weighted by the probability of occurrence of that state. Here, we take Eq. (B.1) as our "reference" state, and the state of the "system" is denoted as $w_s(\theta)$. We consider logarithmic bins in θ, and define the energy of the system as

$$E = \sum_i (w_s(\theta_i) - w(\theta_i))^2 \tag{B.3}$$

where the sum is over all bins. We use 10 bins in the interval $0.01 < \theta_i < 2$. This definition of energy is convenient because ensures that E decreases when the difference between any two bins decreases.

The initial state is a Poisson state. To evolve the system towards $w(\theta)$, we chose a particle and give it randomly in the field. We compute the energy E' of this new configuration and obtain $\Delta E = E' - E$. The move is accepted with probability

$$p(\Delta E) = \begin{cases} 1 & \Delta E \leq 0 \\ \exp(-\Delta E/kT) & \Delta E > 0 \end{cases} \tag{B.4}$$

where kT is the "temperature" of the system. This is chosen to allow the system to evolve as quickly as possible to the minimum state, without getting trapped in local minima. The initial temperature is set to 1. We attempt to move all the particles sequentially and, after a complete round over all the N particles of the system, its temperature is cooled by a factor of 2. The system converges about 30% faster with this cooling schedule.

In Fig. B.1 we show one example, where $A = 0.005$ and $\gamma = 1.7$. This combination of parameters correspond to typical values measured, for instance, on SDSS data up to magnitude limit $r' = 21.5$ (Connolly et al., 2002). The initial system is on the left, the field in the middle is the final state, after 10 rounds over all particles. The plot on the left shows the evolution of the energy of the system. The difference between the initial

and final states is not noticeable by eye and a statistical method must be used to actually measure the two-point correlation function and compute ΔE at each iteration. We use a fast Fourier transform code (Szapudi et al., 2005) to accomplish this. Using this method we have generated 190 fields of 3×3 sq degrees and 1.6×10^4 particles.

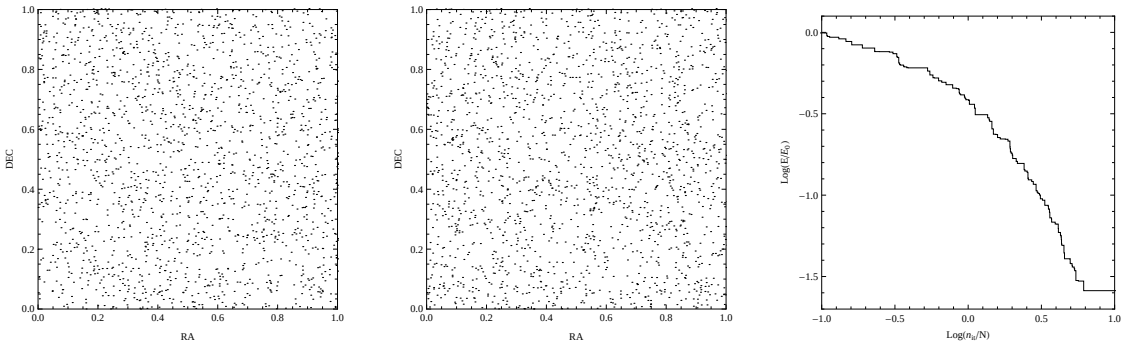

Figure B.1: The left plot shows the initial (Poisson) state of a system meant to evolve towards a configuration with $A = 0.005$ and $\gamma = 1.7$. The final state is the one in the central plot. The left plot is the evolution of the energy of the system (normalized by its initial energy) as a function of the iteration number normalized by the total number of particles in the system. Under this normalization, $n_{it}/N = 1, 2, 3...$ refers to complete rounds over all particles in the field. This simulation was performed in a box of 3×3 sq deg containing 1.6×10^4 particles. Just a 1×1 portion of the field is shown.

B.0.2 Gamma model for the VT cell distribution

We apply the VT code on each of the simulated fields, obtain the distribution of cell normalized cell areas and find the best fit Gamma model (Eq. B.2). Fig. B.2 shows, as an example the VT diagram for the same system featured above. The left and right diagrams correspond to the initial and final state of the system, respectively.

The result of the fit is shown in Fig. B.3, again for the case $A = 0.005$ and $\gamma = 1.7$. For comparison we show as well the traditional Kiang formula (dashed line). The results are $\alpha = 3.89 \pm 0.04$ and $\beta = 3.65 \pm 0.05$. Kiang's formula is more than 5σ away from the best fit.

The results for the ensemble of simulated fields studied are shown in Fig. B.4. The values of α and β fall in the range $3.5 < \alpha < 3.9$ and $3.5 < \beta < 3.8$. The mean error in both is 0.04. There is a noticeable correlation between these two parameters. The difference $\alpha - \beta$ is shown to be 0.26 ± 0.02 all over the parameter space explored.

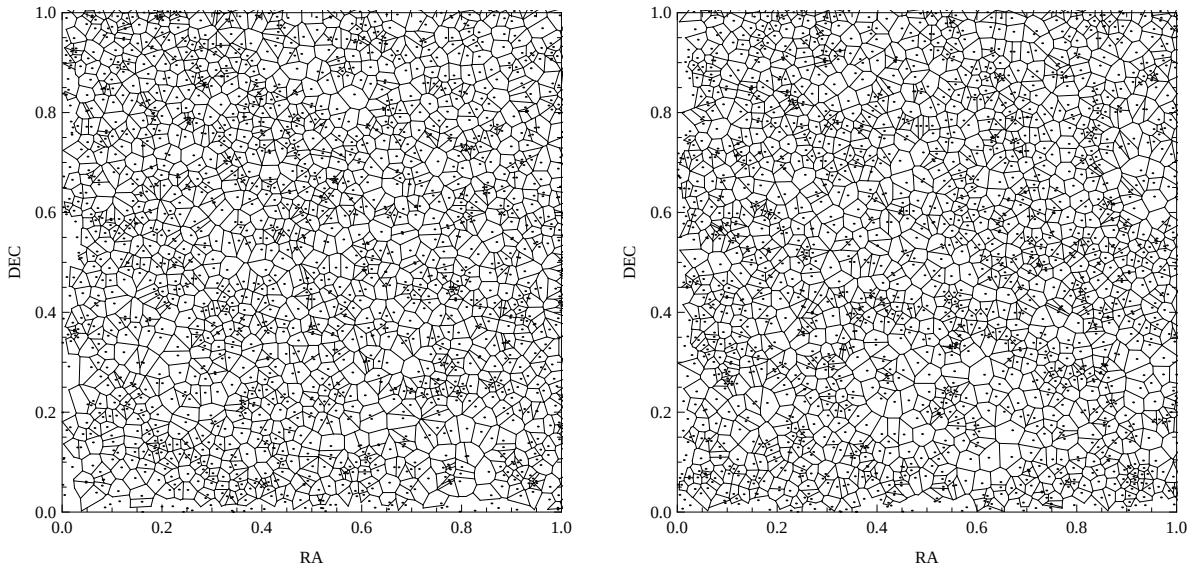

Figure B.2: The Voronoi diagram corresponding to the two fields shown in Fig. B.1. The initial and final states are on the left and right panels, respectively.

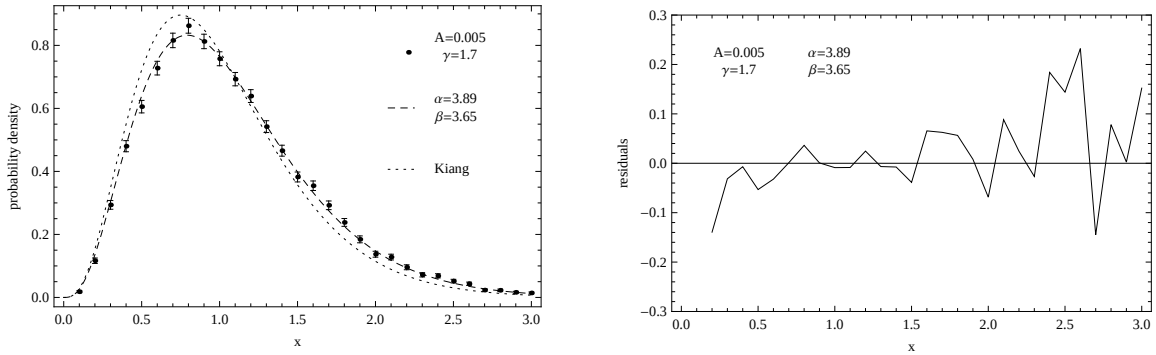

Figure B.3: Right: Best fit model for the distribution of normalized VT cell areas featured in Fig. **??**. The curve for the Poisson case is also shown for comparison (dashed line). Left: Residuals of the fit.

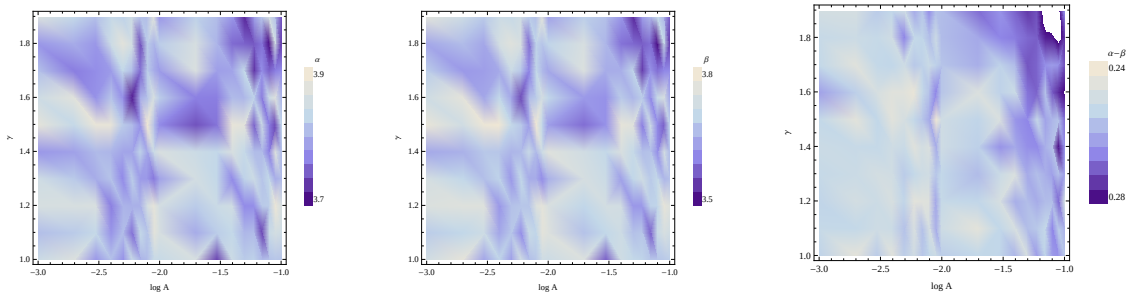

Figure B.4: Density maps showing the results of the fit in the parameter space investigated. There is a noticeable correlation between the two left most maps. The difference of between these two maps is shown in the right.

Table B.1 - VT cell area distribution model parameters

A	γ		α	β	χ^2/ν
0.001	1.0	···	3.91 ± 0.05	3.66 ± 0.05	1.24
0.001	1.1	···	3.86 ± 0.05	3.61 ± 0.05	1.49
0.001	1.2	···	3.93 ± 0.04	3.68 ± 0.03	1.12
0.001	1.3	···	3.92 ± 0.04	3.67 ± 0.03	1.74
0.001	1.4	···	3.81 ± 0.04	3.57 ± 0.04	1.7
0.001	1.5	···	3.96 ± 0.04	3.71 ± 0.03	1.33
0.001	1.6	···	3.91 ± 0.04	3.65 ± 0.03	1.25
0.001	1.7	···	3.86 ± 0.05	3.62 ± 0.05	2.07
0.001	1.8	···	3.81 ± 0.04	3.57 ± 0.03	1.49
0.001	1.9	···	3.94 ± 0.02	3.69 ± 0.01	1.19
0.002	1.0	···	3.94 ± 0.04	3.71 ± 0.03	1.78
0.002	1.1	···	3.87 ± 0.05	3.63 ± 0.04	1.18
0.002	1.2	···	3.93 ± 0.04	3.69 ± 0.05	1.61
0.002	1.3	···	3.87 ± 0.02	3.63 ± 0.03	2.17
0.002	1.4	···	3.83 ± 0.04	3.58 ± 0.03	1.43
0.002	1.5	···	3.9 ± 0.04	3.66 ± 0.03	1.4
0.002	1.6	···	3.95 ± 0.04	3.7 ± 0.03	1.36
0.002	1.7	···	3.79 ± 0.04	3.55 ± 0.04	1.41
0.002	1.8	···	3.84 ± 0.04	3.59 ± 0.03	1.57
0.002	1.9	···	3.89 ± 0.04	3.65 ± 0.05	1.57
0.003	1.0	···	3.81 ± 0.04	3.56 ± 0.04	1.49
0.003	1.1	···	3.9 ± 0.04	3.65 ± 0.05	1.22
0.003	1.2	···	3.94 ± 0.04	3.69 ± 0.03	1.51
0.003	1.3	···	3.84 ± 0.04	3.6 ± 0.03	1.62
0.003	1.4	···	3.86 ± 0.01	3.61 ± 0.01	1.53
0.003	1.5	···	3.97 ± 0.04	3.72 ± 0.03	1.29
0.003	1.6	···	3.81 ± 0.04	3.57 ± 0.04	1.89
0.003	1.7	···	3.8 ± 0.04	3.55 ± 0.05	2.03
0.003	1.8	···	3.81 ± 0.05	3.57 ± 0.04	1.49
0.003	1.9	···	3.86 ± 0.04	3.61 ± 0.05	1.56
0.004	1.0	···	3.86 ± 0.04	3.62 ± 0.05	1.59
0.004	1.1	···	3.81 ± 0.04	3.56 ± 0.05	1.47
0.004	1.2	···	3.79 ± 0.04	3.55 ± 0.04	1.35
0.004	1.3	···	3.87 ± 0.04	3.62 ± 0.03	1.65
0.004	1.4	···	3.85 ± 0.04	3.6 ± 0.05	1.42
0.004	1.5	···	3.97 ± 0.04	3.73 ± 0.03	1.24
0.004	1.6	···	3.87 ± 0.05	3.63 ± 0.05	1.35
0.004	1.7	···	3.82 ± 0.04	3.57 ± 0.04	1.38
0.004	1.8	···	3.91 ± 0.04	3.66 ± 0.03	1.04
0.004	1.9	···	3.9 ± 0.02	3.65 ± 0.01	1.33
0.005	1.0	···	3.81 ± 0.01	3.56 ± 0.03	1.51
0.005	1.1	···	3.86 ± 0.04	3.61 ± 0.03	1.53
0.005	1.2	···	3.85 ± 0.04	3.6 ± 0.05	1.61

Table B.1 - Continued

A	γ		α	β	χ^2/ν
0.005	1.3	···	3.8 ± 0.04	3.55 ± 0.04	1.41
0.005	1.4	···	3.83 ± 0.04	3.58 ± 0.03	1.71
0.005	1.5	···	3.87 ± 0.04	3.63 ± 0.05	1.25
0.005	1.6	···	3.81 ± 0.04	3.57 ± 0.03	1.14
0.005	1.7	···	3.89 ± 0.04	3.65 ± 0.05	1.16
0.005	1.8	···	3.96 ± 0.04	3.69 ± 0.05	1.84
0.005	1.9	···	3.88 ± 0.04	3.64 ± 0.05	1.56
0.006	1.0	···	3.9 ± 0.05	3.66 ± 0.04	1.56
0.006	1.1	···	3.78 ± 0.01	3.54 ± 0.03	1.47
0.006	1.2	···	3.84 ± 0.04	3.61 ± 0.03	1.13
0.006	1.3	···	3.88 ± 0.05	3.63 ± 0.05	1.48
0.006	1.4	···	3.83 ± 0.01	3.59 ± 0.01	2.
0.006	1.5	···	3.86 ± 0.04	3.61 ± 0.05	1.62
0.006	1.6	···	3.71 ± 0.04	3.47 ± 0.04	2.34
0.006	1.7	···	3.86 ± 0.02	3.61 ± 0.03	1.62
0.006	1.8	···	3.92 ± 0.05	3.67 ± 0.05	1.34
0.006	1.9	···	3.91 ± 0.02	3.66 ± 0.01	1.25
0.007	1.0	···	3.85 ± 0.05	3.6 ± 0.04	1.53
0.007	1.1	···	3.9 ± 0.04	3.64 ± 0.05	2.08
0.007	1.2	···	3.84 ± 0.01	3.6 ± 0.03	1.13
0.007	1.3	···	3.82 ± 0.04	3.57 ± 0.03	1.51
0.007	1.4	···	3.89 ± 0.02	3.64 ± 0.001	1.43
0.007	1.5	···	3.81 ± 0.01	3.56 ± 0.01	2.11
0.007	1.6	···	3.84 ± 0.05	3.59 ± 0.05	1.53
0.007	1.7	···	3.77 ± 0.01	3.52 ± 0.001	1.29
0.007	1.8	···	3.75 ± 0.04	3.5 ± 0.04	1.91
0.007	1.9	···	3.86 ± 0.05	3.61 ± 0.04	1.52
0.008	1.0	···	3.86 ± 0.04	3.61 ± 0.05	1.89
0.008	1.1	···	3.87 ± 0.04	3.62 ± 0.05	1.53
0.008	1.2	···	3.92 ± 0.05	3.68 ± 0.05	1.67
0.008	1.3	···	3.82 ± 0.04	3.57 ± 0.03	1.32
0.008	1.4	···	3.83 ± 0.04	3.59 ± 0.03	1.66
0.008	1.5	···	3.99 ± 0.04	3.73 ± 0.03	1.67
0.008	1.6	···	3.86 ± 0.02	3.61 ± 0.03	1.46
0.008	1.7	···	3.88 ± 0.05	3.63 ± 0.05	1.52
0.008	1.8	···	3.88 ± 0.05	3.62 ± 0.05	1.39
0.008	1.9	···	3.86 ± 0.04	3.6 ± 0.03	1.34
0.009	1.0	···	3.9 ± 0.02	3.66 ± 0.03	1.25
0.009	1.1	···	3.96 ± 0.05	3.7 ± 0.05	1.48
0.009	1.2	···	3.96 ± 0.04	3.71 ± 0.05	1.51
0.009	1.3	···	3.78 ± 0.05	3.53 ± 0.04	2.21
0.009	1.4	···	3.91 ± 0.04	3.65 ± 0.05	1.47
0.009	1.5	···	3.86 ± 0.04	3.63 ± 0.05	1.32
0.009	1.6	···	3.93 ± 0.02	3.67 ± 0.01	1.21

Table B.1 - Continued

A	γ		α	β	χ^2/ν
0.009	1.7	···	3.84 ± 0.04	3.59 ± 0.04	1.3
0.009	1.8	···	3.85 ± 0.04	3.6 ± 0.03	1.65
0.009	1.9	···	3.95 ± 0.05	3.69 ± 0.04	1.44
0.01	1.0	···	3.9 ± 0.02	3.65 ± 0.03	1.71
0.01	1.1	···	3.93 ± 0.04	3.69 ± 0.03	1.34
0.01	1.2	···	3.93 ± 0.04	3.68 ± 0.05	1.5
0.01	1.3	···	3.83 ± 0.05	3.58 ± 0.04	1.74
0.01	1.4	···	3.94 ± 0.05	3.69 ± 0.05	1.37
0.01	1.5	···	3.8 ± 0.04	3.56 ± 0.04	1.3
0.01	1.6	···	3.88 ± 0.02	3.63 ± 0.03	1.26
0.01	1.7	···	3.82 ± 0.05	3.57 ± 0.04	1.82
0.01	1.8	···	3.88 ± 0.02	3.62 ± 0.03	1.52
0.01	1.9	···	3.82 ± 0.01	3.56 ± 0.03	1.49
0.02	1.0	···	3.93 ± 0.04	3.68 ± 0.05	1.97
0.02	1.1	···	3.84 ± 0.01	3.6 ± 0.01	1.96
0.02	1.2	···	3.88 ± 0.04	3.63 ± 0.03	1.2
0.02	1.3	···	3.92 ± 0.04	3.67 ± 0.03	1.4
0.02	1.4	···	3.91 ± 0.05	3.66 ± 0.05	1.51
0.02	1.5	···	3.75 ± 0.01	3.5 ± 0.01	1.72
0.02	1.6	···	3.79 ± 0.04	3.55 ± 0.03	1.2
0.02	1.7	···	3.94 ± 0.04	3.68 ± 0.05	1.35
0.02	1.8	···	3.93 ± 0.04	3.67 ± 0.03	1.31
0.02	1.9	···	3.88 ± 0.04	3.62 ± 0.03	1.09
0.03	1.0	···	3.74 ± 0.01	3.49 ± 0.03	1.85
0.03	1.1	···	3.88 ± 0.05	3.63 ± 0.05	1.61
0.03	1.2	···	3.91 ± 0.02	3.66 ± 0.03	1.66
0.03	1.3	···	3.89 ± 0.05	3.65 ± 0.04	1.91
0.03	1.4	···	3.89 ± 0.02	3.63 ± 0.03	1.55
0.03	1.5	···	3.78 ± 0.01	3.54 ± 0.03	1.22
0.03	1.6	···	3.79 ± 0.04	3.54 ± 0.04	1.48
0.03	1.7	···	3.82 ± 0.04	3.57 ± 0.03	1.52
0.03	1.8	···	3.85 ± 0.02	3.59 ± 0.01	0.998
0.03	1.9	···	3.86 ± 0.04	3.6 ± 0.05	1.08
0.04	1.0	···	3.91 ± 0.05	3.66 ± 0.05	1.53
0.04	1.1	···	3.89 ± 0.04	3.65 ± 0.05	1.6
0.04	1.2	···	3.92 ± 0.04	3.66 ± 0.05	1.73
0.04	1.3	···	3.8 ± 0.01	3.56 ± 0.03	1.68
0.04	1.4	···	3.93 ± 0.04	3.68 ± 0.03	1.03
0.04	1.5	···	3.97 ± 0.04	3.71 ± 0.03	1.22
0.04	1.6	···	3.86 ± 0.02	3.61 ± 0.01	1.39
0.04	1.7	···	3.83 ± 0.01	3.57 ± 0.03	1.23
0.04	1.8	···	3.77 ± 0.04	3.51 ± 0.03	1.18
0.04	1.9	···	3.81 ± 0.04	3.54 ± 0.03	1.32
0.05	1.0	···	3.87 ± 0.04	3.63 ± 0.05	0.971

Table B.1 - Continued

A	γ		α	β	χ^2/ν
0.05	1.1	···	3.85 ± 0.01	3.6 ± 0.03	1.33
0.05	1.2	···	3.8 ± 0.04	3.55 ± 0.03	1.18
0.05	1.3	···	3.88 ± 0.04	3.63 ± 0.03	1.39
0.05	1.4	···	3.9 ± 0.04	3.64 ± 0.03	1.29
0.05	1.5	···	3.96 ± 0.04	3.69 ± 0.05	1.27
0.05	1.6	···	3.85 ± 0.04	3.59 ± 0.05	1.42
0.05	1.7	···	3.89 ± 0.05	3.62 ± 0.05	1.19
0.05	1.8	···	3.77 ± 0.04	3.5 ± 0.04	1.27
0.05	1.9	···	3.72 ± 0.01	3.45 ± 0.01	1.45
0.06	1.0	···	3.88 ± 0.04	3.62 ± 0.05	1.36
0.06	1.1	···	3.86 ± 0.04	3.61 ± 0.03	1.62
0.06	1.2	···	3.8 ± 0.04	3.55 ± 0.03	1.37
0.06	1.3	···	3.91 ± 0.02	3.66 ± 0.03	1.72
0.06	1.4	···	3.86 ± 0.04	3.6 ± 0.03	1.19
0.06	1.5	···	3.77 ± 0.04	3.52 ± 0.04	1.6
0.06	1.6	···	3.88 ± 0.04	3.61 ± 0.03	1.31
0.06	1.7	···	3.73 ± 0.04	3.46 ± 0.03	1.52
0.06	1.8	···	3.76 ± 0.01	3.49 ± 0.03	1.14
0.06	1.9	···	3.82 ± 0.04	3.55 ± 0.05	1.29
0.07	1.0	···	3.85 ± 0.04	3.61 ± 0.03	1.58
0.07	1.1	···	3.74 ± 0.01	3.5 ± 0.01	1.28
0.07	1.2	···	3.87 ± 0.02	3.62 ± 0.03	1.67
0.07	1.3	···	3.87 ± 0.04	3.61 ± 0.03	1.47
0.07	1.4	···	3.86 ± 0.05	3.6 ± 0.05	1.8
0.07	1.5	···	3.91 ± 0.05	3.65 ± 0.05	1.34
0.07	1.6	···	3.78 ± 0.04	3.53 ± 0.03	1.36
0.07	1.7	···	3.8 ± 0.04	3.54 ± 0.04	1.49
0.07	1.8	···	3.78 ± 0.01	3.5 ± 0.03	1.72
0.07	1.9	···	3.87 ± 0.02	3.59 ± 0.01	1.01
0.08	1.0	···	3.86 ± 0.04	3.6 ± 0.03	1.42
0.08	1.1	···	3.9 ± 0.05	3.64 ± 0.04	1.28
0.08	1.2	···	3.88 ± 0.04	3.62 ± 0.03	1.03
0.08	1.3	···	3.89 ± 0.05	3.63 ± 0.05	1.47
0.08	1.4	···	3.79 ± 0.04	3.54 ± 0.04	1.53
0.08	1.5	···	3.79 ± 0.05	3.53 ± 0.04	1.94
0.08	1.6	···	3.9 ± 0.04	3.64 ± 0.03	1.29
0.08	1.7	···	3.76 ± 0.04	3.5 ± 0.04	1.54
0.08	1.8	···	3.7 ± 0.04	3.42 ± 0.04	1.98
0.08	1.9	···	3.92 ± 0.05	3.63 ± 0.04	1.15
0.09	1.0	···	3.87 ± 0.02	3.61 ± 0.03	1.34
0.09	1.1	···	3.92 ± 0.05	3.67 ± 0.05	1.44
0.09	1.2	···	3.97 ± 0.04	3.71 ± 0.05	0.92
0.09	1.3	···	3.84 ± 0.05	3.59 ± 0.05	1.63
0.09	1.4	···	3.94 ± 0.04	3.67 ± 0.05	1.87

Table B.1 - Continued

A	γ		α	β	χ^2/ν
0.09	1.5	···	3.75 ± 0.04	3.49 ± 0.04	1.57
0.09	1.6	···	3.95 ± 0.04	3.67 ± 0.03	1.44
0.09	1.7	···	3.79 ± 0.04	3.53 ± 0.05	1.38
0.09	1.8	···	3.92 ± 0.02	3.63 ± 0.03	1.41
0.09	1.9	···	3.86 ± 0.02	3.58 ± 0.03	0.993
0.1	1.0	···	3.87 ± 0.04	3.62 ± 0.05	1.76
0.1	1.1	···	3.91 ± 0.02	3.66 ± 0.03	1.75
0.1	1.2	···	3.93 ± 0.04	3.69 ± 0.05	1.42
0.1	1.3	···	3.85 ± 0.04	3.6 ± 0.05	1.18
0.1	1.4	···	3.86 ± 0.04	3.6 ± 0.03	1.47
0.1	1.5	···	3.88 ± 0.02	3.62 ± 0.03	1.25
0.1	1.6	···	3.85 ± 0.04	3.57 ± 0.03	1.47
0.1	1.7	···	3.8 ± 0.04	3.53 ± 0.04	1.68
0.1	1.8	···	3.74 ± 0.04	3.46 ± 0.03	1.38
0.1	1.9	···	3.76 ± 0.05	3.49 ± 0.04	1.55

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo