# Group-finding with photometric redshifts: The Photo-z Probability Peaks algorithm

by

Bryan Gillis

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Science
in
Physics

Waterloo, Ontario, Canada, 2010

# Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

# *Abstract*

We present a galaxy group-finding algorithm, the Photo-z Probability Peaks (P3) algorithm, optimized for locating small galaxy groups using photometric redshift data by searching for peaks in the signal-to-noise of the local overdensity of galaxies in a 3-dimensional grid. This method is an improvement over similar matched-filter methods in reducing background contamination through the use of redshift information, allowing it to accurately detect groups to a much lower size limit. We present the results of tests of our algorithm on galaxy catalogues from the Millennium Simulation. For typical settings of our algorithm and photometric redshift accuracy of $\sigma_z = 0.05$ it attains a purity of 84% and detects $\sim 83$ groups/deg.$^2$ with an average group size of 5.5 members. With photometric redshift accuracy of $\sigma_z = 0.02$, it attains a purity of 94% and detects $\sim 80$ groups/deg.$^2$ with an average group size of 6.3 members. We also test our algorithm on data available for the COSMOS field and the presently-available fields from the CFHTLS-Wide survey, presenting preliminary results of this analysis.

# *Acknowledgements*

I acknowledge useful discussions with Martha Milkeraitis, thank Simon Lilly and C. Knobel for making the zCOSMOS group catalogue available to us, and thank Hendrik Hildebrandt for making the CFHTLS-Wide photometry available to us. I would also like to thank my adviser, Mike Hudson, for his guidance and contributions.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Most galaxies in the universe are gravitationally bound to one or more other galaxies, forming structures called galaxy groups, or galaxy clusters at the high-mass end. A number of recent studies have indicated that the global mass-to-light (M/L) ratios of groups may be a steep function of the mass of the group (Marinoni and Hudson, 2002, Balogh et al., 2004). This phenomenon may be due to the presence of a critical group mass above which star formation is efficiently quenched. Clearly, it is of interest to improve existing mass-to-light data on group scales, in order to better determine whether such a critical halo mass exists, and to help determine what mechanism may be responsible for the rise in mass-to-light ratios and the quenching of star formation.

In this thesis, we present a new group-finding algorithm, the Photo-z Probability Peaks (P3) algorithm, which detects small galaxy groups with high purity using photometric redshift data. This algorithm was optimized to run on the CFHTLS-Wide (CFHTLS, 2009) catalogue, yielding a group catalogue that will be suitable for mass estimates via gravitational lensing, allowing us to collect further data on the behaviour of the mass-to-light curve as a function of mass.

In the introduction to this thesis, we will discuss the origins of structure in the universe and the typical properties of large-scale structures and then go on to the formation of galaxies and some possible mechanisms through which star formation can be quenched. We will then discuss the techniques relevant to our study: gravitational lensing to estimate group masses, photometric redshifts to estimate the redshifts of galaxies, and various methods that have been used in the past for identifying groups of galaxies.

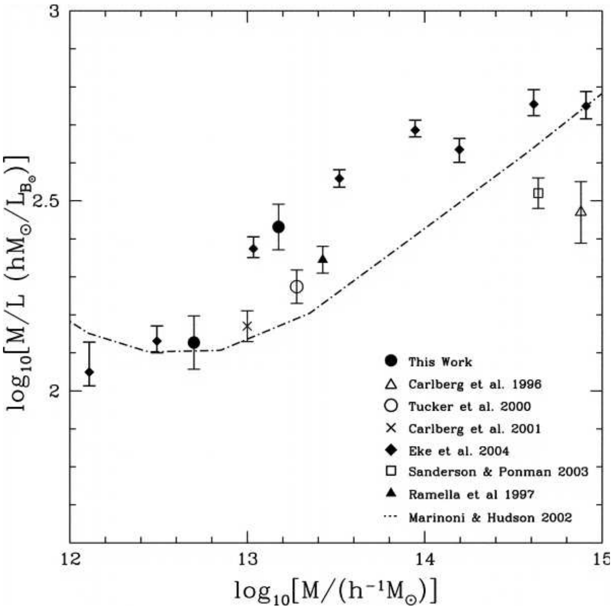


FIGURE 1.1: Figure 6 from Parker et al. (2005), illustrating the sharp rise in mass-to-light ratios at a group mass of around $10^{13}\ M_\odot$.

## 1.1 Structure Formation

Initial density fluctuations in the universe were caused by quantum mechanical uncertainties, which are nearly scale-invariant. At $10^{-36}$ s, inflation stretched these density fluctuations out to cosmological scales, providing the basis for the density distribution observed in the present day. In the presently-accepted Lambda Cold Dark Matter (LCDM or Concordance) paradigm for describing structure formation, density perturbations in the post-inflationary universe grew in magnitude from the time of matter-radiation equality. The smallest-scale perturbations collapsed earliest, and these collapsed structures merged into larger structures as perturbations on these larger scales collapsed. Galaxies formed in the potential wells of collapsed dark matter halos, and as these halos merged together, the galaxies they contained merged or became gravitationally bound to each other in structures known as groups.

To a first-order approximation for low overdensity $\delta$, during the radiation-dominated phase of the universe, density fluctuations evolved as $\delta \propto t$ outside the particle horizon, stagnating inside, and as $\delta \propto t^{2/3}$ in the matter-dominated era of the universe (Carroll and Ostlie, 2007). Structure formation is primarily driven by dark matter, which comprises $\sim 85\%$ of all gravitating matter in the universe. Cold Dark Matter (CDM) models postulate dark matter composed of particles with a non-relativistic velocity and thus a negligible free-streaming scale at the time of matter-radiation equality. The smallest-scale structures have the shortest collapse time. As the epoch for the collapse of larger scales was reached, these smaller structures merged with each other.

Once $\delta$ reaches $\sim 1$, the linear approximation for collapse is insufficiently accurate. In this regime, it is necessary to make some assumptions about the shape of the density distribution in order to make any calculations regarding it. Typically, it is assumed that the distribution is a "spherical top-hat" of overdensity. That is to say, within a distance $R$ of the center, the overdensity is a constant $\delta > 0$, and outside this distance, $\delta = 0$. Since Birchoff's Theorem states that the matter outside a spherical region such as this will exert no gravitational force on it, and CDM models dark matter as being non-interactive and hence pressureless, we can model the change in the size of this region as the being the same as the acceleration of a particle at the edge of the region:

$$\frac{d^2R}{dt^2} = -\frac{GM}{R^2} = -\frac{4\pi G}{3}\rho R \tag{1.1}$$

(White, 2009)

If we compare this to the evolution equation for the cosmological scale factor:

$$\frac{d^2a}{dt^2} = -\frac{GM_0}{a^2} = -\frac{4\pi G}{3}\bar{\rho} a \tag{1.2}$$

(White, 2009)

We can see that this spherical region evolves simply as a positive-curvature universe with a different mean density from the outside universe. Therefore, the same parametrization is valid here as is for the scale factor of a positive-curvature universe:

$$R/R_m = \frac{1}{2}(1 - cos\,\eta);\ t/t_{ta} = (\eta - sin\,\eta)/\pi \tag{1.3}$$

(White, 2009)

where $R_m$ is the maximum scale factor the region reaches, $t_{ta}$ is the time at which it reaches this scale factor, and $\eta$ is a parameter; the region reaches its maximum size when $\eta = \pi$ and collapses to $R = 0$ when $\eta = 2\pi$. The time of collapse is then $t = 2t_{ta}$. If we relate the linearization of this parametrization near $t = 0$ to the linear approximation for the evolution of $\delta$ attained previously, it turns out that the region will collapse when the linear approximation gives $\delta \approx 1.686$. This result is very useful, as it allows the use of the much simpler linear approximation to estimate when structure has collapsed.

It is also notable that the time of collapse is determined only by the initial $\delta$, and not the size of the region. The size of the region instead determines the total mass of the collapsed structure.

Although the spherical top-hat model is a gross oversimplification of the shape and distribution of actual structures, it remains very useful in characterizing their behaviour. Even if the constant-density assumption is dropped, as long as $\delta$ decreases with distance from the center of the structure, the above equations can be used with only slight modifications to the enclosed mass to estimate when a given spherical shell will collapse.

Using this theory for collapse, we can make predictions about the present-day mass function for structures, based on the initial power spectrum of mass fluctuations. Press and Schechter (1974) suggested that the number of collapsed structures could be determined by taking the fraction of points which are surrounded by a sphere of radius R in which the mean density exceeds the threshold value for collapse, $\delta_c = 1.686$. As the initial mass distribution is a Gaussian random field and it evolves self-similarly, this fraction can be calculated as:

$$F(R,\tau) = \int_{\delta_c}^{\infty} d\delta \frac{1}{\sqrt{2\pi}\bar{\rho}(\tau)\Delta_0(R)} exp\left(-\frac{\delta^2}{2\bar{\rho}^2(\tau)\Delta_0^2(R)}\right) \tag{1.4}$$

(White, 2009)

where $\bar{\rho}^2(\tau)\Delta_0^2(R)$ is the mean square density fluctuation in spheres of comoving radius R. We can calculate this quantity from the power spectrum $\delta_k$ by:

$$\bar{\rho}^2(\tau)\Delta_0^2(R) = \int d^3\mathbf{k}\, |\delta_k|^2\, |W(\mathbf{k}R)|^2 \tag{1.5}$$

(White, 2009)

where $W(\mathbf{k}R)$ is the Fourier transform of the normalized top-hat window function for a sphere of radius 1. From this, we can calculate the mass function of collapsed structures:

$$n(M,\tau)dM = -\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}\delta_c}{M\rho\Delta_0^2}\frac{d\Delta_o}{dM} exp\left(\frac{\delta_c^2}{\rho^2\Delta_0^2}\right) dM \tag{1.6}$$

(White, 2009)

If we assume that $|\delta_k|^2 \propto k^n$, then $\Delta_0 \propto M^{-\frac{n+3}{6}}$, giving us:

$$n(M,\tau)dM = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M}\left(1+\frac{n}{3}\right)(M/M_*(\tau))^{\frac{3+n}{6}} exp\left(-(M/M_*(\tau))^{\frac{3+n}{3}}/2\right)\frac{dM}{M} \tag{1.7}$$

(White, 2009)

where $M_*(\tau)$ is defined by $\Delta_0(M_*(\tau)) = \delta_c/\rho(\tau)$.

The simplest model that adequately describes the shape of these collapsed structures is the isothermal sphere, truncated at a maximum radius $r_{max}$ (known as the "virial radius") which is defined as the radius at which the enclosed region has a density of $\rho_{vir} = 178\rho(\tau)$:

$$\rho \propto r^{-2};\ M \propto r;\ V_c^2 = GM/r\,(\text{constant}) \tag{1.8}$$

where $V_c^2$ is the circular velocity of objects in the halo. The factor of 178 comes from conservation of energy and the virial theorem, which determines that the density within this region will be a factor of 8 greater when it's virialized than it was at $t = t_m$, while the background density has decreased by a factor of 4. This gives us the virialized density as being $32\delta_{ta}$, where $\delta_{ta}$ can be calculated to be $\sim 5.55$, resulting in our factor of 178.

A more accurate model for dark matter halos, while still relatively simple analytically, is the NFW (Navarro-Frenk-White) profile: (Navarro et al., 1997)

$$\rho \propto \frac{1}{(r/r_s)(1+r/r_s)^2} \tag{1.9}$$

where the constant of proportionality and $r_s$ are fit to the specific halo. At small $r$, this profile behaves as $r^{-1}$; at large $r$, it behaves as $r^{-3}$; and for $r \sim r_s$, it behaves as $r^{-2}$, similar to the isothermal sphere, which allows for observed flat velocity dispersion curves.

After the time of collapse, virialization keeps the density of the structure constant, while the background density dilutes. The density at a future time will then relate to the background density of the universe by:

$$\rho_{vir} = 178\rho(t_c) = 178\rho_0(1+z_c) \tag{1.10}$$

(Padmanabhan, 1993)

Using this result and the structure of the isothermal sphere, we can derive how its size and velocity dispersion relate to its mass:

$$r_{vir} = 162(1+z_c)^{-1}\left(\frac{M}{10^{12}\,\mathrm{M}_\odot}\right)^{1/3}\ \mathrm{h}^{-2/3}\,\mathrm{kpc} \tag{1.11}$$

$$\left\langle v^2\right\rangle^{1/2} = 126(1+z_c)^{1/2}\left(\frac{M}{10^{12}\,\mathrm{M}_\odot}\right)^{1/3}\ \mathrm{h}^{1/3}\,\mathrm{km/s} \tag{1.12}$$

(Padmanabhan, 1993)

The smallest visible collapsed structures will be the dark matter halos surrounding galaxies. As these halos merge with each other, the galaxies they contain will become bound in galaxy groups. As enough halos merge together, typically including galaxy mergers to form a bright central galaxy (BCG) lying at the bottom of the potential well, the structure is what is more commonly known as a galaxy cluster. The boundary between a group and a cluster is arbitrary, but typically around a mass of $10^{14}M_\odot$.

## 1.2 Galaxy Formation

The transition from dark matter physics to baryonic physics is far from being a solved problem. N-body simulations with dark matter are relatively simple, as gravity is the only force acting on the dark matter particles. The possible interactions of baryons, on the other hand, are nearly limitless and are impossible to model accurately. Additionally, the required length scales for baryonic physics are significantly smaller, as they are not typically scale-invariant as gravity is. Simulations have to make simplifications in order to make any progress in modelling the interactions of baryons. Of particular note, star-formation, supernovae, and black holes are poorly modelled by simulations, but stars are the primary observable we have in the universe. Therefore, theories of stellar and galactic physics must be used to make predictions and then tested against observables in order to build an encompassing model of galaxy formation.

One of the important questions in galaxy physics is why galaxies exhibit the specific mass range they do. While dark matter halos merge easily and wash out most substructure given enough time, galaxies don't seem to merge as readily. Galaxy clusters may have hundreds of galaxies within them, yet typically only a handful of these galaxies will merge, forming a large elliptical galaxy at the center of the cluster. The rest of the galaxies will remain intact. Meanwhile, the dark matter halos of these galaxies will typically all merge into a single halo. Something is happening with baryonic physics which limits the growth of galaxies through mergers.

The primary reason that baryonic matter is clumpier than dark matter is that it is capable of radiative cooling as it contracts, while dark matter must retain its energy. If gas can cool on a shorter timescale than its timescale for collapse, then it will become dense enough for star formation to take place, and the resulting structure will be more tightly gravitationally bound, and thus more resistant to mergers. If the cooling timescale is shorter than the timescale for collapse, then the temperature of the gas will drop to $\sim 10^4\,\mathrm{K}$, and it will contract isothermally until stopped by rotational support. If the timescale for cooling is greater than the timescale for collapse, then the gas will first reach a somewhat steady, pressure-supported state after the timescale of collapse. It will then slowly contract on the timescale for cooling. It is possible that this timescale will be larger than the current age of the galaxy, which will result in us observing it in this slowly-contracting state.

The two primary mechanisms for cooling are Bremsstrahlung and recombination. Bremsstrahlung occurs from the interactions of two charged particles, when the acceleration caused by their interaction results in the release of electromagnetic energy. As electrons are significantly less massive than other charged particles commonly found in collapsing gases, they accelerate much more and thus dominate the radiation. The strength of Bremsstrahlung is then proportional to $n_e n_i$, where $n_e$ is the number density of electrons and $n_i$ is the number density of ions. As the frequency of interactions between particles is proportional to their velocities, which is proportional to $T^{1/2}$, this means that the strength of Bremsstrahlung also varies with $T^{1/2}$. The power per unit density of Bremsstrahlung from an ionized plasma ends up being:

$$\epsilon = 1.4 * 10^{-27}\,\mathrm{erg\,cm^3\,s^{-1}\,K^{-1/2}}\,T^{1/2} n_e n_i Z^2 g_B \tag{1.13}$$

(Rybicki and Lightman, 1986)

where Z is the number of protons in the ion and $g_B$ is the gaunt factor, which is typically of order unity.

Recombination takes place when collisions between atoms in a gas free their electrons, allowing the electrons to later recombine with the atoms and release radiation. The dependence of the power released from recombination on temperature is non-simple, but for our purpose can be approximated well enough by $\epsilon \propto T^{-1/2} n_b^2$. We can then express the total power of radiation by:

$$\Lambda(T) = (A_B T^{1/2} + A_R T^{-1/2})\rho_b^2 \tag{1.14}$$

(Padmanabhan, 1993)

Using $t_{cool} \propto \rho_b T/\Lambda(T)$ and inserting the proper values for the constants, we get:

$$t_{cool} = 8 * 10^6\,\text{year}\left(\frac{n}{\text{cm}^{-3}}\right)^{-1}\left[\left(\frac{T}{10^6\,\text{K}}\right)^{-1/2} + 1.5 f_m \left(\frac{T}{10^6\,\text{K}}\right)^{-3/2}\right]^{-1} \tag{1.15}$$

(Padmanabhan, 1993)

where $f_m$ is a constant that depends on the composition of the gas. $f_m \sim 1$ for a primordial combination of hydrogen and helium, and $f_m \sim 30$ for solar metallicity. From this equation, we can see that recombination will dominate below $T = 10^6\,\text{K}$, while Bremsstrahlung will dominate above this. Galaxies will typically fall within the former regime, though clusters may be in the latter.

The time for collapse can be calculated to be:

$$t_{dyn} = 1.5 * 10^9\,\text{yr}\left(\frac{M}{10^{12}\,\text{M}_\odot}\right)^{-1/2}\left(\frac{R}{200\,\text{kpc}}\right)^{3/2} \tag{1.16}$$

(Padmanabhan, 1993)

Using the relationship between temperature and mass for a virialized gas and an assumption that a fraction $F$ of the mass will be in baryonic matter, we find that the gas will be able to cool efficiently if:

$$M < 6.4 * 10^{11}\,\text{M}_\odot f_m \left(\frac{F}{0.1}\right) \tag{1.17}$$

(Padmanabhan, 1993)

From this equation, we can say that more metallic gases will be able to cool more efficiently, allowing more stars to form in regions of high metallicity. More massive structures will also be able to form in regions with a higher ratio of baryonic matter to dark matter, though this is less likely to occur.

There is no lower limit on the masses of galaxies that can form due to cooling, although the number of low-mass galaxies detected is lower than would be calculated from the Press-Schechter formalism (Press and Schechter, 1974). The firmest lower limit on galaxy mass is the threshold mass necessary for star formation to occur. However, as stars can form in molecular clouds of only $10^4 M_\odot$, this is insufficient to explain the observed lack of lower-mass galaxies, at least on its own. It also doesn't explain why much of the gas in galaxies has not yet collapsed to the point where it can form stars. Rather, what is more likely to be the case is that various "feedback effects" prevent gas within galaxies from accumulating into star-forming clouds.

### 1.2.1 Feedback effects

Feedback effects are not yet fully understood, and they are presently an intense area of study. In all galaxies, supernovae and active galactic nuclei (AGNs) can contribute to feedback which may decrease star formation. For galaxies in groups, it is also possible that interactions between galaxies can result in feedback. Possible processes that may take place in groups include ram pressure stripping, tidal stripping, and mergers. It has been observed that the properties of galaxies change depending on whether or not they're in groups, particularly in an increase of their mass-to-light ratios, implying a decrease in star formation. Investigation of which of these processes may be occurring and to what degree is crucial for our full understanding of galaxy evolution. In particular, our project aims to investigate the possible reasons for this decrease in star formation among galaxies in groups.

Supernovae are the most well-understood form of feedback. When a supernova explodes, it releases an immense amount of energy into the surrounding gas. There are two ways in which this can impede star formation. In smaller galaxies, with shallower potential wells, supernovae could propel gas entirely out of them. This could explain why there is a perceived lower mass limit for galaxies; a handful of supernovae from early star formation could remove all gas from the proto-galaxy and prevent it from forming enough stars to be observable. In addition to this, every supernova explosion injects energy into the surrounding gas. This extra energy will support the gas against collapse, slowing or preventing additional star formation. If the energy added to the gas per unit mass from supernova explosions is $\epsilon_0$, then the decrease in the star formation rate will be:

$$\frac{\Delta \dot{M}}{\dot{M}} = \frac{1}{1 + V_c^2/\epsilon_0} \tag{1.18}$$

(White, 2009)

The maximum possible value for $\epsilon_0$ is $\sim (700\,\mathrm{km/s})^2$, though it can be substantially smaller if the energy from supernovae doesn't efficiently heat the gas. Most star-forming halos have $V_c < 300\,\mathrm{km/s}$, so we can estimate that the effect of supernova feedback will be on the order of a 50% decrease in the star-formation rate. This is very significant, though it depends heavily on how efficiently the energy from supernovae is absorbed by the surrounding gas. If supernova feedback has an effect close to the maximum value of $\epsilon_0$, then stars will typically form in halos with $100\,\mathrm{km/s} < V_c < 300\,\mathrm{km/s}$ (White, 2009).

Other forms of feedback are more difficult to quantify, but they are also important for a full understanding of galactic evolution. AGN feedback, for instance, is important because it can occur without requiring any star formation to take place, allowing it to potentially shut off all star formation. AGNs typically consist of an accretion disk around a black hole. The friction between layers of infalling matter causes a release of energy in the radiation. Energy can also be released in energetic jets of charged particles normal to the plane of the accretion disk. This energy can then reheat gas near the AGN, slowing or stopping star formation. AGNs vary greatly in the wavelength and intensity of the radiation they emit, so developing a coherent model for their effect on star formation in all galaxies is difficult. What models have been made, however, have suggested that AGN feedback may play a key role in affecting the star formation rate in massive galaxies and in suppressing star formation in the cooling flows at the centres of massive clusters (Croton et al., 2006).

### 1.2.2 Environmental effects

Of particular note to our project are effects that take place in a group environment which may quench star formation. Tidal stripping occurs when a smaller group member passes near a larger member or a tighter clustering of members. Gas from the smaller galaxy may be stripped out and fall into the larger galaxy, leaving the smaller galaxy with less gas with which to form stars. In the most dramatic cases, the mass of a galaxy can be reduced by 50% or more through tidal stripping (Kravtsov et al., 2004). Even when gas isn't removed from the smaller galaxy, tidal disruption can inject energy into galaxies, slowing their stellar formation. It is suspected that tidal stripping is a large reason for the perceived decline in star formation among grouped galaxies.

On the scale of clusters, ram pressure stripping can provide a significant amount of quenching. When galaxies move relative to the hot, ionized intracluster medium (ICM), the force of this gas against the gas in the galaxy can strip out the galactic gas. To a first order approximation, the strength of this stripping will be proportional to $\rho_{ICM} v_{gal}^2$. As both the density of the ICM and the velocities of galaxies scale with the size of the cluster, this effect will be far more significant on cluster scales than group scales. As such, it is likely insufficient to explain the initial drop in star-forming rates as galaxies enter groups, though it is necessary to explain the morphology of galaxies within large clusters (Quilis et al., 2000).

## 1.3 Lensing

In order to estimate the masses of groups and the distribution of mass within them, we intend to apply the methodology of weak gravitational lensing. Alternative methods for mass estimates aren't viable in the case of small groups with redshift data determined through photometric redshifts. Kinematics is a useful tool with spectroscopic redshift data, but the errors from photometric redshifts dwarf the peculiar velocities of galaxies, making the method useless. X-ray measurements work much better on cluster scales than on group scales; many smaller groups don't have a significant core which would allow us to estimate the mass through X-rays. Gravitational lensing works much better on these scales, though it does suffer some drawbacks. Notably, it is particularly susceptible to projection effects, so it will be important to ensure that we are actually measuring the mass of the groups we detect rather than the mass of a larger-scale structure along the line of sight.

Gravitational lensing is the principle that, like a moving particle, light will curve in response to a gravitational potential. The angle of deflection is:

$$\theta = \frac{4GM}{rc^2} \tag{1.19}$$

(Schneider et al., 2006)

where $r$ is the distance between the light and the lens in the lensing plane.

In the most extreme cases, known as strong gravitational lensing, light will be able to take multiple paths from the source to the observer, resulting in the appearance of multiple images. In less extreme cases, the image of the source will simply be distorted, resulting in what is known as weak gravitational lensing. In these cases, a circular source image will be distorted into an ellipse. We can define the ellipticity and orientation of the image by $e_1$ and $e_2$ defined as:

$$e_1 = \frac{I_{11} - I_{22}}{I_{11} + I_{22}};\ e_2 = \frac{2I_{12}}{I_{11} + I_{22}} \tag{1.20}$$

(Parker et al., 2005)

where $I_{xy}$ are the components of the weighted quadrupole for the galaxy image. Measuring the galaxy's ellipticity is no simple task, as the point-spread function for the telescope and the atmosphere has to be corrected for, the image is pixelated, and there's noise in each pixel. The accuracy of lensing mass estimates depends heavily on the quality of the data reduction, but it is not an insurmountable task.

In the weak lensing regime, the shear $\gamma \approx e = e_1 + e_2 i$. The shear's relevance is in its orientation relative to the vector from the lens position to the image's position in the sky, so we define the tangential shear $\gamma_t$ as the extent to which the image is sheared along this vector and its perpendicular, and the cross shear $\gamma_x$ as the extent to which the image is sheared along axes tilted $45^o$ relative to the former axes. This can be calculated as:

$$\gamma_t = -\Re\left[\gamma e^{-2i\phi}\right];\ \gamma_x = -\Im\left[\gamma e^{-2i\phi}\right] \tag{1.21}$$

(Schneider et al., 2006)

where $\phi$ is the angle between the vector from the lens center to the image and the shear of the image. In the case of galaxy groups, we expect the dark matter halo to be described approximately by an isothermal sphere, giving the expected form:

$$\gamma_t = \frac{\theta_E}{2\theta} = \frac{4\pi V_c^2}{c^2}\frac{D_{LS}}{D_S} = \frac{4\pi G M(r)}{c^2 r}\frac{D_{LS}}{D_S} \tag{1.22}$$

(Parker et al., 2005)

where $\theta_E$ is the angular radius of an Einstein ring around a point mass, $\theta$ is the angular separation between the center of the lens and the image, $D_{LS}$ is the angular diameter distance between the lens and source, and $D_S$ is the angular diameter distance of the source.

In addition to the noise in estimates of the ellipticity of the image, the unsheared images are not expected to be circular. As observed, galaxies have a range of inherent ellipticities, on the order of $\Delta e \sim 0.2$, with $\langle e \rangle = 0$. For group-galaxy lensing such as this, the typical shear is on the order of 0.01. The result is that it is impossible to accurately estimate the mass of any individual group through weak lensing (though clusters may be massive enough for estimates to be made). Rather, it is necessary to stack the data from multiple groups in order to get an ensemble-averaged mass. The stacked data has a curve for $\gamma_t$ vs. $r$ fit to it, which gives the average mass and velocity dispersion of the dark matter halo, as well as testing how appropriate the assumed model for the halo mass distribution is. A useful test is to also perform this analysis for $\gamma_x$ vs. $r$ to confirm that $\gamma_x \sim 0$ for all $r$, as expected. If this is not the case, then it is likely that the data reduction was of insufficient quality.

## 1.4 Photometric Redshifts

As the CFHTLS-Wide survey is too large to obtain spectra for all galaxies of interest, redshift data must be obtained through the use of photometric redshifts. Photometric redshifts (photo-zs) are a method of estimating the redshifts of galaxies using photometry from a number of different filters. In contrast to spectroscopic redshifts (spectro-zs), which require precise measurements of the spectra of stars or galaxies in order to detect narrow emission lines, photo-zs attempt to fit broader features of galaxy spectra with the photometry.

The accuracy of photo-zs is significantly worse than the accuracy of spectroscopic redshifts (roughly two orders of magnitude), but it has significant practical advantages that make it more useful in many applications. Photo-zs take significantly less telescope time to collect, allowing the decreased quality of redshift data to be compensated by increased quantity. In applications such as lensing where redshift errors aren't the most significant source of error, this is a profitable trade-off.

For the purposes of group-finding, it is more economical in terms of telescope time to collect photometry from a large field than spectroscopy from a small field. The location of a group is much better determined than that of an individual galaxy, resulting in manageable levels of error when using photo-zs. The difficulty comes in identifying groups with photo-zs, but this is not an insurmountable problem, as we hope to demonstrate in this thesis. As long as the purity of the group-finding algorithm isn't unreasonably low, the much greater numbers of groups that can be detected using photo-zs is worth the small decrease in purity and redshift accuracy.

Photo-zs are typically estimated using 4 or 5 standard filters, though there is no upper limit on the number of filters than can be used. Modern photo-zs can achieve redshift errors of $\Delta z \sim 0.05$ for galaxies with $I < 22.5$ when using the standard filter set, and this can be improved to $\Delta z \sim 0.01$ when using photometry of 30 filters as in the COSMOS-30 photo-zs.

There are two main classes of methods for determining photo-zs: neural network programming and template fitting. Neural network methods attempt to use a training set of galaxies with photometry and known redshifts to train their program and create an algorithm that will accurately estimate redshift from photometry. A neural network consists of a number of layers of nodes. The input layer (the photometry) is connected to the output layer (the redshift) through connections between the nodes in each layer. The connections are each assigned a weight, and each of the nodes is assigned an activation function. The result is an algorithm for redshift, for which the connection weights must be optimized to provide the best results.

Neural network methods for obtaining photo-zs have shown to be more accurate than template-fitting methods in situations where the training set greatly resembles the set of galaxies for which the algorithm is estimating redshifts, particularly in redshift and spectral type distributions. When this is not the case - for instance, when the an algorithm trained on galaxies with $0 < z < 0.5$ is used to estimate photo-zs for a selection of galaxies with $0.5 < z < 1$, the accuracy plummets to unusable levels and significant systematic errors are introduced. This limitation makes the application of neural network methods very situational (Collister and Lahav, 2004, Abdalla et al., 2008).

The second main class of methods for estimating photo-zs is template-fitting. This type of method uses theory and expected spectral energy distributions (SEDs) of galaxies to predict what photometry would be observed for a galaxy with a given spectral type at a given redshift. The observed photometry is then compared to the predicted photometry for galaxies of various spectral types and redshifts to find the best match.

Gross features in the SEDs of galaxies are the primary means through which a template-fitting method can estimate the galaxy's redshift. For instance, the 4000 Å break will lie somewhere between the $U$ and $B$ filters for low-redshift galaxies. The more the galaxy is redshifted, the further into the $B$ filter the break will lie. The difference between the photometry from the $U$ and $B$ filters is then correlated with redshift for galaxies of the same spectral type within this range. However, the strength of the break depends on the spectral type of the galaxy. Elliptical galaxies show the strongest break, while starburst galaxies have the break entirely washed out. This makes determinations of the redshift of starburst galaxies nearly impossible. There are other galaxy spectral features which can aid in identification, so a rough estimate can still typically be made.

The strength of template-fitting methods over neural network methods is that they require no prior information about the galaxy set in order to estimate redshifts. They do, however, require universal information about galaxies in the form of the templates they use. Modifications to these templates can have significant effects on the performance of template-fitting methods. The accuracy of the method is thus limited by both the quality of the photometry and of the templates used. Some methods have attempted to get around this limitation by creating an improved set of templates empirically from the observed photometry. The extent to which this is possible is highly limited, but it does show improvement in the results, particularly in correcting for systematics (Abdalla et al., 2008).

A notable danger of photometric redshift methods is the possibility of so-called "catastrophic errors," where the estimated redshift differs from the actual redshift by a factor of many standard deviations. For template-fitting methods, this can be caused by a degeneracy in template-fitting that isn't fully resolved, and for which the photometry from this galaxy favours an incorrect redshift estimate. A common example of this is when the difference in the filter photometry between two filters could be caused by either the 4000 Å in a galaxy at low redshift, or the 912 Å break in a galaxy at high redshift. As template-fitting methods compare a galaxy's observed photometry to the expected photometry from galaxies at a variety of redshifts, it is possible to create a probability distribution function (PDF) for this galaxy's redshift. In cases of catastrophic errors, this PDF will appear bimodal, with the redshift estimated to be near the peak of the larger hump, while the actual redshift lies within the smaller hump. The danger of catastrophic errors can often be averted if galaxies with significantly bimodal PDFs are either avoided or handled carefully.

## 1.5 Group-finding algorithms

In order to perform any analysis on galaxy groups or clusters, it is necessary to first design an algorithm to identify them from galaxy catalogues. If we use the definition of a group as being a collapsed structure of galaxies with $\rho \geq 178\bar{\rho}$, there are a number of methods we can use to detect such regions. When working with exact real-space positions for galaxies, group-finding is essentially a solved problem. The most commonly used group-finding method is known as a friends-of-friends (FoF) algorithm. The FoF algorithm is an iterative process. It starts with one seed galaxy, detecting which other galaxies in the catalogue are within some threshold separation, known as the linking length, from it. Each galaxy within this distance is considered a "friend" and added to a group with the seed galaxy. Each of these friend galaxies is then checked against the catalogue, and every galaxy that lies within the linking length is also declared a friend. This process is repeated until no more friends are added to the group, at which point the algorithm proceeds to the next galaxy in the catalogue that is not already assigned a group and repeats the process.

The linking length is typically assigned a value equal to 1/5 the mean interparticle separation of the catalogue. A group of galaxies with separations under this would then have $\rho > 125\bar{\rho}$, which is roughly the expected density of a virialized group. Galaxy catalogues derived from real-world data have a few limitations, however, which require modifications to the standard method. The first limitation is that most galaxy catalogues are apparent-magnitude-limited, not including any galaxies past the detection limit of the survey. The effect of this is that the galaxy density of the catalogue decreases as a function of redshift. This is typically compensated for by increasing the linking in inverse proportion to the galaxy density at a given redshift, as estimated either by the Schechter function or directly from the data.

The second complication is the fact that peculiar velocities of galaxies will cause their apparent redshifts to not correspond perfectly to their distances from us. The result is that actual groups of galaxies will appear to be stretched out along the line-of-sight direction in what are termed "Fingers of God." The typical method of compensating for this face is using a larger linking length in the line-of-sight direction than in the transverse directions. This linking length is typically taken from the expected velocity dispersion for the group size of interest, though it must be adjusted to an appropriate value which minimizes the possibility of overlinking and possibly merging groups, while also being large enough to link most members of the group.

When galaxy positions are determined with spectroscopic redshifts, these complications are more easily overcome, providing high-accuracy catalogues. However, the redshift errors of photometric redshifts are two orders of magnitude greater than the errors caused by peculiar velocities of galaxies. FoF methods can be applied to photo-z catalogues by simply increasing the redshift linking length even more, but this has been found to achieve unsuitable levels of accuracy. At the linking lengths required to compensate for the errors of photo-zs, there is too much numerical instability with the method, as the required linking lengths are almost guaranteed to link to at least one background galaxy from any point in the field.

Various methods have been developed to handle this problem. Some of these methods are modifications of FoF algorithms designed to work better with photometric redshifts, typically by incorporating data from the galaxies' PDFs. For instance, Li and Yee (2008)'s probability friends-of-friends (pFoF) algorithm works by developing a probability distribution for each group. It starts with a seed galaxy like the traditional FoF algorithm, which is initially considered to be a group of one galaxy. It determines friends of this group by which galaxies are both within a chosen transverse linking length and through a probability comparison, starting with those galaxies that have the best probability match. The probability comparison is done by calculating a probability value:

$$P_{ij} = \frac{\int_{z_{min}}^{z_{max}} P_i(z) P_j(z) dz}{P_{ij,max}} \tag{1.23}$$

Where $P_i(z)$ and $P_j(z)$ are the PDFs of the group and the galaxy and $P_{ij,max}$ is the maximum value of $\int_{z_{min}}^{z_{max}} P_i(z - z_0) P_j(z) dz$ for $z_{min} < z_0 < z_{max}$, used to normalize $P_{ij}$. If $P_{ij}$ is above a chosen threshold value, the galaxy is added to the group. After each galaxy is added to the group, the group's PDF is updated to be:

$$P_i(z) = \frac{\Pi_k P_k(z)}{\int_{z_{min}}^{z_{max}} \left(\Pi_k P_k(z')\right) dz'} \tag{1.24}$$

Where $P_k(z)$ are the PDFs of all galaxies currently assigned to this group. The galaxy list is cycled through until no more galaxies are added to the group. The next galaxy in the list (whether or not it has been added to a previous group) is then used as a seed, and the process repeats until each galaxy has been used as a seed. This group catalogue is then examined for duplicate groups, which are merged into one group with the most likely membership, and groups of low significance, which are removed from the final catalogue.

As this method works primarily on a friends-of-groups basis, the risk of numerical instability is much lower than for FoF. Real groups tend to end up with tight PDFs, while spurious groups have very broad PDFs and are detected and discarded. Li and Yee claim a 90% purity and a $> 99\%$ completeness for their method for groups of at least 8 members.

Aside from FoF-based methods, many group-finding algorithms operate by detecting overdensities in the galaxy distribution. Galaxies that appear tightly clustered on the sky are likely to be grouped, even if their redshifts are unknown. Matched filter techniques apply knowledge of the typical density profile of clusters and attempt to find overdensities in the sky that fit this profile (Postman et al., 1996). Red-sequencing techniques search for overdensities specifically of large, red, elliptical galaxies, which are known to reside primarily within clusters (Gladders and Yee, 2001). Some methods look for bright central galaxies (BCGs) which are often at the centres of groups (Koester et al., 2009).

These different methods each use their own assumptions about the nature and composition of clusters to detect clusters which match these properties. This has the risk of resulting in a biased catalogue which misses clusters that don't fit assumptions. For instance, if there were a large population of clusters that contained few or no large, red, elliptical galaxies, red-sequencing techniques would fail to detect these. Most of these methods also have significant limitations on the scale of groups, as well, which show less significant overdensities and density profiles, along with having fewer large, red, elliptical galaxies and often lacking BCGs.

As our project is focused on investigating groups, it is important that we do not use a method that is significantly biased toward clusters. It is also important that our method has as few false detections as possible. Too low purity will introduce large errors into our lensing estimates, even with a significant number of galaxies in the sample. If the false detections are due to projection effects, this has the result of introducing systematic errors into lensing calculations, which estimate the mass along the line of sight. The result of using projected overdensities such of this would be that we would treat an ensemble of field galaxies as if they were a group, which would suppress any change in the M/L curve. As the data provided by the CFHTLS-Wide survey (CFHTLS, 2009), which we plan to use for our analysis, comprises 170 deg.$^2$ of sky, we will have significant statistical power even with low completeness. We have designed the P3 algorithm with these goals in mind. It searches for overdensities both in the sky and in the redshift dimension, using the PDFs of galaxies to determine the density at each redshift tested, marking the first use of photo-z PDFs for an overdensity-based group-finding algorithm.

## 1.6 Thesis summary

In Chapter 2 of this thesis we explain the methodology behind the P3 algorithm for identifying galaxy groups. Chapter 3 gives the results of testing the algorithm on simulated and real fields, varying some parameters of the method to show which work best and showing how the accuracy of the method is affected by the quality of the photometry used. Chapter 4 discusses the applicability of this method to the galaxy catalogues from the CFHTLS-Wide survey, including preliminary results and comparisons with other group catalogues made from these fields. Chapter 5 discusses our conclusions from testing this method and our plans for future research.

We use a cosmology with the following parameters: $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70\,\mathrm{km/s/Mpc}$. All magnitudes are in the AB system unless stated otherwise.

# Chapter 2

# Group-Finding Method

The methodology behind the P3 algorithm involves searching for significant overdensities in the distribution of galaxies in 3D. Specifically, to search for overdensities, we construct a 3-dimensional grid of points within the lightcone of our field, and at each of these points, we calculate the local overdensity of galaxies in a circular aperture surrounding the point compared to the nearby background in an annulus surrounding this point. As certain areas of the field are masked out, we use the positions of unmasked galaxies to determine the mask. From this, we calculate the portions of our local aperture and annulus which fall into masked regions and use this to properly calculate the local and background densities.

The 3D grid has a spacing of $\sim 0.2$ comoving Mpc in the transverse direction and redshift slices have a thickness of $\Delta z = 0.02$. The typical photo-z errors are $\sim 0.04$, so are resolved with this spacing. High quality photometric redshifts, such as those provided by Ilbert et al. (2009) may have lower errors and require a finer grid-spacing. A small galaxy group will have a radius of $\sim 0.25$ Mpc, so are also resolved with this spacing.

Our calculation for the galaxy surface density within the aperture (represented by $\rho_{\text{ap}}$, though note that the calculated density is only pseudo-3D, as we use a probability density in the z-dimension) can be illustrated by the following procedure:

- For each galaxy, use the photometric redshift probability density function (PDF) of to determine the probability ($p_i$) that it is within the same redshift slice of thickness $\Delta z$.
    - Our algorithm approximates the PDF as a Gaussian distribution to decrease computation time required. We multiply this weight by the ODDS parameter

  provided by the photometric redshift method. In BPZ, the ODDS parameter gives the probability that the true redshift lies within the primary peak of the PDF, and so is necessary in normalizing the Gaussian height.

- Determine an angular weight ($w_i$) for this galaxy, which will depend on the transverse distance between it and the test point. Our method uses a weight of 1 if it falls within the circular aperture of radius $r_{\text{ap}}$ and a weight of zero if it doesn't. The weight can also be set as a function of the transverse distance, resulting in a matched-filter method.
- Sum the weighted probabilities for all galaxies within the aperture and divide by the area of the aperture which falls outside any masked regions ($A_{\text{ap}}$). This gives us the density within the aperture, as shown in Equation (2.1).
  - The area of the aperture which falls outside any masked regions is determined by calculating it for points in a 3-dimensional grid in the lightcone of our field. The area is determining by testing a 2-dimensional grid of points around each of the former points to determine which lie within the aperture and also do not lie in any masked region. This number can be divided by the total number of points of the latter grid which lie within the aperture and then multiplied by the full area of the circle to give $A_{\text{ap}}$. The area of the annulus which lies outside any masked regions, $A_{\text{annu}}$, is calculated in the same manner.

$$\rho_{\text{ap}} = \frac{\Sigma_{i=1}^{n} w_i p_i}{A_{\text{ap}}} \tag{2.1}$$

This procedure can also be used with only minor modifications to determine the density within the annulus surrounding the test point, which will give us the local background galaxy density. We can then calculate the overdensity ($\delta$) from:

$$\delta = \frac{\rho_{\text{ap}}}{\rho_{\text{annu}}} - 1 \tag{2.2}$$

In order to determine a pure sample of galaxies, we select only galaxies with a sufficiently high signal-to-noise ratio (S/N) of $\delta$. In order to calculate the noise in our measurement of $\delta$, we model the number of galaxies which make a significant contribution to the density as a Poisson distribution. In the procedure above, we include only galaxies that have a probability of being within this redshift slice of at least 0.1%, rounding other probabilities down to zero. Using Poisson statistics, a sample which finds $n$ contributing galaxies would give us a standard error of $\sqrt{n}$. We can then estimate the Poisson error of our density as:

$$\sigma_{\text{ap,Poisson}} = \frac{\langle w_i p_i \rangle * \sqrt{n}}{A_{\text{ap}}} = \frac{\rho_{\text{ap}}}{\sqrt{n_{\text{ap}}}} \tag{2.3}$$

There will also be a component to the error contributed by the inaccuracy in determining the average weighted probability from the sample. This would be calculated as the standard error in the weighted probabilities multiplied by the number of galaxies contributing. However, calculating the error in the probabilities is no simple matter, as we would expect a highly non-Gaussian distribution of probabilities, particularly near groups. Fortunately, the error we would expect to see for each probability result is significantly less than the probabilities themselves, so this component can be safely disregarded in our final calculations.

We can then repeat these calculations for the annulus, giving the error in its density. In the end, we combine these errors in quadrature to give the final error in $\delta$. This allows us to calculate the S/N for each test point.

With our 3D grid of S/N, we then proceed to detect the peaks, as these are most likely to correspond to the centres of galaxy groups. In order to not identify multiple peaks with the same group, we apply a threshold distance - typically 0.5 Mpc, the size of a large group - in which a peak must be the highest point, rather than simply requiring that the peak must be higher than the points immediately surrounding it in the grid. We then extract only those peaks which have a S/N $> 3$, in order that we'll have a very pure sample remaining. This leaves us with our ultimate group catalogue.

# Chapter 3

# Tests of the Group-Finding Algorithm

In order to test the P3 algorithm, we compared its results to group catalogues obtained through a friends-of-friends algorithm applied to fields extracted from the Millennium simulation by Kitzbichler and White (2007) using their actual redshifts, and to a friends-of-friends sprectroscopic group catalogue generated by Knobel et al. (2009) using the zCOSMOS 10k sample covering the COSMOS field, which overlaps with the CFHTLS D2 field. Our comparison method aimed primarily to assess the purity of our samples, so we tested how many of our detected groups fell within $\Delta z = 0.04$ and $\Delta r = 0.05$ Mpc of at least one spectroscopic group. We used these parameters, as the redshift threshold distance is approximately twice the uncertainty in the mean photometric redshift for a group of 5 members, and the transverse distance threshold is approximately the upper size limit for a group. This comparison was run for various S/N limits on our detected peaks, along with a control field which consisted of positions generated from a uniform random distribution in R.A., Dec., and z. Although completeness was not our goal, we also measured it by calculating the number of spectroscopically-identified groups in the field which had at least one photometrically-identified group matched to it.

## 3.1 Simulated Data

To assess the accuracy of the P3 algorithm against an ideal catalogue, we used six simulated 2 deg.$^2$ lightcones extracted from the Millennium simulation (Springel et al., 2005, De Lucia and Blaizot, 2007) by Kitzbichler and White (2007). Given the resolution limits of the Millennium simulation, the catalogue is complete for Johnson $I < 24$ in the AB system. We also used a magnitude limit of 22.5 in the $I$-band for most of the testing,

TABLE 3.1: Summary of the purity and completeness of the P3 algorithm when its results are matched to the FoF catalogue of groups, for various cuts on the data and a control ("rand"). The data covers a total of 12 deg.$^2$ of simulated sky. Nhit is the number of detected peaks which match to a FoF group; Ntot is the total number of detected peaks; p is the portion of peaks which match to a FoF group; C is the portion of FoF groups which match to a detected peak; and $\langle N_\mathrm{m} \rangle$ is the average number of members in the detected peaks which match to a FoF group.

| | | CFHTLSpz errors | | | | | COSMOS30pz errors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cut | $r_\mathrm{ap}$ | Nhit | Ntot | p | C | $\langle N_\mathrm{m} \rangle$ | Nhit | Ntot | p | C | $\langle N_\mathrm{m} \rangle$ |
| rand | 0.5 | 1492 | 2988 | 0.499 | N/A | 3.458 | 1492 | 2988 | 0.499 | N/A | 3.458 |
| S/N > 2 | 0.5 | 1541 | 2070 | 0.744 | 0.169 | 4.559 | 1681 | 2049 | 0.820 | 0.219 | 4.736 |
| S/N > 3 | 0.5 | 996 | 1181 | 0.843 | 0.132 | 5.526 | 963 | 1026 | 0.939 | 0.170 | 6.315 |
| S/N > 4 | 0.5 | 491 | 531 | 0.925 | 0.085 | 7.652 | 413 | 416 | 0.993 | 0.107 | 10.017 |
| none | 0.5 | 1974 | 2887 | 0.684 | 0.194 | 4.108 | 2068 | 2806 | 0.737 | 0.240 | 4.308 |
| rand | 0.25 | 1492 | 2988 | 0.499 | N/A | 3.458 | 1492 | 2988 | 0.499 | N/A | 3.458 |
| S/N > 2 | 0.25 | 6699 | 10906 | 0.614 | 0.449 | 3.583 | 5578 | 7162 | 0.779 | 0.450 | 4.026 |
| S/N > 3 | 0.25 | 2775 | 3537 | 0.785 | 0.265 | 4.948 | 1643 | 1715 | 0.958 | 0.227 | 7.007 |
| S/N > 4 | 0.25 | 866 | 986 | 0.878 | 0.130 | 8.234 | 445 | 448 | 0.993 | 0.102 | 13.924 |
| none | 0.25 | 9761 | 20735 | 0.471 | 0.557 | 3.196 | 9560 | 18735 | 0.510 | 0.591 | 3.315 |

as this matches the spectroscopic catalogue of Knobel et al. (see Section 3.3 below). We also tested including galaxies with $I$-band magnitudes between 22.5 and 24 to assess how this affected our accuracy (see Section 3.2.1, also below).

To simulate photometric redshifts for this dataset, for simplicity we applied a Gaussian deviate to the redshifts of the galaxies. We generated two mock datasets, corresponding to two different levels of photo-z error ranges. The first set, hereafter CFHTLSpz, simulated the accuracy of the photometric redshifts in the CFHTLS Deep fields (Ilbert et al., 2006), with a redshift error of 0.05 for $I < 22.5$, and 0.15 for $22.5 < I < 24$. The second set, hereafter COSMOS30pz, mimicked the accuracy of the COSMOS-30 (Ilbert et al., 2009) photometric redshifts: 0.02 for $I < 22.5$, and 0.06 for $22.5 < I < 24$.

We note that, after these tests were done, a recent analysis of photo-z's in the CFHTLS-Wide survey (Hildebrandt et al., 2009) suggests that these are closer to those adopted for the simulated COSMOS30pz set.

Fig. 3.1 and Fig. 3.2 show a graphical representation of the S/N calculated by the P3 algorithm for a selection of redshift slices, to illustrate how the detected peaks correspond to actual groups. A summary of the accuracy of the P3 algorithm when compared to a FoF algorithm can be seen in Table 3.1. The P3 algorithm typically detects around 98 groups/deg.$^2$ of the simulated data. Of these detected groups, approximately 84% match to at least one spectroscopically-identified group with at least two bright members when we apply CFHTLSpz errors to our galaxies for the photometric data. This will give us approximately 83 correct detections per deg.$^2$. Our completeness is very low, however, picking up at best 40% of groups with 10 or more bright members, and less than 15%

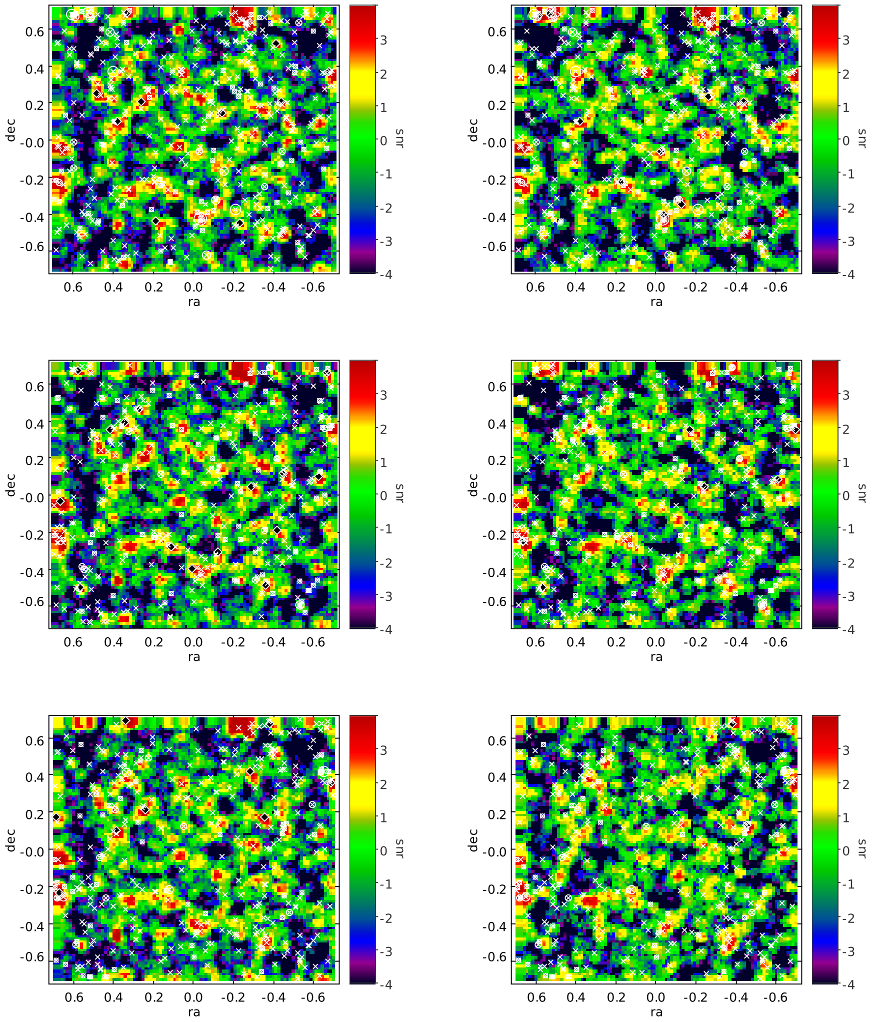


Figure 3.1: Plots of the calculated S/N for the $\delta$ of galaxies on a grid of points in R.A. and Dec., sliced at different values of the redshift, for a field drawn from the Millennium simulation by Kitzbichler and White (2007). S/N is indicated by the colour. Locations of real groups detected through a friends-of-friends algorithm are indicated by white circles, with their sizes indicating the sizes of the groups. White crosses indicate the location of a circle in a nearby lair, within $\Delta z = 0.04$. Detected peaks with a S/N $> 3$ are indicated by the black diamonds. Peaks are detected in three dimensions, so what appear to be peaks in the individual plots may actually be detected on another slice. Additionally, peaks have a threshold radius within which they must be the highest point to count as a peak, so some peaks may not be detected if they are sufficiently close to another peak. Left column contains plots using CFHTLSpz errors with $r_{\mathrm{ap}} = 0.5\,\mathrm{Mpc}$, right column contains plots using COSMOS30pz errors with $r_{\mathrm{ap}} = 0.5\,\mathrm{Mpc}$. Redshift slices, from top to bottom: 0.58, 0.60, 0.62

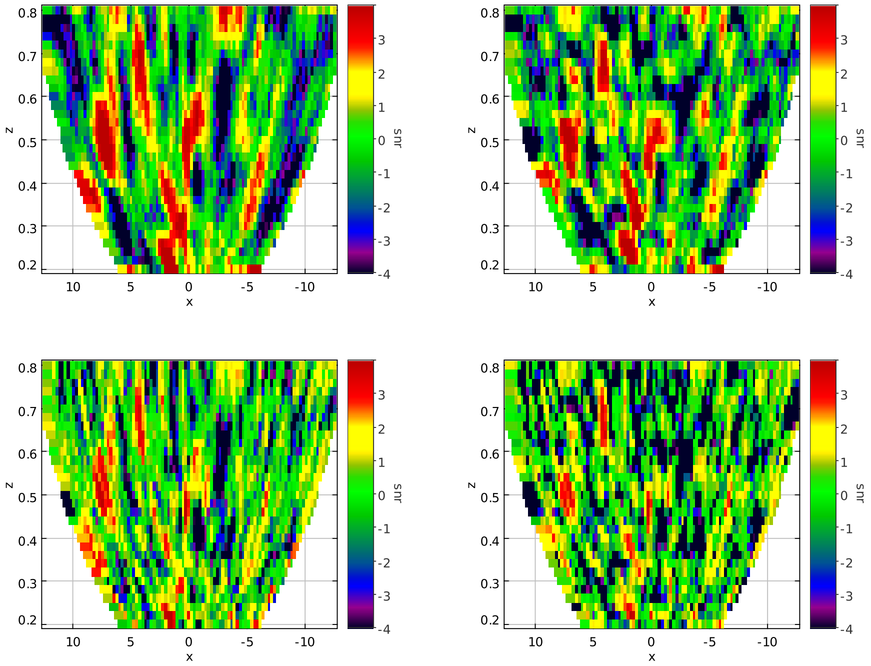


Figure 3.2: A side view of the plots from Fig. 3.1, showing how far groups extend in the redshift dimension, sliced at a Dec. of 0.25. The left column uses CFHTLSpz errors, and the right column uses COSMOS30pz errors. The top row uses $r_{\mathrm{ap}} = 0.5\,\mathrm{Mpc}$, and the bottom row uses $r_{\mathrm{ap}} = 0.25\,\mathrm{Mpc}$.

of all groups with 2 or more bright members. For our purposes, this isn't a concern, as the total number of groups we detect will be enough for our purposes, given the size of the CFHTLS-Wide survey.

When we use COSMOS30pz errors, our purity goes up to 94%, with 86 groups/deg.$^2$ detected and 80 of these being real. This is extremely good, and we hope that our accuracy with the CFHTLS-Wide data will approach this.

The typical size of the groups we detected is around 3-7 members, though this number depends on what signal-to-noise cut and which dataset we used. Even though it might seem that a high signal-to-noise cut would significantly bias us toward larger groups, the fact that there are many more small groups than large groups means that many of these groups will, by chance, have a large signal-to-noise and be detected by our algorithm. This effect can easily be seen in Fig. 3.3, Fig. 3.4, and Fig. 3.5.

In order to estimate the size of our detected groups, we investigated whether we could use the local $\delta$ to estimate the number of members contained in the group. For each photometrically-detected group that matched to a spectroscopically-identified group, we plotted its $\delta$ versus the number of bright ($I < 22.5$) members, determined through a friends-of-friends algorithm. We also calculated the $\delta$ at the location of every spectroscopically-identified group for further data. These plots can be seen in Fig. 3.3.

Although there is a weak correlation between the number of members in a group and its $\delta$, using this to estimate the number of members is problematic. This is primarily due to the fact that there are many more groups with smaller numbers of members, so on this end there are more groups that appear to have a relatively high (or low) $\delta$ by chance. For instance, a group with $\delta = 5$ (measured with a 0.25 Mpc aperture size) has a roughly equal chance to have 2 members as it does to have 10 or more. Given this, a simple mapping of $\delta$ to number of members would be unwise. A statistical approach which uses a probability distribution for the number of members a galaxy might have, given its $\delta$, might show better results. However, as Fig. 3.5 shows, the histogram of number of group members for different $\delta$ cuts only shows a significant change in shape at the very high-mass end, where it is dominated almost entirely by groups with high $\delta$. A successful estimate of the number of group members will require an alternative method, such as a friends-of-groups algorithm using our catalogue of group centres.

One potential concern was that the above-random match rate of our photometric galaxy catalogue to the spectroscopic catalogue might have been due primarily to a very high match rate among larger groups averaged with a lower match rate to smaller groups. If this were the case, then our method could in actuality be little better than random for identifying small groups. To test this, we took our galaxy catalogues from our simulated

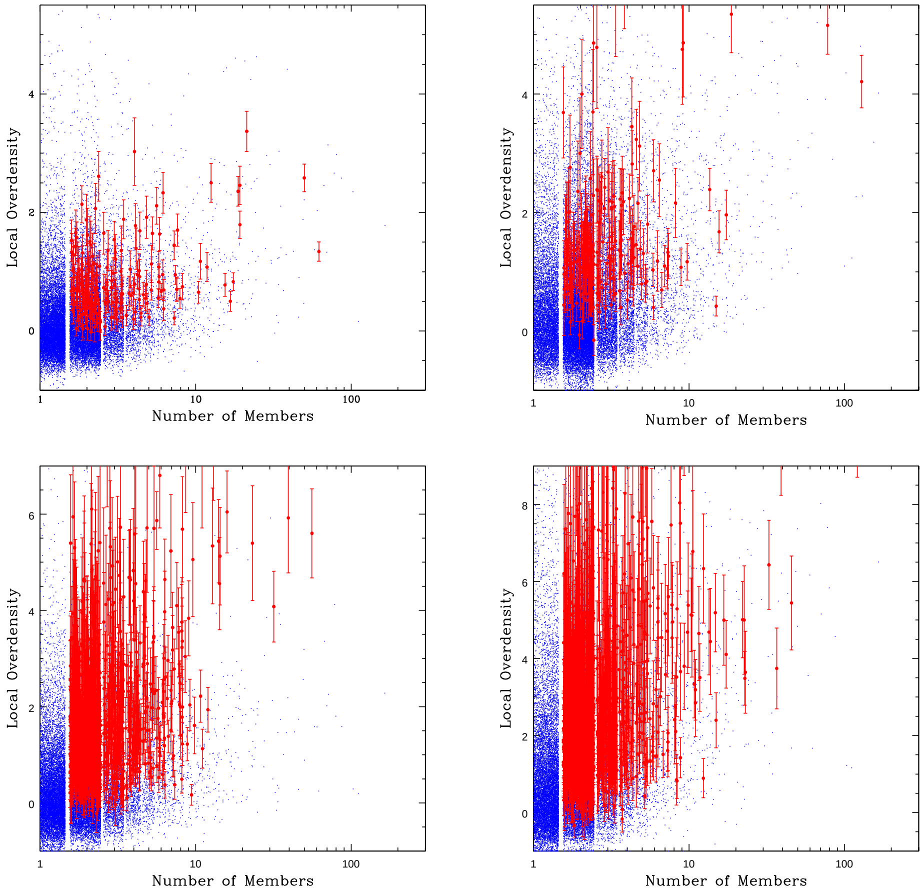


Figure 3.3: Plots of the $\delta$ calculated by our algorithm at the location of every FoF-identified group (small dots) versus the number of members in those groups for the Millennium fields. Also shown are the $\delta$ of all estimated groups and the number of bright ($I < 22.5$) members of the FoF-identified group they match to (large dots with error bars). The numbers of members for all data points have a random component of less than 1 included in order to aid viewing. The left column uses CFHTLSpz errors, and the right column uses COSMOS30pz errors. The top row uses $r_{\mathrm{ap}} = 0.5\,\mathrm{Mpc}$, and the bottom row uses $r_{\mathrm{ap}} = 0.25\,\mathrm{Mpc}$.

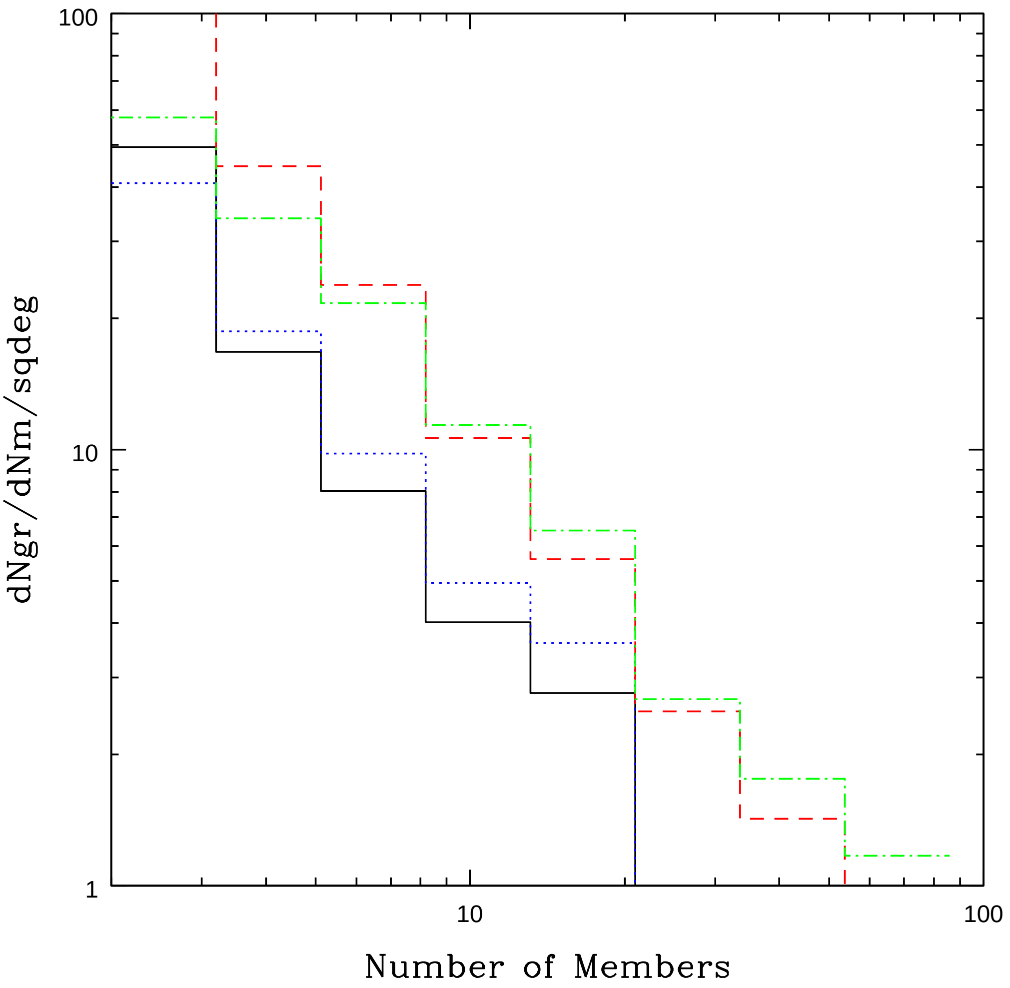


Figure 3.4: Histograms of the number of groups detected with a given number of members for the Millennium fields, using a minimum S/N cut of 3. Solid black: CFHTLSpz errors, $r_{\rm ap} = 0.5\,\rm Mpc$. Dotted blue: COSMOS30pz errors, $r_{\rm ap} = 0.5\,\rm Mpc$. Dashed red: CFHTLSpz errors, $r_{\rm ap} = 0.25\,\rm Mpc$. Dash-dotted green: COSMOS30pz errors, $r_{\rm ap} = 0.25\,\rm Mpc$.

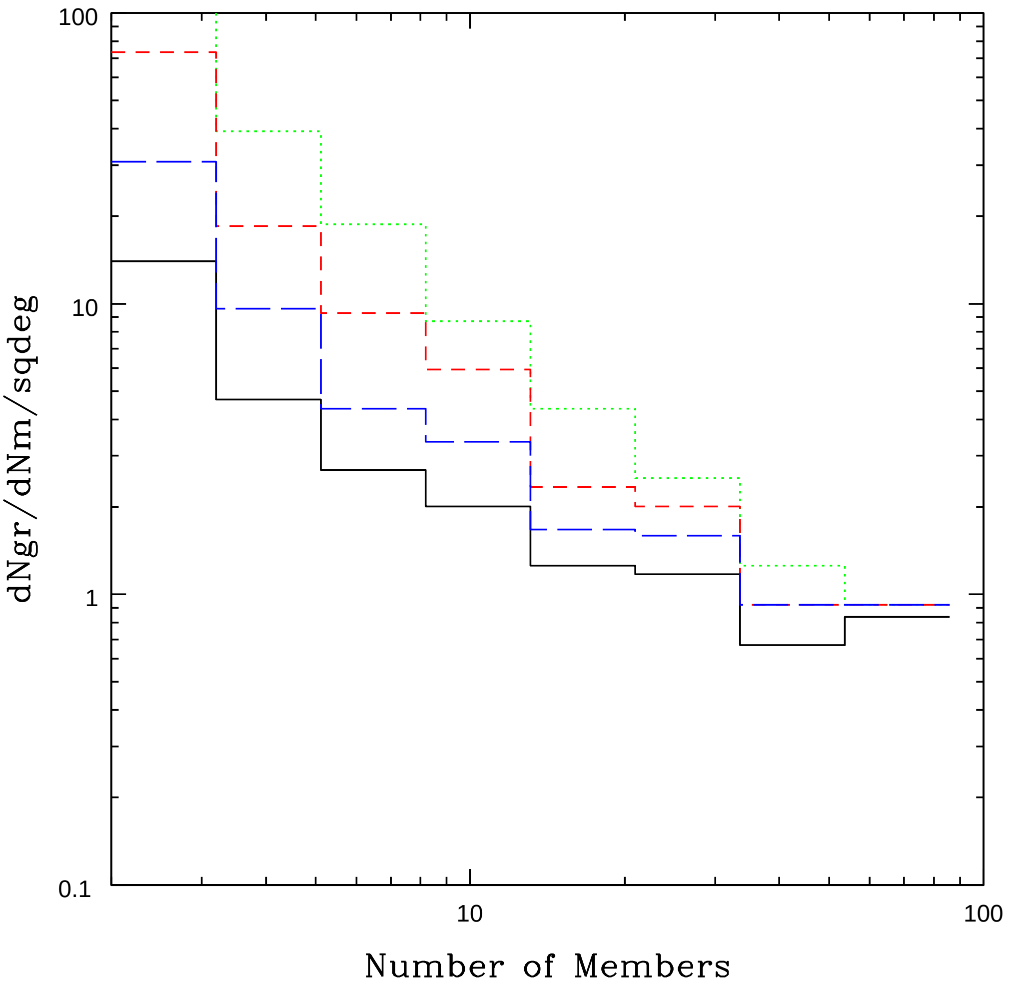


Figure 3.5: Histograms of the number of groups detected with a given number of members for the Millennium fields, using all peaks detected, CFHTLSpz errors, and $r_{\mathrm{ap}} = 0.25$ Mpc, for various cuts on $\delta$. Dotted green: $\delta > 2$. Short dashed red: $\delta > 3$. Long dashed blue: $\delta > 4$. Solid black: $\delta > 5$.

Table 3.2: Summary of the purity and completeness of the P3 algorithm using $r_{\text{ap}} = 0.5$ Mpc, CFHTLSpz errors (comparable to Table 3.1, top row, left column), and including all galaxies with $I < 22.5$ that were not found by our friends-of-friends algorithm to lie within a large ($N_{gal} > 10$) group, matched to the FoF catalogue of small ($N_{gal} < 10$) groups, for various cuts on the data and a control ("rand"). The purity shows no statistically significant change relative to the catalogue used for Table 3.1, showing that large groups are not significantly biasing our purity and completeness upwards. The data covers a total of 12 deg.$^2$ of simulated sky. Columns are as Table 3.1

| Cut | Nhit | Ntot | p | C | $\langle N_{\text{m}} \rangle$ |
|---|---|---|---|---|---|
| rand | 1492 | 2988 | 0.499 | N/A | 3.458 |
| S/N > 2 | 1630 | 2179 | 0.748 | 0.180 | 3.863 |
| S/N > 3 | 1029 | 1222 | 0.842 | 0.136 | 4.361 |
| S/N > 4 | 484 | 528 | 0.917 | 0.082 | 5.514 |
| none | 2097 | 3054 | 0.687 | 0.206 | 3.584 |

data and removed all galaxies within them that were found to be a member of a large group (which we considered any group having more than 10 members to be). We then ran our algorithm on this pruned catalogue and assessed its accuracy through the same method as before. The results of this test are summarized in Table 3.2, which shows that there is in fact very little effect on the accuracy of our algorithm when the larger groups are removed from consideration.

## 3.2 Optimization of the algorithm

### 3.2.1 Magnitude limit

Although galaxies with $I > 22.5$ have less accurate redshifts, they should still provide some information that could be useful for identifying groups. There is no known lower limit for the luminosity of substructure in groups; even around the Milky Way, we have found substructure that can only be identified statistically from the distribution of stars. In theory, however poor the precision of the redshift estimates for fainter groups might be, in sufficient quantity they should assist in increasing the overall precision of group estimates.

In practise, this turned out not to be the case. We ran the P3 algorithm on our simulated data, using all galaxies with $I < 24$, assigning errors to galaxies with $22.5 < I < 24$ of 3x as much as for galaxies with $I < 22.5$, which were 0.15 for the CFHTLSpz errors and 0.06 for the COSMOS30pz errors. In the end, using this extra data showed a decrease in the purity of group-finding for both the simulated and real data of up to 10%, which can be seen through comparison Table 3.3 to Table 3.1.

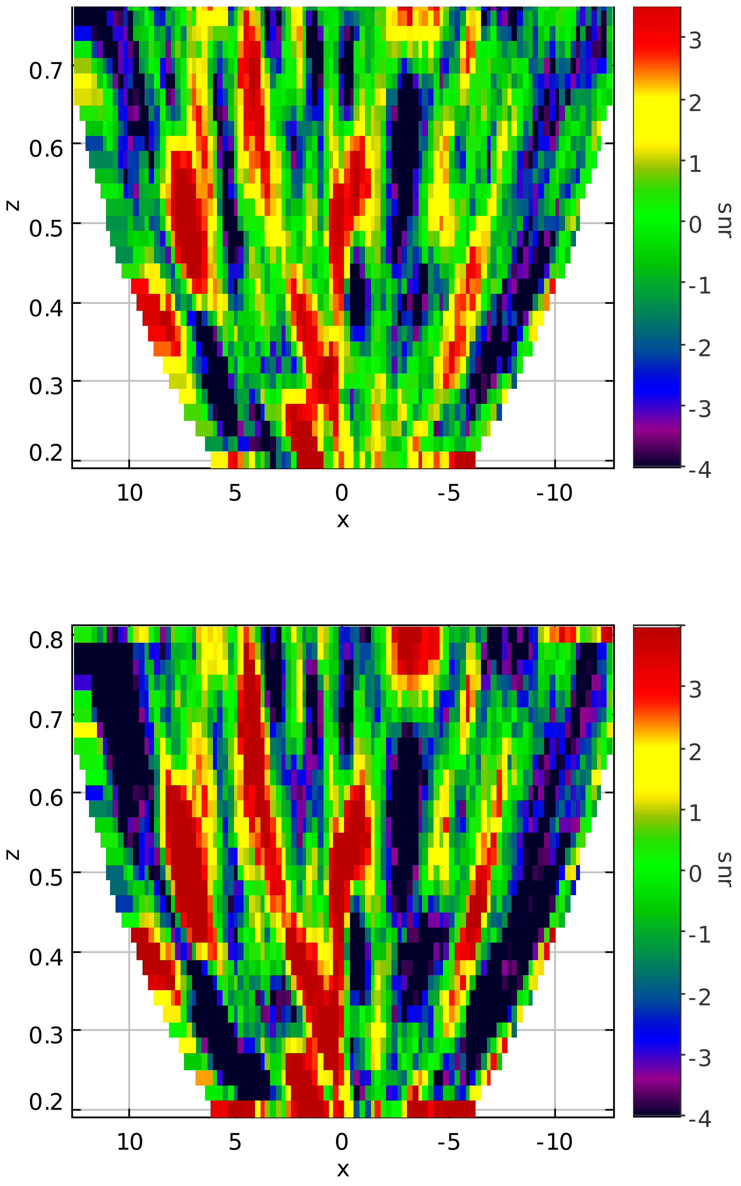


Figure 3.6: Top plot shows S/N map of one of the Millennium fields, using CFHTLSpz errors for galaxies with $I < 22.5$. The bottom plot shows the same, except also including galaxies with $22.5 < I < 24$, which were assigned errors of 3x as much as the brighter galaxies. Both plots use $r_{\mathrm{ap}} = 0.5$ Mpc. It can be seen here that including fainter galaxies tends to sharpen the distribution, but there is no significant change to the shape of it.

Table 3.3: Summary of the purity and completeness of the P3 algorithm applied to the Millennium fields with $r_{\mathrm{ap}} = 0.5$ (comparable to Table 3.1, top row), and including all galaxies with $I < 24$ when matched to the FoF catalogue of groups, for various cuts on the data and a control ("rand"). The data covers a total of 12 deg.$^2$ of simulated sky. Columns are as Table 3.1.

| | CFHTLSpz errors | | | | | COSMOS30pz errors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cut | Nhit | Ntot | p | C | $\langle N_{\mathrm{m}} \rangle$ | Nhit | Ntot | p | C | $\langle N_{\mathrm{m}} \rangle$ |
| rand | 1492 | 2988 | 0.499 | N/A | 3.458 | 1492 | 2988 | 0.499 | N/A | 3.458 |
| S/N $> 2$ | 1583 | 2278 | 0.695 | 0.168 | 4.513 | 1700 | 2148 | 0.791 | 0.212 | 4.630 |
| S/N $> 3$ | 1340 | 1840 | 0.728 | 0.153 | 4.875 | 1402 | 1696 | 0.827 | 0.192 | 5.059 |
| S/N $> 4$ | 1007 | 1324 | 0.761 | 0.130 | 5.465 | 1011 | 1150 | 0.879 | 0.164 | 5.942 |
| none | 1894 | 2901 | 0.653 | 0.186 | 4.204 | 1900 | 2537 | 0.749 | 0.223 | 4.389 |

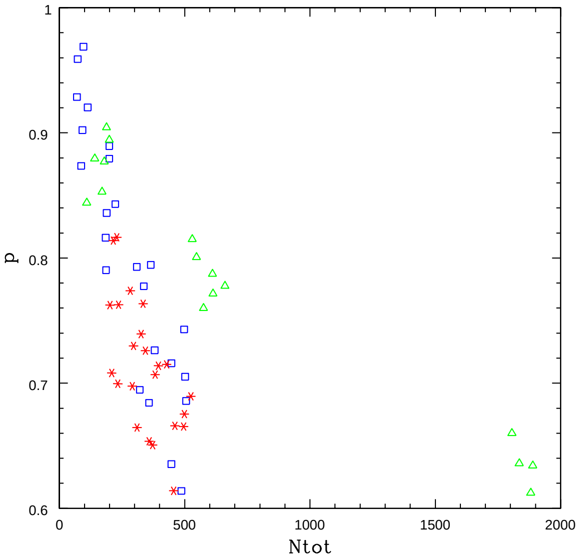
The issue seems to be that, as the locations of fainter galaxies are highly correlated with the locations of bright galaxies, the effect of adding their data is to sharpen the existing S/N distribution, as can be seen in Fig. 3.6. It is possible that multiple groups along the line of sight may become blended into a single structure, with the peak of this structure lying between the two groups. When the peak catalogue is compared to the actual locations of the groups, it is possible that it is within our threshold matching distance of neither.

### 3.2.2 Aperture size

Although decreasing the aperture size typically resulted in decreasing the purity of our catalogue at a given S/N cut, it also greatly increased the number of peaks detected. The important measurement is whether the decreased aperture size results in increased purity when the same number of groups are detected, or similarly, whether the decreased aperture size results in more groups detected at the same purity level. Fig. 3.7 shows a graphical representation of how the purity relates to the number of groups detected for both aperture sizes, along with the results of changing to a fainter magnitude limit. From this graph, it is clear that the smaller aperture size is beneficial, though it may require a larger S/N cut to attain sufficient purity.

## 3.3 zCOSMOS Data

In addition to simulated data, we tested the P3 algorithm on data from the COSMOS/CFHTLS D2 field. This field has spectroscopic redshifts a large number of the galaxies, in addition to photometric redshifts based on *ugriz* from Ilbert et al. (2006) with errors of around 0.05 for $I < 22.5$, and also much smaller errors from the COSMOS 30 band data (Ilbert et al., 2009) (around 0.02 for $I < 22.5$). This allows us to better see



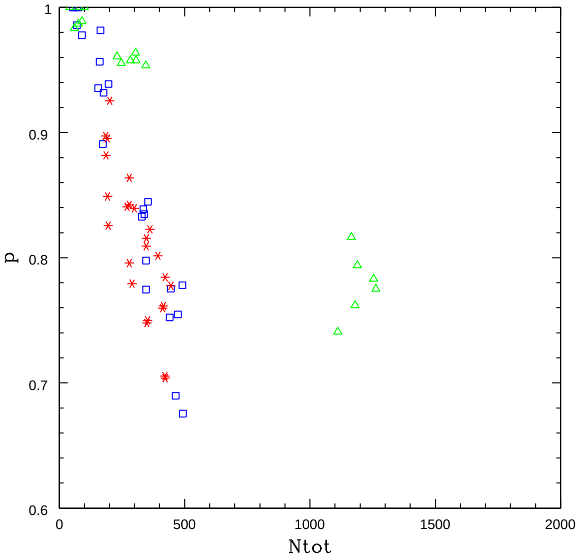


FIGURE 3.7: Plots showing purity as a function of the total number of peaks detected. Top plot uses CFHTLSpz errors, bottom plot uses COSMOS30pz errors. Blue squares use $r_{\mathrm{ap}} = 0.5$ and $I < 22.5$. Green triangles use $r_{\mathrm{ap}} = 0.25$ and $I < 22.5$. Red stars use $r_{\mathrm{ap}} = 0.5$ and $I < 24$. Each field from the Millennium simulation is represented by three points, for cuts of S/N $> 2$, S/N $> 3$, and S/N $> 4$.

Table 3.4: Summary of the purity and completeness of the P3 algorithm applied to the galaxies in the D2 field. The data covers approximately 1 deg.$^2$ of sky. Columns are as Table 3.1.

| | | CFHTLS photometry | | | | | COSMOS-30 photometry | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cut | $r_{\mathrm{ap}}$ | Nhit | Ntot | p | C | $\langle N_{\mathrm{m}} \rangle$ | Nhit | Ntot | p | C | $\langle N_{\mathrm{m}} \rangle$ |
| rand | 0.5 | 55 | 200 | 0.275 | 0.144 | 3.636 | 55 | 200 | 0.275 | 0.139 | 3.636 |
| S/N $> 2$ | 0.5 | 46 | 77 | 0.597 | 0.170 | 4.261 | 75 | 98 | 0.765 | 0.315 | 4.653 |
| S/N $> 3$ | 0.5 | 35 | 50 | 0.700 | 0.129 | 4.486 | 35 | 40 | 0.875 | 0.176 | 5.914 |
| S/N $> 4$ | 0.5 | 16 | 21 | 0.762 | 0.057 | 5.125 | 13 | 13 | 1.000 | 0.076 | 6.154 |
| none | 0.5 | 53 | 100 | 0.530 | 0.199 | 4.170 | 94 | 157 | 0.599 | 0.380 | 4.191 |
| rand | 0.25 | 55 | 200 | 0.275 | 0.144 | 3.636 | 55 | 200 | 0.275 | 0.139 | 3.636 |
| S/N $> 2$ | 0.25 | 156 | 321 | 0.486 | 0.413 | 4.032 | 112 | 147 | 0.762 | 0.343 | 4.393 |
| S/N $> 3$ | 0.25 | 59 | 91 | 0.648 | 0.184 | 4.305 | 19 | 20 | 0.950 | 0.085 | 5.737 |
| S/N $> 4$ | 0.25 | 9 | 12 | 0.750 | 0.029 | 5.778 | 0 | 1 | 0.000 | 0.008 | 0.000 |
| none | 0.25 | 230 | 605 | 0.380 | 0.608 | 3.557 | 322 | 1148 | 0.280 | 0.794 | 3.357 |

what purity we can expect from when we run the P3 algorithm on the CFHTLS-Wide survey, and how the accuracy might improve in surveys with better photometric redshift accuracy. We also have a friends-of-friends identified spectroscopic group catalogue provided by Knobel et al. (2009) which we used to assess the accuracy of our method.

In working with the photometric data, we found that there appeared to be a small, but significant, offset between the photometric redshifts provided by Ilbert et al. (2006) and the spectroscopic redshifts. To correct for this, we performed a sky match on galaxies present and both of the catalogues and fit a linear correction function to their redshifts, of the form $z_{real} = 0.957 z_{phot} + 0.00843$. This correction allowed us to properly match our detected groups to those from Knobel et al.. Although this correction had a significant effect on the apparent quality of our group-matching, it is unlikely to have any significant effect on lensing measurements. The reason for this is that redshift comes into the equations for lensing in the form of a $D_{LS}/D_S$ factor applied to each lensed image. Our expected redshift offsets are $\sim 0.02$ at most, which is large for the purpose of group matching, but proportionally smaller for the purpose of lensing.

Fig. 3.8 and Fig. 3.9 show a graphical representation of the S/N in selected redshift slices for the D2 field. Although the catalogue of FoF-identified groups is significantly sparser, we still do see most of our detected peaks corresponding to a FoF-identified group.

As can be seen in Table 3.4, our results using real photometric redshifts and spectroscopically-identified groups were somewhat worse that our simulated data with CFHTLSpz errors, as would be expected. The real data has a number of catastrophic errors, which we did not simulate. Additionally, there are incompleteness factors in the real data which means that some groups may not be detected spectroscopically due to some of their

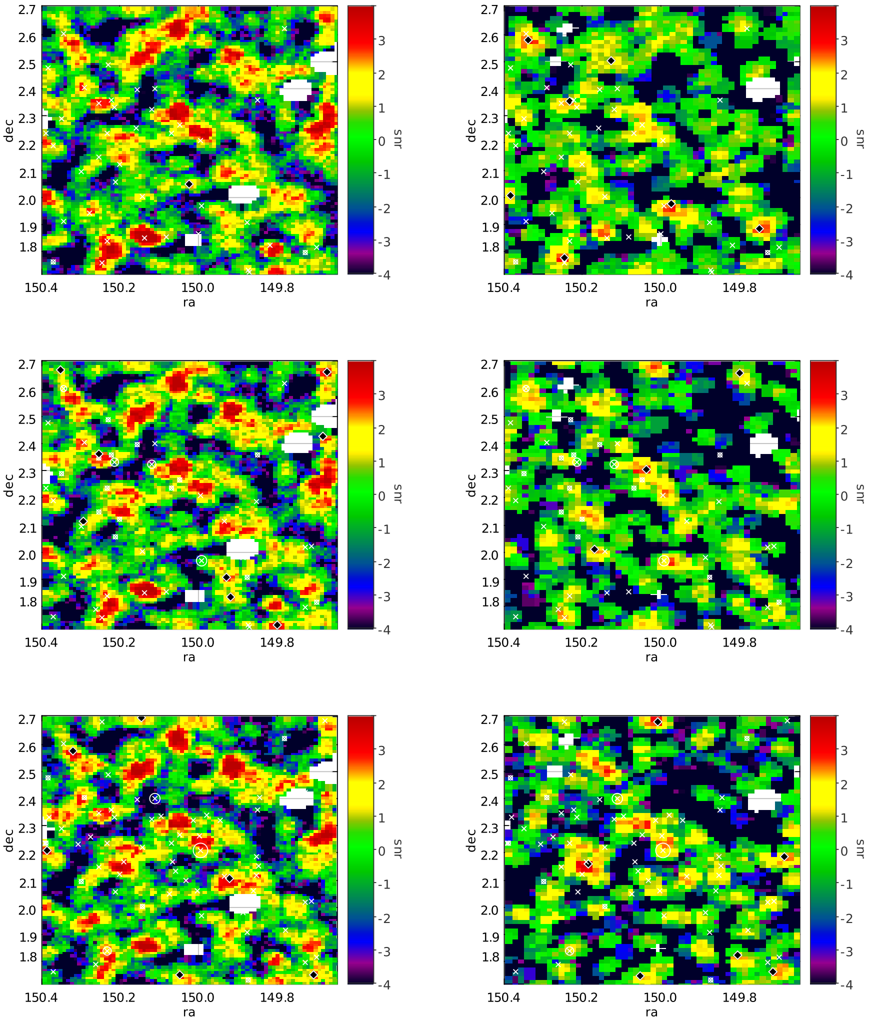


Figure 3.8: Plots of the calculated S/N for the $\delta$ of galaxies on a grid of points in R.A. and Dec., sliced at different values of the redshift, for the D2 field, using $r_{\mathrm{ap}} = 0.5$ Mpc. S/N is indicated by the colour. Locations of real groups detected through Knobel et al. friends-of-friends algorithm are indicated by white circles, with their sizes indicating the sizes of the groups. White crosses indicate the location of a circle in a nearby layer, within our threshold redshift for being considered a match. Detected peaks with a S/N of more than 2 are indicated by the black diamonds. Peaks are detected in three dimensions, so what appear to be peaks in the individual plots may actually be detected on another slice. Additionally, peaks have a threshold radius within which they must be the highest point to count as a peak, so some peaks may not be detected if they are sufficiently close to another peak. Left column shows data using the Ilbert et al. (2006) photometry, with errors similar to the CFHTLSpz errors used previously, and the right column uses the COSMOS-30 data, with errors similar to the COSMOS30pz erros used previously. Redshift slices, from top to bottom: 0.58, 0.60, 0.62. The COSMOS plot shows the interesting effect that many galaxies are individually resolvable, as their redshift errors are smaller than the width of the slices. These galaxies appear as solid circles in the plot.

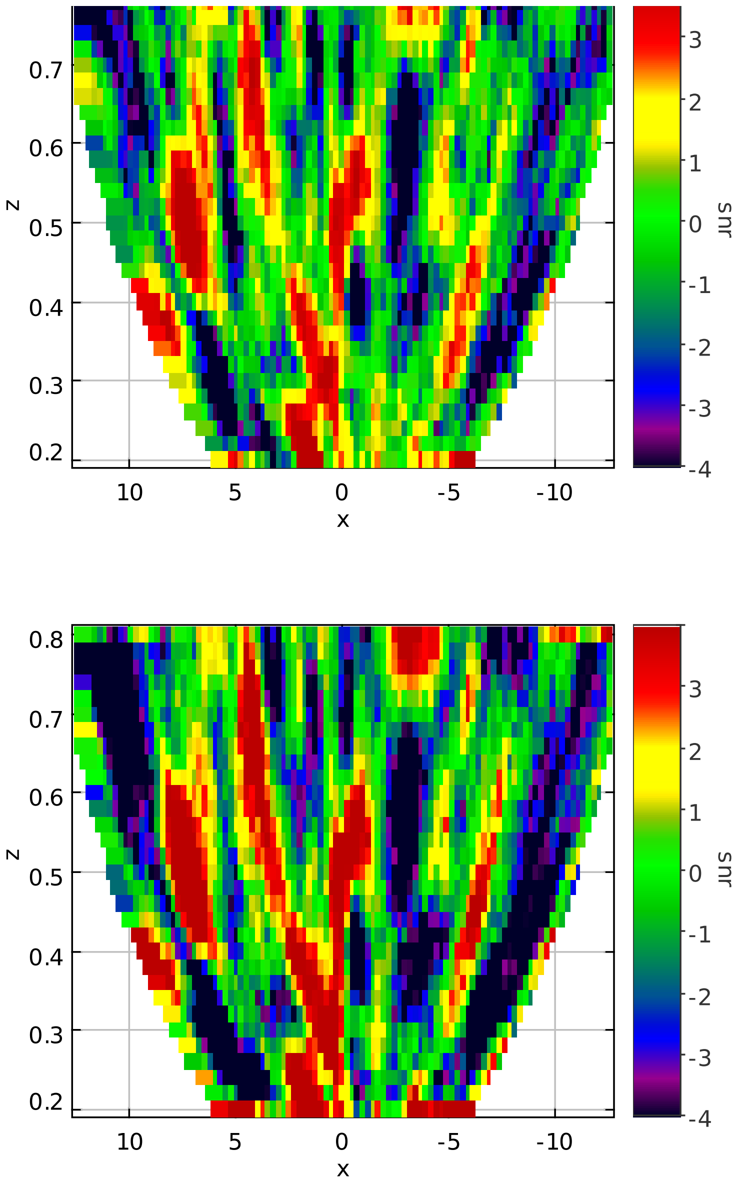


FIGURE 3.9: A side view of the plots from Fig. 3.8, showing how far groups extend in the redshift dimension, sliced at a Dec. of 0.25. The top plot uses Ilbert et al. (2006) photometry and $r_{\mathrm{ap}} = 0.5$ Mpc; the bottom uses the COSMOS-30 photometry with $r_{\mathrm{ap}} = 0.5$ Mpc. In this case, we can see much more clearly in the COSMOS-30 plot the contraction of groups in the redshift dimension.

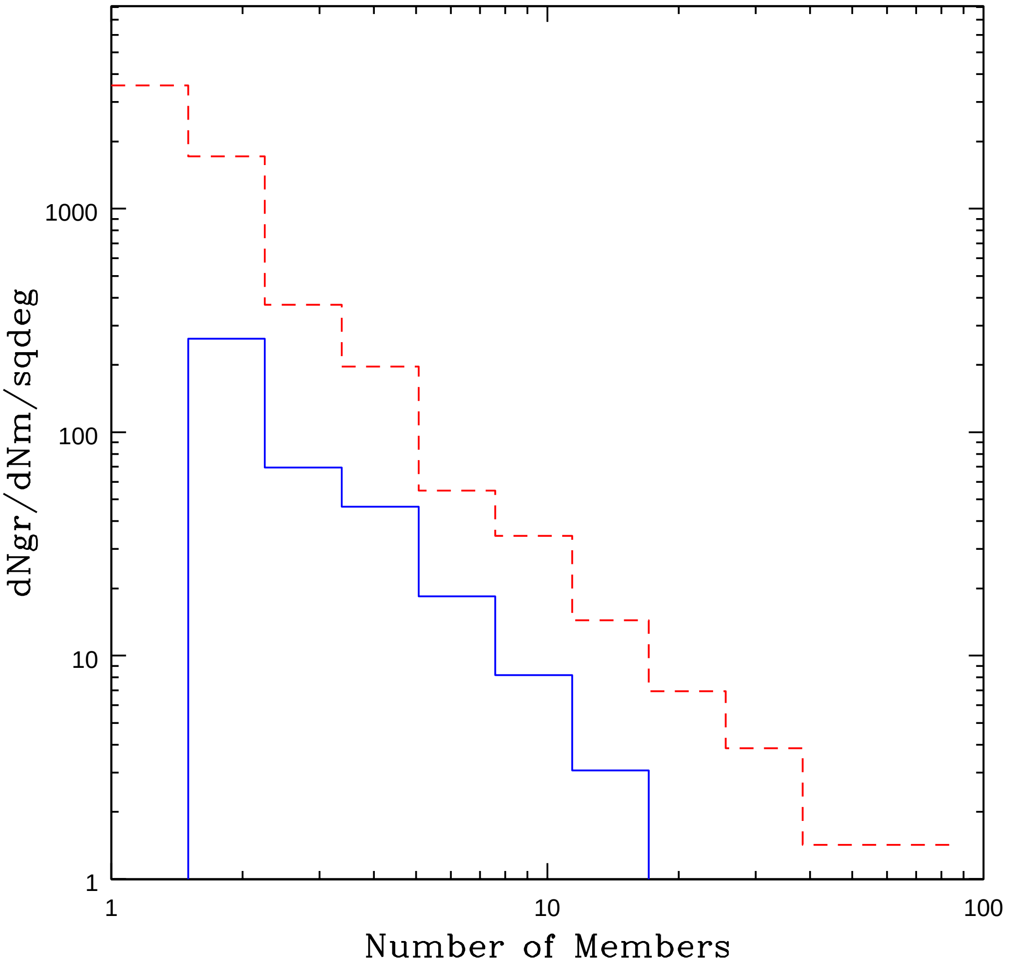


FIGURE 3.10: This plot shows the number of groups of varying numbers of members identified in the magnitude range $I < 22.5$ for both the Millennium simulation (red dashed) and the zCOSMOS FoF-groups of Knobel et al. (2009) (blue solid). The vertical axis is normalized by the field size to aid comparison. We can assess completeness by measuring how far the zCOSMOS plot lies to the left of the Millennium plot. This gives us an estimate of 40-50% completeness.

members not having spectra. These completeness effects are evident in the smaller average number of members in our detected groups, compared to the averages for the simulated data. They are also evident in the fact that our control catalogue ("rand") shows lower purity for this field than for the Millennium fields. Fig. 3.10 compares the group counts for various memberships of the D2 and Millennium fields, directly illustrating the incompleteness of the D2 galaxy and group catalogues.

We also attempted to assess how accurately we will be able to determine the size of a group from its local $\delta$ using this data, illustrated in Fig. 3.11. Although the data is significantly sparser than our simulated data, a positive correlation between $\delta$ and the number of members in a group can still be seen. However, the trend is not significant enough to allow us to make future estimates of the size of groups from their $\delta$.

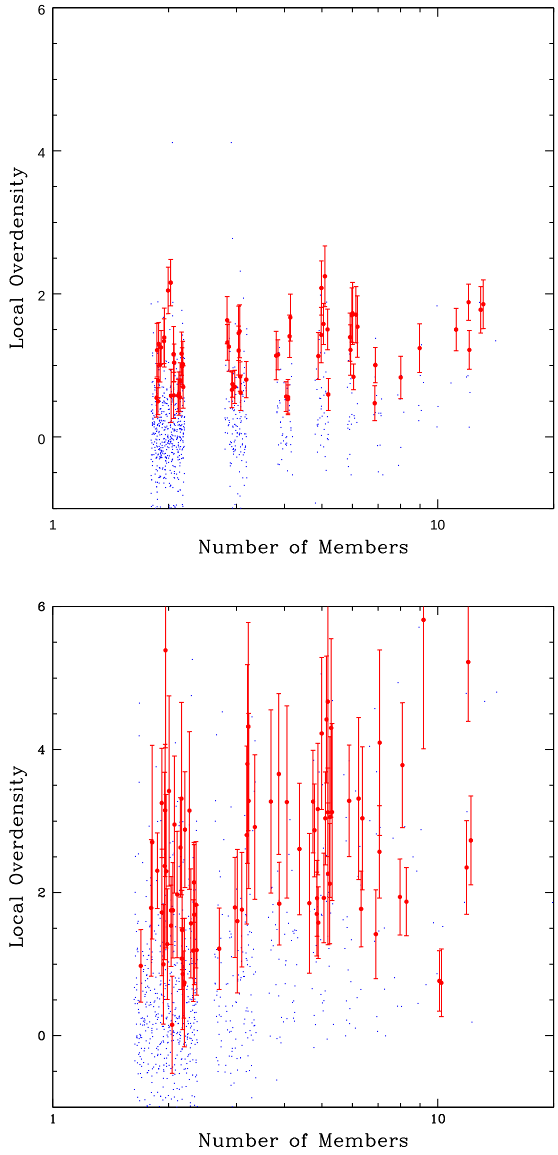


FIGURE 3.11: Plots of the $\delta$ calculated by the P3 algorithm at the location of every FoF-identified group (small dots) versus the number of members in those groups for the D2 field. Also shown are the $\delta$ of all estimated groups and the number of bright ($I < 22.5$) members of the FoF-identified group they match to (large dots with error bars). The numbers of members for all data points have a random component of less than 1 included in order to aid viewing. Top plot uses photometry from Ilbert et al. (2006), bottom plot uses photometry from the COSMOS-30 survey.

# Chapter 4

# Application to the CFHTLS-Wide

The CFHTLS-Wide survey is a 170 deg.$^2$ survey over four patches of sky, taken by the Canada-France-Hawaii Telescope. The patches are all far from the galactic plane in order to minimize extinction and bright star contamination. Photometric redshifts were prepared by Hildebrandt et al. (2009) using the publicly-available code BPZ (Benítez, 2000), with accuracy $\sigma_z \sim 0.03$ for $I < 24$. Lensing-quality shear measurements for galaxies within the survey are currently in preparation.

We've been able to do some preliminary analysis on the CFHTLS-Wide fields currently available using the P3 algorithm. We used the following settings for this run:

- $r_{\mathrm{ap}} = 0.5$ Mpc. The larger aperture size seemed to work better for the D2 field.
- $I < 22.5$. Going fainter showed no benefit, and possibly some harm.
- S/N $> 3$. This cut provided the best balance of purity and number of groups detected.

With these settings, we detected a total of 6038 groups over the 78 fields available, giving an average of 77 groups/deg$^2$. We've plotted the distribution of groups over redshift in Fig. 4.1. Although our low completeness results in the normalization of this plot not matching previous results, such as those from Milkeraitis et al. (2009), the shape of the distribution is consistent.

Milkeraitis et al. (2009)'s results arrived too late to run extensive comparisons, but we have been able to compare the results of our algorithm to Lu et al. (2009)'s cluster catalog, finding a purity of 54%. Although this result is low compared to our previous results, this is to be expected. As Lu et al.'s catalogue was derived through searching for galaxies on the red sequence, it consists primarily of clusters of 10 or more large

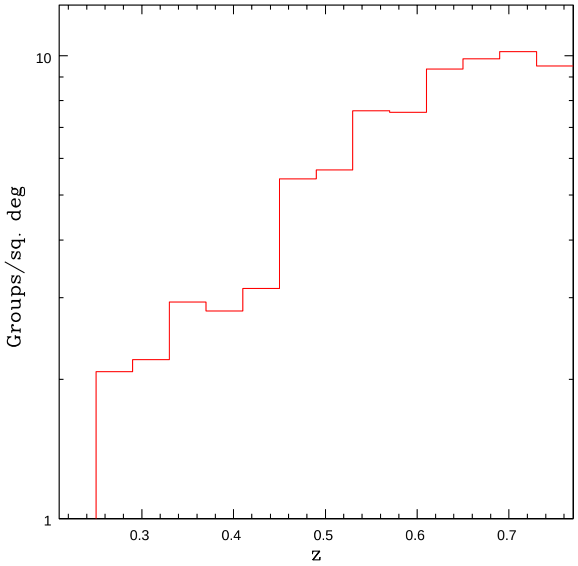


Figure 4.1: Histogram of the number of groups detected over redshift. Vertical axis has been scaled to give groups per deg.$^2$

Table 4.1: Summary of the purity and completeness of the P3 algorithm applied to the available CFHTLS-Wide fields with $r_{\mathrm{ap}} = 0.5$ Mpc to the cluster catalogue provided by Lu et al. Our catalogue shows an above random correlation to Lu et al.'s catalogue, with approximately 40% completeness, comparable to the completeness achieved at the high-mass end of groups in our simulated data. Columns are as Table 3.1, with $\langle N_{\mathrm{m}} \rangle$ referring to the average number of red sequence galaxies in detected and matched groups.

| Cut | Nhit | Ntot | p | C | $\langle N_{\mathrm{m}} \rangle$ |
|---|---|---|---|---|---|
| rand | 971 | 8661 | 0.112 | 0.348 | 6.578 |
| S/N > 2 | 1018 | 2256 | 0.451 | 0.464 | 7.715 |
| S/N > 3 | 914 | 1701 | 0.537 | 0.408 | 7.951 |
| S/N > 4 | 829 | 1225 | 0.677 | 0.298 | 9.007 |
| none | 1284 | 3039 | 0.423 | 0.452 | 7.767 |

red galaxies, while our catalogue contains a large number of smaller groups that we would expect to not match to anything in Lu et al.'s catalogue, resulting in a decreased apparent purity. The important point is that our purity is significantly above the 11% attained with our control catalogue. Our completeness here is 40%, which was the completeness level we found when comparing our algorithm to groups in the Millennium simulation with more than 10 members. From this, we are confident that our algorithm is working properly on this dataset.

We've also been able to estimate the distribution of galaxy spectral types within each group, and these are plotted versus redshift in Fig. 4.2. Each galaxy was assigned a best-fit spectral type by the photometric redshift algorithm. To estimate the composition of each group, we used the weighted probability ($w_i p_i$, as in Equation (2.1)) of each galaxy within a transverse distance of $r_{\rm ap}$ as its contribution to the composition of the group. The sum of the weighted probabilities of all galaxies with a given spectral type was then our predicted number of galaxies with that type in the group. These type compositions were then normalized so that, for each group, the type compositions summed to 1.

The results of this analysis are consistent with past results, such as Balogh et al. (2004), showing that elliptical galaxies are more likely to be in groups than other types of galaxies.

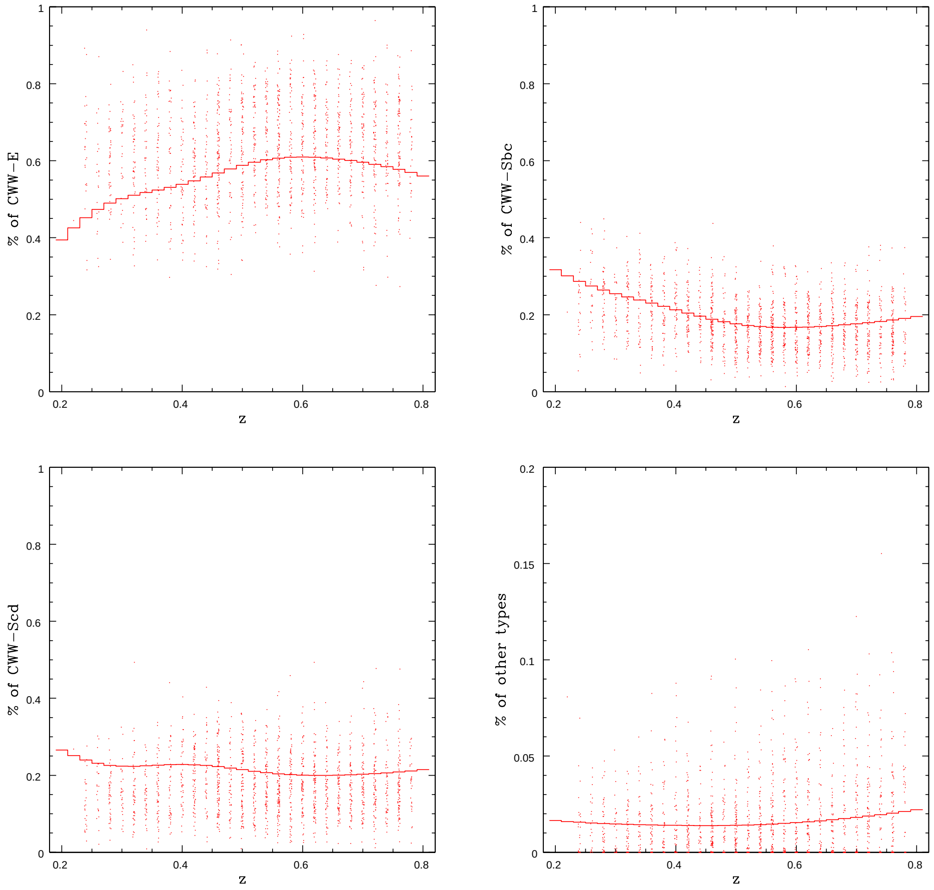


Figure 4.2: Plots showing the estimated percent of group members for each of six spectral types of galaxies. Solid line is the percent of all galaxies that are of each type at a given redshift, points are data for detected groups. As Im, SB2, and SB3 galaxies are uncommon at these redshifts, these spectral types have been combined into one plot. Redshifts have had a small random component added to aid viewing. These plots show little redshift dependence for the composition of groups. The Elliptical and Sbc galaxy-types show a weak dependence with redshift for $0.2 < z < 0.5$, while other types show no significant correlation with redshift. The composition of a typical group detected by our algorithm is $\sim 60\%$ Elliptical, $\sim 20\%$ Sbc, $\sim 20\%$ Scd, and $< 10\%$ Im, SB2, and SB3. This shows a small preference for Elliptical galaxies to be grouped and Scd galaxies to be found in groups.

# Chapter 5

# Conclusion

In this work we developed and tested the Photo-z Probability Peaks algorithm, a method for identifying a very pure sample of galaxy using photometric redshifts. We predict that the method will result in a purity of $\sim 94\%$ for the quality of photometry present in the CFHTLS-Wide survey. Running our algorithm on the available fields, we detected an average of 77 groups per square degree field in the redshift range $0.2 < z < 0.8$, which will give a predicted 13000 groups once photometry from the entire, 170 square degree survey is available, assuming the masked portion of the remaining fields doesn't differ significantly from that of the present fields. This presents a significant improvement over the work of Parker et al. (2005), which used only a sample of 192 groups to analyze the shape of the M/L curve. From simulated data, we estimate that our groups have an average membership of $\sim 6$ bright ($I < 22.5$) galaxies.

The P3 algorithm shows a limited ability to estimate the size of detected groups through their local $\delta$. There is a positive correlation between the number of members in a group and its $\delta$, although the significantly larger number of smaller groups makes a direct estimate of the number of members impractical. An alternative method, such as a Friends-of-Groups algorithm which uses our detected group centres, will be necessary to estimate the membership of detected groups.

Our detected distribution of groups over redshift is consistent with past results (Milkeraitis et al., 2009), as are our estimates of the spectral types of galaxies that compose these groups. Our group catalogue of the presently-available CFHTLS-Wide fields is consistent with Lu et al. (2009)'s catalogue, to the extent that would be expected given the differences in our methods.

For the next step of our research, we'll need to develop a method to estimate the specific membership of groups. This will be necessary in order to calculate the luminosities of groups for the purpose of M-L calculations. A Friends-of-Groups algorithm, which uses our group centres and then searches for galaxies that are likely members of those groups, will likely work well enough. The method will require testing against simulated data to determine the best matching parameters and the amount of background contamination that is to be expected.

Once we have our group catalogue with estimates of the groups' membership and size, we can proceed to apply weak lensing calculations to the catalogue. We will bin the data by luminosity, the number of bins to be determined by the size of the lensing errors, and make mass estimates for each luminosity bin. This will allow us to fill in more data points on a plot such as Fig. 1, which will help to illustrate the true nature of the M-L curve. We hope that this will lead to insight into which factor is responsible for the decrease in star formation among galaxies in groups.

In addition to this, we plan to use gravitational lensing to see what we can learn about the distribution of mass within galaxy groups. It is commonly accepted that galaxies have separate dark matter halos in loose groups and live in a shared halo in clusters, but it is unknown at what group size this transition takes place. We can attempt to determine this by binning groups by richness, then applying weak lensing to galaxies that lie directly in the background of groups. We can compare two lensing signals: The lensing signal relative to the center of the group, and the stacked lensing signal relative to the galaxy positions. The relative strengths of these signals will help us determine how dark matter is distributed within groups.

The P3 algorithm promises to be a strong tool for identifying high-purity samples of galaxy groups using photometric redshifts. It is capable of accurately identifying samples of poor groups, which no other group-finding algorithm is able to accomplish with photo-zs. This will allow us to perform precision measurements such as lensing with minimal fear of contamination.

# Bibliography

Filipe B. Abdalla, Manda Banerji, Ofer Lahav, and Valery Rashkov. A Comparison of Six Photometric Redshift Methods Applied to 1.5 Million Luminous Red Galaxies. *ArXiv Astrophysics e-prints*, December 2008.

M. L. Balogh, I. K. Baldry, R. Nichol, C. Miller, R. Bower, and K. Glazebrook. The Bimodal Galaxy Color Distribution: Dependence on Luminosity and Environment. *ApJL*, 615:L101–L104, November 2004. doi: 10.1086/426079.

N. Benítez. Bayesian Photometric Redshift Estimation. *ApJ*, 536:571–583, June 2000. doi: 10.1086/308947.

Bradley W. Carroll and Dale A. Ostlie. *An Introduction to Modern Astrophysics*. Addison-Wesley, 2007.

CFHTLS. Terapix - CFHTLS-Wide, 2009. URL `http://terapix.iap.fr/cplt/oldSite/Descart/summarycfhtlswide.html`.

Adrian A. Collister and Ofer Lahav. ANNz: estimating photometric redshifts using artificial neural networks. *Publ.Astron.Soc.Pac.*, 116:345–351, 2004.

Darren J. Croton, Volker Springel, Simon D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon.Not.Roy.Astron.Soc.*, 365:11–28, 2006.

Gabriella De Lucia and Jeremy Blaizot. The hierarchical formation of the brightest cluster galaxies. *Mon.Not.Roy.Astron.Soc.*, 2007:2–14, 2007.

Michael D. Gladders and H. K. C. Yee. The Red-Sequence Cluster Survey. *The New Era of Wide Field Astronomy, ASP Conference Series*, 232:126, 2001.

H. Hildebrandt, J. Pielorz, T. Erben, L. van Waerbeke, P. Simon, and P. Capak. CARS: The CFHTLS-Archive-Research Survey II. Weighing dark matter halos of Lyman-break galaxies at z=3-5. *Astronomy and Astrophysics*, 498:725–736, March 2009.

O. Ilbert, S. Arnouts, H. J. McCracken, M. Bolzonella, E. Bertin, O. Le Fevre, Y. Mellier, G. Zamorani, R. Pello, A. Iovino, L. Tresse, D. Bottini, B. Garilli, V. Le Brun, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodeggio, G. Vettolani, and A. Zanichelli ... Accurate photometric redshifts for the CFHT Legacy Survey calibrated using the VIMOS VLT Deep Survey. *Astronomy and Astrophysics*, 457:841–856, 2006.

O. Ilbert, P. Capak, M. Salvato, H. Aussel, H. J. McCracken, D. B. Sanders, N. Scoville, J. Kartaltepe, S. Arnouts, E. Le Floc'h, B. Mobasher, Y. Taniguchi, F. Lamareille, A. Leauthaud, S. Sasaki, D. Thompson, M. Zamojski, and G. Zamorani ... COSMOS Photometric Redshifts with 30-bands for 2-deg2. *The Astrophysical Journal*, 690: 1236–1249, September 2009.

Manfred G. Kitzbichler and Simon D. M. White. The high redshift galaxy population in hierarchical galaxy formation models. *Mon.Not.Roy.Astron.Soc.*, 376:2–21, 2007.

C. Knobel, S. J. Lilly, A. Iovino, C. Porciani, K. Kovac, O. Cucciati, A. Finoguenov, M. G. Kitzbichler, C. M. Carollo, T. Contini, J. P. Kneib, O. Le Fevre, V. Mainieri, A. Renzini, M. Scodeggio, G. Zamorani, S. Bardelli, M. Bolzonella, A. Bongiorno, K. Caputi, and G. Coppa ... An optical group catalogue to z = 1 from the zCOSMOS 10k sample. *The Astrophysical Journal*, 697:1842–1860, March 2009.

Benjamin P. Koester, Timothy A. McKay, James Annis, Risa H. Wechsler, August E. Evrard, Eduardo Rozo, Lindsey Bleem, Erin S. Sheldon, and David Johnston. MaxBCG: A Red Sequence Galaxy Cluster Finder. *The Astrophysical Journal*, 660: 221–238, 2009.

Andrey V. Kravtsov, Oleg Y. Gnedin, and Anatoly A. Klypin. The tumultuous lives of galactic dwarfs and the missing satellites problem. *Astrophys.J.*, 609:482–497, 2004.

I-hui Li and Howard K. C. Yee. Finding Galaxy Groups In Photometric Redshift Space: the Probability Friends-of-Friends (pFoF) Algorithm. *The Astronomical Journal*, 135: 809–822, January 2008.

Ting Lu, David G. Gilbank, Michael L. Balogh, and Adam Bognat. Recent Arrival of Faint Cluster Galaxies on the Red-sequence: Luminosity Functions from 119 square degrees of CFHTLS. *Monthly Notices of the Royal Astronomical Society*, 399:1858–1876, May 2009.

C. Marinoni and M. J. Hudson. The Mass-to-Light Function of Virialized Systems and the Relationship between Their Optical and X-Ray Properties. *ApJ*, 569:101–111, April 2002.

M. Milkeraitis, L. Van Waerbeke, C. Heymans, H. Hildebrandt, J. P. Dietrich, and T. Erben. 3D-Matched-Filter Galaxy Cluster Finder I: Selection Functions and CFHTLS Deep Clusters. *ArXiv Astrophysics e-prints*, December 2009.

J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490:493–+, December 1997.

T. Padmanabhan. *Structure Formation in the Universe.* Cambridge University Press, 1993.

L. C. Parker, M. J. Hudson, R. G. Carlberg, and H. Hoekstra. Mass-to-Light Ratios of Galaxy Groups from Weak Lensing. *ApJ*, 634:806–812, December 2005. doi: 10.1086/497117.

M. Postman, L. M. Lubin, J. E. Gunn, J. B. Oke, J. G. Hoessel, D. P. Schneider, and J. A. Christensen. The Palomar Distant Cluster Survey: I. The Cluster Catalog. *Astron.J.*, 111:615, 1996.

William H. Press and Paul Schechter. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *Astrophysical Journal*, 187:425–438, 1974.

Vicent Quilis, Ben Moore, and Richard Bower. Gone with the wind: the origin of S0 galaxies in clusters. *Science*, 288:1617–1620, 2000.

G. B. Rybicki and A. P. Lightman. *Radiative Processes in Astrophysics.* Wiley-VCH, 1986.

P. Schneider, C. S. Kochanek, and J. Wambsganss. *Gravitational Lensing: Strong, Weak and Micro.* Springer, 2006.

Volker Springel, Simon D. M. White, Adrian Jenkins, Carlos S. Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, John A. Peacock, Shaun Cole, Peter Thomas, Hugh Couchman, August Evrard, Joerg Colberg, and Frazer Pearce. Simulating the joint evolution of quasars, galaxies and their large-scale distribution. *Nature*, 2005:629–636, 2005.

Simon D. M. White. Formation and Evolution of Galaxies: Les Houches Lectures. *ArXiv Astrophysics e-prints*, 2009.