

DOCTORAL THESIS IN PHYSICS

---

**Precise photometric redshifts with  
narrow-band filters, quality cuts  
and their impact on the measured  
galaxy clustering**

---

Pol Martí Sanahuja

*Supervisors:*

Ramon Miquel Pascual

Enrique Fernández Sánchez

Universitat Autònoma de Barcelona

Institut de Física d Altes Energies

Barcelona March 18, 2014



*“Simplicity is the ultimate sophistication.”*

Leonardo da Vinci



# Acknowledgements

Many thanks to:

- Ramon Miquel for accepting to be my thesis adviser;
- Enrique Fernández for accepting to be my thesis co-adviser;
- the members of the thesis committee for their time;
- my office mates, Carles Sánchez and Lluís Galbany, for joining me during this time;
- Enrique Gaztanaga, Francisco J. Castander, Anne Bauer, Martin Eriksen, Linda Ostman, Ricard Casas, Jorge Carretero, Stephanie Jouvel and Laia Cardiel for their help;
- the PAU@WHT team, for being very enjoyable;
- my partner, Sandra Aguilera, for being patient with me;
- and, my family and friends, for their support.

This work was partially supported by the Spanish Ministerio de Economía y Competitividad (MINECO) under projects AYA2009-13936, FPA2012-39684, and Consolider-Ingenio 2010 CSD2007-00060.



# Contents

<b>Acknowledgements</b>	<b>5</b>
<b>1 Introduction</b>	<b>9</b>
<b>2 Framework</b>	<b>11</b>
2.1 Modern cosmology . . . . .	11
2.2 Photometry . . . . .	23
2.3 Photometric Redshifts (Photo- $z$ s) . . . . .	29
2.4 Galaxy Redshift Surveys . . . . .	36
<b>3 Precise photo-<math>z</math>s with narrow-band filters: The PAU@WHT Survey</b>	<b>41</b>
3.1 Default filter set-up . . . . .	42
3.2 The mock catalog . . . . .	45
3.3 Photo- $z$ performance . . . . .	50
3.4 Optimization of the PAU filter set . . . . .	65
3.5 Discussion and Conclusions . . . . .	71
<b>4 Photo-<math>z</math> quality cuts and their impact on the measured galaxy clustering</b>	<b>73</b>
4.1 Data Samples . . . . .	74
4.2 Photometric redshifts . . . . .	77
4.3 Galaxy clustering and the effect of the photo- $z$ quality cuts . . . . .	83
4.4 Correcting the effect of photo- $z$ quality cuts on galaxy clustering . . . . .	89
4.5 Effects on the extraction of the BAO scale . . . . .	94
4.6 Discussion and Conclusions . . . . .	98
<b>5 Summary and Outlook</b>	<b>101</b>
<b>Bibliography</b>	<b>105</b>



# Chapter 1

## Introduction

Since ancient times humankind has wondered about the nature of the night sky and its celestial objects, to which it always has attributed a mystical or religious character. However, with the advent of the Modern era, and in turn the scientific method, we gradually realized that the laws of nature that govern everyday phenomena also explain the celestial world. Stars are objects just like the Sun but at much greater distance, and planets are worlds like Earth, also trapped in the Sun's gravitational field. Early last century, it was thought that the Universe was basically composed of stars uniformly spread out over space, which had occupied the same place since ever. Later, astronomers realized that in fact stars were grouped into large islands called galaxies and these again were grouped forming galaxy clusters and superclusters placed between huge voids. This is known as the large-scale structure of the Universe. Moreover, these galaxies seemed to be moving away from each other the further the faster, as if the space between them was growing. This led to the undeniable conclusion that the Universe is expanding, and it had not always been like now. In fact, about 13,700 million years ago, everything should have been confined in a much smaller space that eventually expanded and evolved into the Universe we know today. This is referred to as the theory of the Big Bang, which gave birth to a new branch of Astronomy called Cosmology, studying the matter-energy content and evolution of the Universe as a whole. A brief introduction to the astronomical and cosmological concepts necessary to understand this work is presented in the first half of chapter 2.

Einstein's gravitational field equations predict that the evolution of the Universe and the growth of the large-scale structure are directly related to its matter-energy content. Nowadays, by measuring the evolution of the 3D position of a large number of distant galaxies, together with other cosmological probes cosmologists have determined that only about five percent of the content of the universe is ordinary matter (molecules, atoms, plasma, neutrinos and photons) while the rest should be in two types of matter-energy whose origin and properties are unknown; Dark Matter and Dark Energy. This presents a challenge for modern Cosmology. There are several theories that attempt to respond to it, however many more galaxy surveys are needed to reach the necessary precision to rule out or confirm some of them. Spectroscopic surveys (2dF, Colless et al. (2001); VVDS, Le Fèvre et al. (2005); WiggleZ, Drinkwater et al. (2010); BOSS, Dawson et al. (2013)) provide a 3D image

of the galaxy distribution in the near universe, but most of them suffer from limited depth, incompleteness and selection effects. Imaging surveys (SDSS, York et al. (2000); PanSTARRS, Kaiser, Tonry & Luppino (2000); LSST, Tyson et al. (2003); DES, ) solve these problems but, on the other hand, do not provide a true 3D picture of the universe, due to their limited resolution in the position along the line of sight, which is obtained measuring the galaxy redshift through photometric techniques using a set of broad-band filters, a technique known as photometric redshift (photo- $z$ ). The Physics of the Accelerated Universe (PAU) survey at the William Herschel Telescope (WHT) in the Roque de los Muchachos Observatory (ORM) in the Canary island of La Palma (Spain) will use narrow-band filters to try to achieve a quasi-spectroscopic precision in the redshift determination that will allow it to map the large-scale structure of the universe in 3D using photometric techniques, overcoming the limitations of spectroscopic surveys (Benítez et al., 2009). A detailed explanation of photo- $z$  methods and an overview of the galaxy redshift surveys used in this work are given in the second half of chapter 2.

The photo- $z$  performance capabilities of a narrow-band filter system, like that of the PAU@WHT survey, constitute the main topic of this thesis and are detailed in chapter 3. The stringent photo- $z$  precision requirements of this survey, defined in Gaztanaga et al. (2012), can only be achieved by applying photo- $z$  quality cuts, that is, by not considering galaxies with unreliable photo- $z$  estimations. However, if these galaxies are not removed homogeneously over the sky, this can severely bias the large-scale structure measurements. Since photo- $z$ s depend mainly on the photometry whose quality depends on the atmospheric conditions of the observed region of the sky, the distribution of the photo- $z$  quality over it will not be homogeneous. Therefore, one must be careful when applying this kind of cuts. In chapter 4 we study the impact of these cuts on the angular clustering of a real luminous red galaxy based catalog (Collister et al., 2007) from the Sloan Digital Sky Survey (SDSS) and propose a method to correct for it.

Finally, in chapter 5, we conclude the thesis with a brief summary of the two previous chapters, how they relate to each other, and providing a first look at their implications and possible future studies.

# Chapter 2

## Framework

In this chapter we introduce the concepts and tools necessary to follow the next two chapters. The outline is as follows. In section 2.1 we present the  $\Lambda$ CDM cosmological model, the main probes supporting it, the concepts of cosmological redshift and distances, and a brief introduction to the inhomogeneous Universe. In section 2.2 we give basic notions underlying the science of photometry such as: fluxes, apparent and absolute magnitudes, photometric systems, etc. In section 2.3 we introduce the concept of photometric redshift, the main topic of this thesis, we explain its origin and usefulness on observational cosmology and enumerate different methods to compute it, emphasizing the BPZ method (Benitez, 2000) used in this thesis. Finally, in section 2.4 we present the three galaxy redshift surveys whose data are used in this thesis, when it is available (SDSS and 2dFGRS), or simulated, when it is not (PAU@WHT Survey).

### 2.1 Modern cosmology

Modern cosmology, or simply cosmology, is the study of the largest-scale structures and dynamics of the Universe as a whole and is concerned with fundamental questions about its formation, evolution, and ultimate fate. Cosmology as a science originated with the Copernican principle, which states that celestial bodies obey identical physical laws to those on Earth. During the 20th century, Cosmology developed widely with Albert Einstein's General Theory of Relativity, and with better astronomical observations of extremely distant objects. Currently, the most accepted cosmological model is called  $\Lambda$  Cold Dark Matter ( $\Lambda$ CDM).

#### 2.1.1 The geometry of the Universe

The cosmological principle asserts that the distribution of matter-energy in the universe at large scales ( $\gtrsim 100$ Mpc) is homogeneous and isotropic, so there is no preferred direction on the sky. According to the General Theory of Relativity, the unique metric compatible with this principle is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, whose space-time line element is:

$$ds^2 = -c^2 dt^2 + a(t)^2 dx^2 \tag{2.1}$$

where  $-cdt$  is the temporal contribution with  $c$  the speed of light and  $a(t)dx$  the spatial, which is the product of a time-dependent scale factor  $a(t)$  and a comoving space line element  $dx$ . Since only the product  $a(t)dx$  has a physical meaning of distance, the absolute value of  $a$  is arbitrary and it is usually taken equal to 1 today. If the scale factor grows in time ( $a > 0$ ), spatial distances will grow as well, so in that case, we will say that the Universe is expanding. The FLRW metric admits three different kinds of spatial curvature for the whole Universe: hyperbolic, flat and spherical. In the case that it is flat, which seems to be in good agreement with the current Cosmic Microwaves Background (CMB) radiation observations (Ade & Others, 2013, Komatsu & Others, 2009), it can be written in spherical coordinates  $r \theta$  as:

$$dx^2 \equiv dr^2 + r^2 d\Omega^2 \quad (2.2)$$

where  $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$  is the differential solid angle. From now on, we will consider that the Universe is spatially flat and all equations we will show will be valid under this assumption.

### 2.1.2 Cosmological Redshift

When light (or any electromagnetic radiation) coming from a distant celestial object, such as a galaxy, travels across the space of an expanding Universe, all its wavelengths suffer an elongation due to the growth of distances. This effect is called cosmological redshift and it is equivalent to the elongation of the wavelengths that any wave suffers when it comes from a source which is moving away (Doppler effect). It can be quantified as:

$$z \equiv \frac{\eta_{obs} - \eta_{em}}{\eta_{em}} \quad (2.3)$$

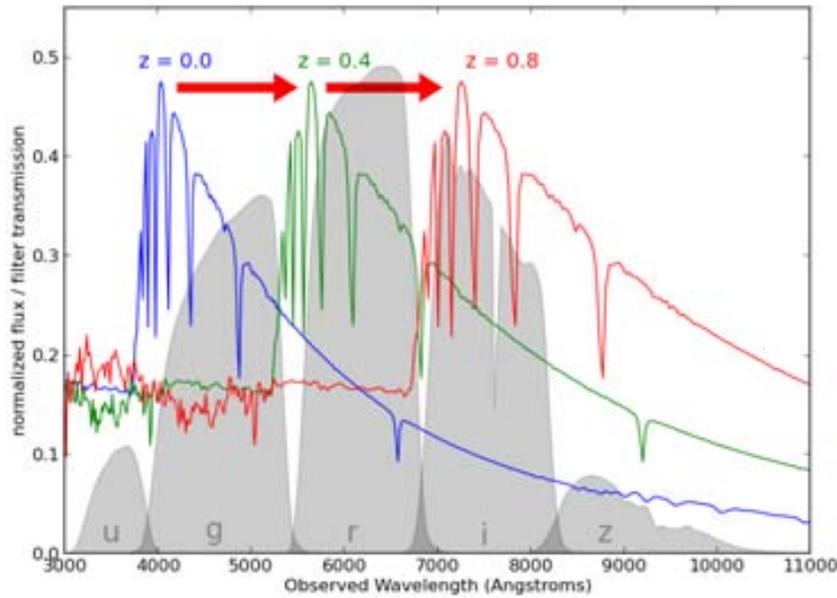
where  $\eta_{obs}$  is the wavelength of the light when is received by the observer and  $\eta_{em}$  when it was emitted by the source. From now on, we will simply refer to redshift when we actually refer to cosmological redshift. It is easy to show through the geodesic equation (movement equation) in a FLRW metric that there is a relation between the scale factor  $a$  at the time when light was emitted and how redshifted the light is today.

$$1 + z = \frac{1}{a(t)} \quad (2.4)$$

In Fig. 2.1 we show the spectrum of the star Vega ( $\alpha$ -Lyr) as reference at three different redshifts: 0, 0.4, 0.8 (note that such redshifts for a star in the Milkyway are not real, but only illustrative). As it is redshifted, its spectral features, such as the prominent absorption valleys, shift towards higher (redder) wavelength and they are stretched.

### 2.1.3 The evolution of the universe

Assuming that the matter-energy content of the universe at large scales can be described as a perfect fluid, which follows the state equation  $P = \xi$ , where  $P$  is the pressure,  $\xi$  the density and  $w$  is the state parameter of the fluid that relates



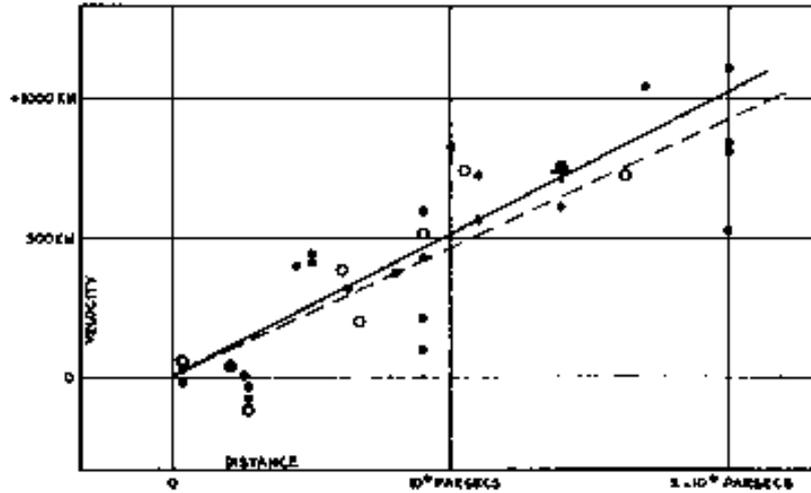
**Figure 2.1:** On the foreground, the spectrum of the star Vega ( $\alpha$ -Lyr) as a reference at three different redshifts: 0 (blue), 0.4 (green) and 0.8 (red). As it is redshifted, its spectral features, such as the prominent absorption valleys, move towards higher wavelength and they are stretched. The spectra have been normalized to have the same height. On the background, the throughput  $R(\lambda)$  of the SDSS broad-band photometric system *ugriz* (Fukugita et al., 1996). (Plot generated with the public script at [http://www.astroml.org/sklearn\\_tutorial/\\_downloads/plot\\_sdss\\_filters.py](http://www.astroml.org/sklearn_tutorial/_downloads/plot_sdss_filters.py).)

them, the gravitational field equation gives us the temporal evolution  $a(t)$  of the FLRW metric in its differential form called the Friedman equation:

$$H \equiv \frac{\dot{a}}{a} = H_0 \sqrt{\sum_{\omega} \Omega_{\omega} a^{-3(1+w)}}, \quad (2.5)$$

where  $H_0 \sim 70(km/s)/Mpc$  is the Hubble constant, which is the value of the so called Hubble parameter  $H$  today. The sum inside the root is over all the different perfect fluids that fill the universe, such as non-relativistic matter (dust, gas, cold dark matter<sup>1</sup>) ( $\omega \sim 0$ ), radiation ( $\omega = 1/3$ ), etc., and  $\Omega_{\omega} \equiv \rho/\rho_c$  are their densities relative to the critical density  $\rho_c \equiv 3H_0^2/8\pi G$  today, where  $G \sim 6.67 \cdot 10^{-11} N \cdot (m/kg)^2$  is the gravitational constant. Relative densities must fulfill  $\sum_{\omega} \Omega_{\omega} = 1$  in a flat Universe. We refer to the collection of parameters  $\{\Omega_{\omega}, H_0\}$  as the cosmology of the Universe, which determines the evolution and destiny of its space-time metric. Using the spatial part of the FLRW space-time line element at (2.1) and the definition of  $H$ , it can be deduced that the receding velocity  $v$  of a galaxy in an expanding universe ( $\dot{a} > 0$ ) is  $v = Hd$ , where  $d$  is the physical distance to the galaxy. Edwin Hubble proved, by measuring the distance and the receding velocity of 22 galaxies (see Fig. 2.2), that this is the case for our Universe (Hubble, 1929). Many times

<sup>1</sup>Cold dark matter (or CDM) is a hypothetical form of matter that interacts very weakly with electromagnetic radiation (dark) and most of whose particles move slowly compared to the speed of light (cold). Its presence can be only detected through its gravitational or weak interactions with ordinary matter and radiation.



**Figure 2.2:** Original scatter plot from [Hubble \(1929\)](#) of the velocity  $v$  (km/s) vs. distance  $d$  (pc) of 22 galaxies that proves that the universe is expanding. According to this, galaxies follow a lineal relation  $v = Hd$ .

cosmological parameters are given in terms of the normalized Hubble parameter  $h \equiv H_0 / 100 (\text{km s}^{-1} \text{Mpc}^{-1}) \sim 0.7$ . Also note that when  $c > Hd$ , where  $c$  is the speed of light, the receding galaxy becomes causally disconnected to the observer, which defines a causal horizon.

Besides the expansion rate of the universe given by  $H$ , another interesting quantity is the deceleration parameter  $q$  defined as:

$$q \equiv -\frac{\ddot{a}}{aH^2} = \frac{1}{2} \Omega (1 + 3w) \quad (2.6)$$

According to this definition, a universe basically made of ordinary matter ( $w \sim 0$ ) would have a deceleration parameter  $q > 0$ , that is, the expansion rate would be decelerating due to the gravitational attraction. However, any unknown component with  $w < -1/3$  would do the opposite, accelerate the expansion rate. In the next subsection, we will see that this is actually the case.

### 2.1.4 Cosmological distances

The radial comoving distance  $r$  that light traveled since it left the source until it arrives to us can be computed by equating the space-time line element in (2.1) to 0. Note that in spherical coordinates light only travels radially respect to us ( $d\Omega = 0$  and  $dx = dr$ ). Using the definition of the Hubble parameter in (2.5) and the relation between the redshift and the scale factor in (2.4),  $r$  can be expressed as

$$r(z) = c \int_0^z \frac{dz}{H(z)} \quad (2.7)$$

where the integral goes from redshift 0, when light is observed, to redshift  $z$ , when it was emitted. If we want to know the equivalent current physical distance, we only have to multiply by the scale factor today.

In astronomy, distances to celestial objects have been usually determined by measuring either its apparent brightness (flux  $F$ ) on the sky when its absolute luminosity  $L$  is well-known (standard candles), or the apparent size (angle subtended  $\theta$ ) on the sky when its actual size is also well-known (standard rulers). It is easy to see through geometrical and energy conservation arguments that the luminosity distance defined as follows is related to the redshift  $z$  as

$$D_L \equiv \frac{L}{4\pi F} = r(z)(1+z) \quad (2.8)$$

while the angular distance is

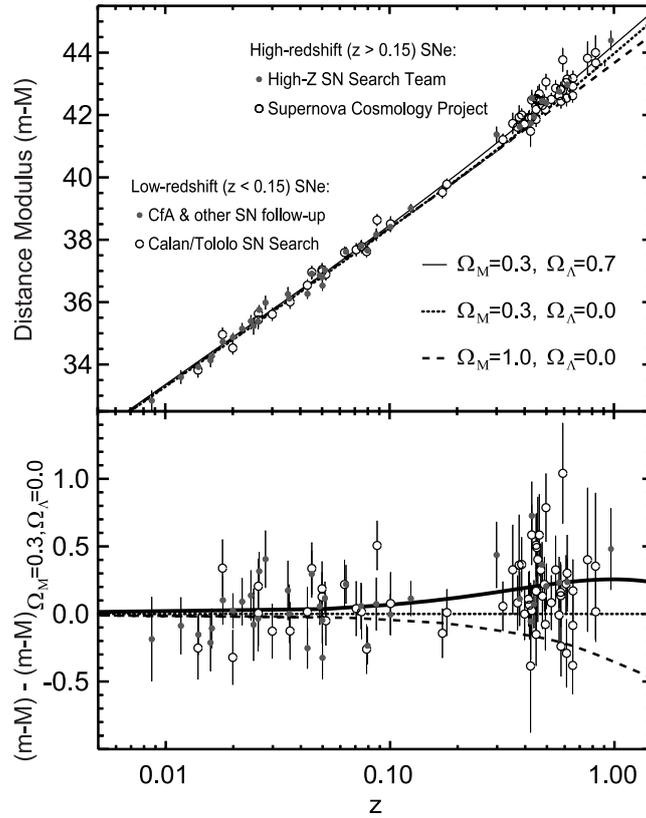
$$D_A \equiv \frac{r(z)}{\theta} = \frac{r(z)}{(1+z)} \quad (2.9)$$

These two relations are very important because they allow us to constrain the cosmology of our Universe by simply measuring the relation of two physical quantities such as the flux  $F$  or subtended angle  $\theta$  of standard candles or rulers respectively with its redshift  $z$ . Moreover, if we have enough redshift precision, we will be able to measure the radial size (the difference in redshift  $\delta z$  between the nearest and the farthest part) of a standard ruler that is aligned with the line of sight. With this, we see by differentiating (2.7) that we can directly measure the Hubble parameter as  $H(z) = c\delta z$ . If the standard ruler is perfectly spherical with diameter  $d$ , we can measure  $D_V(z)$  (Eisenstein et al., 2005), a hybrid distance that takes into account both contributions:

$$D_V \equiv \left( (1+z)^2 D_A^2 \frac{cz}{H(z)} \right)^{1/3} \quad (2.10)$$

Typical standard candles in observational cosmology are type Ia Supernovae (SNIa), which are very energetic nuclear explosions that occur when a white dwarf star in a binary system accretes material from the companion until surpasses the Chandrasekhar's limit. When they are produced in another galaxy, they can shine brighter than the host galaxy itself, so they can be seen at very large distances (or very high redshifts). Moreover, after some corrections, they all appear to have very similar absolute luminosities. In the late 1990s two independent teams, the Supernova Cosmology Project (Perlmutter et al., 1999) and the High- $z$  SN Search (Riess et al., 1998) measured the relation between the luminosity distance  $D_L$  of high- $z$  SNIa ( $z > 0.15$ ) and their redshift (see Fig. 2.3). They found out that (2.8) fits best the data when the cosmology is  $\Omega_{CDM} \sim 0.3$ ,  $\Omega_\Lambda \sim 0.7$ , where  $\Lambda$  is an unknown component of the universe, such as a cosmological constant or dark energy, with state parameter  $w_\Lambda \sim -1$ . According to this result the deceleration parameter  $q$  defined in (2.6) is negative, so the universe expansion rate is not decelerating, as was originally thought, but accelerating.

A typical standard ruler in observational cosmology is the Baryon Acoustic Oscillation (BAO). Back ( $z > 1100$ ) when Universe was denser and hotter ( $T \sim 13.6$  eV) enough that hydrogen atoms could not form, the plasma of baryons and photons continuously interacting through Thomson scattering ( $p + e^- \rightleftharpoons H + \gamma$ ) created a pressure on that medium through which acoustic waves traveled at the speed  $c_s = \frac{c}{\sqrt{3(1+R)}}$ , where  $R \equiv 3\xi_b/4\xi_\gamma$  and,  $\xi_b$  and  $\xi_\gamma$  are the densities

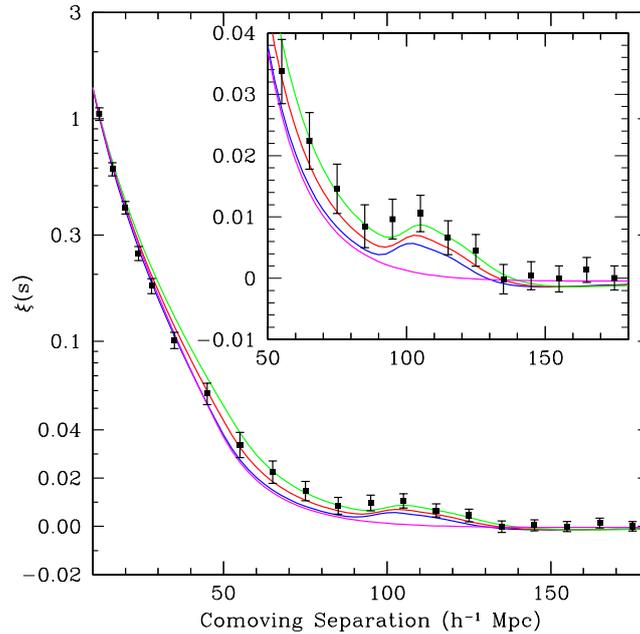


**Figure 2.3:** Top: The scatter plot of the distance modulus, defined in subsection 2.2.5 as  $m-M \equiv 5 \log_{10}(D_L / 10 \text{ pc})$ , vs. the redshift of low- $z$  SNIa ( $z < 0.15$ ) from the Calan/Tololo SN Search (Hamuy et al., 1993) and other SN surveys, and high- $z$  SNIa ( $z > 0.15$ ) from the High- $z$  SN Search (Riess et al., 1998) and the Supernova Cosmology Project (Perlmutter et al., 1999). Bottom: The residuals of the distances relative to a  $\Omega_{CDM} = 0.3$ ,  $\Omega_{\Lambda} = 0.0$  Universe. The best fit to data occurs with  $\Omega_{CDM} = 0.3$  and  $\Omega_{\Lambda} = 0.7$ , which implies that the Universe expansion rate is accelerating. Figure from Perlmutter & Schmidt (2003).

of baryons and photons respectively at that time. The Universe was expanded and cooled down until the recombination era when hydrogen atoms could start to form ( $p + e^- \rightarrow H + \gamma$ ), so the pressure on the medium vanished and traveling waves stalled at a comoving distance

$$r_d \sim c_s \int_{1100}^{\infty} \frac{dz}{H(z)} \sim 150 \text{ Mpc} \quad (2.11)$$

from the initial overdensity where they were created, where  $\infty$  denotes the beginning of the Universe. Additionally, photons started to move freely across space creating what we now observe as the Cosmic Microwave Background (CMB) radiation. Recent accurate measurements of the CMB with the Planck mission show that, in fact,  $r_d = 147.49 \pm 0.59 \text{ Mpc}$  at  $z_d = 1059.25 \pm 0.58$  (Ade & Others, 2013). Later, these secondary overdensities, as well as the initial ones, seeded the formation of galaxies resulting in an excess of galaxies separated by  $\sim 150 \text{ Mpc}$  from each other. Eisenstein et al. (2005) was the first to detect this preferred distance by measuring the 3D redshift-space correlation function  $\sigma(s)$  (black points in Fig. 2.4) with the



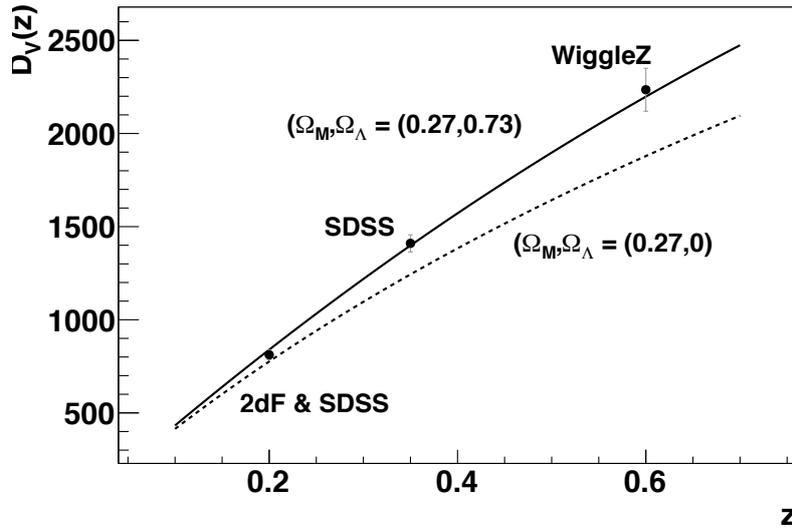
**Figure 2.4:** The large-scale redshift-space correlation function  $\xi(s)$  of the SDSS LRG spectroscopic sample (Eisenstein et al., 2005). Note that the vertical axis mixes logarithmic and linear scales. The inset shows an expanded view with a linear vertical axis. Points correspond to measured data, while lines correspond to different models. The BAO peak is clearly seen at  $100 \text{ Mpc}/h \sim 150 \text{ Mpc}$ . The magenta line is a model which does not include the BAO signature.

SDSS survey (see subsection 2.4.2) LRG<sup>2</sup> spectroscopic galaxy sample. Note, from Fig. 2.4, that besides the exponential decreasing shape of  $\sigma(s)$  there is a small bump at  $\sim 150 \text{ Mpc}$  corresponding to the BAO signature. During the last fifteen years, detections of BAO with other samples have been carried out. In Fig. 2.5 we show a Hubble diagram  $D_V$  vs.  $z$  for different measurements at different redshifts with the spectroscopic samples: 2dFGRS (Colless et al. (2001),  $z = 0.2$ ), SDSS (York et al. (2000),  $z = 0.2$  and  $z = 0.35$ ) and WiggleZ (Drinkwater et al. (2010),  $z = 0.6$ ). Two predicted curves for two different cosmologies,  $\Lambda\text{CDM} \equiv \Omega_{\text{CDM}} \sim 0.3 \Omega_{\Lambda} \sim 0.7$  (solid line) and a pure CDM Universe  $\Omega_{\text{CDM}} \sim 0.3 \Omega_{\Lambda} \sim 0$  (dashed), are also shown. Data are in agreement with the  $\Lambda\text{CDM}$  model, consistent with the SNIa results.

Nowadays, most of the observational cosmology is focused into determine precisely the cosmology of our Universe. Current cosmological probes agree that the Universe is very close to flat and basically made of cold dark matter ( $\rho_{\text{CDM}} = 0$ ) with relative density  $\Omega_{\text{CDM}} \sim 0.3$  and some unknown ingredient called Dark Energy ( $\rho_{\Lambda} \sim -1$ ) with  $\Omega_{\Lambda} \sim 0.7$ , responsible of the expansion rate acceleration. The Friedman equation for a Universe like this is reduced to:

$$H(z) = H_0 \sqrt{\Omega_{\text{CDM}}(1+z)^3 + \Omega_{\Lambda}} \quad (2.12)$$

<sup>2</sup>Luminous Red Galaxies (LRGs) are luminous intrinsically red galaxies consistent with large/giant elliptical galaxies.



**Figure 2.5:** The Hubble diagram  $D_V(z)$ , defined in (2.10), vs. redshift obtained by measuring the BAO signature in a combination of the transversal and radial directions with the spectroscopic samples: 2dF (Colless et al. (2001),  $z = 0.2$ ), SDSS (York et al. (2000),  $z = 0.2$  and  $z = 0.35$ ) and WiggleZ (Drinkwater et al. (2010),  $z = 0.6$ ). The data points agree with the prediction corresponding to a cosmology  $\Lambda$ CDM  $\equiv \Omega_{CDM} \sim 0.3$   $\Omega_\Lambda \sim 0.7$  (solid line). Figure from Astier & Pain (2012).

where  $a$  on the right of equality (2.5) has been replaced by (2.4). It is usually referred as the  $\Lambda$ CDM Universe. In this paradigm, ordinary matter (baryonic matter) would also account for  $\lesssim 5\%$  of the total content.

### 2.1.5 Large-Scale Structure (LSS)

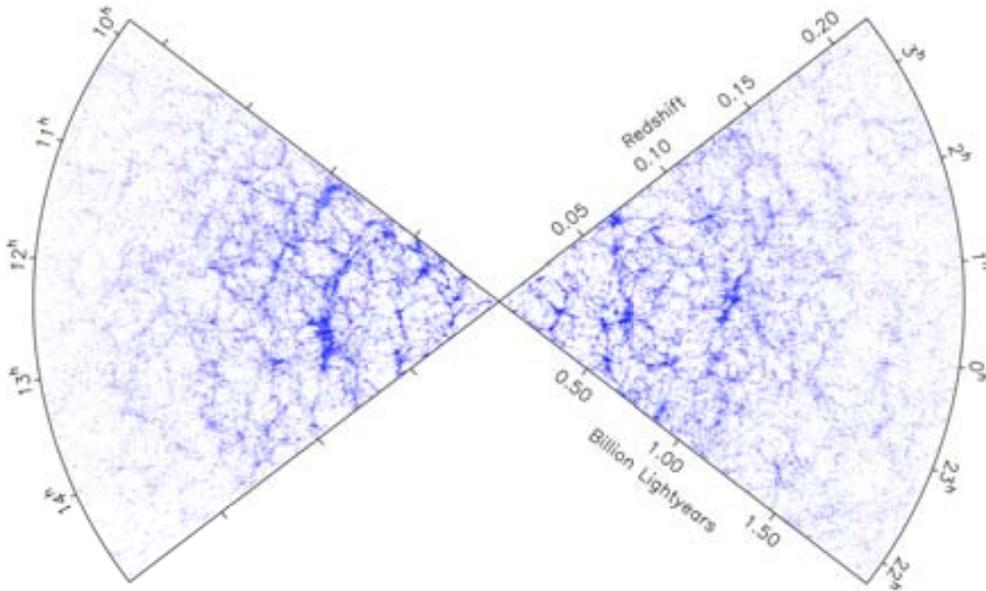
At scales  $< 100$  Mpc Universe can not be considered homogeneous since it presents structures such as galaxy clusters, voids and filaments, as shown in Fig. 2.6. At those scales the FLRW metric is no longer valid however, the Newtonian description of gravity is a good approximation. Consider that small deviations at different positions and times on the density  $\xi(x, t)$  of the perfect fluid can be quantified as:

$$\delta(x, t) \equiv \frac{\xi(x, t)}{\xi(t)} - 1 \quad (2.13)$$

which is called the density contrast, with  $\xi$  the average density. According to the Newtonian potential and mass conservation, the evolution in time of these fluctuations in an expanding universe is given by the second order harmonic differential equation:

$$\delta + 2H\dot{\delta} - \frac{3}{2}H^2\delta = 0 \quad (2.14)$$

where  $H$  is the Hubble parameter defined in (2.5). The solution,  $\delta(x, t) = D^\pm(t)\delta(x)$ , consists of a growing (+) and a decaying (-) mode, with a completely decoupled temporal and spatial parts. Since eventually the growing mode dominates, we simply



**Figure 2.6:** The distribution of  $\sim 0.2$  million galaxies (blue) observed with the 2dFGRS survey (Colless et al., 2001) in two slices 4 deg thick, centered at declination  $-2.5$  deg in the North Galactic Pole and  $-27.5$  deg in the South Galactic Pole. Although the distribution of matter at large scales ( $> 100$  Mpc) can be considered as homogeneous and isotropic, at lower scales galaxies form structures such as galaxy clusters, voids and filaments.

write the solution as

$$\delta(\vec{x}, z) = D(z)\delta(\vec{x}), \quad (2.15)$$

where  $D(z)$  ( $t$  has been replaced by  $z$  through (2.4)) is called the linear growth factor and gives the evolution in redshift (time) of the initial spatial perturbation  $\delta(\vec{x})$ . Clearly, the evolution of these perturbations and, therefore, the evolution of the large-scale structure is not only driven by the gravitational pull of the perturbations themselves, but also by the evolution of the space-time metrics of the whole Universe. This allows cosmologists trace to the story of  $a(t)$  and determine the cosmology of the Universe by observing how structures have formed and evolved. Since the formation of galaxies is seeded by cold dark matter overdensities (the most dominant component of matter), the easiest way to trace the distribution of cold dark matter is by looking at the three-dimensional position of galaxies on the sky. However, in Fry & Gaztanaga (1993) it was shown that there is a bias  $b$  between the fluctuations of the galaxy density contrast  $\delta_G$  and the CDM density contrast  $\delta$  defined as follows

$$\delta(\vec{x}, z)_G = b(z)\delta(\vec{x}, z), \quad (2.16)$$

where  $b$  can depend on the redshift.

Cosmologists are not actually interested in the value of  $\delta(\vec{x}, z)$ , but in their statistical properties. For this purpose, they compute the two-point correlation function in a given volume  $V$

$$\xi(\vec{r}) = \langle \delta(\vec{y})\delta(\vec{y} + \vec{r}) \rangle_V = \frac{1}{V} \int \delta(\vec{y})\delta(\vec{y} + \vec{r})dV_y, \quad (2.17)$$

which, assuming isotropy, gives the differential probability  $dP = \bar{\rho}^2(1+\xi(r))dV_ydV_{y+r}$  to find a fluctuation of size  $\delta(\vec{y})$  a distance  $r = |\vec{r}|$  away from another fluctuation of

size  $\delta(y+r)$ . Even more interesting is the Fourier transform of the 2-point correlation function  $\sigma(r)$ :

$$P(k) = \int \sigma(r) e^{ikr} d^3k \quad (2.18)$$

which is called the matter power spectrum. Note that the density contrast,  $\delta(x)$ , can also be Fourier transformed to  $\delta(k)$ , so that the matter power spectrum can be computed simply as  $P(k) = V \delta(k)^2$ . Most inflationary<sup>3</sup> theories predict a primordial density contrast whose amplitude  $\delta(r)$  is isotropic and homogeneously Gaussian distributed (observationally confirmed in Komatsu & Others (2009)), as well as a power spectrum  $P(k) \propto k$  (Harrison-Zeldovich power spectrum) that is scale-invariant. However, the shape of  $P(k)$  at low scales is strongly modified by the radiation dominated era ( $\Omega_{rad} = \Omega_{matter}$ ) where perturbations whose wavelength scale  $\eta$  was smaller than the causal horizon had a linear growth factor  $D(z)$  that remains constant until the matter dominated era due to the suppression caused by the pressure of radiation (see Dodelson (2003) for a more detailed explanation). This results in the power spectrum, shown in Fig. 2.7, which asymptotes to  $P(k) \sim k$  for small  $k$ , and behaves as  $P(k) \sim k^{-3}$  for large  $k$ , with a peak at  $k^* \sim 2 \times 10^{-2} h \text{ Mpc}^{-1}$  corresponding to  $\eta^* \sim 350 h^{-1} \text{ Mpc}$ .

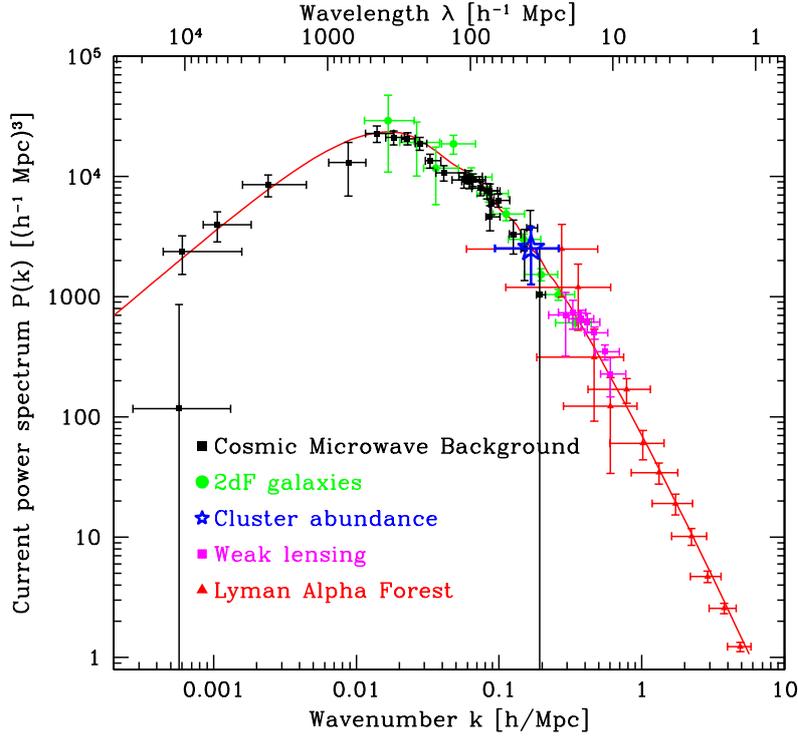
We typically measure distances to galaxies through the relation between their redshift and distance given by the Hubble law. However, peculiar radial velocities of galaxies not associated with the expansion flow can cause distortions on the measurement of the cosmological redshift. The most obvious example of this is the Fingers-of-God effect (Bahcall, Soneira & Burgett, 1986), where long thin filaments in the distribution of galaxies in redshift space point directly back at the observer due to the addition of the random peculiar velocities of galaxies, given by the virial theorem<sup>4</sup>, in the radial direction within a galaxy cluster. Another important redshift-space distortion is the Kaiser effect (Kaiser, 1984), which is more subtle and difficult to quantify. The Kaiser effect describes the peculiar velocities of galaxies bound to a central mass as they undergo infall. This differs from the Fingers-of-God in that the peculiar velocities are coherent, not random, towards the central mass. According to this effect, in the large-scale linear regime ( $\delta \ll 1$ ) and the plane-parallel approximation (where galaxies are taken to be sufficiently far away from the observer that the displacements induced by peculiar velocities are effectively parallel), the distortion caused by coherent infall velocities takes a particularly simple form in Fourier space:

$$\delta^s(k) = (b + f\phi_k^2)\delta(k) \quad (2.19)$$

where  $\phi$  is the cosine of the angle between  $k$  and the line-of-sight,  $b$  is the bias described in (2.16), the subscript  $s$  indicates redshift space, and  $f$  is the velocity

<sup>3</sup>Inflation is a cosmological model for the early universe that predicts an extremely rapid exponential expansion driven by a negative-pressure vacuum energy density. Initial quantum fluctuations become macroscopic on this process seeding the current inhomogeneities observed on the matter distribution.

<sup>4</sup>The virial theorem relates the average over time of the total kinetic energy of a stable system consisting of  $N$  particles, bound by potential forces, with that of the total potential energy.



**Figure 2.7:** Matter power spectrum  $P(k)$  versus wavenumber extrapolated to  $z = 0$ , from various measurements of cosmological structure. The best fit  $\Lambda$ CDM model is shown as a solid line. Figure from Tegmark & Zaldarriaga (2002).

growth factor defined as

$$f(z) \equiv \frac{d \ln D}{d \ln a} = \frac{\delta}{\delta} \equiv \Omega_{CDM}^\gamma(z) \quad (2.20)$$

and  $\gamma$ , the gravitational growth index (Linder, 2005).

In many cases what is measured is the angular galaxy-galaxy cross-correlation function of the galaxies at different redshifts. It can be predicted by considering the projected spatial galaxy contrast in a redshift bin  $i$

$$\delta_{G_i}(n) = \int dz N_i(z) \delta(n, z) \quad (2.21)$$

where  $N_i(z)$  is the redshift distribution in the redshift bin  $i$  and  $n$  is some unitary vector pointing to some direction on the sky. Then, the angular cross-correlation between two redshift bins  $i$  and  $j$  is

$$G_i G_j(\theta) \equiv \delta_i(n) \delta_j(n + \theta) = \int dz_1 N_i(z_1) \int dz_2 N_j(z_2) \sigma^s(r_{12}) \quad (2.22)$$

where  $\sigma^s(r_{12})$  is the redshift space correlation of the pairs of galaxies at redshift  $z_1$  and  $z_2$  subtending an angle  $\theta$  with the observer, which is related to the redshift-space dark matter correlation function  $\sigma^s(r)$  through

$$\sigma^s(r_{12}) \equiv \delta_g(n, z_1) \delta_g(n + \theta, z_2) = b(z_1) b(z_2) D(z_1) D(z_2) \sigma^s(r) \quad (2.23)$$

where  $b(z)$  is the bias at redshift  $z$  described in (2.16) and  $D(z)$  the linear growth factor described in (2.15). The function  $\sigma^s(r)$  is obtained by inverse Fourier transforming the power spectrum  $P^s(k) = V \delta^s(k)^2 = (b + f\phi_k^2)^2 P(k)$  as described in Hamilton (1992), where in the second equality we have used (2.19).

In the case that perturbations in the perfect fluid are not small  $\delta \sim 1$ , linear theory is no longer valid. Therefore, in order to study the evolution of these non-linear structures further, which are the seeds for galaxies and clusters, large numerical simulations are carried out. Starting from a Gaussian random field, a realization is evolved in time. The resulting dark matter structures can be investigated to find fitting formulas for the non-linear power spectrum, as done in Smith et al. (2003). These non-linear structures begin to be important in at  $k \gtrsim 0.2 h \text{ Mpc}^{-1}$  in the matter power spectrum, which at typical redshifts correspond to an angular scale of  $\theta \lesssim 0.1 \text{ deg}$ .

When two redshift bins  $i$  and  $j$ , one in the foreground with mean redshift  $z_i$  and the other on the background with mean redshift  $z_j$  ( $z_j > z_i$ ), are cross-correlated, the measured signal comes mainly from the galaxies in the overlap of these two redshift bins. However, a smaller contribution can come from the gravitational lensing effect that the foreground produces onto the background, which basically consists on changing the area and the brightness of the background sources behind the foreground lenses. This phenomenon is called weak-lensing magnification and it is characterized by inducing an extra term of fluctuations on the galaxy density contrast,  $\delta_G = \delta_G + \alpha \delta_\mu$ , where  $\alpha \equiv 2.5s - 1$  is the amplitude of the weak lensing magnification effect (with  $s \equiv d \log_{10} N(< m) / dm$ , the slope of the galaxy number counts  $N$  at the magnitude limit  $m$  of the galaxies in the bin). This extra term generates three extra terms on the galaxy-galaxy cross-correlation,  $\langle \delta_{G_i} + \alpha_i \delta_{\mu_i} \rangle \langle \delta_{G_j} + \alpha_j \delta_{\mu_j} \rangle = \langle \delta_{G_i} \delta_{G_j} \rangle + \langle \delta_{G_i} \delta_{\mu_j} \rangle + \langle \delta_{\mu_i} \delta_{G_j} \rangle + \langle \delta_{\mu_i} \delta_{\mu_j} \rangle$ . The first one is already shown in (2.22), while the other three involve weak-lensing magnification:

$$\langle \delta_{G_i} \delta_{\mu_j} \rangle(\theta) = \alpha_j \int dz_1 N_i(z_1) \int dz_2 N_j(z_2) b_i(z_1) p_{\mu_j}(z_2) \sigma^s(r_{12}) \quad (2.24)$$

$$\langle \delta_{\mu_i} \delta_{G_j} \rangle(\theta) = \alpha_i \int dz_1 N_i(z_1) \int dz_2 N_j(z_2) p_{\mu_i}(z_1) b_j(z_2) \sigma^s(r_{12}) \quad (2.25)$$

$$\langle \delta_{\mu_i} \delta_{\mu_j} \rangle(\theta) = \alpha_i \alpha_j \int dz_1 N_i(z_1) \int dz_2 N_j(z_2) p_{\mu_i}(z_1) p_{\mu_j}(z_2) \sigma^s(r_{12}) \quad (2.26)$$

where we have used the relation (2.16) for the bias  $b(z)$ , and  $p_\mu(z)$  is the efficiency of the weak-lensing magnification effect between the two galaxy bins:

$$p_{\mu_i}(z) = \frac{3\Omega_{CDM} H_0 r(z)}{2H(z) a(z) r_0} \int_z dz \frac{r(z; z)}{r(z)} N_i(z) \quad (2.27)$$

where  $r(z)$  is the radial comoving distance defined in (2.7),  $r(z; z)$  is the angular diameter comoving distance between  $z$  and  $z$ ,  $r_0 \equiv c H_0$ , and  $N_i(z)$  is the redshift distribution of the galaxies in redshift bin  $i$ .

## 2.2 Photometry

Photometry, in the context of astronomy, is the technique concerned with measuring the flux of an astronomical object's electromagnetic radiation over large wavelength intervals ( $\sim 100\text{nm}$ ). This is in contrast to spectroscopy in which the same electromagnetic radiation is split into small wavelength intervals ( $\sim 1\text{nm}$ ) with a chromatic disperser such as a prism. Nowadays, most of the astronomical photometry is carried out using Charge-Coupled Devices (CCDs) as detectors that count individually the number of photons coming from the celestial sources.

### 2.2.1 Photometric system

A photometric system is a set of well-defined passbands (or filters) with a known sensitivity to incident radiation. The first known standardized photometric system is the *UBV* photometric system (Johnson & Morgan, 1953), widely used for star classification, which is composed of three broad bands that cover the ultraviolet (U), blue (B) and visual (V) regions of the spectrum (from  $\sim 300\text{ nm}$  to  $\sim 650\text{ nm}$ ). At present, there are more than 200 photometric systems. One of the most famous is the *ugriz* system (Fukugita et al., 1996) used in the Sloan Digital Sky Survey (SDSS) instrument, whose five bands, shown in gray on Fig. 2.1, cover from  $\sim 300\text{ nm}$  to  $\sim 1000\text{ nm}$ . The three bands *ugr* are roughly equivalent to the *UBV* ones, and the remaining three reach into the infrared range, which is very useful to detect and measure light coming from distant galaxies whose spectrum has been substantially redshifted.

Photometric systems can be classified according to the widths of their passbands as: broad band, whose passbands are wider than  $\sim 30\text{ nm}$ , such as the *UBV* or the *ugriz* systems, and narrow bands, whose passbands are less than  $\sim 30\text{nm}$  wide, such as the PAU (subsection 2.4.3) or ALHAMBRA (Moles et al., 2008) photometric systems.

### 2.2.2 Flux and Apparent magnitude

The flux  $F$  of an astronomical object is the amount of energy as electromagnetic radiation that we receive from the object per unit area and time. If the electromagnetic radiation passes through a bandpass then the received flux is:

$$F = \int_0 f_\nu(\alpha) R_\nu(\alpha) d\alpha = \int_0 f_\lambda(\eta) R_\lambda(\eta) d\eta \quad (2.28)$$

where  $\alpha$  and  $\eta$  are the frequency and wavelength of the radiation respectively, which satisfy  $\eta\alpha = c$ ,  $f_\nu(\alpha)$  and  $f_\lambda(\eta)$  are the spectral density fluxes of the object, these are the fluxes per unit frequency and wavelength respectively, and  $R_\nu(\alpha)$  and  $R_\lambda(\eta)$  are the transmission profiles of the bandpass as a function of frequency and wavelength respectively, which give the probability that a photon with associated frequency  $\alpha$  or wavelength  $\eta$  passes through the filter.

Fluxes of different astronomical objects, or even of the same object but at different passbands, can differ by several orders of magnitude between them. The logarithmic response of the human eye is well adapted to this wide range of fluxes. In

fact, ancient astronomers classified stars according to how many times they seemed to shine with more intensity than other stars. This classification has persisted over time and has been formalized through:

$$m \equiv -2.5 \log \frac{F}{F_0} \quad F = F_0 10^{-0.4m} \quad (2.29)$$

where  $m$  is called the apparent magnitude of the object and  $F$  is the actual flux received from it. The zero subindex indicates that the value of the apparent magnitude is defined with respect to another object with flux  $F_0$ , which defines the zero point. Different zero points define different magnitude systems, such as, for example, the Vega system in which the star Vega is defined to have an apparent magnitude of zero as measured through all filters (that is,  $F_0$  is Vega's flux).

### 2.2.3 The AB magnitude system

Another widely used system is the AB magnitude system, which is not defined by flux ratios as in (2.29) but spectral density flux  $f_\nu(\alpha)$  ratios, in such a way that

$$m_{AB}^R \equiv -2.5 \log \frac{\int_0^\infty f_\nu(\alpha) R_\nu(\alpha) \frac{d\nu}{\nu}}{\int_0^\infty f_0^\nu R_\nu(\alpha) \frac{d\nu}{\nu}} = -2.5 \log \frac{\int_0^\infty f_\lambda(\eta) R_\lambda(\eta) (\eta/c) d\eta}{\int_0^\infty f_0^\lambda(\eta) R_\lambda(\eta) (\eta/c) d\eta} \quad (2.30)$$

where the reference flux is  $f_0^\nu = 3631 \text{ Jy}^5$  (Oke & Gunn, 1983), which is constant in terms of frequency  $\alpha$ , or  $f_0^\lambda(\eta) = (c/\eta^2) f_0^\nu$ , which depends on  $\eta$ ,  $c$  being the speed of light. In the second equality we have used the relations  $\alpha\eta = c$ ,  $d\alpha/\alpha = -d\eta/\eta$  and  $\alpha f(\alpha) = \eta f(\eta)$ . In the AB magnitude system a source with constant flux per unit frequency will have the same magnitude value in all bands. If we consider a bandpass infinitely narrow such that  $R(\alpha) = \delta(\alpha)$ , the previous definition becomes

$$m_{AB}(\alpha) = -2.5 \log f(\alpha) - 48.6 \quad (2.31)$$

### 2.2.4 Color index

In the same way that magnitudes quantify how brighter is a celestial object with respect to another, the color index quantifies how bluer or redder it is. It is defined as the difference between the magnitude  $m_a$  of the celestial object in one passband  $a$  with mean wavelength  $\eta_a$  and the magnitude  $m_b$  of the same celestial object in a different passband  $b$  with mean wavelength  $\eta_b$  such that  $\eta_a < \eta_b$ . According to the magnitude definition in (2.29), if  $m_a - m_b < 0$ , the ratio of fluxes between the celestial object  $F$  and the object of reference  $F_0$  in the two different bands fulfill  $(F/F_0)_a > (F/F_0)_b$ , so that more flux is coming from the bluer part of the object's spectrum with respect to the reference spectrum and, therefore, the object is bluer than the reference. For example, in the Vega photometric system, the Vega star has  $m_V = m_B = 0$ , so obviously its color index is  $B - V \equiv m_B - m_V = 0$ . Therefore, a star with  $B - V < 0$  is bluer than Vega while a star with  $B - V > 0$  is redder.

<sup>5</sup>1 Jy =  $10^{-23} \text{ erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1} \cdot \text{Hz}^{-1}$  or  $1.51 \cdot 10^7 \cdot \text{photons} \cdot \text{m}^{-2} \cdot \text{s}^{-1} \cdot d \log^{-1} \lambda$  in terms of photons.

Color indices are very useful in galaxy surveys to classify galaxies according to their color, which, in a similar way as stars, tell us about its activity, temperature and age. For example, red galaxies, also known as early-type galaxies or ellipticals, are older, with less star-formation activity and with a higher metallicity than bluer ones, which are known as late-type galaxies or starburst galaxies.

### 2.2.5 Absolute magnitude and K-correction

The absolute magnitude  $M$  is defined to be the apparent magnitude that a source would have if it were 10 pc away, at rest (i.e., not blue- or red-shifted), and compact. In [Hogg et al. \(1996\)](#) it is shown that its relation with the apparent magnitude is:

$$M = m - \phi - K \quad (2.32)$$

where  $\phi$  is the distance modulus

$$\phi \equiv 5 \log \left[ \frac{D_L}{10 \text{pc}} \right] \quad (2.33)$$

where  $D_L$  is the luminosity distance, defined in [\(2.8\)](#), to the source in parsecs, and the last term the  $K$ -correction

$$K \equiv -2.5 \log \left[ \frac{1}{1+z} \frac{\int_0^\infty f_\lambda(\eta) R(\eta) d\eta}{\int_0^\infty f_\lambda(\eta) R(\eta) (1+z) d\eta} \right] \quad (2.34)$$

which corrects for the fact that in an expanding universe the spectrum  $f(\eta)$  of a source at a certain distance is redshifted by  $1+z$  (so that  $f(\eta) = f(\eta(1+z))$  according to [\(2.3\)](#)) and therefore, the passband  $R(\eta)$  does not see the same part of the spectrum.

The absolute magnitude is a measure of the celestial object's intrinsic brightness. Note that the distance modulus  $\phi = m - M - K$  only depends on the redshift  $z$  and the profile of the object's spectrum  $f_\lambda(\eta)$  for a given cosmology  $(\Omega, H_0)$ . This is specially useful to constrain cosmology with standard candles such as SNIa (section [2.1.4](#)), since their absolute magnitude is known beforehand.

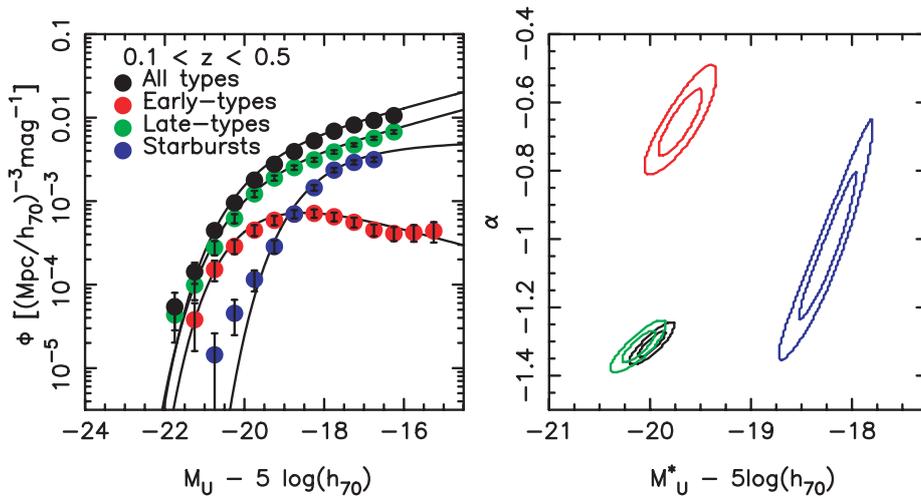
### 2.2.6 Luminosity function

A quantity closely related to the absolute magnitude  $M$  is the Luminosity Function (LF), which is the cumulative distribution of stars or galaxies up to a certain luminosity  $L$  (energy per unit time). The Schechter function ([Schechter, 1976](#)) provides a parametric description of the LF, which in differential form is

$$n(x)dx = \rho x^\alpha e^{-x} dx \quad (2.35)$$

$$\text{with } x \equiv \frac{L}{L^*} \quad (2.36)$$

Since  $L = F = 10^{-0.4M}$  we can write  $x$  in terms of the absolute magnitude,  $x = 10^{-0.4(M-M^*)}$ . The function is parametrized by  $(\rho, \alpha, M^*)$ , where  $\rho$  has units of number density and provides the normalization,  $\alpha$  is the power-law slope at high



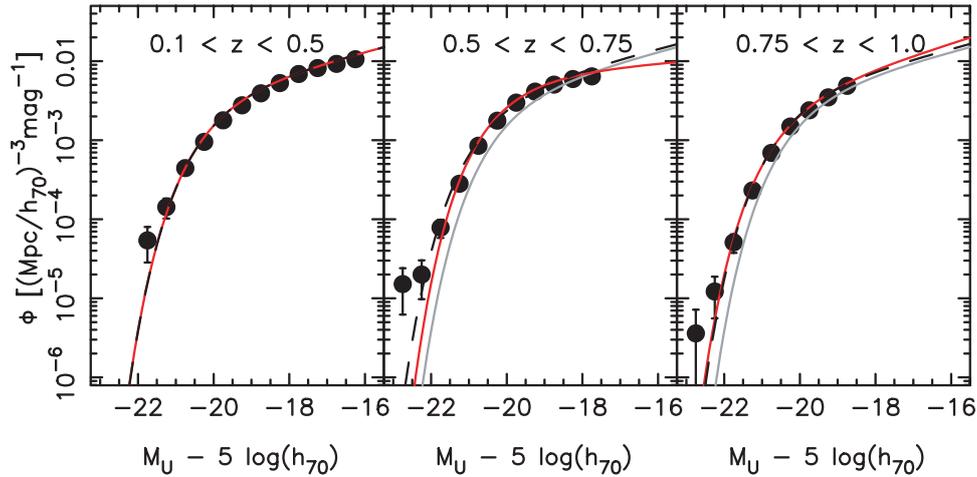
**Figure 2.8:** Left: Rest-frame U-band luminosity function in the redshift range  $0.1 < z < 0.5$ , showing the composite LF (black circles) and the type-specific LFs for early-type (red circles), late-type (green circles), and starburst (blue circles) galaxies. The types are determined from best-fitting spectral templates. Right: The 68.3% and 95% error ellipses for  $M^*$  and  $\alpha$  from Schechter function fits to different populations. Figures from [Dahlen et al. \(2005\)](#).

magnitudes and  $M^*$  the characteristic absolute magnitude where the power-law profile cuts off.

The galaxy LF may have different parameters for different populations and environments; it is not a universal function. In Figs. 2.8 and 2.9 we show examples of LF in the  $U$ -band calibrated with data from the Hubble Space Telescope (HST) and the Great Observatories Origins Deep Survey (GOODS) ([Dahlen et al., 2005](#)). They will be used in chapter 3 in order to create galaxy mock catalogs for the PAU survey (subsection 2.4.3). While in Fig. 2.8 galaxies have been grouped in different spectral types such as early-type (Ellipticals), late-type and starburst (spirals and irregulars) galaxies, in Fig. 2.9 we show the evolution in redshift of the composite (LF) through three different  $z$ -bins:  $0.1 < z < 0.5$ ,  $0.5 < z < 0.75$  and  $0.75 < z < 1.0$ .

## 2.2.7 Measuring magnitudes

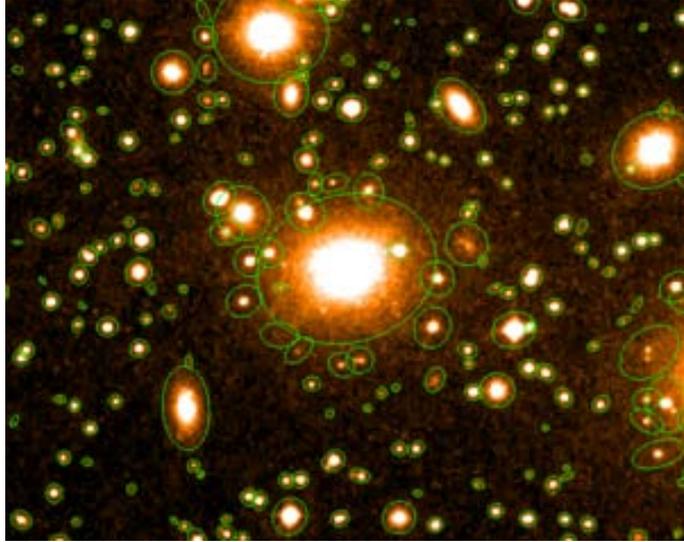
We mentioned that photometry is a technique concerned with measuring the flux of an astronomical object's electromagnetic radiation. In fact, astronomers do not measure fluxes directly, but the number of incident photons  $N_{obj}$  that the telescope collects in a certain period of time. These photons are detected and counted by a CCD (Charge-Coupled Device) imaging camera. CCDs were invented in 1969 at AT&T Bell Labs by Willard Boyle and George E. Smith (Nobel Laureates in Physics in 2009), and rapidly began to be widely used for imaging. A CCD is a two-dimensional matrix of pixels that convert infalling photons into electric charge through the photo-electric effect. Then, by applying weak electrical fields, one can move this charge throughout the pixels until it reaches the edge of the matrix where the cumulated charge is measured and digitalized. In observational astronomy, the



**Figure 2.9:** Rest-frame U-band luminosity function for the redshift intervals  $0.1 < z < 0.5$  (left),  $0.5 < z < 0.75$  (middle), and  $0.75 < z < 1.0$  (right). Solid red lines show the best-fit Schechter function, while dashed black lines show the best-fitting Schechter function for which the faint-end slope,  $\alpha$ , is fixed to the value measured in the lowest redshift bin. For comparison, we show with a gray line in the mid- and high-redshift bins the best-fitting Schechter function found in the low-redshift bin. Figures from [Dahlen et al. \(2005\)](#).

Focal Plane (FP) of the telescope where the images of the objects are formed, is covered with several CCDs. Usually the image of a celestial object covers several pixels, even if it is a point source such as a star or a quasar. The focal plane scale  $\alpha$  translates the angular size of a portion of the sky to the number of pixels the image occupies. The integrated flux of the source can be computed directly by summing up the photon counts in a certain region over the FP that encloses the object image, as shown with the green circles in Fig. 2.10. However, often it is difficult to know where the limits of the object are, so on those cases more sophisticated ways are used, such as integrating a model for the 2D-image profile whose parameters are obtained by fitting it previously to the image. Focusing on the photometry of the SDSS survey (subsection 2.4.2), which will be used in chapter 4, and depending on how they are measured, several magnitudes are defined:

- **PSF magnitudes:** For ground based optical telescopes, point source targets, such as stars or quasars, appear on the FP with a certain size due to aberrations introduced by atmospheric turbulence (seeing), diffraction caused by the optics, etc. On those cases, the 2D distribution of light of the image is well described by the Point Spread Function (PSF), which usually would be an Airy pattern from diffraction theory. However, in most cases it can be approximated by a Gaussian profile.
- **Model magnitudes:** Just as the PSF magnitudes are optimal measures of the fluxes of stars, the optimal measure of the flux of a galaxy would use a matched galaxy model. Typically, two models are fitted to the 2D image of



**Figure 2.10:** Illustration of how a software for source extraction, such as the SExtractor (Bertin & Arnouts, 1996), detects galaxies from an image and selects an effective area (green circles) around them where the flux will be integrated.

each object: a pure De Vaucouleurs profile (de Vaucouleurs, 1948)

$$I(r) = I_0 e^{-7.67(r/r_e)^{1/4}} \quad (2.37)$$

and a pure exponential profile

$$I(r) = I_0 e^{-1.68(r/r_e)} \quad (2.38)$$

where  $I(r)$  is the brightness at a distance  $r$  from the center of the image,  $I_0$  is the brightness at the center and  $r_e$  an effective radius. The model magnitudes in all the bands will be computed by applying the model (exponential or De Vaucouleurs) of higher likelihood in some reference band filter (the  $r$ -band in SDSS) and then the same model with the same  $r_e$  to the rest of the bands, allowing only the amplitude to vary. This procedure ensures that the color index between different bands is measured through a consistent aperture.

- **De Vaucouleurs magnitudes:** The De Vaucouleurs profile is specially good to describe the brightness of elliptical galaxies, therefore the best estimate of the magnitudes for these kind of galaxies is obtained by fitting a pure De Vaucouleurs profile to their image in each band independently. The value of the  $r_e$  parameter in some band is called the De Vaucouleurs radius in that band, and is specially useful for galaxy-star separation in catalogs of elliptical galaxies, since they present a  $r_e$  value substantially larger than stars.
- **Petrosian, Fiber, Detmodel, Cmodel, Aperture, Auto magnitudes** and many others whose importance is beyond the scope of this work.

Once the counts of photons of the celestial object  $N_{obj}$  are estimated, the value of the magnitude is given by Eq. (2.29) replacing  $F$  by  $N_{obj}$  and  $F_0$  by the number

of counts of the reference object (when observed in the same conditions, i.e. same exposure time). We can also estimate the error of the magnitude as  $\sigma_m = (m \pm \sigma_m) - m = -2.5 \log_{10}(F \pm \sigma_F) + 2.5 \log_{10}(F) = 2.5 \log_{10}(1 \pm \sigma_F/F)$ , where if  $F \gg \sigma_F$  then the two solutions are approximately the same. Defining the signal-to-noise (S/N) ratio as the inverse of  $\sigma_F/F$ , we have that:

$$\sigma_m = 2.5 \log_{10} \left( 1 + \frac{1}{(S/N)} \right) \quad (2.39)$$

The noise, for ground based telescopes, is basically given by adding quadratically three components: the Poisson noise of the signal coming from the object  $\overline{N_{obj}}$ , the Poisson noise of the signal coming from the intrinsic sky brightness  $\overline{N_{sky}}$  (which is integrated within the same area of the object's image) and the Gaussian noise  $RN$  introduced by the electronics in the read-out system. The resulting signal-to-noise ratio expression is:

$$\frac{S}{N} = \frac{N_{obj}}{\overline{N_{obj} + N_{sky} + RN^2}} \quad (2.40)$$

Besides photometric distortions introduced on the measured flux of a celestial object by the Earth's atmosphere (e.g. the sky brightness and atmospheric absorption) and by the optics of the telescope+camera system, galactic gas and dust between an astronomical object and the observer also absorb and scatter part of the electromagnetic radiation. Since bluer light is more strongly attenuated than redder, dust extinction causes objects to appear redder than expected, a phenomenon referred to as reddening (not to be confused with redshift). In order to measure magnitudes properly, they must be corrected for galactic (Milky way) extinction in each band. In any photometric system, interstellar reddening can be described by a color index excess, defined as the difference between an object's color index and its intrinsic color index (the color index it would have if it was not affected by extinction).

## 2.3 Photometric Redshifts (Photo-*zs*)

A photometric redshift (photo-*z*) is an estimate for the redshift of an astronomical object, such as a galaxy or quasar, using its photometry  $m_i$  in various standard filters  $i$ . This is in contrast to the spectroscopic redshift (spec-*z*) where the shift in a spectral feature, such as an emission or absorption line, along the spectrum is determined, allowing the determination of  $z$  directly from its definition in (2.3).

The technique is an old idea that originated in the 1960s (Baum (1962), Connolly et al. (1995), Koo (1985), Loh & Spillar (1986), Puschell, Owen & Laing (1982)), but was later largely replaced by spec-*zs*. However, it has made a come-back in the last two decades, as a result of new large sky surveys (section 2.4), since it provides galaxy redshifts substantially faster than the spectroscopic technique, especially when the observed objects are faint. Moreover, photo-*z* based surveys do not suffer the incompleteness of the target selection needed in spectroscopic surveys, since all objects in the telescope's field-of-view are registered at once (in contrast to spectroscopic surveys where the spectrograph must be pointed to each target individually and some objects may be ignored). The disadvantage is that the precision

for these sorts of measurements is typically  $\sigma_z \sim 0.05 - 0.10$ , and are less reliable than spectroscopic determinations.

There are basically two types of methods to determine photometric redshifts: the template-fitting methods (e.g. Hyperz, Bolzonella, Miralles & Pell (2000); BPZ, Benitez (2000) & Coe et al. (2006); LePhare, Ilbert et al. (2006); EAZY, Brammer, van Dokkum & Coppi (2008); LRT, Assef et al. (2008); GALEV and GAZELLE, Kotulla & Fritze (2009)), and the training-set methods (e.g. ANNz, Collister & Lahav (2004); ArborZ, Gerdes et al. (2010); TPZ, Carrasco Kind & Brunner (2013); Li, Li & Yee (2008); Carliles, Carliles et al. (2010)). In some cases a method can belong to the two previous groups at the same time (e.g. Kernelz; Wolf (2009)), so they are called hybrid methods. In this thesis we will focus on the template-fitting methods leaving the training-set and hybrid methods as alternatives.

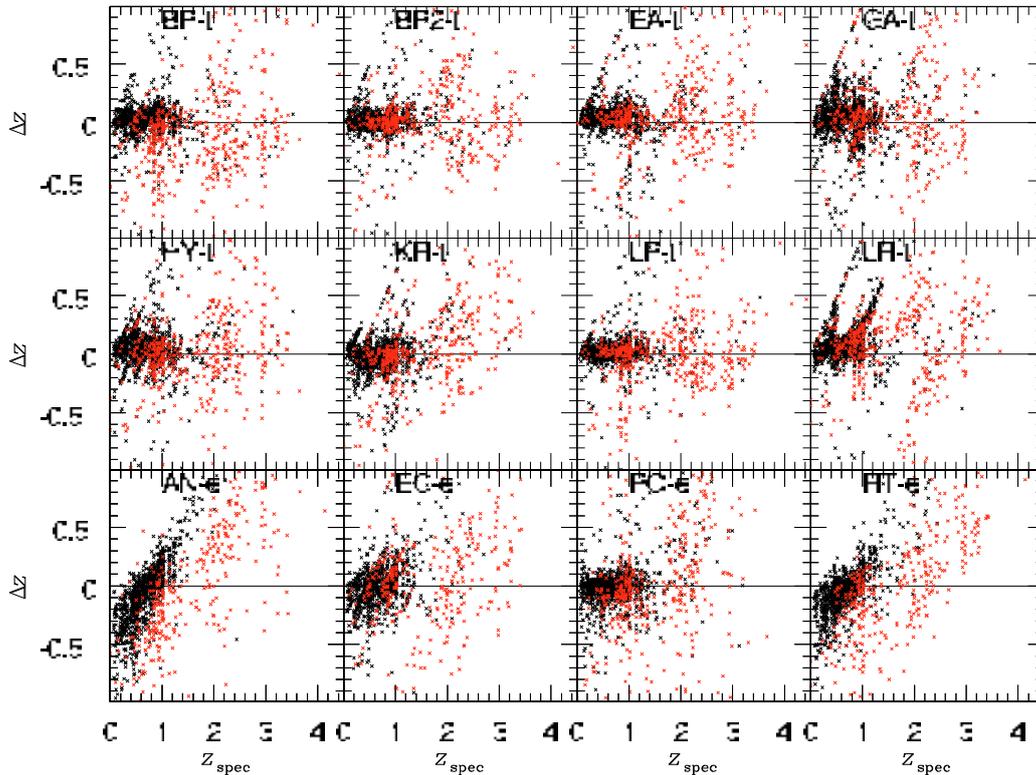
In order to calibrate the goodness of the photo- $z$  determinations one typically uses a spectroscopic sample with known spec- $z$  redshifts, taken as fiducial. In Fig. 2.11 we show several scatter plots of  $\Delta z \equiv z_{ph} - z_{sp}$  vs.  $z_{sp}$ , where  $z_{ph}$  is the photometric redshift and  $z_{sp}$  the spectroscopic one, for different photo- $z$  methods applied on the same galaxy sample. If the photo- $z$  determination was perfect, all the points would be at  $\Delta z = 0$  (over the black line). However, they are actually spread out around this value. The problem is that they are not typically spread out according to a Gaussian profile, but rather to a more complex distribution with asymmetries, bias, double maxima, etc., which needs to be characterized carefully. Moreover, these  $\Delta z$  distributions can rapidly change their shape with redshift, spectral type, magnitude, etc. Usually, these complex distributions are characterized using three different metrics: the bias (mean, median, etc.), which tell us how the overall shape of the distribution is offset from  $\Delta z = 0$ , the precision (RMS,  $\sigma_{68}$ , etc.), which tells us how dispersed are the photo- $z$  determinations from these offsets, and the outlier fraction, which tells us the rate of determinations much worse than expected (that is, the extend of the tails of the  $\Delta z$  distribution).

### 2.3.1 Template-Fitting Methods

The template-fitting methods compare the observed fluxes  $F_i$  of the celestial object in different passbands  $i$  with the predicted fluxes  $F_i^t(z)$  obtained by redshifting a set of spectral templates,  $f^t(\eta) = f^t(\eta / (1+z))$ , and integrating them in the passbands  $R_i(\eta)$  through Eq. (2.28). The photometric redshift value will be the redshift  $z_{ph}$  that, for example, minimizes the following  $\omega^2$ :

$$\omega^2(z; F_i) = \sum_i \frac{(F_i - F_i^t(z))^2}{\sigma_{F_i}^2} \quad (2.41)$$

where  $\sigma_{F_i}$  is the error of the measured flux  $F_i$ . Figure 2.1 provides a visual explanation of what this  $\omega^2$  minimization means. According to the shape of the Vega spectrum, which abruptly starts growing from  $\sim 4000 \text{ \AA}$ , it is easy to see that the less flux will pass through bluer filters, the more shifted to redder wavelenghts the spectrum will be. Therefore, if we know beforehand the shape of the spectrum of the object we are observing, we only have to find how redshifted it would have to be in order to recover the observed fluxes. That is why prominent spectral features with



**Figure 2.11:** Scatter plots  $\Delta z \equiv z_{ph} - z_{sp}$  vs.  $z_{sp}$ , where  $z_{ph}$  is the photometric redshift and  $z_{sp}$  is the spectroscopic redshift, of galaxies from the Great Observatories Origins Deep Survey northern field (GOODS-N, [Giavalisco et al. \(2004\)](#)) in the PHAT1 test ([Hildebrandt et al., 2010](#)) for different photo- $z$  methods: BPZ (BP-t or BP2-t, [Benitez \(2000\)](#)), EASY (EA-t, [Brammer, van Dokkum & Coppi \(2008\)](#)), GALEV and GAZELLE (GA-t, [Kotulla & Fritze \(2009\)](#)), Hyperz (HY-t, [Bolzonella, Miralles & Pell \(2000\)](#)), Kernelz (KR-t), Le Phare (LP-t, [Arnouts et al. \(1999\)](#) & [Ilbert et al. \(2006\)](#)), LRT (LR-t, [Assef et al. \(2008\)](#)), ANNz (AN-e, [Collister & Lahav \(2004\)](#)), Wolf (EC-e, [Wolf \(2009\)](#)), Li (PO-e, [Li & Yee \(2008\)](#)) and Carliles (RT-e, [Carliles et al. \(2010\)](#)). Faint objects with  $R \geq 24$  are shown in red. Figure from [Hildebrandt et al. \(2010\)](#).

an abrupt change on the spectral density flux, such the  $4000\text{\AA}$  break in elliptical galaxies (see Fig. 3.4), are especially good for photo- $z$  determination (this is also valid for training-set methods).

The goodness of the photo- $z$  determination will depend on how well the spectral templates represent the observed galaxy population. If the sample of galaxies to be studied has a wide spectral-type diversity, it is desirable to use a set of templates that covers all possible spectral types. If, on the contrary, the sample only contains galaxies of a single spectral type, it is desirable to use only templates corresponding to the same spectral type in order to avoid spectral type confusions. Commonly used template sets are the [Coleman, Wu & Weedman \(1980\)](#) Spectral Energy Densities (SEDs) which are derived observationally, or those of [Bruzual A. & Charlot \(1993\)](#), derived from population synthesis models. Some examples of galaxy templates can be found in Fig. 3.4, which contains spectral types ranging from early-types to late-type galaxies, and Fig. 4.4, which is more focused on early-type galaxies. Note

that galaxy templates evolve smoothly through different spectral types starting from elliptical galaxies, then to spirals, irregulars and finally reaching starburst galaxies. For this reason, the accuracy with which we want to resolve different spectral types can be tuned by adding more or less interpolated templates between the original ones.

### 2.3.2 Training-Set Methods

The training-set methods derive a parametrization  $z = f(m_i; p_j)$  for the redshift, where  $p_j$  are some parameters, as a function of the photometry  $m_i$  in the different bands  $i$ . The values of the parameters  $p_j$  are deduced by fitting  $z = f(m_i; p_j)$  to a training set of galaxies, for which we have both photometry and spectroscopic redshifts, representative of the target set. The training set could be obtained empirically or derived from a set of spectral templates (as in the template-fitting methods) or from simulated catalogs (e.g. [Vanzella et al. \(2004\)](#)). The training method will only be reliable when it is applied to galaxies within the ranges of redshift and spectral type of the training set.

The parametrization  $z = f(m_i; p_j)$  can be expressed throughout a polynomial (e.g. [Connolly et al. \(1995\)](#); [Sowards-Emmerd et al. \(2000\)](#)), an artificial neural network ([Collister & Lahav, 2004](#)), boosted decision trees ([Gerdes et al., 2010](#)), random forests ([Carliles et al., 2010](#)), etc. Additionally, it can be generalized to any other properties of the galaxies besides the redshift, such as the spectral type, the ellipticity or even the probability to be either a galaxy or a star.

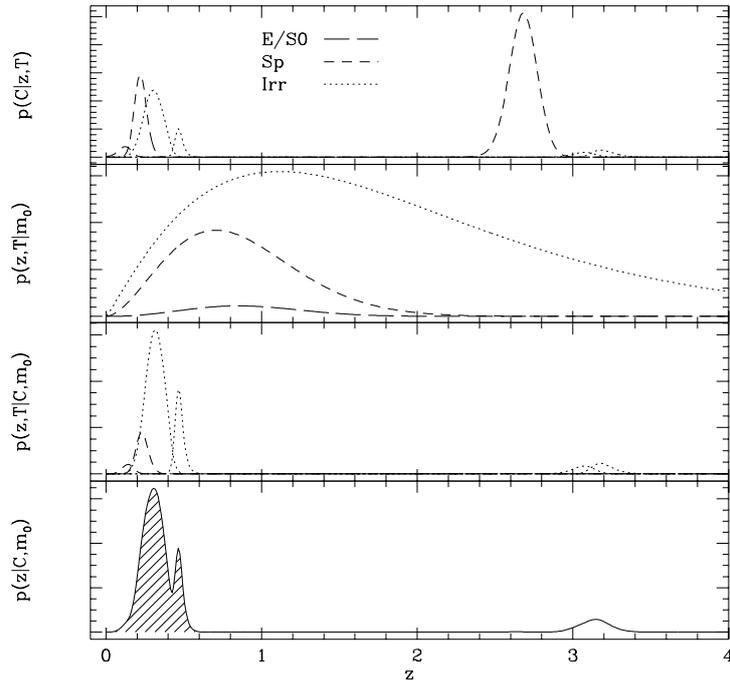
### 2.3.3 Bayesian photometric redshifts

The presence of some similar prominent spectral features at significantly different wavelengths can lead to degeneracies in color space and, consequently to confusions in the photo- $z$  determination. When this happens, photo- $z$ s can result much higher or lower than actual redshifts, which are known as catastrophic redshift determinations.

As in the training-set methods, a representative subset of the photometric target sample for which we know their spectroscopic redshifts could tell us beforehand how galaxies are actually distributed in redshift, and mitigate the catastrophic redshift determinations. According to [Benitez \(2000\)](#) this can be done in template-fitting methods by using Bayesian statistics as follows:

$$p(z | m_j) = \frac{p(z | t, m_j)}{t} = \frac{L(m_j | z, t) \Pi(z | t)}{t} \quad (2.42)$$

where  $p(z | m_j)$  is the posterior probability density distribution (pdf) that the galaxy has redshift  $z$  when its photometry is  $m_j$ ,  $L(m_j | z, t) = \exp -1/2\omega^2(z | t; F_j)$  is the likelihood that the galaxy has photometry  $m_j$  when the redshift is  $z$  and the spectral type is  $t$ , and  $\Pi(z | t)$  is the prior probability that the galaxy has redshift  $z$  and spectral type  $t$ . The prior probability  $\Pi(z | t)$  must be an empirical function that describes the redshift and spectral type distributions of the spectroscopic training set.

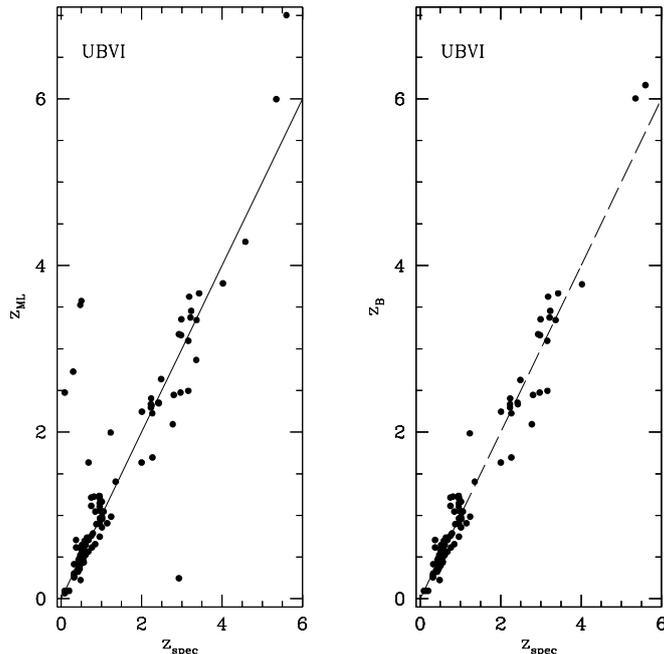


**Figure 2.12:** In descending order: the likelihood  $L(m_j | z, t)$ , the prior probability  $\Pi(z | t)$ , the product of both  $p(z | t, m_j)$  and the posterior probability  $p(z | m_j)$  of the photo- $z$  determination of a galaxy at  $z = 0.28$  using the BPZ code. The prior probability helps suppress the high redshift peak at  $z \sim 2.7$  in the likelihood and moves the galaxy closer to its actual redshift. Figure from Benitez (2000).

In Fig. 2.12 and in descending order, we show the likelihood  $L(m_j | z, t)$ , the prior probability  $\Pi(z | t)$ , the product of both  $p(z | t, m_j)$  and the posterior probability  $p(z | m_j)$  of the photo- $z$  determination of an irregular galaxy at  $z = 0.28$  and magnitude  $I \sim 26$  using BPZ (Benitez, 2000). According to the likelihood alone the galaxy should be a spiral galaxy at redshift  $z_{ph} \sim 2.7$ . However, the prior probability function suppresses the high redshift peak, moving the posterior pdf to a more reasonable redshift with the right spectral type.

In the scatter plot  $z_{ph}$  vs.  $z_{zp}$  on the left of Fig. 2.13, we can see examples of catastrophic redshift determinations: four dots at  $z_{sp} \sim 0.5$  whose photo- $z$  is around 3, as well as a dot at  $z_{sp} \sim 3$  when  $z_{ph}$  is close to 0.5. Once the prior probability is used (scatter plot on the right), all these points get closer to the diagonal.

Besides reducing the catastrophic redshift determination rate, the Bayesian approach returns a probability density function  $p(z | m_j)$  which contains much more information than a single photo- $z$  value. In subsection 2.1.5 we saw that spatial galaxy-galaxy correlations at different redshifts are useful to trace the evolution of the large scale structure in the Universe over time and, finally, determine cosmological parameters. These different redshifts are selected through photo- $z$  bins  $i$  of width comparable to the photo- $z$  precision whose galaxies' true redshift distribution is  $N_i(z)$ . These distributions must be known in order to compute predictions of, for example, the angular correlations  $G_i G_j(\theta)$  through Eq. 2.22. The easiest way to obtain the shape of these distributions is by splitting galaxies of a representative



**Figure 2.13:** The  $z_{ph}$  vs.  $z_{sp}$  scatter plots for the photo- $z$ s obtained with the BPZ code when run on 101 galaxies in the Hubble Deep Field (HDF, Williams et al. (1996)) spectroscopic sample with photometry on the Hubble Space Telescope (HST) bands  $UBVI$ . The left plot corresponds to photo- $z$ s obtained by maximizing only the likelihood in (2.42), while in the right plot the whole posterior probability is maximized. Clearly, Bayesian statistics help reduce the presence of catastrophic redshift determinations (the points far away from the diagonal). Figure from Benitez (2000).

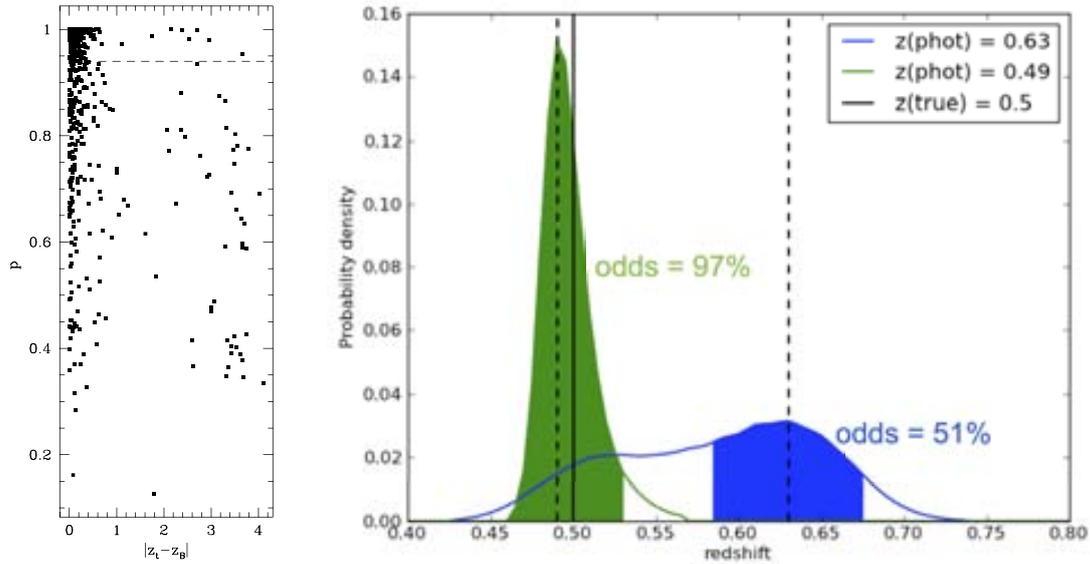
spectroscopic sample into the same photo- $z$  bins used for the target sample, and then compute the resulting spec- $z$  distribution in each bin. However, sometimes such spectroscopic sample is not available. On those cases, the stacking of all the  $p(z | m_j)$  of the galaxies inside the photo- $z$  bin can give a shape very close to the actual  $N_i(z)$  (Bonnett, 2013).

### 2.3.4 Photo- $z$ quality estimator (*odds*)

Photo- $z$  codes, besides returning a best estimate for the redshift, typically also return an indicator of the photo- $z$  quality. It can be simply an estimation of the error on  $z_{ph}$ , or something more complex, but the aim is the same. In BPZ this indicator is called *odds*, and it is defined as

$$odds = \frac{\int_{z_{ph}-\delta z}^{z_{ph}+\delta z} p(z | m_j) dz}{\int_{z_{ph}-\delta z}^{z_{ph}+\delta z} p(z | m_j) dz} \quad (2.43)$$

where  $\delta z$  determines the redshift interval where the integral is computed. *Odds* can range from 0 to 1, and the closer to 1, the more reliable is the photo- $z$  determination, since  $p(z | m_j)$  becomes sharper and most of its area is enclosed within  $z_{ph} \pm \delta z$ . On the right plot of Fig. 2.14 we illustrate this by showing two examples of  $p(z | m_j)$  for two LRGs in the 2SLAQ sample (see subsection 2.4.2), whose spectroscopic



**Figure 2.14:** Left: Scatter plot of odds (shown as  $p$ ) vs.  $|z_{sp} - z_{ph}|$  for galaxies of the Hubble Deep Field (HDF). Points under the dashed line comprise 25% of the catalog. Right: Two examples of  $p(z|m_j)$  for two luminous red galaxies in the 2SLAQ sample whose spectroscopic redshift is roughly the same  $z_{sp} = 0.5$  (solid black vertical line). Colored regions under the curves represent the area enclosed by the interval where the odds are computed. Dashed vertical lines show the best estimate of photo- $z$  given by BPZ (the mode of  $p(z|m_j)$ ). Left plot from Benitez (2000).

redshift is roughly the same,  $z_{sp} \sim 0.5$ , but the returned photo- $z$  by BPZ is clearly distinct. While for one galaxy (green curve) the measured photo- $z$  is very close to the spectroscopic redshift, for the other (blue curve) it differs by  $\sim 0.1$ , something that we were already warned of by the colored areas under the curves, which show the interval where the odds is computed. Note how in the first case the relative area is clearly larger (97%) than in the second (51%). Also note that in the second case the pdf shows a small bump pointing to the right redshift, which tells us, that there may have been a confusion on the template matching similar to that in Fig. 2.12 causing a catastrophic redshift determination, something that could be fixed by using a prior.

Since *odds* are a proxy for the photo- $z$  quality, we should expect a correlation between the *odds* and  $|\Delta z|$ , in the sense that higher *odds* should correspond to lower  $|\Delta z|$ . On the left plot of Fig. 2.14 there is a scatter plot of *odds* (shown as  $p$ ) vs.  $|\Delta z|$  for galaxies in the Hubble Deep Field (HDF). We see that catastrophic outliers, defined as  $|\Delta z| > 1$ , have rather low *odds* values. Points under the dashed line comprise 25% of the catalog. Most of the catastrophic outliers are in this region, so they can be removed by setting an appropriate threshold on the *odds* value and removing all those galaxies below it.

## 2.4 Galaxy Redshift Surveys

Many of the cosmological probes (e.g. SNIa, BAO, LSS, etc., see section 2.1) involve measuring the 3D position on the sky of large amounts of celestial objects such as galaxies. Therefore, observations of large areas of the sky must be carried out to provide photometry or spectroscopy of those objects and, thus, measure their cosmological redshift. These observation campaigns are known as galaxy redshift surveys and, nowadays, there are a wide range of them depending on the area they survey, the depth, etc. In this thesis we refer to several of them: the 2-degree Field Galaxy Redshift Survey (2dFGRS), the Sloan Digital Sky Survey (SDSS) and the Physics of the Accelerating Universe at the William Herschel Telescope (PAU@WHT) Survey.

### 2.4.1 2-degree Field Galaxy Redshift Survey (2dFGRS)

The 2dF Galaxy Redshift Survey (2dFGRS) (Colless et al., 2001) was a spectroscopic survey conducted by the Anglo-Australian Observatory (AAO) with the 3.9 m Anglo-Australian telescope (left of Fig. 2.15) between 1997 and 2002. At the time, it was the second largest redshift survey, next to the Sloan Digital Sky Survey (next subsection). The 2dF survey covered an area of about 1500 sq. deg. of the sky, surveying regions in both the north and the south galactic poles.

The 2dF instrument (right of Fig. 2.15) was installed at the primary focus permitting the observation of a field of 2 deg per pointing, what gives the name to the survey. The instrument possessed a spectrograph equipped with two banks each of 200 optical fibers<sup>6</sup>, permitting the simultaneous measurement of 400 spectra.

The data from this survey were made public on 30 June 2003 and consist of photometry of 382323 objects, from which 245591 include spectra. Out of spectra, 232155 are galaxies (221414 with good quality spectra), 12311 are stars, and 125 are quasi-stellar objects (quasars). The survey needed 272 nights of observation, during 5 years.

In chapter 4 we use a derived galaxy catalog called the 2dF-SDSS LRG and QSO catalog (2SLAQ) (Cannon et al., 2006) of about 13000 luminous red galaxies with good spectroscopic redshift quality as a training catalog where to calibrate our photo-zs.

### 2.4.2 Sloan Digital Sky Survey (SDSS)

The Sloan Digital Sky Survey (SDSS) (York et al., 2000) has been one of the more ambitious galaxy surveys carried out. It was performed in several phases: SDSS-I (2000-2005) and SDSS-II (2005-2008). Later the survey was extended until 2014 with a program called SDSS-III that lies beyond this work.

The SDSS used a dedicated 2.5-meter telescope (Gunn et al., 2006), shown on the left of Fig. 2.16, at Apache Point Observatory (APO), New Mexico, equipped with

---

<sup>6</sup>Optical fibers are used to provide multiobject spectroscopy simultaneously in a single exposure. Each fiber is placed on the focal plane to capture the light of the projected image of a single target. Then, this light is redirected to a dispersing element that produces the spectra.



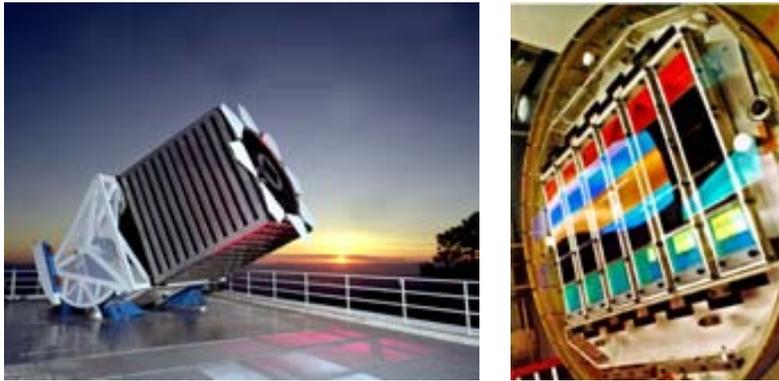
**Figure 2.15:** The Anglo-Australian telescope (left) at the Australian Astronomical Observatory (AAO) sheltered the 2DF instrument which consisted on a multi-fiber spectrograph called AAOmega. On the right is shown the focal plane where each fiber is moved individually to the image of the celestial object whose spectrum is to be extracted.

two special-purpose instruments. The first one is a 120-megapixel (24  $2048 \times 2048$  CCDs) camera (Gunn et al., 1998), shown on the right of Fig. 2.16, that imaged in drift-scan mode<sup>7</sup> 1.5 sq. deg. of sky at a time in five broad bands *ugriz* (Fukugita et al., 1996), whose throughput is shown in Fig. 2.1. The photometry is calibrated to an AB system (Oke & Gunn, 1983), and the zero points of the system are known to 1%-2% (Abazajian et al., 2004) for the Data Release 7 (DR7) used in this thesis. The 95% completeness magnitude limits of the images are 22.0, 22.2, 22.2, 21.3 and 20.5, respectively for *ugriz* (Abazajian et al., 2004), although these values depend, as expected, on seeing and sky brightness. The second instrument is a 640-fiber-fed pair of multiobject double spectrographs, giving coverage from 3800 Å to 9200 Å at a spectral power resolution<sup>8</sup> of  $R \approx 2000$ , which measured spectra of more than 600 galaxies and quasars in a single observation.

The SDSS-II final Data Release 7 (DR7) (Abazajian et al., 2009) has provided multi-color images in a high-latitude region in the Northern Galactic Cap for a total of  $\sim 230$  million celestial objects within a sky area of 8423 sq. deg., together with spectroscopy of complete samples of galaxies and quasars covering about 8200 sq. deg., what is referred as the Legacy survey. Additionally, it has also imaged 3240 sq. deg. of sky at lower Galactic latitudes providing data for 127 million distinct objects (including many stars at low latitude) to study the structure of the Milky Way, what is referred as the SEGUE (Sloan Extension for Galactic Understanding and Exploration) survey (Yanny et al., 2009). The catalog derived includes more than 350 million celestial objects, and spectra of 930,000 galaxies, 120,000 quasars, and 460,000 stars. The data are fully calibrated and reduced, carefully checked for quality, and publically accessible on the web through the Catalog Archive Server (CAS) (Thakar et al., 2008).

<sup>7</sup>The drift-scan technique makes use of the movement of the sky at the focal plane of a telescope when its drive is switched off, to image long continuous strips of the sky. Quite a big field can be explored almost automatically in this way.

<sup>8</sup>The spectral power resolution of a spectrograph is a measure of its ability to resolve features in the electromagnetic spectrum. It is quantified as  $R \equiv \lambda / \Delta\lambda$  where  $\Delta\lambda$  is the smallest difference in wavelengths that can be distinguished at a wavelength of  $\lambda$ .



**Figure 2.16:** The SDSS 2.5m telescope (left) at the Apache Point Observatory (APO) shelters the SDSS CCD imaging camera (right) covered by five broad bands *ugriz* whose throughput is shown in Fig. 2.1.

Some derived galaxy catalogs have been built from the SDSS data, such as the Mega-Z LRG galaxy catalog (Collister et al., 2007), that we use in chapter 4, which consists of a selection of  $\sim 1$  million LRGs in the redshift range  $0.4 < z < 0.7$  with limiting magnitude  $i_{AB} < 20$ . A much reduced version of the Mega-Z catalog called the 2dF-SDSS LRG and QSO catalog (2SLAQ) (Cannon et al., 2006) with matched spectroscopic information from the 2-degree Field Galaxy Redshift Survey (2dFGRS) has been produced and it can be used to calibrate and train photo- $z$  methods for the photo- $z$  determination of the Mega-Z galaxies. It is composed of a total of 13100 luminous red galaxies along a strip of 2 deg width along the celestial equator. A map of the Mega-Z and 2SLAQ galaxies on the sky can be found in Fig. 4.1 in blue and red respectively.

### 2.4.3 Physics of the Accelerating Universe at the William Herschel Telescope (PAU@WHT) Survey

The Physics of the Accelerating Universe survey at the William Herschel Telescope (PAU@WHT) will start mapping a small area of the sky ( $\sim 200$  sq. deg) in the fall of 2014 using a new technique to determine redshifts semi-spectroscopically. It consists of a photometric system composed of 40 narrow bands of  $\sim 125\text{\AA}$  width (bottom of Fig. 3.1), plus six broad bands *ugrizY* (top of Fig. 3.1), similar to the ones on the SDSS instrument.

The PAU instrument (PAUCam; Castander et al. (2012)), shown on the right of Fig. 2.17, will be located at the prime focus of the 4-m diameter William Herschel Telescope (WHT) (left of Fig. 2.17), in the Observatorio del Roque de los Muchachos (ORM observatory) in La Palma. Simulations indicate that PAUCam at the WHT will be able to image about 2 square degrees per night with the 40 narrow and broad bands up to an AB magnitude depth of  $i \sim 22.5$ , providing low-resolution spectra ( $R \sim 50$ ) for around 30000 galaxies, 5000 stars and 1000 quasars per night.

Originally, in Benítez et al. (2009), the PAU Survey was intended to be a larger survey covering about 8000 sq. deg. of the sky to measure Baryonic Acoustic Oscillations (BAO) on the transversal (or angular) and radial (or the line-of-sight)



**Figure 2.17:** The William Herschel Telescope (WHT) dome (left) which will shelter PAUCam (right) whose photometric system is composed of 40 narrow bands.

directions. Measuring BAO transversally does not suppose any challenge, since only the angular coordinates of the galaxies on the sky must be known with a precision not necessarily better than  $\sim 0.1$  deg (any telescope has a better angular resolution). However, detecting the BAO signature on the radial direction involves measuring the distance to galaxies, and therefore the redshift, with a precision of at least  $\Delta z \sim 0.003(1+z)$ , which cannot be achieved with conventional photometric surveys. [Benítez et al. \(2009\)](#) shows that a precision better than this does not result in a significant improvement in the measurement of BAO, since the intrinsic width of the BAO signature of this size in redshift.

The PAU@WHT survey will, instead, study the properties of dark energy by cross-correlating measurements of Redshift Space Distortions (RSD), on a bright sample ( $i_{AB} < 22.5$ ), and Weak-Lensing Magnification (MAG), on a faint sample ( $22.5 < i_{AB} < 23.7$ ), over a reduced area of the sky. [Gaztanaga et al. \(2012\)](#) shows that such measurements, even done in a more moderate area, can provide useful information about the nature of dark energy. Measurements of RSD also require a redshift precision of at least  $\sim 0.003(1+z)$ .

When not in use by the PAU team, PAUCam will be used as a community instrument, able to provide spectral energy distributions (SED) of moderate resolution for a very large sample of objects, allowing the study of a variety of scientific topics beyond cosmology.



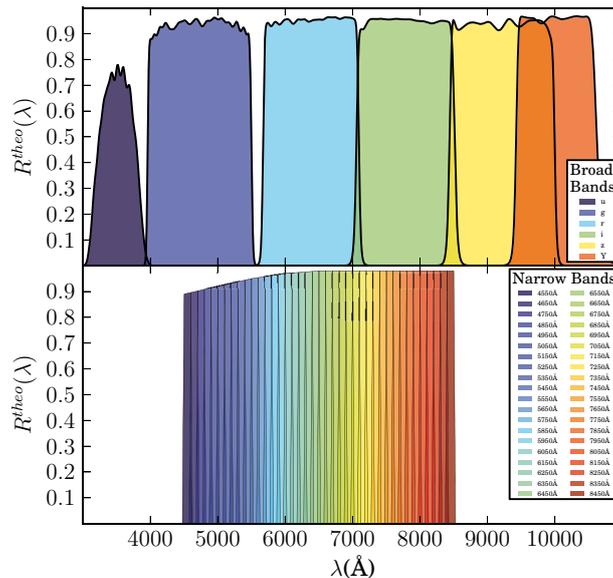
# Chapter 3

## Precise photo- $z$ s with narrow-band filters: The PAU@WHT Survey

In this chapter we present the study of the photo- $z$  performance and its impact on clustering measurements expected in the PAU survey in a sample consisting of all galaxies of all types with  $i_{AB} < 22.5$ . Based on detailed simulation studies (Gaztanaga et al., 2012), the requirement for the precision is set at  $\sigma(z) / (1+z) = 0.0035$ . The PAU survey will observe many galaxies beyond the  $i_{AB} = 22.5$  limit that play a crucial role in the PAU science case (Gaztanaga et al., 2012). We will also study the performance in a fainter galaxy sample with  $22.5 < i_{AB} \lesssim 23.7$ , expecting to reach a photo- $z$  precision not worse than  $\sigma(z) / (1+z) = 0.05$  (Gaztanaga et al., 2012). Finally, we will also study the impact on the photo- $z$  performance of small variations on the default filter set, in terms of width, wavelength coverage, etc.

Throughout the chapter we will be using the Bayesian Photo-Z (BPZ) template-based code from Benitez (2000), after adapting it to our needs. We have also tried several photo- $z$  codes based on training methods. We have found that, because of the large,  $\mathcal{O}(50)$ , number of filters, some of them run into difficulties due to the combinatorial growth of the complexity of the problem, while others confirm the results presented here. The results obtained with training methods will be described in detail elsewhere (Bonnett et al., in preparation).

The outline of the chapter is as follows. In section 3.1 we present the default PAU filter set. Section 3.2 discusses the mock galaxy samples that we use in our study, the noise generation, and the split into a bright and a faint galaxy samples. In section 4.2 we introduce the BPZ original code and our modifications, with special emphasis on the prior redshift probability distributions and the *odds* parameter. We also show the results obtained when running BPZ on the mock catalog using the default filter set. Furthermore, we compute the so-called migration matrix  $r_{ij}$  (Gaztanaga et al., 2012), corresponding to the probability that a galaxy at a true redshift bin  $j$  is actually measured at a photo- $z$  bin  $i$ , and its effect on the measurement of galaxy auto- and cross-correlations. In section 3.4 we try several modifications to the filter set (wider/narrower filters, bluer/redder filters, etc.), study their performance on the brighter and fainter galaxy samples and find the optimal set. Finally, in section 4.6, we discuss the results and offer some conclusions.



**Figure 3.1:** The nominal (or theoretical) response  $R^{theo}(\lambda)$  of the  $ugrizY$  PAU Broad Bands (top) and the 40 Narrow Bands (bottom). The  $u$  band is the same as in the USNO 40-in telescope at Flagstaff Station (Arizona), while the  $grizY$  are the same as in the DECam mounted in the Blanco Telescope (CTIO, Chile). The narrow-band filters have a  $125\text{\AA}$   $FWHM$  and overlap by 25% with the adjacent bands. They are labeled on the plot through their central wavelength. Their overall response is set to match that of the ALHAMBRA survey bands.

### 3.1 Default filter set-up

In this section we construct the effective filter response  $R(\lambda)$  of the PAU bands and compute their  $5\text{-}\sigma$  limiting magnitudes,  $m_{AB}(5\sigma)$ .

#### 3.1.1 Nominal response

PAUCam will mount two sets of filters: the Broad-Band (BB) filters, composed of 6 bands  $ugrizY$ <sup>1</sup>, whose nominal (or theoretical) response  $R^{theo}(\lambda)$  is shown on the top of Fig. 3.1, and the Narrow-Band (NB) filters, shown on the bottom, which are composed of 40 top-hat adjacent bands with a rectangular width of  $100\text{\AA}$  ranging from  $4500\text{\AA}$  to  $8500\text{\AA}$ . Since there is a technical limitation to construct such narrow top-hat bands, we relax the transition from 0 to the maximum response by adding two lateral wings of  $25\text{\AA}$  width, resulting in a  $FWHM$  of  $125\text{\AA}$ . This induces an overlap of  $\sim 25\%$  between contiguous bands. Additionally, we set the overall NB response to match that from the ALHAMBRA survey instrument (Moles et al., 2008), which are comparable in technical specifications (although with a wider wavelength range,  $\sim 310\text{\AA}$ , transmission) and have similar coatings.

<sup>1</sup>The  $u$  band is assumed to be the same as the used in the USNO 40-in telescope at Flagstaff Station (Arizona) and its transmission can be obtained from <http://www.sdss.org/dr7/algorithms/standardstars/Filters/response.html>, while the rest are assumed to be the same as in the DECam mounted in the Blanco Telescope (CTIAO, Chile).

### 3.1.2 Effective response

The filter responses  $R^{theo}(\eta)$  in Fig. 3.1 are the nominal: this is the response that we would measure if light went only through the filter. Light also goes through the Earth's atmosphere, which absorbs part of the light, and then, also goes through the optics (mirror and corrector) of the telescope before getting into the filter. Moreover, the CCD detectors behind filters also are affected by a Quantum Efficiency (QE) response curve. Therefore, if we want to know the effective response of the filters, we will have to take into account all the transmission curves  $T_i(\eta)$  of these effects  $i$ . In our case, these curves are shown in the top plot of Fig. 3.2. The QE curve (blue) corresponds to the measured QE of CCDs provided by Hamamatsu, the measured transmission curve of the telescope's optics (mirror + corrector) (green) corresponds to that from the William Herschel Telescope (WHT) optics, and the atmospheric transmission curve (red) is taken from the Apache Point Observatory (APO) at New Mexico. We assume that the APO atmosphere transmission is close enough to the ORM for the purpose of this study. The resulting effective response  $R(\eta)$  is derived with the expression:

$$\begin{aligned} R(\eta) &= R^{theo}(\eta) \prod_i T_i(\eta) \\ &= R^{theo}(\eta) \cdot T_{CCD}(\eta) \cdot T_{opt}(\eta) \cdot T_{atm}(\eta) \end{aligned} \quad (3.1)$$

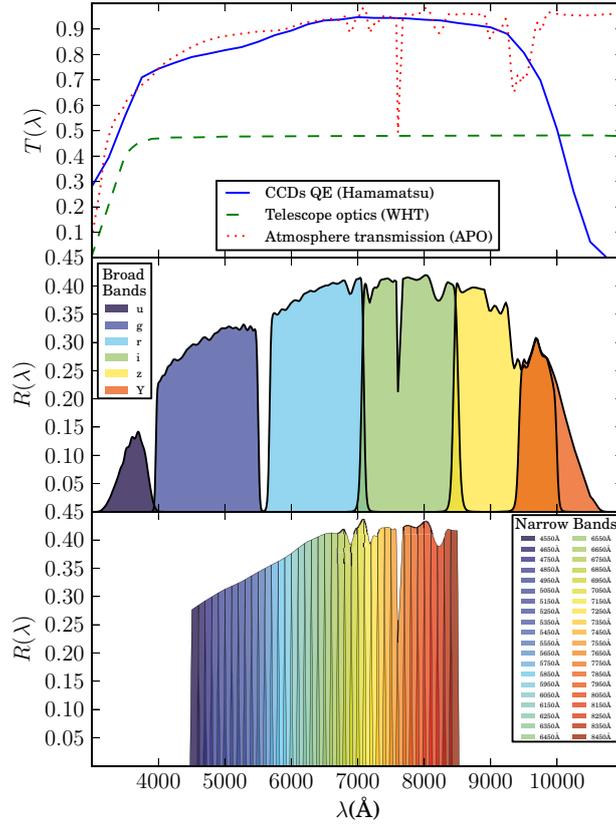
The transmission of the WHT optics is less than 50% in all the wavelength range, so that the resulting effective responses are significantly reduced. On the other hand, the three transmission curves  $T_i(\eta)$  begin to fall when they enter the ultraviolet region ( $\sim 3800\text{\AA}$ ). Similarly, the CCDs QE drops as we approach the infrared region above  $\sim 9000\text{\AA}$ . Overall, the  $u$  and  $Y$  broad bands are less efficient than the rest. This does not affect the NB, since their wavelength range are within these limits. Atmospheric telluric absorption bands, located between  $\sim 700\text{nm}$  and  $\sim 1\mu\text{m}$ , are also imprinted in the final response of the filters. This is particularly relevant for the NB since their typical width is similar to the width of these valleys. In particular, the profile of the narrow band with central wavelength at  $\sim 7550\text{\AA}$  (orange) is drastically changed by the telluric absorption  $A$ -band.

### 3.1.3 5- $\sigma$ limiting magnitudes

Next, we compute the 5- $\sigma$  limiting magnitudes,  $m_{AB}(5\sigma)$ , for all the PAU bands in the AB photometric system<sup>2</sup> (Oke & Schild, 1970). This is the apparent magnitude whose Signal-to-Noise ratio, given by

$$\frac{S}{N} = \frac{\overline{A}}{\alpha^2} \frac{N_{gal}}{N_{gal} + N_{sky} + nRN^2} \quad (3.2)$$

<sup>2</sup>According to Hogg et al. (1996), the apparent magnitude  $m_{AB}$  in the AB system in a band with response  $R(\lambda)$  for a source with spectral density flux  $f(\nu)$  (energy per unit time per unit area per unit frequency) is defined as  $m_{AB} \equiv -2.5 \log_{10} \left[ \int f(\nu) R(\nu) \frac{d\nu}{\nu} \right] - (3631 \text{Jy}) R(\nu) \frac{d\nu}{\nu}$ , where  $1 \text{Jy} = 10^{-23} \text{erg} \cdot \text{s}^{-1} \cdot \text{cm}^{-2} \cdot \text{Hz}^{-1}$  or  $1.51 \cdot 10^7 \text{photons} \cdot \text{m}^{-2} \cdot \text{s}^{-1} \cdot \frac{\lambda}{\text{\AA}}$  in wavelength space.



**Figure 3.2:** Top: the QE curve (blue) of the PAUCam CCDs, the transmission curve of the WHT optics (green), and the atmospheric transmission (red) at the APO (Apache Point Observatory), that affect the final response of the PAU bands. The two lower plots are the same as in Fig. 3.1, but after taking into account these additional transmission curves  $T_i(\lambda)$  through Eq. (3.1). For the sake of clarity, we have rescaled the  $y$ -axes due to the low efficiency of the mirror's reflexion.

is equal to 5, where

$$N_{gal} = 3631 \cdot 1.51 \cdot 10^7 \cdot 10^{-0.4m_{AB}} \cdot \left(\frac{\alpha^2}{A}\right) \cdot \pi \left(\frac{\phi}{2}\right)^2 \cdot nt_R \cdot \int_0^\infty R(\lambda) \frac{d\lambda}{\lambda}, \quad (3.3)$$

$$N_{sky} = \alpha^2 \cdot \pi \left(\frac{\phi}{2}\right)^2 \cdot nt_R \cdot \int_0^\infty f_{sky}(\lambda) R(\lambda) d\lambda, \quad (3.4)$$

are the photons per pixel coming from the galaxy and the sky brightness respectively,  $\{\phi, \alpha, RN, A, n\}$  are the parameters of both the WHT and PAUCam instrument, whose values and description are given in Table 3.1,  $f_{sky}(\lambda)$  is the spectral density flux per unit of aperture of the sky brightness, whose curve is on top of Fig. 3.3, and  $t_R$  is the exposure time for the filter  $R(\lambda)$ . All the filters intended for the photometry are arranged over the central part of the Focal Plane (FP) where vignetting is practically negligible. NB are distributed through 5 interchangeable trays. From the bluest to the reddest, each tray carries a group of 8 consecutive NB. This gives

**Table 3.1:** Description and values of the WHT and PAUCam parameters used in (3.2), (3.3) and (3.4), to compute the Signal-to-Noise ratio ( $S/N$ ).

$\rho$	Telescope mirror diameter	4.2 m
$\alpha$	Focal Plane Scale	0.265 arcsec/pix
$RN$	Read-out Noise	5 electrons/pix
$A$	Galaxy Aperture	2 arcsec <sup>2</sup>
$n$	# of Exposures	2

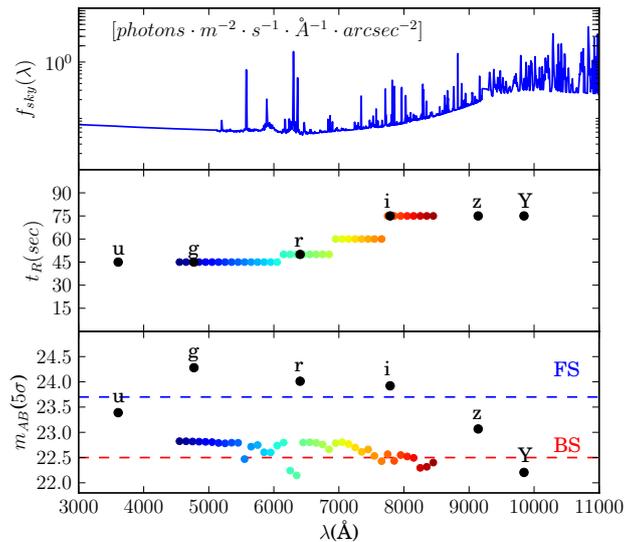
**Table 3.2:** Left: Exposure times  $T_i$  for each PAUCam NB filter tray. The individual NB exposure times are equal to those of the tray where they are. Right: The BB exposure times. Exposure times  $t_R$  per filter are also shown in the middle plot of Fig. 3.3.

NB tray $T_i$		BB $t_R$	
$T_1$	45 sec	u	45 sec
$T_2$	45 sec	g	45 sec
$T_3$	50 sec	r	50 sec
$T_4$	60 sec	i	75 sec
$T_5$	75 sec	z	75 sec
		Y	75 sec

5 trays  $\times$  8 NB = 40 NB. Values for the exposure times  $T_i$  of each tray are shown on the left column of Table 3.2. On the other hand, each BB filter is mounted into a dedicated tray with its particular exposure time, in such a way that NB and BB exposure times are completely decoupled. Values for the exposure times  $t_R$  of each BB filter are shown on the right column of Table 3.2. The exposure times  $t_R$  and the derived limiting magnitudes  $m_{AB}(5^-)$  for each filter are also shown on the middle and bottom plots of Fig. 3.3 respectively, in a color degradation for NB and in black for BB. Since  $f_{sky}(\eta)$  increases with wavelength, exposure times  $t_R$  for redder filters are also set to increase in order to compensate the noise introduced by the sky (note that the exposure times for the NB increase in steps, due to their arrangement in groups per tray). However, this increment is not enough to compensate the sky brightness as we can see with the descending  $m_{AB}(5^-)$  with wavelength. Note that the  $u$  band has a lower limiting magnitude compared with  $g$  even being at shorter wavelengths. This is because the  $u$  response is strongly affected by the transmission curves  $T_i$ . Furthermore, there are large drops in  $m_{AB}(5^-)$  for the NB with central wavelength 5550Å, 6250Å and 6350Å. This is due to emission lines in the sky spectrum  $f_{sky}(\eta)$  at these wavelengths.

## 3.2 The mock catalog

In this section we generate a photometric mock catalog  $m_j \pm \vec{m}_j, z, t$  with observed magnitudes  $m_j \pm \vec{m}_j$  in each PAU band  $j$  for galaxies at redshift  $z$  and with spectral type  $t$ .



**Figure 3.3:** Top: a model of the spectral density flux of the sky brightness  $f_{sky}(\lambda)$  at La Palma (day 7 on the lunar cycle) assuming an airmass of 1.0, used in (3.4). Middle: the exposure times  $t_R$  for each PAU band used in (3.3) and (3.4). Bottom: the resulting limiting magnitudes  $m_{AB}(5\sigma)$  for each band computed through (3.2), (3.3) and (3.4). Colored points correspond to Narrow Bands and black to Broad Bands.

### 3.2.1 Noiseless magnitudes

We use a method similar to the one described in Jouvel et al. (2009), which consists on sampling the cumulative Luminosity Function (LF):

$$N(z, t) = \frac{M_{lim}(z, t)}{\int_{x \equiv 10^{-0.4(M-M^*)}} \rho x^\alpha e^{-x} dx} \quad (3.5)$$

in the redshift range  $z = [0, 6]$ , for a total of  $N \sim 10^6$  galaxies.  $M$  is the absolute magnitude and  $M_{lim}(z, t)$  the absolute magnitude limit at redshift  $z$  and spectral type  $t$  for a given apparent magnitude limit of the catalog  $m_{lim}$  in some reference band. In our case  $m_{lim} < 26.0$  in the  $r_{SDSS}$  band. Finally,  $M^*, \rho, \alpha$  are the parameters of the LF, which also depend on  $z$  and  $t$ . We assume that their redshift dependency is:

$$M^* \log_{10} \rho \alpha = a \exp[-(1+z)^b] + c \quad (3.6)$$

where  $a, b, c$  are type-dependent parameters whose values are in Table 3.3. These values are chosen to match the LFs and their evolution from Dahlen et al. (2005), where three different spectral types, 1 = Elliptical (Ell), 2 = Spiral (Sp), 3 = Irregular (Irr), are distinguished.

The relation between the absolute magnitude  $M$  and the apparent magnitude  $m$  for a galaxy at redshift  $z$  and with spectral type  $t$ , used in the magnitude limit of (3.5), is:

$$M = m - 5 \log_{10} D_L(z) - 25 - K(z, t) \quad (3.7)$$

**Table 3.3:** Parameters  $a$   $b$   $c$  that, through (3.6), give the values and evolution in redshift of the LF parameters  $\log_{10} \lambda M^* \alpha$  for a given spectral type  $t$ . These values are based on the LFs in Dahlen et al. (2005), where three galaxy types are distinguished: 1 = Ell, 2 = Sp and 3 = Irr.

t	$\log_{10} \rho$			$M^*$			$\alpha$		
	a	b	c	a	b	c	a	b	c
1	2.4	1.1	-2.7	5.0	1.6	-21.90	1.7	1.6	-1.00
2	0.5	0.1	-2.28	3.2	2.5	-21.00	0.7	-0.9	-1.50
3	1.0	-3.5	-3.1	5.0	1.3	-20.00	1.8	0.9	-1.85

where  $D_L(z)$  is the luminosity distance of the galaxy at redshift  $z$  in Mpc, which in a  $\Lambda$ CDM universe is expressed as:

$$D_L(z) \equiv (1+z) \frac{c}{H_0} \int_0^z \frac{dz}{\Omega_M(1+z)^3 + \Omega_\Lambda} \quad (3.8)$$

with cosmological parameters chosen to be:  $H_0 = 75$  (km/s)/Mpc,  $\Omega_M = 0.25$  and  $\Omega_\Lambda = 0.75$ , and  $K(z, t)$  is the  $K$ -correction:

$$K(z, t) \equiv -2.5 \log_{10} \left[ \frac{1}{1+z} \frac{\int_0^z f_t(\eta) R_0(\eta) \eta d\eta}{\int_0^z f_t((1+z)\eta) R_0(\eta) d\eta} \right] \quad (3.9)$$

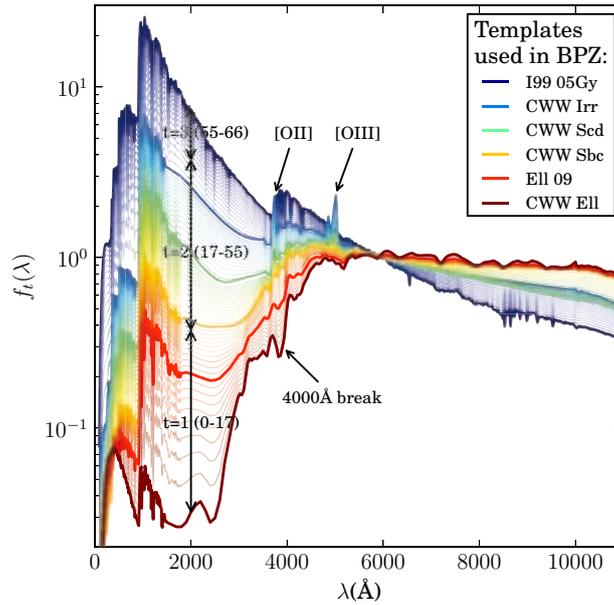
where  $R_0(\eta)$  is the response of the reference band,  $f_t(\eta)$  is the Spectral Energy Density (SED) of the galaxy with spectral type  $t$  in the rest frame, and  $f_t((1+z)\eta)$  the same SED at redshift  $z$ . As a representation of these SEDs, we use the CWW (Coleman, Wu & Weedman, 1980) extended template library from the LePhare<sup>3</sup> photo- $z$  code. It contains 66 templates ranging through Ell Sp Irr and shown on Fig. 3.4. We split them in three groups: Ell = (0-17), Sp = (17-55) and Irr = (55-66). Then, a specific template within one of these groups is randomly selected and assigned to the galaxy. Actually, we allow the spectral type  $t$  to range from 1 to 66 with a resolution of 0.01 by interpolating between templates.

After assigning  $z, t$  values for all galaxies, we also assign them an absolute magnitude  $M$  randomly within the range  $[-M_{lim}(z, t)]$  following the LF probability distribution in (3.5). Then, the apparent magnitude  $m_0$ , in our reference band  $r_{SDSS}$ , is computed from (3.7). The other magnitudes  $m_j$  at any band  $j$  are obtained through:

$$m_j = m_0 + 2.5 \log_{10} \frac{\int_0^z f_t((1+z)\eta) R_0(\eta) \eta d\eta}{\int_0^z f_t((1+z)\eta) R_j(\eta) \eta d\eta} \frac{R_j(\eta) \frac{d\lambda}{\lambda}}{R_0(\eta) \frac{d\lambda}{\lambda}} \quad (3.10)$$

where  $R_j(\eta)$  is the response of some PAU band  $j$ .

<sup>3</sup>The extended CWW library can be found in the folder /lephare\_dev/sed/GAL/CE\_NEW/ of the LePhare package at [http://www.cfht.hawaii.edu/~arnouts/LEPHARE/DOWNLOAD/lephare\\_dev\\_v2.2.tar.gz](http://www.cfht.hawaii.edu/~arnouts/LEPHARE/DOWNLOAD/lephare_dev_v2.2.tar.gz)



**Figure 3.4:** The 66 spectral templates in the rest frame extracted from the extended CWW library. They are used, in (3.7) and (3.10), to generate the photometry of the PAU mock catalog. They evolve from Ellipticals (0-17, in red) to Spirals (17-55, in yellow, green and cyan), and finally to Irregulars (55-66, in blue and violet). Wider and deeper curves highlight the five templates used in BPZ to compute the photo-*z*s.

### 3.2.2 Noisy magnitudes

The resulting magnitudes are noiseless, so we have to transform them to observed magnitudes by adding a random component of noise as follows:

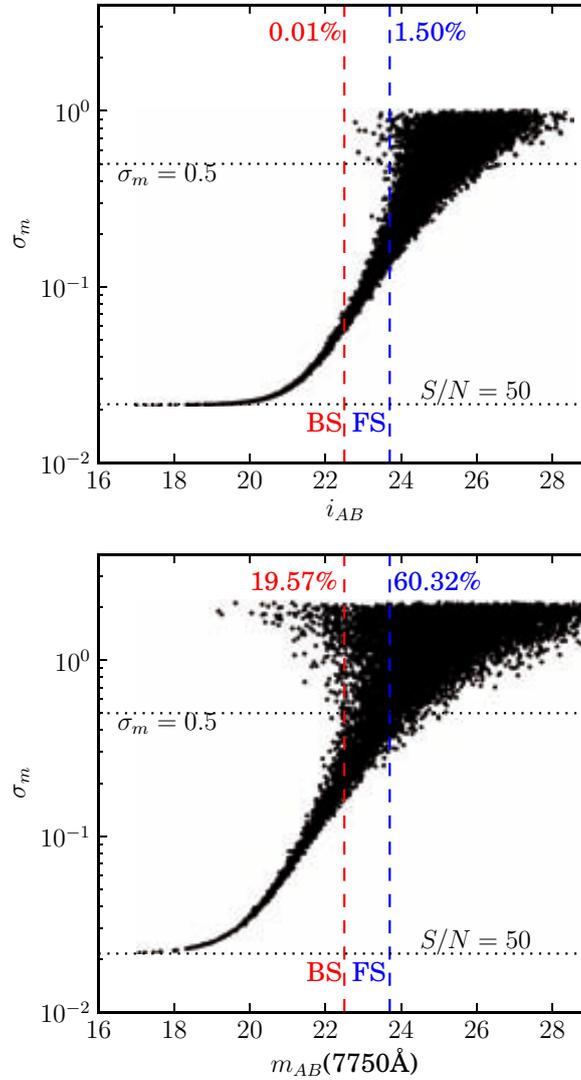
$$m_j \rightarrow m_j + \eta(0, 1)\sigma_{m_j}, \quad (3.11)$$

where  $\eta(0, 1)$  is a normal variable and  $\sigma_{m_j}$  the expected magnitude error which is related to the signal-to-noise in Eq. 3.2 as follows:

$$\sigma_{m_j} = 2.5 \log_{10} \left( 1 + \frac{1}{(S/N)_j} \right) \quad (3.12)$$

Additionally, we add an extra component of noise of size  $\sim 0.022$ , corresponding to a  $S/N = 50$ , in quadrature to  $\sigma_m$ , which takes into account some possible photometric calibration issues. Finally, we obtain the mock catalog  $\{m_j \pm \sigma_{m_j}, z, t\}$ .

The resulting  $\sigma_m$  vs.  $m$  scatter plots, in the  $i$  BB and the  $7750\text{\AA}$  NB, are shown in Fig. 3.5, where for the sake of clarity we only plot 10000 randomly selected galaxies. The  $7750\text{\AA}$  band is chosen because its central wavelength is very similar to that of the  $i$  band. Note how on both bands  $\sigma_m$  starts being flat at  $\sim 0.022$  (the calibration error), and then, at fainter magnitudes, when the sky brightness and the CCD read-out noise become important, it grows and the scatter becomes wider.



**Figure 3.5:** Scatter plots of  $\sigma_m$  vs.  $m$  for the  $i$  BB (top) and the  $7750\text{\AA}$  NB (bottom). For the sake of clarity only 10000 randomly selected galaxies are plotted. The magnitude limits of the BS (red) and FS (blue) are also plotted as vertical-dashed lines. The bottom dotted line in both plots shows the calibration error ( $S/N = 50$ ) added in quadrature to  $\sigma_m$ , while the top dotted line shows the threshold where magnitudes are considered as non-observed. The proportions of non-observed magnitudes ( $\sigma_m > 0.5$ ) in each sample are also shown at the top with their correspondent color.

### 3.2.3 Bright and Faint Samples

The PAU survey science will be mostly focused on Large Scale Structure (LSS) studies such as cross measurements of Redshift Space Distortions (RSD) and Magnification Bias (MAG) between two galaxy samples: the Bright Sample (BS) on the foreground and the Faint Sample (FS) on the background (see [Gaztanaga et al. \(2012\)](#)). The BS should contain galaxies bright enough to have a large signal-to-noise in all bands, including the narrow bands, and, therefore, reach the necessary photo- $z$  accuracy to measure RSD. We see in [Fig. 3.3](#) that the 5- $\sigma$  limiting magnitudes for the NB are close to 22.5, so we define the BS as all those galaxies with  $i_{AB} \equiv m_{AB}(i) < 22.5$ . The FS will contain the rest of the galaxies within  $22.5 < i_{AB} < 23.7$ . The upper limit has been chosen to roughly match the 5- $\sigma$  limiting magnitudes of the broad bands (see [Fig. 3.3](#)). We consider that a magnitude is not observed in one band if its correspondent error is  $\sigma_m > 0.5$ . We find that a  $\sim 0.01\%$  of galaxies in the BS are not observed in the  $i$  band, with the fraction increasing to  $\sim 1.5\%$  in the FS. Similarly, in the BS  $\sim 19.57\%$  are not observed in the 7750Å band, increasing to  $\sim 60.32\%$  in the FS (see [Fig. 3.5](#)). This tells us that, while most of the BB information will be present in both samples, the presence of NB information in the FS will be rather limited, degrading considerably the photo- $z$ s.

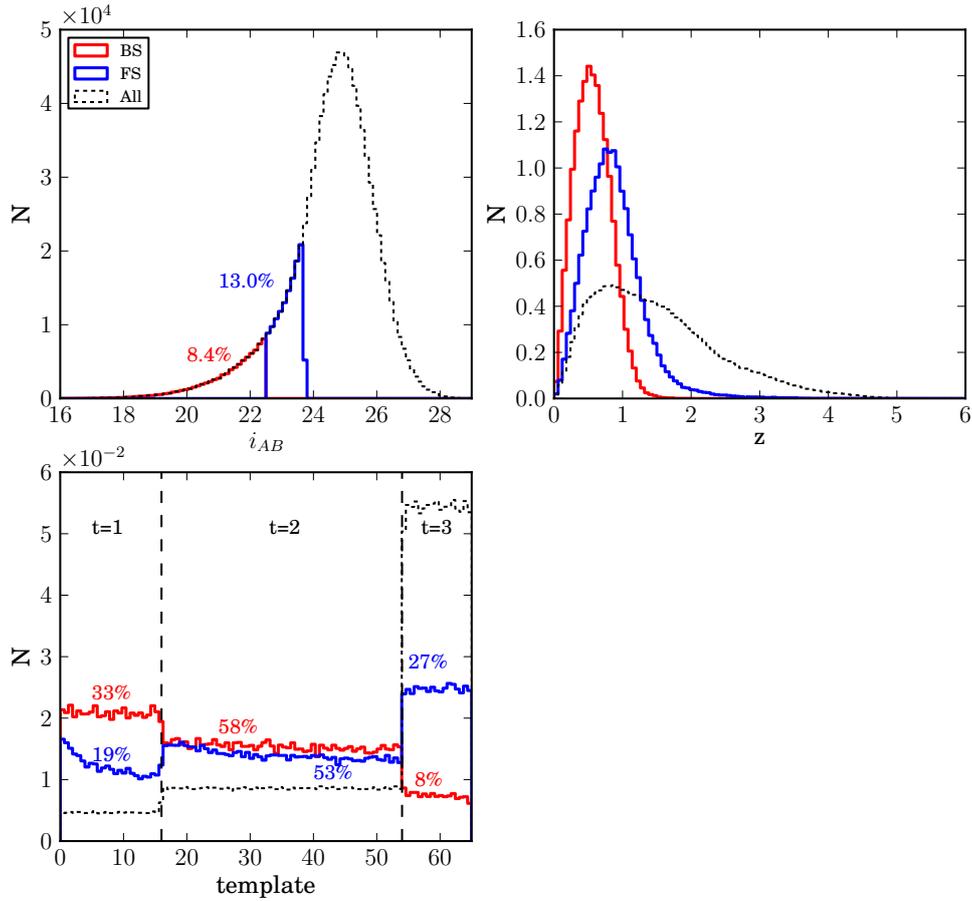
In [Fig. 3.6](#) we show the resulting distributions of the magnitude  $i_{AB}$  (top-left), the true redshift  $z$  (top-right) and the spectral type  $t$  (bottom) of the galaxies in the whole catalog (black-dotted), the BS (red-solid) and the FS (blue-solid). The magnitude distribution of the whole catalog has its maximum at  $\sim 25.0$ , so that the BS and FS are on the brighter tail of the distribution and account for  $\sim 8.4\%$  and  $\sim 13\%$  of the whole catalog respectively. However, this also helps both samples to have a very good completeness up to their magnitude limit. We can also see that, while the whole catalog extends up to  $z \sim 5$ , the BS only goes up to  $z \sim 1.5$  and the FS up to  $z \sim 3$ . Finally, we see that both BS and FS have a similar proportion of Spiral galaxies ( $t = 2$ ),  $\sim 55\%$ ; however the BS contains more elliptical galaxies (33%) than the FS (19%), and consequently, the FS contains more irregular galaxies.

## 3.3 Photo- $z$ performance

In this section we compute the photometric redshifts  $z_{ph}$  of the galaxies in the Bright Sample (BS) and the Faint Sample (FS) generated in the previous section and analyze their photo- $z$  performance through different statistical metrics: bias, photo- $z$  precision and outlier fraction. We also apply some photo- $z$  quality cuts (*odds cuts*) on the results and analyze how the performance improves. We investigate how many galaxies with poor photo- $z$  quality we need to remove in order to achieve the photo- $z$  precision requirements defined in [Gaztanaga et al. \(2012\)](#).

The photo- $z$ s  $z_{ph}$  are obtained using the Bayesian Photometric Redshifts<sup>4</sup> (BPZ) template-fitting code described in [Benitez \(2000\)](#). It uses Bayesian statistics to produce a posterior probability density function  $p(z | m_j)$  that a galaxy is at redshift

<sup>4</sup>BPZ can be found at <http://www.its.caltech.edu/~coe/BPZ/>.



**Figure 3.6:** The observed magnitude in the  $i$  band (top-left), true redshift (top-right) and spectral type distributions for the whole catalog (black), the BS  $i_{AB} < 22.5$  (red) and the FS  $22.5 < i_{AB} < 23.7$  (blue). We also show the fraction of the total galaxies in each sample in the magnitude distribution plot, in their corresponding color. Similarly, we show the proportions of spectral types in each sample. The redshift and type distributions have been normalized.

$z$  when its magnitudes in the different bands  $j$  are  $m_j$ :

$$p(z | m_j) \propto \int_t L(m_j | z, t) \Pi(z | t | m) dz \quad (3.13)$$

where  $L(m_j | z, t)$  is the likelihood that the galaxy has magnitudes  $m_j$ , if its redshift is  $z$  and its spectral type  $t$ , and  $\Pi(z | t | m)$  is the prior probability that the galaxy has redshift  $z$  and spectral type  $t$  when its magnitude in some reference band is  $m$ . The proportionality symbol is a reminder that  $p(z | m_j)$  must be properly normalized in order to be a probability density function. The photometric redshift  $z_{ph}$  of the galaxy will be taken as the position of the maximum of  $p(z | m_j)$ .

We have modified the BPZ code in order to increase its efficiency when estimating photo- $z$ s using a large number of narrow-band filters. Instead of estimating the photo- $z$  for each galaxy, the calculations are done in blocks of hundreds of galaxies using linear operations. Details will be presented in Eriksen et al. (in preparation).

### 3.3.1 Templates

The likelihood  $L(m_j | z | t)$  is generated by comparing the observed magnitudes with the ones that are predicted through a collection of galaxy templates that span all the possible galaxy types  $t$ . BPZ includes its own template library; however, we use a subset of 6 templates from the same library used in the previous section for the mock catalog generation. They are highlighted in Fig. 3.4 and correspond to the templates with file name: `CWW_E11.sed`, `E11_09.sed`, `CWW_Sbc.sed`, `CWW_Scd.sed`, `CWW_Irr.sed` and `I99_05Gy.sed`. Additionally, we also include two interpolated templates between each consecutive pair of the six by setting the BPZ input parameter `INTERP=2`. This results in a total of 16 templates. However, we will see later in Fig. 3.10 that the number of interpolated templates does not affect much the photo- $z$  performance.

### 3.3.2 Prior

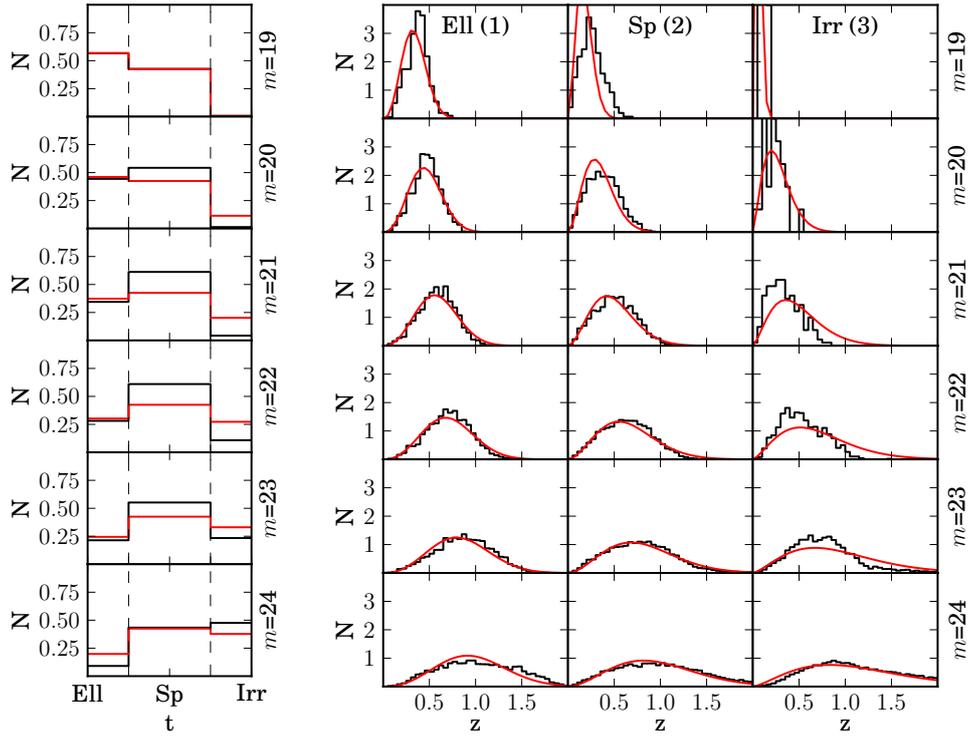
An important point of BPZ is the prior probability  $\Pi(z | t | m)$  that helps improve the photo- $z$  performance. Benitez (2000) proposes the following empirical function:

$$\Pi(z | t | m) = \Pi(t | m) \cdot \Pi(z | t | m) \left( f_t e^{-k_t(m-m_0)} \cdot z^{\alpha_t} \exp \left[ - \left( \frac{z}{z_{mt}(m)} \right)^{\alpha_t} \right] \right) \quad (3.14)$$

where  $z_{mt}(m) = z_{0t} + k_{mt}(m - m_0)$  and  $m_0$  is a reference magnitude, in our case equal to 19 in the  $i$ -band. Each spectral type  $t$  has associated a set of five parameters  $f_t, k_t, \alpha_t, z_{0t}, k_{mt}$  that determine the shape of the prior. In order to calibrate the prior  $\Pi(z | t | m)$  and determine the value of these parameters, we construct a training sample consisting of 10000 galaxies randomly selected from the mock catalog with  $i_{AB} < 24$ . We only need to know their observed magnitude  $i_{AB}$  in our reference band, their true redshift  $z_{tr}$  and their true spectral type  $t_{tr}$ . Originally,  $t$  ranged from 0 to 66 which is the number of templates used to generate the mock catalog; however, as we did for the Luminosity Functions (LF) galaxy types in the previous section, we group all these galaxy types in three groups:  $t = 1$  (ellipticals),  $t = 2$  (spirals) and  $t = 3$  (irregulars), whose correspondence is 1 (0-17), 2 (17-55) and 3 (55-66). From now on, we will use  $t$  for either the 66 templates or these 3 galaxy type groups. Finally, we fit Eq. 4.10 to the training sample and recover the prior parameters. We show the resulting values in Table 4.1. Parameters  $f_3$  and  $k_3$  do not appear in the table because  $\Pi(t = 3 | m_0)$  is deduced by imposing the proper normalization. In Fig. 3.7, we show a comparison between the fitted curve and the actual distribution of the prior.  $\Pi(t | m)$  on the left and  $\Pi(z | t | m)$  on the right are shown at different magnitudes (rows) from  $i_{AB} = 19$ , the reference magnitude, to  $i_{AB} = 24$ . On the one hand, the fitted  $\Pi(t | m)$  agrees by definition with the actual distribution at the reference magnitude (top row), since we use those values as a starting point. However, at higher magnitudes a significant mismatch appears for the spiral and irregular galaxies. This is related to the fact that the  $k$  parameters, which control the migration of galaxies from one spectral type to another across

$t$	$f$	$k$	$\alpha$	$z_0$	$k_m$
1	0.565	0.186	2.456	0.312	0.122
2	0.430	0.000	1.877	0.184	0.130
3	-	-	1.404	0.047	0.148

**Table 3.4:** The resulting values of the prior parameters obtained by fitting (4.10) to a subset of 10000 randomly selected galaxies from the BS and FS together. The three galaxy types  $t = 1\ 2\ 3$  are the same that were defined in Section 3.2 for the Luminosity Functions (LF).  $f_3$  and  $k_3$  do not appear because  $\Pi(t = 3\ m_0)$  is deduced by normalization.



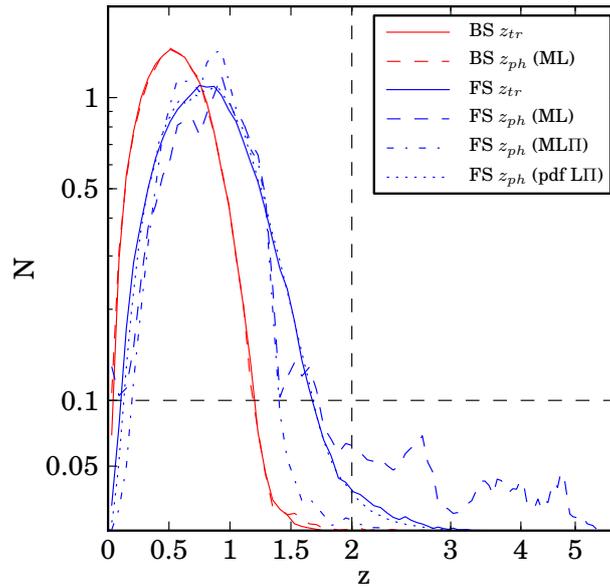
**Figure 3.7:** Left: Comparison between the fitted  $\Pi(t|m)$  prior (red) of Eq. (4.10) and the actual distribution (black). Right: The same for  $\Pi(z|t|m)$ . We only differentiate between 3 galaxy types  $t$  (as we did for the LFs in section 3.2): 1 = Elliptical, 2 = Spiral and 3 = Irregular. Rows correspond to  $i_{AB}$  magnitude bins of width  $\Delta m = 0.2$  centered at values from  $m = 19$  to 24 in steps of 1. All curves have been normalized. Resulting prior parameters are on Table 4.1.

magnitude, are positive definite. With this, elliptical and spiral galaxies should turn to Irregulars as the magnitude increases. However, in Fig. 3.7 we observe that actually the spiral abundance grows slightly before starting to decrease at  $m \sim 21$ , and this forces the fit to  $k_2 = 0$ . Consequently, elliptical galaxies migrate directly to irregulars, causing a mismatch on the pace of growth of this galaxy type abundance. On the other hand, the fit of  $\Pi(z | t, m)$  is particularly good for Ell and Sp galaxies at higher magnitudes (the eight bottom-left panels on the right plot of Fig. 3.7), but for Irr and magnitudes close to the 19 (the reference magnitude) it is less accurate.

In Fig. 3.8 we show a comparison between the  $z_{tr}$  (solid) and  $z_{ph}$  (discontinuous) distributions for the BS and the FS. For the sake of clarity, the x- and y-axis have been set to be linear below  $z = 2$  and  $N = 0.1$  respectively, and logarithmic elsewhere. Dashed lines correspond to  $z_{ph}$  obtained by maximizing only the likelihood  $L(m_j | z, t)$  in (4.9). This gives a  $z_{ph}$  distribution in the BS very close to the actual, while in the FS a residual long tail towards much higher redshift ( $z \sim 5$ ) appears. In Fig. 3.9 we show the equivalent  $\Delta z = (1 + z_{tr})$  distributions, where  $\Delta z \equiv z_{ph} - z_{tr}$ . Once again, the x- and y-axis have been set to be linear below  $\Delta z = (1 + z_{tr}) = 1$  and  $N = 0.05$  respectively, and logarithmic elsewhere. Note that the tail is also present on the right-hand side of the blue curve. If we define as *catastrophic outliers* those galaxies with  $\Delta z = (1 + z_{tr}) > 1$ , we find that they account for  $\sim 7.7\%$  in the FS (the blue region under the curve) and  $\sim 0.2\%$  in the BS. Catastrophic outliers are typically caused by degeneracies in color space, which cause confusions in the template fit and result in a much larger  $\Delta z$ . The blue-dotted line in Fig. 3.9 shows that when the prior is included almost all of the catastrophic outliers in the FS are removed, leaving only a small fraction of  $\sim 0.1\%$ . In fact, we see in Fig. 3.8 that the  $z_{ph}$  distribution after applying the prior (blue dot-dashed) decays at high redshifts faster than the  $z_{tr}$  distribution. This is because we are only using the maximum of  $p(z | m_j)$  for the  $z_{ph}$  value. If we use the whole pdf information (blue-dotted line), the resulting  $z_{ph}$  distribution is much closer to the true. Defining the photo- $z$  precision  $\vec{\tau}_z$  as half of the symmetric interval that encloses the 68% of the  $\Delta z = (1 + z_{tr})$  distribution area around the maximum, we find that  $\vec{\tau}_z$  almost does not change in the BS (0.72% to 0.70%), while it improves by a factor  $\sim 1.8$  in the FS by going from  $\vec{\tau}_z \sim 16\%$  to  $\sim 8.86\%$  when adding the prior.

### 3.3.3 Performance vs. template interpolation

At this point, we want to explore how the number of interpolated templates used in BPZ changes the  $z_{ph}$  performance. In Fig. 3.10 we show the  $\Delta z = (1 + z_{tr})$  distribution only for the BS when we use: 9 (blue), 2 (red) and 0 (green), interpolated templates. Solid lines correspond to the  $z_{ph}$  obtained when the input magnitudes are noiseless (without applying Eq. 3.11), while dotted lines include the noise. The  $\vec{\tau}_z$  of each distribution is shown in the legend. We see that, while for noiseless magnitudes the number of interpolated templates has a significant impact on the width of the distributions and so, on their  $\vec{\tau}_z$ , which gets worse by a factor of  $\sim 3$  at each step, for noisy magnitudes these differences are smaller. In fact, going from 9 to 2 interpolated templates the differences are negligibly small and going from 2 to 0 interpolations the difference is less than a factor of 2.



**Figure 3.8:** Comparison between the true redshift (solid) and the photo- $z$  (discontinuous) distributions for the BS (red) and FS (blue). For the sake of clarity, we have set the  $x$ - and  $y$ -axis scales to be linear below the black dashed lines and logarithmic above them. All distributions have been normalized to equal area. Dashed lines correspond to the  $z_{ph}$  obtained by maximizing only the likelihood  $L(m_j | z, t)$  (ML) of (4.9), the dash-dotted line includes the prior (MLII), and the dotted line corresponds to the case when all the  $p(z | m_j)$  are stacked (pdf LII).

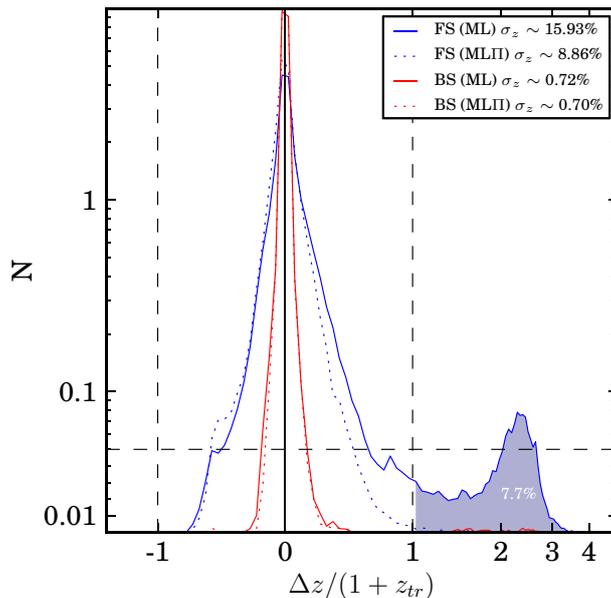
### 3.3.4 Performance vs. *Odds*

Photo- $z$  codes, besides returning a best estimate for the redshift, typically also return an indicator of the photo- $z$  quality. It can be simply an estimation of the error on  $z_{ph}$ , or something more complex, but the aim is the same. In BPZ this indicator is called *odds*, and, it is defined as

$$odds = \frac{\int_{z_{ph}-\delta z}^{z_{ph}+\delta z} p(z | m_j) dz}{\int_{z_{ph}-\delta z}^{z_{ph}+\delta z} p(z | m_j) dz} \quad (3.15)$$

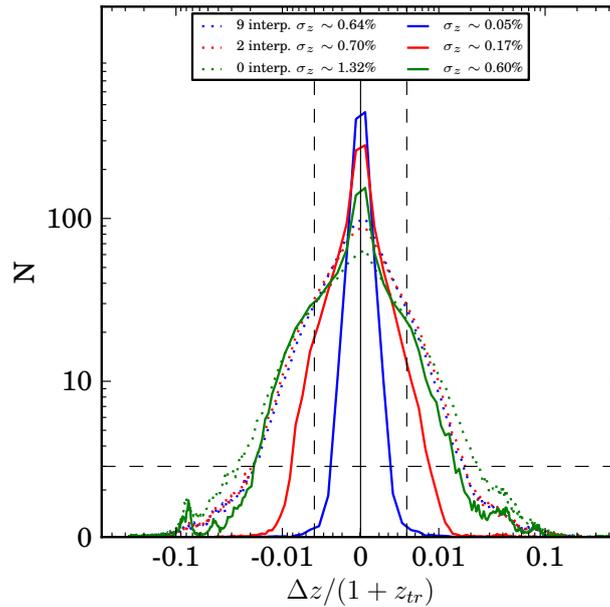
where  $\delta z$  determines the redshift interval where the integral is computed. *Odds* can range from 0 to 1, and the closer to 1, the more reliable is the photo- $z$  determination, since  $p(z | m_j)$  becomes sharper and most of its area is enclosed within  $z_{ph} \pm \delta z$ . We choose  $\delta z = 0.0035$  in the BS and 0.05 in the FS, which is close to the expected  $\vec{z}$  in these samples for the PAU Survey (see the  $\vec{z}$  plots in Fig. 3.11). A bad choice of  $\delta z$  could lead to the accumulation of all *odds* close to either 0 or 1. Since *odds* are a proxy for the photo- $z$  quality, we should expect a correlation between the *odds* and  $\Delta z$  in the sense that higher *odds* should correspond to lower  $\Delta z$ .

At the top of Fig. 3.11, we show the  $\Delta z (1 + z_{tr})$  distributions in a color degradation for subsets of the BS (left) and the FS (right) with increasingly higher cuts on the *odds* parameter. In fact, the exact *odds* values are quite arbitrary, since they depend on the size of  $\delta z$ . Therefore, we have translated these *odds* cuts into the fraction of the galaxy sample remaining after a certain cut has been applied, in such



**Figure 3.9:**  $\Delta z/(1+z_{tr})$  distributions, where  $\Delta z \equiv z_{ph} - z_{tr}$ , for the BS (blue) and FS (red) with the  $z_{ph}$  obtained by maximizing only the likelihood  $L(m_j|z, t)$  (ML) (solid) or when also including the prior (MLII) (dotted). Photo- $z$  precision  $\sigma_z$  values, defined as half of the symmetric interval that encloses the 68% of the distribution area around the maximum, are shown for each case in the legend. The x-axis scale is linear between the two dashed vertical lines and logarithmic on the sides. Similarly, the y-axis is linear below the horizontal dashed line and logarithmic above. The prior makes no difference on the BS, but in the FS it removes the long tail (blue region) of catastrophic outliers ( $|\Delta z|/(1+z_{tr}) > 1$ ) that accounts for a  $\sim 7.7\%$  of the sample. This reduces  $\sigma_z$  by a factor of  $\sim 1.8$ .

a way that the bluest curve corresponds to 100% completeness while the reddest corresponds to 5% completeness, with 5% steps. We can clearly see how the harder are the *odds* cuts, the narrower and peaky become the distributions in both samples. On the bottom plots of the same figure we show how some statistical metrics of these distributions (the bias (median), the photo- $z$  precision ( $\sigma_z$ ) and the  $3\sigma$ -outlier fraction) evolve with each *odds* cut of completeness given in the x-axis. The  $3\sigma$ -outlier fraction is defined as the fraction of galaxies with  $|\Delta z|/(1+z_{tr}) > 3\sigma_z$ . For the sake of clarity, each point has been colored as its correspondent distribution. Errors are computed by bootstrap (Efron, 1979) for the bias and  $\sigma_z$ , and, by computing the  $\sigma_{68}$  of a binomial distribution with mean  $n_{outlier}/N$  for the outlier fraction. As we expected,  $\sigma_z$  decreases as the *odds* cuts get more stringent. In the BS (left), it goes from  $\sim 0.7\%$  at 100% of completeness to  $\sim 0.1\%$  at 5%, and in the FS (right), from  $\sim 9\%$  to  $\sim 1\%$ . The photo- $z$  precision requirements, as defined in Gaztañaga et al. (2012), are  $\sigma_z < 0.35\%$  in the BS and  $\sigma_z < 5\%$  in the FS. They are fulfilled when  $\sim 50\%$  of each catalog is removed. We find a very small bias of a few percent of  $\sigma_z$  in both samples towards negative  $\Delta z$  values. It practically vanishes when high *odds* cuts are applied. The  $3\sigma$ -outlier fraction in the BS starts at  $\sim 13\%$ , drops to  $\sim 3\%$  at  $\sim 40\%$  completeness and then, starts increasing again up to  $\sim 4.5\%$  at

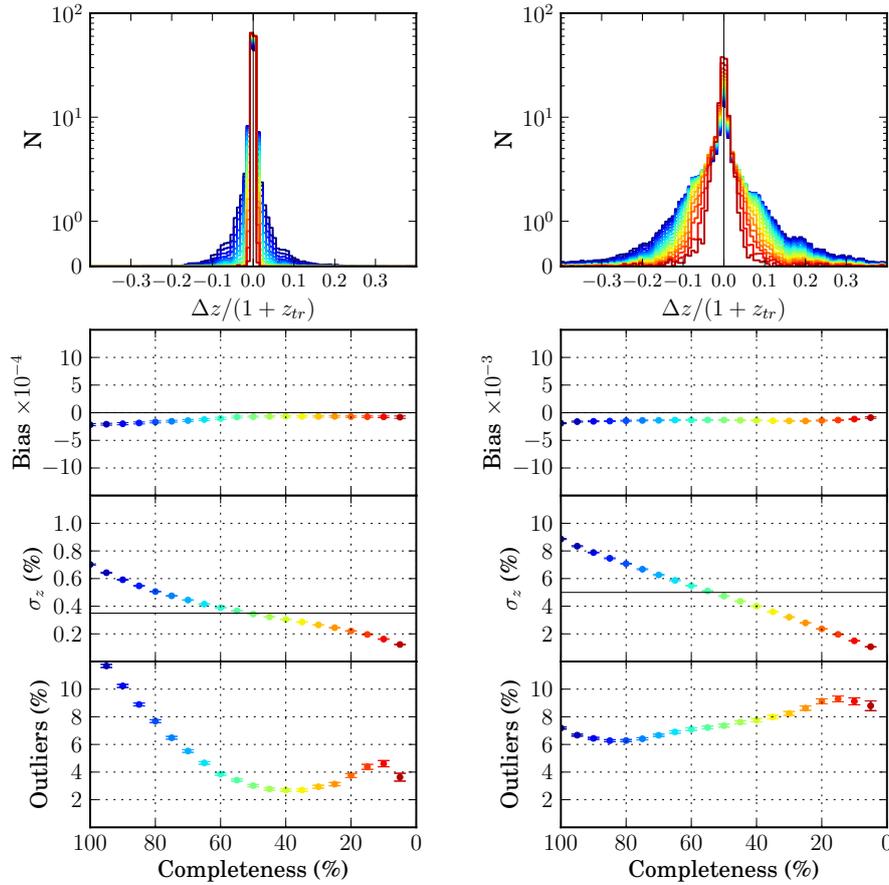


**Figure 3.10:** The  $\Delta z (1 + z_{tr})$  distributions for the BS using different number of interpolated templates in BPZ, when input magnitudes are noiseless (solid) and when they are noisy (dotted). Photo- $z$  precision  $\sigma_z$  values for each case are shown on the legend. The x-axis scale is linear between the two dashed vertical lines and logarithmic on the sides. Similarly, the y-axis is linear below the horizontal dashed line and logarithmic above.

$\sim 10\%$  completeness. Therefore, we deduce that the gain on  $\vec{z}$  with the *odds* cut occurs basically through the cleaning of outliers. However, in the FS, even if  $\vec{z}$  decreases with the *odds* cuts, the outlier fraction increases from  $\sim 7\%$  to  $\sim 9\%$  at  $\sim 15\%$  completeness.

On the last three rows of plots in Figs. 3.12 (BS) and 3.13 (FS), we show how these statistical metrics evolve with respect to the  $i_{AB}$  observed magnitude (left), the true spectral type  $t_{tr}$  (center) and the true redshift  $z_{tr}$  (right), after each photo- $z$  quality cut shown in Fig. 3.11 in the same color. In the first three rows and by order, we also show the scatter plot  $\Delta z (1 + z_{tr})$ , the number of galaxies and the completeness after the same photo- $z$  quality cuts with respect to the same variables ( $i_{AB}, t_{tr}, z_{tr}$ ) on the x-axis.

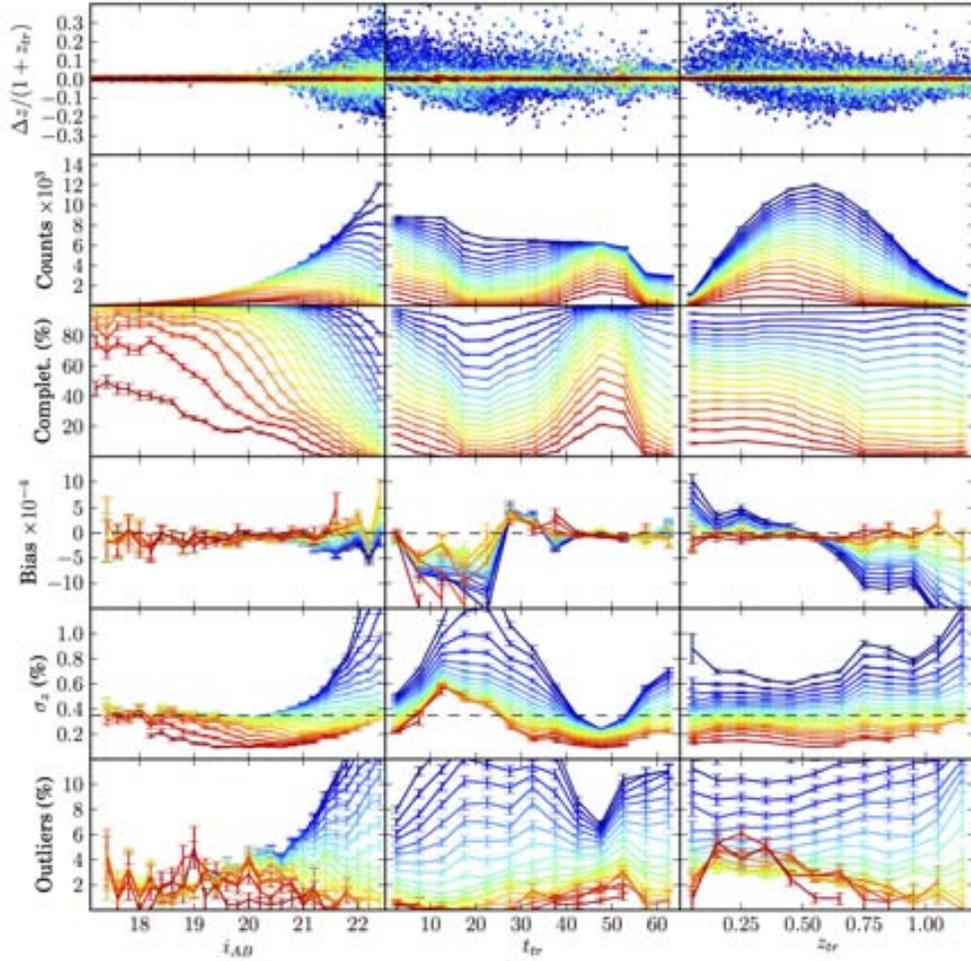
In the BS (Fig. 3.12), we can see that the low photo- $z$  quality galaxies (blue points in the scatter plot) are mostly faint galaxies with  $i_{AB} > 21$ , the magnitude where the noise coming from the sky brightness plus the CCD read-out starts to be comparable to the Poisson noise in signal. This is reflected in Fig. 3.5 as a turning point on the slope of the  $\vec{m}$  vs.  $m_{AB}$  scatter. In fact, these galaxies represent most of the outliers and the principal source of bias seen in Fig. 3.11. As the *odds* cuts are applied, these bad photo- $z$  faint galaxies are removed. The *odds* cut removes the bias, reduces  $\vec{z}$  from  $\sim 2.2\%$  to  $\sim 0.35\%$  and the outlier fraction from  $> 10\%$  to  $\sim 1\%$  at magnitudes close to the limit  $i_{AB} = 22.5$ . Looking at the scatter plots of the next two columns in Fig. 3.12, we realize that these low-*odds* galaxies at faint magnitude are spread out over the whole  $t_{tr}$  and  $z_{tr}$  ranges. Moreover, after the hardest *odds* cut, only galaxies of types  $t \sim 0$  (elliptical) and  $t \sim 50$  (irregular) survive and the



**Figure 3.11:** Top:  $\Delta z (1 + z_{tr})$  distributions for the BS (left) and the FS (right) at different photo- $z$  quality cuts whose completeness, shown in a color degradation, range from 100% (bluest) to 5% (reddest) in 5% steps. On the bottom and by rows: the bias (median), the photo- $z$  precision ( $\sigma_z$ ) and the 3 $\sigma$ -outlier fraction of the distributions at the top as a function of the completeness. The colors of the points match the colors in the distributions. Note that the vertical scale in the bias and the  $\sigma_z$  panels changes by an order of magnitude between the two samples. The black-dashed horizontal lines on the  $\sigma_z$  panels show the photo- $z$  precision requirements defined in Gaztanaga et al. (2012):  $\sigma_z < 0.35\%$  (BS) and  $< 5\%$  (FS).

mean of  $z_{tr}$  is shifted from  $\sim 0.6$  to  $\sim 0.4$ . The worst bias,  $\vec{\tau}_z$  and outlier fraction are obtained for Spiral galaxies ( $t \sim 10-30$ ). The *odds* cuts mitigate these results, but even after applying them, spiral galaxies still have the worst bias and  $\vec{\tau}_z$ . The worst bias is located at low and high  $z_{tr}$ , with opposite sign and it is largely reduced with the *odds* cuts. The value of  $\vec{\tau}_z$  gets flatter over all  $z_{tr}$  as the *odds* cuts are harder.

Regarding the  $z_{ph}$  precision requirement  $\vec{\tau}_z < 0.35\%$  (black solid horizontal line), we find that, when no *odds* cut is applied, it is achieved only for galaxies with  $i_{AB} < 21$  and galaxy type around  $t \sim 50$  (irregulars). However, it is not fulfilled at any  $z_{tr}$ . Once we apply a 50% completeness *odds* cut, which gives an overall  $\vec{\tau}_z$  equal to the requirement, as we saw in Fig. 3.11, the requirement is fulfilled in all the  $i_{AB}$  and  $z_{tr}$  ranges. Only for spiral galaxies the requirement is not fulfilled even after the hardest *odds* cut. Originally, in Benítez et al. (2009), it was assumed



**Figure 3.12:** Statistics showing the PAU-BS photo- $z$  performance as a function of the observed  $i_{AB}$  magnitude (left), true galaxy type  $t_{tr}$  (center) and true redshift  $z_{tr}$  (right). In the first row we show the scatter  $\Delta z$  ( $1 + z_{tr}$ ). Then, in a descending order of rows, we show the number of galaxies, the completeness, the bias (median), the photo- $z$  precision ( $\sigma_z$ ) and the  $3\sigma$ -outlier fraction for all the *odds* cuts shown in Fig. 3.11 (same color). Also as in Fig. 3.11, the black-solid horizontal lines on the  $\sigma_z$  panels show the photo- $z$  precision requirement defined in Gaztanaga et al. (2012).

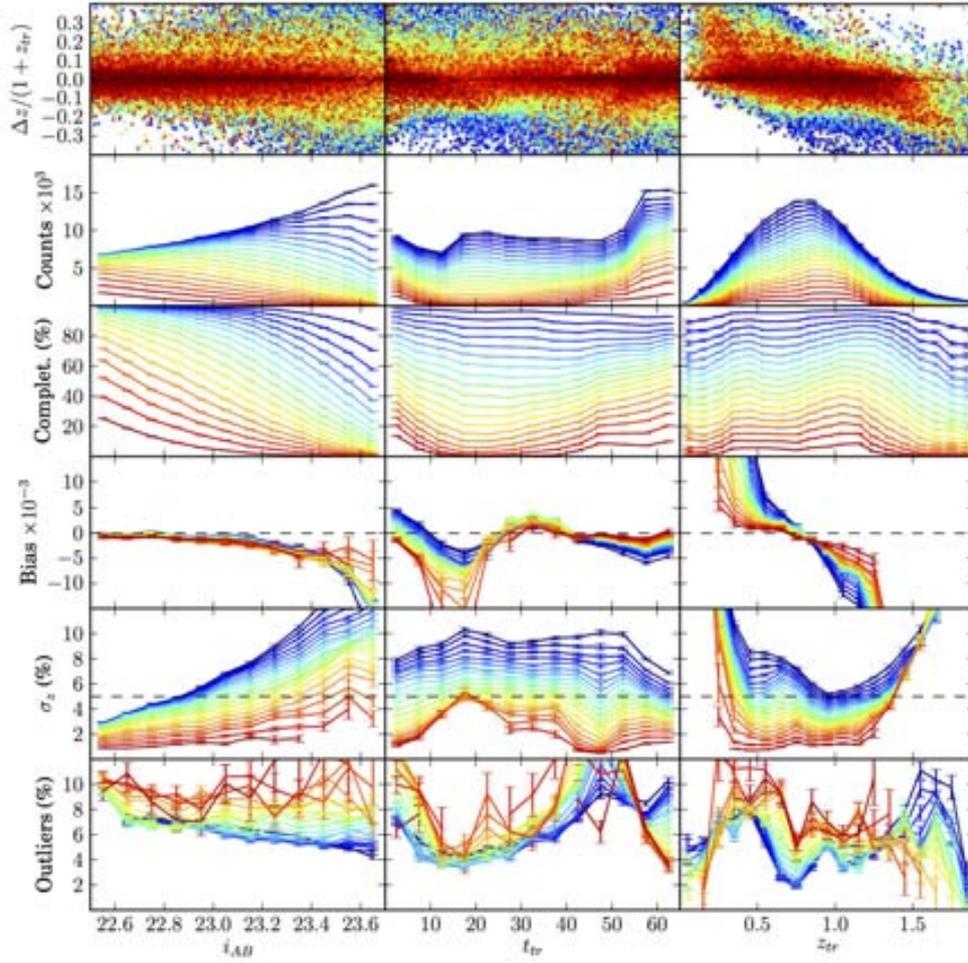
that elliptical galaxies (or rather Luminous Red Galaxies) yield the best photo- $z$  precision in the PAU Survey, with the narrow bands tracking the  $\sim 4000\text{\AA}$  break spectral feature (Fig. 3.4). This is partially true, since we actually see that elliptical galaxies give better  $\sigma_z$  than spirals, but our analysis shows that in fact irregulars with  $t \sim 50$  give the best photo- $z$  performance. Before any *odds* cut, their photo- $z$  precision is almost twice better than the requirement. Probably this is due to the fact that, in contrast to elliptical galaxies where a single spectral feature is tracked, irregulars have the two emission lines [OII]  $\sim 3737\text{\AA}$  and [OIII]  $\sim 5000\text{\AA}$  (Fig. 3.4) with intrinsic widths narrower than the width of the NB filters.

In the FS (Fig. 3.13), we see that the scatter of  $\Delta z$  ( $1 + z_{tr}$ ) is much larger, as expected from the wider histograms in Fig. 3.11. However, a wider but still tight core close to  $\Delta z = 0$  with high photo- $z$  quality (red points) remains. We recognize

behaviors similar to those in the BS in most aspects of the  $z_{ph}$  performance, although they are substantially larger. For example, the highest magnitude as well as lowest and highest  $z_{tr}$  galaxies are the most biased. The *odds* cuts also mitigate this bias, but a residual bias of opposite sign persists at the extremes of  $z_{tr}$ . Spiral galaxies ( $t \sim 10$ -30) are also the ones with the highest bias, and the *odds* cuts even aggravates this. We also see that elliptical ( $t \sim 0$ ) and irregular ( $t > 50$ ) galaxies are initially biased, but, in contrast to spirals, the *odds* cuts help to reduce the bias. Unlike in the BS,  $\vec{\tau}_z$  increases along all the magnitude range since at those magnitudes the noise from the sky brightness dominates over the signal (Fig. 3.5). However, we see that the slope of the  $\vec{\tau}_z$  increase is smaller the harder the *odds* cuts. This is because the gain in photo- $z$  precision is at the expense of keeping only brighter galaxies each time. The mean  $i_{AB}$  magnitude goes from  $\sim 23.2$  to  $\sim 22.8$  with the *odds* cuts, and the shift in the  $z_{tr}$  mean is from  $\sim 0.86$  to  $\sim 0.81$ . Unlike in the BS, we see that the best  $\vec{\tau}_z$  is obtained at the extremal spectral types:  $t \sim 0$  (elliptical) and  $t \sim 66$  (irregular). However, once the hardest *odds* cuts are applied, irregular galaxies with  $t \sim 50$  are again the once with the best  $\vec{\tau}_z$ . In fact, the hardest *odds* cuts also remove all spiral galaxies. Note that the large bias seen at the extremes of  $z_{tr}$  make  $\vec{\tau}_z$  take values much larger at these redshifts. The photo- $z$  precision requirement,  $\vec{\tau}_z < 5\%$ , is fulfilled when the *odds* cut of 50% completeness is applied up to magnitude  $\sim 23.1$ , for all galaxy types except spirals, and at the  $z_{tr}$  interval from  $\sim 0.4$  to  $\sim 1.3$ . As we already saw on Fig. 3.11, the 3 $\vec{\tau}$ -outlier fraction grows with the *odds* cuts. In general, its values are higher where  $\vec{\tau}_z$  is lower, since the outliers criterion becomes more stringent.

### 3.3.5 Narrow bands vs. Broad bands

We want to quantify the improvement that the NB bring to the photo- $z$  performance. For this purpose, we run BPZ on the BS and the FS using the NB and the BB separately. Then, in Fig. 3.14 and Table 3.5, we compare results between these runs and also with the original ones when the BB and the NB are used together (BB+NB). Figure 3.14 shows normalized  $\Delta z (1 + z_{tr})$  distributions for the BS (red) and the FS (blue) using only the BB (dashed), only the NB (dotted) and both together BB+NB (solid). We see that the resulting distributions when using only BB (dashed) show overall shapes close to Gaussian with perhaps larger tails on both sides. However, when the NB are also included (solid) the peaks of the distributions become clearly sharper. This is more noticeable in the BS than in the FS, because the non-observed condition ( $\vec{\tau}_m < 0.5$ ) defined in Section 3.2 implies that most of the NB are not used in the photo- $z$  determination for the FS. Table 3.5 shows bias (median),  $\vec{\tau}_z$  ( $\vec{\tau}_{68}$ ) and 3 $\vec{\tau}$ -outlier fraction of each distribution.  $\vec{\tau}_z$  in the BS is reduced  $\sim 4.8$  times going from  $\sim 3.34\%$  to  $\sim 0.7\%$  when the NB are included, while the improvement is much less significant in the FS. We also see that bias is reduced by an order of magnitude when the NB are included in both samples. On the contrary, the outlier fraction increases, but as mentioned, this is due to the fact that improvements on  $\vec{\tau}_z$  penalize the outlier fraction. On the other hand, we see that using NB alone slightly degrades all metrics in the BS and the FS, except the outlier fraction in the FS which is improved for the same reason. In fact,  $\vec{\tau}_z$  in the



**Figure 3.13:** Statistics showing the PAU-FS photo- $z$  performance in the same layout as in Fig. 3.12.

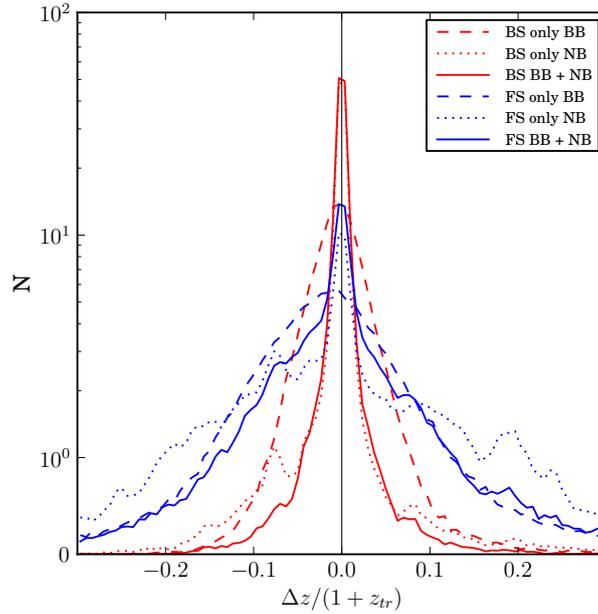
FS gets almost twice worse than when only using BB or BB+NB. It seems that in the FS NB by themselves only help the bias, while if they are used together with the BB, the improvement also extends to  $\sigma_z$ .

### 3.3.6 Impact of the photo- $z$ s on clustering

We want to study the impact of the PAU photo- $z$  performance on the measurements of angular clustering. In Gaztanaga et al. (2012) it is shown that galaxy cross-correlation measurements  $\xi_{ij}$  between two photo- $z$  bins  $i$  and  $j$  are related to their cross-correlation between real redshift bins  $\xi_{ij}$  as

$$\xi_{ij}^{A \times B} = \sum_{kl} r_{ik}^A \xi_{kl}^{A \times B} r_{jl}^B = r_A \cdot \xi^{A \times B} \cdot r_B^T \quad (3.16)$$

where  $r_{ij}$  is called the migration matrix and gives the probability that a galaxy observed at the photo- $z$  bin  $i$  will be actually at the true redshift bin  $j$ , while  $A$  and  $B$  denote different galaxy samples. We compute the migration matrices from the PAU photo- $z$  simulations and show them on Fig. 3.15 for the BS (top) and the



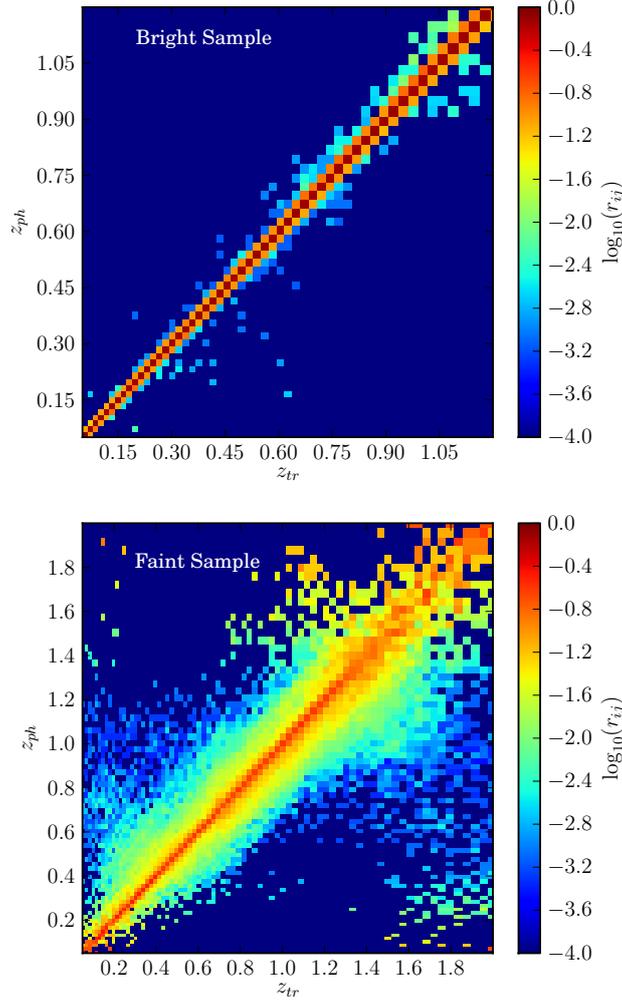
**Figure 3.14:** Normalized  $\Delta z (1 + z_{tr})$  distributions for the BS (red) and the FS (blue) using only BB (dashed), only NB (dotted) or both together BB+NB (solid). Bias,  $\sigma_z$  and  $3\sigma$ -outlier fraction of each distribution are shown in Table 3.5.

	Bright Sample		
	BB	NB	BB + NB
Bias $\times 10^{-4}$	-31.64	-3.31	-2.18
$\vec{z}(\%)$	3.34	0.83	0.70
Outliers(%)	4.41	18.19	13.28
	Faint Sample		
	BB	NB	BB + NB
Bias $\times 10^{-4}$	-152.66	-41.19	-19.01
$\vec{z}(\%)$	9.38	16.17	8.86
Outliers(%)	6.79	4.90	7.18

**Table 3.5:** Bias (median),  $\sigma_z$  ( $\sigma_{68}$ ) and  $3\sigma$ -outlier fraction when using only BB, only NB or both together BB+NB for the bright and faint samples.

FS (bottom) in photo- $z$  bins of width  $0.014(1 + z)$ , which is four times the photo- $z$  precision  $\vec{z}$  in the BS once the photo- $z$  quality cut that leaves a 50% completeness is applied. Note that the matrices are normalized row-wise by definition. The FS migration matrix has quite a thick diagonal, i.e. around  $z_{tr} = 1$  the width is  $\Delta z = 0.1$  for 10% probabilities and  $\Delta z = 0.4$  for 1%. There are also outliers going from very large true redshifts  $z_{tr} = 1.8$  to lower photo- $z$  redshifts and some from low true redshifts up to  $z_{ph} = 1$ . The BS has a thinner diagonal with  $\Delta z < 0.04$  at 1% probability and with fewer outliers. These values are of course in agreement with previous results in Figs. 3.9 and 3.11.

To illustrate how these matrices affect angular clustering measurements, we use a simple model to predict  $i_j$  at an arbitrary reference angular scale,  $\theta$ , of 1 arcmin.



**Figure 3.15:** The resulting migration matrices  $r_{ij}$  from the PAU photo- $z$  simulations after applying the photo- $z$  quality cut that leaves 50% completeness. The top plot corresponds to the BS and the bottom plot to the FS. For a higher contrast and clarity we plot the logarithm of the matrix values. These matrices give the probability that a galaxy observed at the photo- $z$  bin  $i$ , will be actually at the true redshift bin  $j$ . The bin widths are  $0.014(1+z)$ , four times the photo- $z$  precision expected in the BS.

We include intrinsic galaxy clustering and weak lensing magnification:

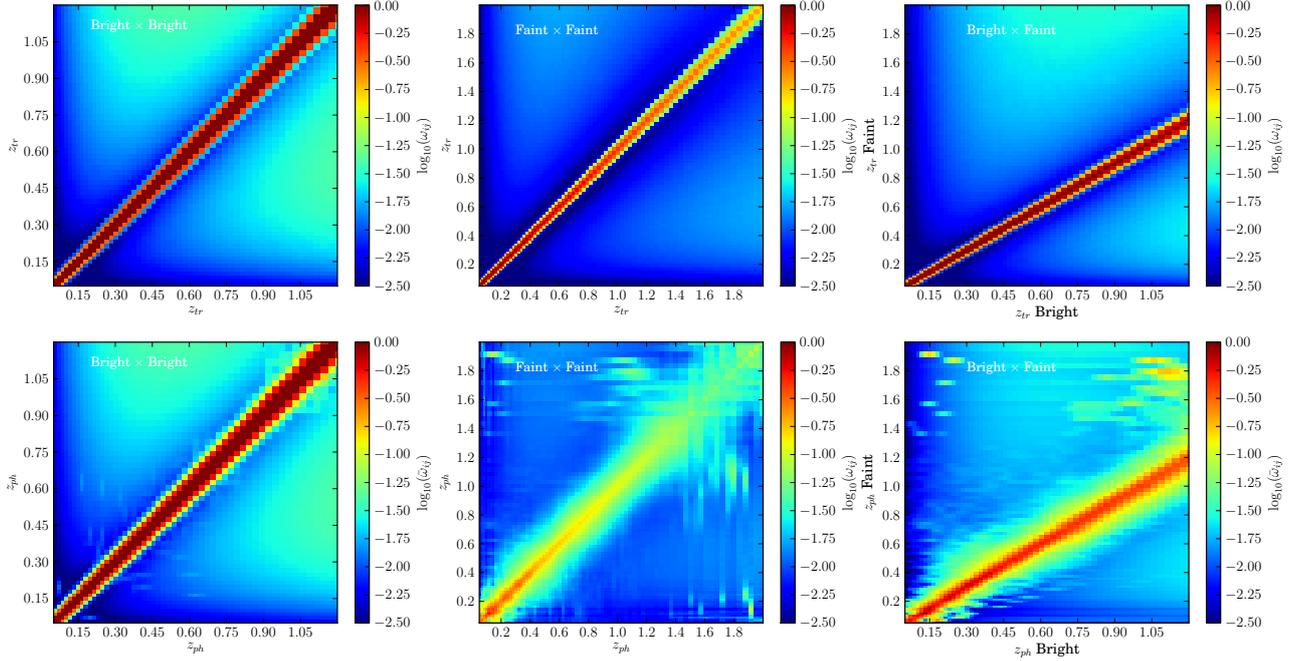
$$w_{ij} = w_{G_i G_j} + w_{G_i \mu_j} + w_{\mu_i G_j} + w_{\mu_i \mu_j} \quad (3.17)$$

$$w_{G_i G_j} = b_i b_j \int dz_1 dz_2 \rho_{G_i}(z_1) \rho_{G_j}(z_2) \sigma(r_{12}) \quad (3.18)$$

$$w_{G_i \mu_j} = b_i \alpha_j \int dz_1 dz_2 \rho_{G_i}(z_1) p_{\mu_j}(z_2) \sigma(r_{12}) \quad (3.19)$$

$$w_{\mu_i \mu_j} = \alpha_i \alpha_j \int dz_1 dz_2 p_{\mu_i}(z_1) p_{\mu_j}(z_2) \sigma(r_{12}) \quad (3.20)$$

where  $\sigma(r_{12})$  is the non-linear matter 2-point correlation between the positions of two galaxies separated in 3D space by  $r_{12} = r_2 - r_1$ , where the angular separation between  $r_1$  and  $r_2$  is fixed to be the reference angle (i.e.  $\theta = 1$  arcmin), while the



**Figure 3.16:** Top panels show the angular auto (diagonal) and cross- (off-diagonal) correlations  $c_{ij}$  at 1 arcmin between the true redshift bins  $i$  and  $j$  of the BS (left), the FS (middle) and the crossing of both of them (right). Bottom plots show the same correlations but measured with photo- $z$  bins, which can be computed by using Eq. (3.16) with the migration matrices in Fig. 3.15. The bin widths are  $0.014(1+z)$ , four times the photo- $z$  precision expected in the Bright Sample.

radial separation is integrated out via  $z_1$  and  $z_2$ . We have that  $\rho_{G_i}(z)$  is a top-hat distribution for galaxies in the redshift bin  $i$  and  $p_{\mu_j}(z)$  is the efficiency of weak lensing effect for lenses at  $z$  and sources at  $z_j$  (following the notation in Gaztanaga et al. (2012)). The coefficient  $b_i$  is the effective galaxy bias at  $z_i$  and  $\alpha_j \equiv 2.5s_j - 1$  is the amplitude of the weak lensing magnification effect (with  $s_j$  the slope of the galaxy number counts at the flux limit of the sample at  $z_j$ ). The first equation above has 4 terms corresponding to galaxy-galaxy (intrinsic clustering), galaxy-magnification, magnification-galaxy and magnification-magnification correlation. In our test, we use  $\alpha_i = 1$  and  $b_i = 1$  to generate the starting point for  $c_{ij}$ . We then apply the following transformation:

$$\begin{matrix} B \times B \\ ij \end{matrix} \quad b_B(z_i)b_B(z_j) \quad c_{ij} \quad (3.21)$$

$$\begin{matrix} F \times F \\ ij \end{matrix} \quad b_F(z_i)b_F(z_j) \quad c_{ij} \quad (3.22)$$

$$\begin{matrix} B \times F \\ ij \end{matrix} \quad b_B(z_i)b_F(z_j) \quad c_{ij} \quad (3.23)$$

depending on which galaxy samples we are cross-correlating (FS or BS), where

$$b_B(z_i) = 2 + 2(z_i - 0.5) \quad (3.24)$$

$$b_F(z_i) = 1.2 + 0.4(z_i - 0.5) \quad (3.25)$$

are the biases in the BS and FS respectively in the photo- $z$  bin  $i$  (at mean redshift  $z_i$ ). This corresponds to using linear bias for the intrinsic correlation (as in Gaztanaga

et al. (2012)) and some particular evolving slope ( $s_i \sim 1$ ) for the magnification cross-correlations. In Fig. 3.16 we show  $\frac{B \times B}{ij}$  (left),  $\frac{F \times F}{ij}$  (middle) and  $\frac{B \times F}{ij}$  (right) before (top) and after (bottom) being transformed by the migration matrices  $r_{ij}$  in Fig. 3.15 through Eq. (3.16). Note that for the  $B \times B$  and  $F \times F$  cases the correlations matrices are symmetrical. Also note that the redshift ranges in both samples are different, so that the  $B \times F$  correlation matrix is not squared.

In the top panels, we can see the intrinsic galaxy-galaxy clustering in the diagonal of the matrix, which has an amplitude of order unity and decreases rapidly to zero for separated redshift bins. The galaxy-magnification correlation appears as a diffused off-diagonal cloud with an amplitude  $< 0.05$  (clear colors). The magnification-magnification contribution is negligible. In the bottom panel, we see the effect of the photo- $z$  migration. The diagonal (auto-correlations) becomes thicker and diluted. The off-diagonal galaxy-magnification cloud becomes more diffused, specially for the FS. The cross-correlation  $F \times B$  produces results that are intermediate between  $B \times B$  and  $F \times F$ .

We can invert the migration matrices to go from the bottom panels (which are the observations  $^{-A \times B}$ ) to the top panels (i.e. true correlations  $^{A \times B}$ ) by inverting Eq. (3.16):

$$^{A \times B} = r_A^{-1} \cdot ^{-A \times B} \cdot (r_B^T)^{-1} \quad (3.26)$$

This should work perfectly well if we can calibrate the  $r_{ij}$  matrices properly. For a large fiducial survey with about 5000 sq. deg. we need about 1% accuracy in  $r_{ij}$  (Gaztanaga et al., 2012). In practice, the accuracy of the above reconstruction can be used to put requirements on the photo- $z$  calibration.

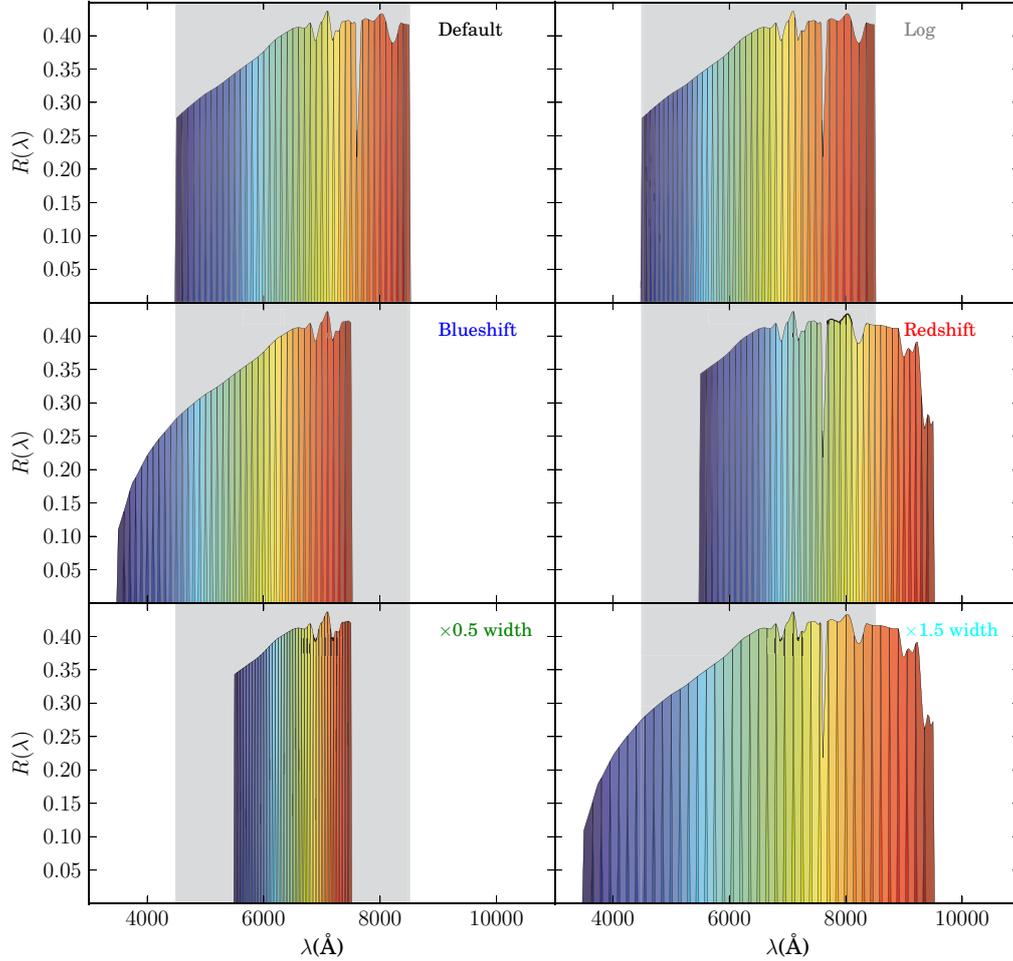
## 3.4 Optimization of the PAU filter set

In this section we want to explore how the photo- $z$  performance changes under variations of the PAU NB filter set.

### 3.4.1 NB filter set variations

We study five variations of the original NB filter set whose response is shown in Fig. 3.17. All the variations conserve the number of filters. In the order that appear in Fig. 3.17, the proposed filter sets are:

- **Default:** This is the default filter set already shown in Fig. 3.2.
- **Log:** In this filter set, band widths increase in wavelength logarithmically, so that they fulfill  $\eta_0 \Delta\eta = const$ , where  $\Delta\eta$  is the width of the rectangular part of the band (without taking into account the lateral wings) and  $\eta_0$  is the central wavelength of the band. We impose the overall wavelength range covered by the set of bands to be the same as for the **Default** filter set. Given that the total number of bands is kept at 40, we obtain that the bluest filter has a width of 97Å, while the reddest is 159Å wide. The reason for this filter set is that, when spectra are redshifted, their spectral features are moved to



**Figure 3.17:** On the top-left, the original PAU NB filter set (same as in Fig. 3.2). The rest are the five variations to be compared in terms of photo- $z$  performance. Grayed areas show the covered wavelength range by the Default filter set. In descending order from left to right we have: the **Log** filter set with the same overall range as the **Default** but with band widths that increase logarithmically; the **Blueshift** filter set, which is the same as the **Default** but with the bands shifted  $1000\text{\AA}$  towards bluer wavelengths; the **Redshift** filter set which is the same as **Default** but shifted towards redder wavelengths; the  $\times 0.5$  width filter set whose band widths are half those of the **Default** ones; and the  $\times 1.5$  width filter set whose bands are 1.5 times wider. The overall wavelength ranges of these last two set-ups are chosen to be centered respect to the range of the **Default** set.

**Table 3.6:** Global photo- $z$  performance results for each filter set shown in Fig. 3.17. Photo- $z$  performance is characterized through the three metrics: bias (median),  $\sigma_z$  ( $\sigma_{68}$ ) and the  $3\sigma$ -outlier fraction. Photo- $z$  quality cuts resulting in a 50% overall completeness are applied in all cases. We show results for the Bright Sample (BS) and Faint Sample (FS).

	Bias	$\sigma_z$ (%)	Outliers(%)
Default BS	$-0.71 \cdot 10^{-4}$	0.34	3.02
Default FS	$-1.33 \cdot 10^{-3}$	4.73	7.36
Blueshift BS	$-2.11 \cdot 10^{-4}$	0.38	3.23
Blueshift FS	$-3.11 \cdot 10^{-3}$	5.19	7.05
Redshift BS	$-0.74 \cdot 10^{-4}$	0.35	3.31
Redshift FS	$-0.65 \cdot 10^{-3}$	4.99	7.21
Log BS	$-0.69 \cdot 10^{-4}$	0.35	2.80
Log FS	$-1.46 \cdot 10^{-3}$	4.73	7.43
x1.5 width BS	$-3.18 \cdot 10^{-4}$	0.45	3.00
x1.5 width FS	$-2.76 \cdot 10^{-3}$	3.87	7.76
x0.5 width BS	$-0.00 \cdot 10^{-4}$	0.32	5.22
x0.5 width FS	$-0.99 \cdot 10^{-3}$	6.91	5.52

redder wavelengths, but also their widths are stretched as  $\Delta\eta = (1+z)\Delta\eta$ . If the photo- $z$  determination depends strongly on the tracking of any spectral feature, such as the 4000Å break in elliptical galaxies, a filter set like **Log** will continue to enclose the same part of the spectral feature in a single feature independently of how redshifted is the spectrum.

- **Blueshift:** This is the same as the **Default** filter set, however bands have been shifted 1000Å towards bluer wavelengths. We expect to get better photo- $z$  performance at low redshift and for late-type galaxies. The down side of this filter set is that the overall response turns out to be very inefficient in the ultraviolet zone (middle-left of Fig. 3.17), like it was for the  $u$  band.
- **Redshift:** This is the same variation as before but shifting bands 1000Å towards redder wavelengths. We expect to get better photo- $z$  performance at high redshift and for early-type galaxies. This filter set does not suffer from the problem of the ultraviolet, so its band responses are much more uniform over the covered range (middle-right of Fig. 3.17). On the other hand, the sky brightness on this region is higher.
- **x0.5 width:** This is a filter set whose band widths are half of the **Default** ones. Lateral wings are also reduced to half of their size, from 25Å to 12.5Å, in order to avoid an excessive overlap between adjacent bands. We expect to improve the photo- $z$  precision, at least for galaxies with good Signal-to-Noise ratio on their photometry. The down side of this filter set is that, since the number of bands is kept, the overall wavelength range covered is also reduced by a half. We choose it to be centered with respect to the **Default**, so that it covers from 5500Å to 7500Å, roughly spanning only from the bluest filter

of the `Redshift` set to the reddest of the `Blueshift` set. This can lead to a degradation of the photo- $z$ s at very low and high redshift, although the broad bands may attenuate this effect.

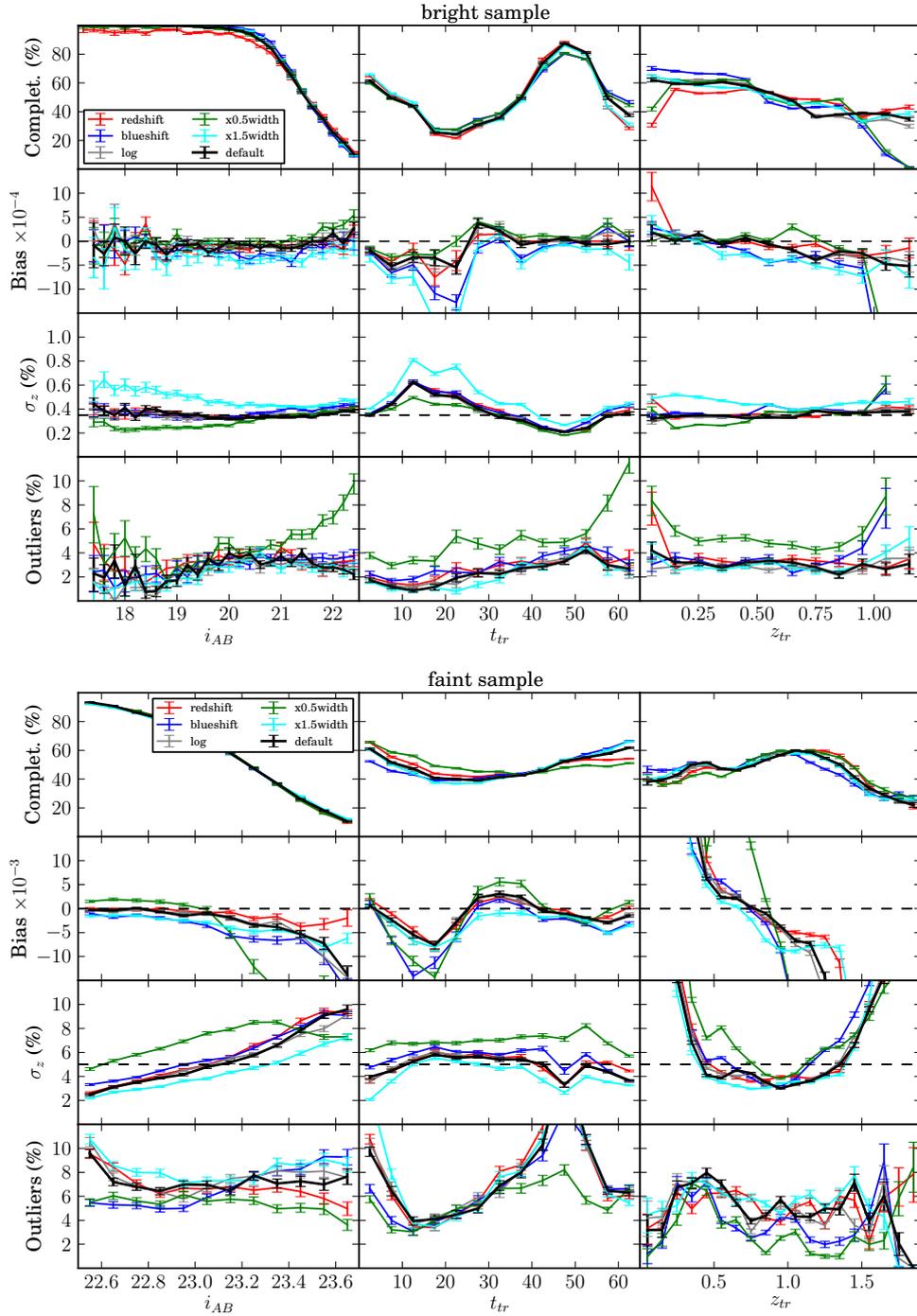
- **$\times 1.5$  width:** This is a filter set whose band widths are 1.5 times wider than the `Default` ones. Because of this, we expect a significant degradation of the photo- $z$  precision for galaxies with good Signal-to-Noise ratio on their photometry. However, the increase in  $S/N$  may help. Moreover, the covered wavelength range also increases by 50%. We choose the new range to be centered with respect to the `Default` set, so that it covers from 3500Å to 9500Å, roughly from the bluest edge of the `Blueshift` set to the reddest edge of the `Redshift` set, so we expect to see a more uniform photo- $z$  performance over the whole redshift range.

We generate magnitudes for each filter set band as described in Section 3.2 using the same exposure times per NB filter tray and BB as in the `Default` filter set. We do not try to optimize the exposure times for each filter set. The aim of this study is to see how, in spite of this, the photo- $z$  performance changes. Once the new photometric mock catalogs are created, they are also split into a Bright Sample ( $i_{AB} < 22.5$ ) and Faint Sample ( $22.5 < i_{AB} < 23.7$ ). We run BPZ on each catalog using the same settings as for the `Default` filter set. There is no need to calibrate a different prior for each filter set, since the prior was initially calibrated on the broad band  $i$ , which is shared by all these filter sets. Photo- $z$  quality cuts resulting in an overall completeness of  $\sim 50\%$  are applied in all cases.

### 3.4.2 Global photo- $z$ performance

Global photo- $z$  performance results for each filter set are shown on Table 3.6, using the same metrics as in Section 4.2: bias (median),  $\vec{\tau}_z$  ( $\vec{\tau}_{68}$ ) and the 3 $\vec{\tau}$ -outlier fraction. We find that the  $\times 0.5$  width set gives the best bias (it completely vanishes), and  $\vec{\tau}_z$  ( $\sim 6\%$  better than `Default`) in the BS, while in the FS it is the `Redshift` set which gives the best bias ( $\sim 54\%$  better) and the  $\times 1.5$  width set which gives the best  $\vec{\tau}_z$  ( $\sim 18\%$  better). On the other hand, the  $\times 1.5$  width set gives the worst bias (a factor 4.6 worse) and  $\vec{\tau}_z$  ( $\sim 32\%$  worse) in the BS, while in the FS it is the `Blueshift` set which gives the worst bias (a factor 2.4 worse) and the  $\times 0.5$  width which gives the worst  $\vec{\tau}_z$  ( $\sim 46\%$  worse). Regarding the outlier fraction, its direct comparison is trickier since it depends on the value of  $\vec{\tau}_z$ . Even so, we see that in the BS the `Log` set gives the best value, while in the FS the  $\times 0.5$  width set gives the worst.

The general conclusions are that the `Log` set gives almost the same photo- $z$  performance as the `Default` set, with a slight increase of 3% in  $\vec{\tau}_z$  in the BS. Therefore, we see that the logarithmic broadening of the band widths does not suppose any global improvement. On the other hand, and as we expected, if the Signal-to-Noise ratio in the photometry is good enough, the narrower the bands, the better the photo- $z$  performance results. On the other hand, wider bands are the ones that give better photo- $z$  precision in the FS, because there are more bands that pass the cut  $\vec{\tau}_m < 0.5$  introduced in Section 3.2.



**Figure 3.18:** Photo- $z$  performance metrics using the different filter sets shown in Fig. 3.17. Rows: Completeness after applying a photo- $z$  quality cut leading to a 50% global completeness; bias (median);  $\sigma_z$  ( $\sigma_{68}$ ) and  $3\sigma$ -outlier fraction, as a function of  $i_{AB}$ ,  $t_{tr}$  and  $z_{tr}$  (columns), in the BS (top) and the FS (bottom).

### 3.4.3 Results as a function of $i_{AB}$ , $t_{tr}$ and $z_{tr}$

In Fig. 3.18 we show plots similar to those in Figs. 3.12 and 3.13 with the photo- $z$  performance metrics as a function of  $i_{AB}$ ,  $t_{tr}$  and  $z_{tr}$  for the BS (top) and the FS (bottom) when using the different filter sets of Fig. 3.17. Black curves correspond to the photo- $z$  results of the **Default** filter set when the 50% completeness photo- $z$  quality cut is applied, and we will treat them as the reference results. The rest of curves in different colors correspond to the variations of the **Default** filter set. As a general trend, we see that these curves do not deviate much from the reference. Even so, we will discuss each case separately.

In the BS, we see that redshifting the bands (red curves) slightly degrades the completeness at low magnitudes up to  $i_{AB} < 21.5$ . As was expected, all the metrics also degrade at low  $z_{tr}$ . In contrast, blueshifting the bands (blue curves) shows the opposite behavior, a degradation of all the metrics at high  $z_{tr}$ . This is due to the lack of coverage at blue and red wavelengths respectively of each filter set, as we have already mentioned before. Something similar happens for the  **$\times 0.5$  width** set, where band widths, and consequently the covered wavelength range, is reduced by a half (green curve). The resulting photo- $z$  performance is worse at both low and high  $z_{tr}$ . However, the photo- $z$  precision  $\vec{\tau}_z$  is slightly better at intermediate redshifts ( $0.15 < z_{tr} < 0.5$ ), for spiral galaxies ( $10 < t_{tr} < 30$ ) and at bright magnitudes ( $i_{AB} < 20.5$ ). Increasing the band width by a factor 1.5 (cyan curve) does not result in an improvement in any case. Bias and  $\vec{\tau}_z$  degrade all over the range of the three variables,  $i_{AB}$ ,  $t_{tr}$  and  $z_{tr}$ . This is in full agreement with the results shown in Table 3.6, where this filter set was seen as giving the worst photo- $z$  performance. Also in agreement with Table 3.6, we see that the **Log** filter set practically does not introduce any change from the **Default** filter set.

In the FS we do not observe big differences between the completeness curves, but for example the **Blueshift** filter set shows slightly lower completeness for elliptical galaxies than for irregulars, unlike the **Redshift** and  **$\times 0.5$  width** filter sets, which show the opposite behavior. In the  $z_{tr}$  range we also recognize similar behaviors as in the BS, as for example the fact that the **Blueshift** filter set shows better completeness at low  $z_{tr}$  and worse at high, as well as the opposite behavior of the **Redshift** and  **$\times 0.5$  width** filter sets. The **Blueshift** filter set seems to cause a significant degradation in the bias for faint, spiral and high redshift galaxies, and also delivers a considerably worse  $\vec{\tau}_z$  than the **Default** over all the ranges. In return, the **Redshift** filter set shows better bias at high magnitudes and redshifts. On the other hand, we observe that the  **$\times 0.5$  width** filter set shows much more pronounced trends on the bias, in particular at  $i_{AB} > 23.1$ , spiral galaxies and over all the  $z_{tr}$  range, where values are substantially worse than for the **Default** filter set. As in Table 3.6, we observe that the worst  $\vec{\tau}_z$  is found for the  **$\times 0.5$  width** filter set, while the best is for the  **$\times 1.5$  width** filter set over all the ranges. This is exactly the opposite to the behavior seen for the BS. In general, narrower bands are useful in the BS, but not in the FS.

## 3.5 Discussion and Conclusions

In the previous sections, we have seen that, at the level of simulated data, a photo- $z$  precision of  $\sigma_z \sim 0.0035(1+z)$  can be achieved for  $\sim 50\%$  of galaxies at  $i_{AB} < 22.5$  by using a photometric filter system of 40 narrow bands of  $125\text{\AA}$  width together with the  $ugrizY$  broad bands. The precision degrades to  $\sigma_z \sim 0.05(1+z)$  when we move to the magnitude range  $22.5 < i_{AB} < 23.7$ . These coincide with the two photo- $z$  precision requirements defined in Gaztanaga et al. (2012) needed to simultaneously measure Redshift Space Distortions (RSD) and Magnification bias (MAG) on two samples, one on the foreground and one on the background, over the same area of the sky. The galaxies removed are the ones with the worst photo- $z$  quality according to our photo- $z$  algorithm used. In chapter 4 it is shown that this kind of cuts, when they remove a substantial fraction of galaxies, can grossly bias the measured galaxy clustering. However, in the same chapter we propose a way to correct for it.

On the other hand, we found that spiral galaxies are the ones that give the worst photo- $z$  performance. Moreover, quality cuts mostly remove them. Contrary to what was assumed in Benítez et al. (2009), elliptical galaxies do not provide the best photo- $z$  performance, but irregular galaxies with prominent emission lines at  $\sim 3737\text{\AA}$  [OII] and  $\sim 5000\text{\AA}$  [OIII] are actually the ones that give the best performance. A possibility is that these two emission lines are better traced by the narrow bands than a single feature as the  $4000\text{\AA}$  break of elliptical galaxies, making the photo- $z$  determination more robust.

We also studied the effect of including the 40 narrow bands in a typical broad band filter set  $ugrizY$ . We find that the  $\Delta z$  ( $1+z_{tr}$ ) distributions become more peaky around the maximum moving away from Gaussianity. Bias improves by an order of magnitude. Precision also improves in a factor of  $\sim 5$  below  $i_{AB} \sim 22.5$ . However, the low signal-to-noise in narrow bands makes the improvement very small within  $22.5 < i_{AB} < 23.7$ , concluding that narrow bands at faint magnitudes are useful to improve the bias but not the precision.

We have also estimated the photo- $z$  migration matrices  $r_{ij}$  which correspond to the probability that a galaxy observed at the photo- $z$  bin  $i$  is actually at the true redshift bin  $j$ . These are shown in Fig. 3.15 for both the BS and FS. We then show (in Fig. 3.16) how this photo- $z$  migration matrix  $r$  distorts the observed auto and cross-correlation of galaxies in narrow redshift bins. We show results for both the intrinsic clustering, which dominates the diagonal in the measured angular cross-correlation matrix,  $\xi$ , and the magnification effect, which appears as a diffused off-diagonal cloud in  $\xi_{ij}$ . The true cross-correlation matrix  $\xi_{ij}$  can be obtained from the inverse migration  $r^{-1}$  with a simple matrix operation:  $\xi_{ij} = r^{-1} \cdot \xi \cdot (r^T)^{-1}$ . This is a standard deconvolutive problem, and it is only limited by how well we know the migration matrix.

We find that introducing slight variations on the 40 narrow band filter set, such as: shifting bands to the higher/lower wavelengths, narrowing/broadening or increasing logarithmically band widths, do not introduce significant changes on the final photo- $z$  performance. Even so, general trends are that narrowing/broadening bands improve/worsens the photo- $z$  performance below/above  $i_{AB} \sim 22.5$ , while red/blueshifting bands improve the photo- $z$  performance at high/low redshifts or

the quality-cuts efficiency for early/late-type galaxies. Logarithmically growing band widths do not turn into any measurable improvement. Therefore, we conclude that the initial proposed filter set of 40 narrow bands seems to be close to optimal for the purposes of the PAU Survey at the WHT.

Additionally, we have also tried doubling the exposure times in all bands. The magnitude limits become  $\sim 0.5$  mag deeper, so that the new bright sample goes down to  $i_{AB} < 23$  and the new faint sample down to  $i_{AB} < 24.1$ . Within these magnitude limits, the photo- $z$  performance remains almost unchanged, and we still reach the requirement in  $\tau_{68}$  after removing  $\sim 50\%$  of the galaxies in both the new bright and the new faint catalogs. On the other hand, if we stick to the initial magnitude limits while still doubling the exposure times, we find that in this case the requirements are reached by removing only 30% of the galaxies in the bright sample and none in the faint.

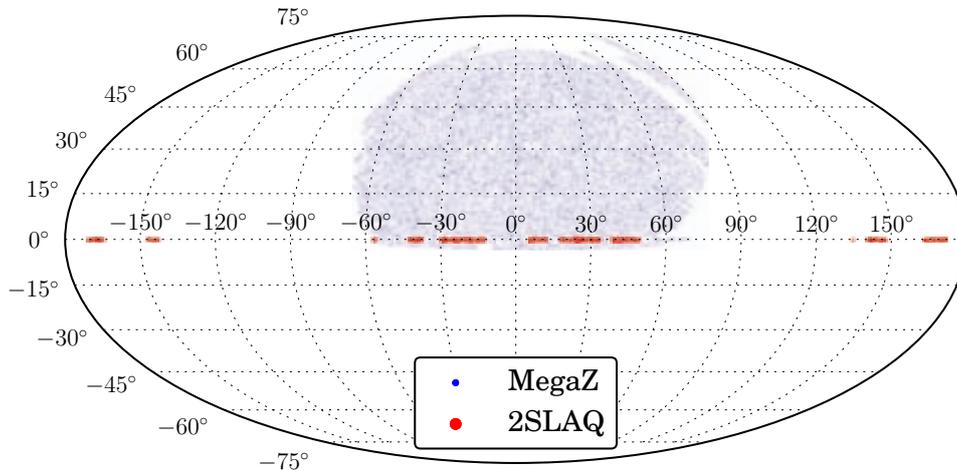
# Chapter 4

## Photo- $z$ quality cuts and their impact on the measured galaxy clustering

Most if not all photo- $z$  algorithms provide not only a best estimate for the galaxy redshift but also an estimate of the quality of its determination, be it simply an error estimation or something more sophisticated like the *odds* parameter in the BPZ code (Benitez, 2000). Applying cuts on the value of this quality parameter, one can clean up the sample from galaxies with an unreliable photo- $z$  determination (Benitez, 2000), or even select a smaller sample of galaxies with significantly higher photo- $z$  precision (chapter 3).

However, we will show in this chapter that these quality cuts can affect very significantly the observed clustering of galaxies in the sample retained after the cuts, thereby biasing the cosmological information that can be obtained. Therefore, this effect needs to be corrected. Fortunately, this can be readily achieved using a technique similar to the one presented in Ho et al. (2012) to deal with, among others, the effect of stars contaminating a galaxy sample.

The outline of the chapter is as follows. Section 4.1 discusses the galaxy samples that we use in our study: the Mega-Z photometric galaxy sample (Collister et al., 2007) and its companion, the 2SLAQ spectroscopic sample (Cannon et al., 2006), which we use to characterize the redshift distribution of the galaxies in Mega-Z. We will also describe the photo- $z$  algorithm used (BPZ) and the resulting redshift distributions in four photometric redshift bins in Mega-Z. In section 4.3 we present the measurement of the galaxy-galaxy angular correlations within, and cross-correlations across, the four photo- $z$  bins in Mega-Z, comparisons with the theoretical expectations, and the effects of applying several photo- $z$  quality cuts to the data. Section 4.4 introduces the correction we have devised for the effect of the photo- $z$  quality cut and applies it to the data, resulting in corrected angular correlation functions that we then compare with the predictions. In section 4.5 we extract the Baryon Acoustic Oscillation (BAO) angular scale from the corrected data and compare it with the result obtained without any photo- $z$  quality cuts. Finally, in section 4.6 we discuss the relevance of our results, particularly for previous studies that applied photo- $z$  quality cuts while ignoring their effects on clustering, and we offer some conclusions.



**Figure 4.1:** Mega-Z and 2SLAQ maps in Mollweide projection plotted in blue and red respectively. For the sake of clarity only a hundred thousand galaxies randomly selected from Mega-Z and five thousand from 2SLAQ have been plotted. The Mega-Z sample covers a total of 7750 deg<sup>2</sup>, while 2SLAQ covers 180 deg<sup>2</sup>. The 2SLAQ area is divided into several fields inside a 2°-wide strip that extends along the celestial equator.

## 4.1 Data Samples

The Mega-Z LRG DR7<sup>1</sup> catalog (Collister et al., 2007) includes  $\sim 1.4$  million Luminous Red Galaxies from the SDSS Data Release 7 in the redshift range  $0.4 < z < 0.7$ , with limiting magnitude  $i_{AB} < 20$ . It covers an area of  $\sim 7750$  deg<sup>2</sup> of the sky that is displayed in Fig. 4.1. This is the sample in which we will investigate the effect of the photo- $z$  quality cuts on the observed galaxy clustering.

In order to calibrate the photometric redshifts of the Mega-Z galaxies, a representative galaxy sample with known redshifts is needed. Fortunately, such a sample exists: the 2dF-SDSS LRG and Quasar<sup>2</sup> (2SLAQ) catalog (Cannon et al., 2006) was obtained using the same selection criteria as the Mega-Z catalog and includes  $\sim 13100$  LRGs with spectroscopic redshifts. Its sky coverage of only 180 deg<sup>2</sup> can be seen in Fig. 4.1 in red. The galaxies are located on a strip of 2 deg along the celestial equator, the area subtended by the 2dF spectrograph. Only non-repeated objects ( $ind = 1$ ) with high spectroscopic redshift confidence level ( $hqs \geq 3$ ) are used in this analysis.

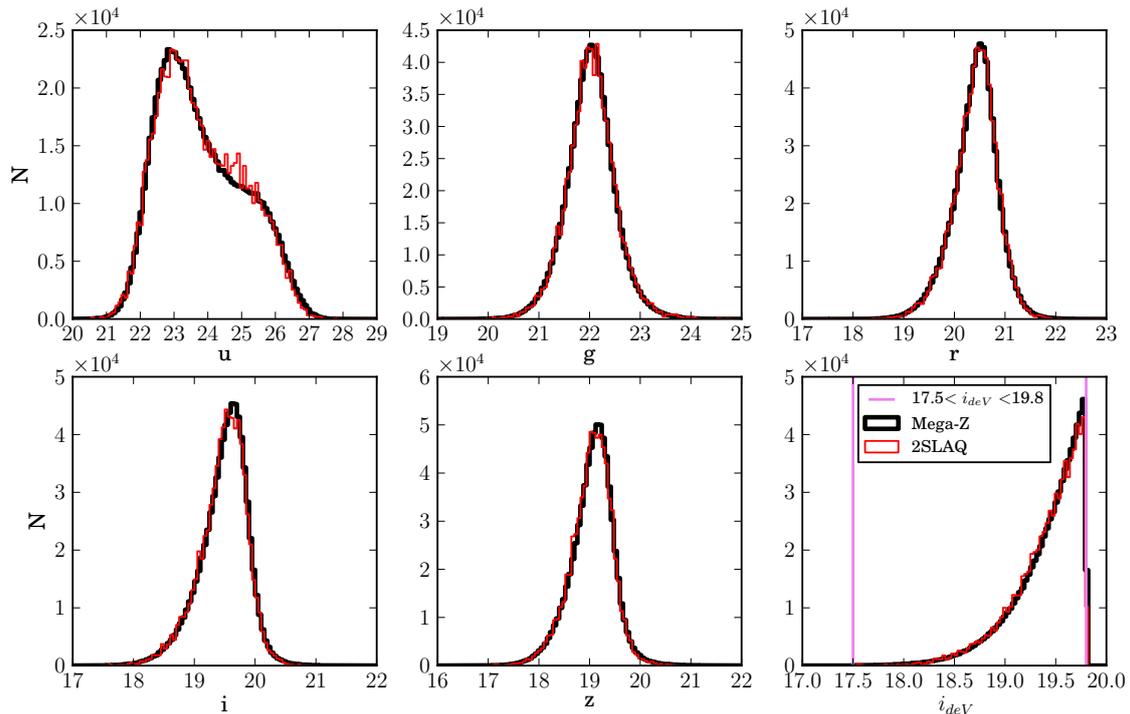
The selection criteria in both catalogs consist of a magnitude cut and several color cuts. All magnitudes have been corrected for galactic extinction; we use model magnitudes for the color cuts and to compute the photometric redshifts (section 4.2). The magnitude cut

$$17.5 < i_{deV} < 19.8 \quad (4.1)$$

is motivated by the limiting magnitude of the 2dF spectrograph and to ensure completeness of the 2SLAQ catalog. While the Mega-Z catalog is complete up to mag-

<sup>1</sup>An ASCII version of the Mega-Z LRG DR7 catalog can be found at <http://zuserver2.star.ucl.ac.uk/~sat/Mega-Z/Mega-ZDR7.tar.gz>.

<sup>2</sup>The whole 2SLAQ data can be downloaded from [http://www.2slaq.info/query/2slaq\\_LRG\\_webcat\\_hdr.txt](http://www.2slaq.info/query/2slaq_LRG_webcat_hdr.txt).



**Figure 4.2:** From top-left to bottom-right, the model magnitude distributions in the *ugriz* bands for 2SLAQ in red and Mega-Z in black, normalized to each other. The last plot corresponds to the *i* band de Vaucouleurs magnitude. All magnitudes are corrected for galactic extinction. The purple lines in the last plot show the magnitude cut in (4.1), which is the nominal limiting magnitude for the 2SLAQ sample. The agreement is excellent, except in the *u* band, where the low signal-to-noise produces a small disagreement around magnitude  $\sim 25$ .

nitude  $i_{deV} = 20$ , the 2SLAQ completeness drops off sharply beyond  $i_{deV} = 19.8$ , so the cut forces us to cut the Mega-Z sample at this limit. It eliminates  $\sim 32\%$  of the Mega-Z galaxies leaving a total of  $\sim 950000$ . The  $i_{deV}$  magnitude distribution for both samples is plotted on the bottom-right of Fig. 4.2. The purple lines represent the cut in (4.1).

The color cuts applied are:

$$0.5 < g - r < 3 \quad (4.2)$$

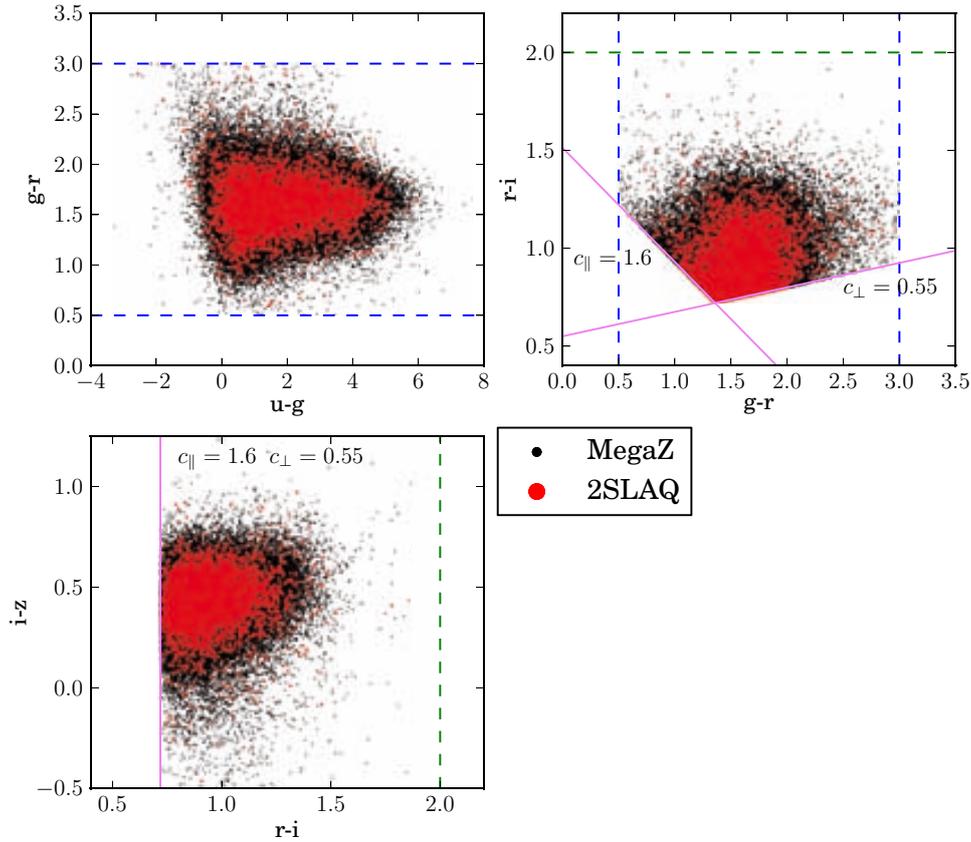
$$r - i < 2 \quad (4.3)$$

$$c \equiv 0.7(g - r) + 1.2(r - i - 0.18) > 1.6 \quad (4.4)$$

$$d \equiv (r - i) - (g - r) - 0.8 > 0.55 \quad (4.5)$$

They are used to isolate the LRGs from the rest of galaxies. In particular, (4.4) separates later-type galaxies from LRGs, and (4.5) acts as an implicit photo- $z$  cut of  $z \gtrsim 0.45$ , as we will see in the next section. In Collister et al. (2007) the  $d$  cut is set to 0.5 for the Mega-Z catalog, but, once again, the 2SLAQ completeness within  $0.5 < d < 0.55$  is very poor: once all other cuts are applied, they represent 3.6% of the galaxies, instead of 27% in Mega-Z. Therefore, we choose  $d > 0.55$ .

These magnitude and color cuts leave a total of 749152 objects in the Mega-Z



**Figure 4.3:** Color-color diagrams for both the Mega-Z catalog in black and the 2SLAQ catalog in red. For the sake of clarity, we have only plot a hundred-thousand galaxies for Mega-Z and five thousand for 2SLAQ. We can see that 2SLAQ covers the same color area as Mega-Z, so that we can conclude that it is a good representative spectroscopic sample. The blue and green dashed lines show the color cuts in (4.2) and (4.3) respectively, while the purple solid lines show the cuts in (4.4) and (4.5), used to select LRGs and high- $z$  galaxies, respectively. The two last cuts shown in the top-right plot translate into an implicit cut of  $r - i > 0.72$  in the bottom plot.

catalog and 11810 in the 2SLAQ catalog. In Figs. 4.2 and 4.3, we plot the model magnitude distributions and the color-color scatters, respectively, for all the  $ugriz$  bands, after applying all cuts. 2SLAQ is shown in red and Mega-Z in black. Solid and dashed lines represent the cuts. The 2SLAQ magnitude distributions have been normalized up to the total amount of galaxies in Mega-Z. The agreement between both catalogs is excellent, so that we can conclude that 2SLAQ is a representative spectroscopic sample of Mega-Z.

Additionally, some extra cuts have been applied in the Mega-Z catalog to reduce the star contamination:

$$i_{psf} - i_{model} > 0.2 \times (21.0 - i_{dev}) \quad (4.6)$$

$$i\text{-band de Vaucouleurs radius} > 0.2 \quad (4.7)$$

$$\delta_{sg} > 0.2 \quad (4.8)$$

As explained in Collister et al. (2007), the first two cuts separate galaxies from stars

leaving a residual  $\sim 5\%$  contamination of M-type stars, which cannot be trivially separated either using *gri* colors or through cuts on the subtended angular diameter. Because of this, the last cut is applied. First, the photo-*z* neural network ANNz (Collister & Lahav, 2004) is trained on the 2SLAQ catalog that contains reliable information about whether objects are stars or galaxies. The trained network is then run on the whole Mega-Z catalog to compute the probability  $\delta_{sg}$  that objects be galaxies. Removing all objects with probability below 0.2 reduces the stellar contamination from 5% to 2% (Collister et al., 2007). In 2SLAQ, we can remove stars by simply getting rid of all those objects with redshift less than 0.01.

## 4.2 Photometric redshifts

We will perform the galaxy clustering study in several photometric redshift (photo-*z*) bins. Therefore, we will need to estimate the photo-*z* of each galaxy in the Mega-Z sample. Furthermore, the theoretical predictions for the clustering need the true-redshift distribution of the galaxies in each photo-*z* bin  $i$ ,  $N_i(z)$ , which we can obtain from the 2SLAQ sample. For this purpose, we need to compute photometric redshifts of the 2SLAQ galaxies, split them into several photo-*z* bins, and, finally, recover the spectroscopic redshift distribution in each bin. Additionally, we will study the photo-*z* performance using 2SLAQ, and apply photo-*z* quality cuts to improve it. The impact of these cuts on  $N_i(z)$  will also be studied.

We use the Bayesian Photometric Redshifts<sup>3</sup> (BPZ) template-fitting code described in Benitez (2000) to compute the photometric redshift of galaxies in both catalogs. It uses Bayesian statistics to produce a posterior probability density function  $p(z | m_i)$  that a galaxy is at redshift  $z$  when its magnitudes are  $m_i$ :

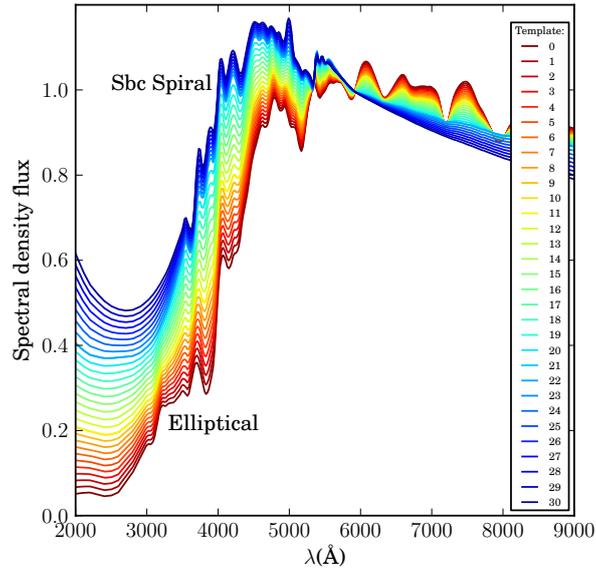
$$p(z | m_i) = \int_t L(m_i | z, t) \Pi(z | t | m_i) dt \quad (4.9)$$

where  $L(m_i | z, t)$  is the likelihood that the galaxy has magnitudes  $m_i$ , if its redshift is  $z$  and its spectral type  $t$ , and  $\Pi(z | t | m_i)$  is the prior probability that the galaxy has redshift  $z$  and spectral type  $t$ . Finally, the photometric redshift  $z(\text{phot})$  of the galaxy will be taken as the position of the maximum of  $p(z | m_i)$ .

Each spectral type  $t$  can be represented by a galaxy template. BPZ includes its own template library, but we prefer to use the new CWW library from LePhare<sup>4</sup>, another template-based photo-*z* code described in Arnouts et al. (1999), Ilbert et al. (2006). Both libraries are based on Coleman, Wu & Weedman (1980), Kinney et al. (1996), but BPZ contains only 8 templates compared to 66 in LePhare. The large number of templates allows us to focus on the LRG templates, which correspond to the genuine galaxy type of our catalogs. In particular, we select four: E11\_01, E11\_09, E11\_19 and Sbc\_06, and then we create nine interpolated templates between consecutive templates, giving a total of 31 templates, shown in Fig. 4.4.

<sup>3</sup>BPZ can be found at <http://www.its.caltech.edu/~coe/BPZ/>.

<sup>4</sup>The new CWW library can be found in the folder /lephare\_dev/sed/GAL/CE\_NEW/ of the LePhare package at [http://www.cfht.hawaii.edu/~arnouts/LEPHARE/DOWNLOAD/lephare\\_dev\\_v2.2.tar.gz](http://www.cfht.hawaii.edu/~arnouts/LEPHARE/DOWNLOAD/lephare_dev_v2.2.tar.gz).



**Figure 4.4:** The spectral galaxy templates used in the determination of the 2SLAQ and Mega-Z photo-zs. There are a total of 31 templates that range from elliptical galaxies, the vast majority in both samples, to Sbc spiral galaxies.

Template-based photo- $z$  codes require the knowledge of the filter band-passes of the survey instrument in order to compute the predicted photometry that will be compared with the observations to produce the likelihood  $L(m_i | z, t)$ . The SDSS instrument carries five broad-band filters,  $ugriz$ , described in Fukugita et al. (1996), whose throughputs are obtained from <http://home.fnal.gov/~annis/astrophys/filters/filters.new>.

A crucial point of BPZ is the prior probability  $\Pi(z | t, m)$  that helps improve the photo- $z$  performance. Benitez (2000) proposes the following empirical function:

$$\Pi(z | t, m) = f_t e^{-k_t(m-m_0)} \cdot z^{\alpha_t} \exp\left(-\left[\frac{z}{z_{mt}(m)}\right]^{\alpha_t}\right) \quad (4.10)$$

where  $z_{mt}(m) = z_{0t} + k_{mt}(m - m_0)$ . Every spectral type  $t$  has associated a set of five parameters  $f, k, \alpha, z_0, k_m$  that determine the shape of the prior. These parameters are determined from the spectroscopic data themselves. In principle, we could assign one prior  $\Pi(z | t, m)$  to each one of the 31 templates in Fig. 4.4, but this would give a total of  $31 \times 5 = 155$  parameters, too many for the spectroscopic data sample available. Instead, we split the 31 templates in two groups:  $t = 1$ , with the 10 first templates that make up the group of pure elliptical galaxies, and  $t = 2$  with the rest. Then, running BPZ on 2SLAQ a first time, without priors, we find the group each galaxy belongs to. Fitting (4.10) to this output, together with the spectroscopic redshifts and the observed magnitudes in one band (we choose the  $i$  band), we find the values of the prior parameters. Results are given in Table 4.1.

The  $k$  parameters are related to the migration of galaxies from one spectral type to another at different magnitudes. The fit gives values of  $k$  very close to 0, so we impose explicitly not having type migration by setting them exactly to 0.  $f$  gives the fraction of galaxies of each type at magnitude  $m_0$ , which we choose to be  $m_0 = 18.5$ .

$t$	$f$	$k$	$\alpha$	$z_0$	$k_m$
1	0.72	0.0	8.679	0.477	0.078
2	0.14	0.0	7.155	0.488	0.064

**Table 4.1:** The values of the prior parameters of (4.10) for the 2SLAQ galaxies.

Since  $k = 0$ , we find that the 72% of the galaxies belong to the spectral type group 1 independently of the magnitude, confirming that most of the galaxies are purely elliptical.

We define galaxies with catastrophic redshift determinations as those with  $\Delta z \equiv z(\text{phot}) - z(\text{spec}) > 1$ , where  $z(\text{spec})$  is the spectroscopic redshift and  $z(\text{phot})$  the photo- $z$ . With the help of the prior, we are able to remove all these catastrophic redshift determinations, which account for  $\sim 4.2\%$  of the 2SLAQ sample. They are typically galaxies with degeneracies in their color space, which cause confusions in the template fit and result in a photo- $z$  much larger than the real redshift. Defining the photo- $z$  precision  $\vec{\sigma}_z$  as half of the symmetric interval that encloses the 68% of the  $\Delta z$  distribution area around the maximum, we also find that its value for the non-catastrophic determinations improves by a factor 1.7, down to  $\vec{\sigma}_z \sim 0.042$ , in agreement with Collister et al. (2007), Padmanabhan et al. (2005), Thomas, Abdalla & Lahav (2011b).

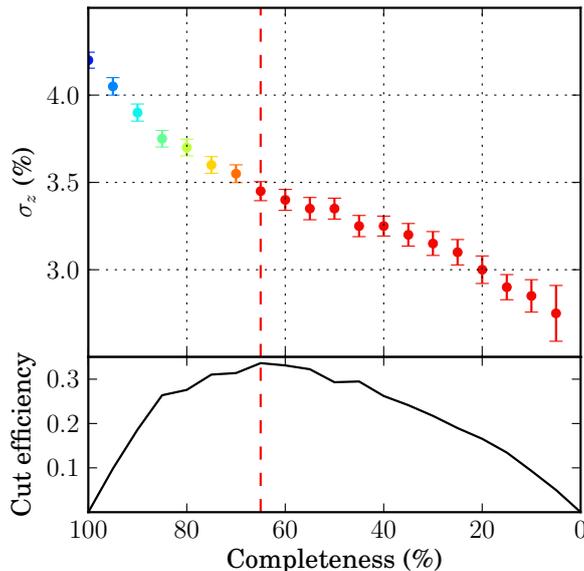
We also apply a cut on the quality of the photometry, consisting of not using any band, for each galaxy, with magnitude error  $> 0.5$ . This cut mostly removes the information from the  $u$  band for many galaxies, since the signal-to-noise tends to be lower in this band. The overall precision improves slightly to  $\vec{\sigma}_z \sim 0.041$ .

Photo- $z$  codes, besides returning the best estimate for the redshift, typically also return an indicator of the photo- $z$  quality. It can be simply an estimation of the error on  $z(\text{phot})$ , or something more complex, but the aim is the same. In BPZ, this indicator is called *odds*, and, it is defined as

$$\text{odds} = \frac{\int_{z(\text{phot})-\delta z}^{z(\text{phot})+\delta z} p(z | m_i) dz}{\int_{z(\text{phot})-\delta z}^{z(\text{phot})+\delta z} p(z | m_i) dz} \quad (4.11)$$

where  $\delta z$  determines the redshift interval where the integral is computed. *Odds* can range from 0 to 1, and the closer to 1, the more reliable is the photo- $z$  determination, since  $p(z | m_i)$  becomes sharper and most of its area is enclosed within  $z(\text{phot}) \pm \delta z$ . In our case, we choose  $\delta z = 0.03$ , which is close to the photo- $z$  precision in 2SLAQ and Mega-Z. A bad choice of  $\delta z$  could lead to the accumulation of all *odds* close to either 0 or 1. Since *odds* are a proxy for the photo- $z$  quality, we should expect a correlation between the *odds* and  $\Delta z$ , in the sense that higher *odds* should correspond to lower  $\Delta z$ . In Fig. 4.5, we show  $\vec{\sigma}_z$  for subsets of the Mega-Z sample with increasingly higher cuts on the *odds* parameter. In fact, the exact *odds* values are quite arbitrary, since they depend on the size of  $\delta z$ . Therefore, we have translated these *odds* cuts into the fraction of the galaxy sample remaining after a certain cut has been applied. The abscissa in Fig. 4.5 corresponds to this completeness for increasingly tighter *odds* cuts.

We see that, by removing the galaxies with low *odds* in steps of 5% in completeness, we are able to reduce the photo- $z$  dispersion from  $\vec{\sigma}_z \sim 0.042$  to  $\sim 0.028$ , a



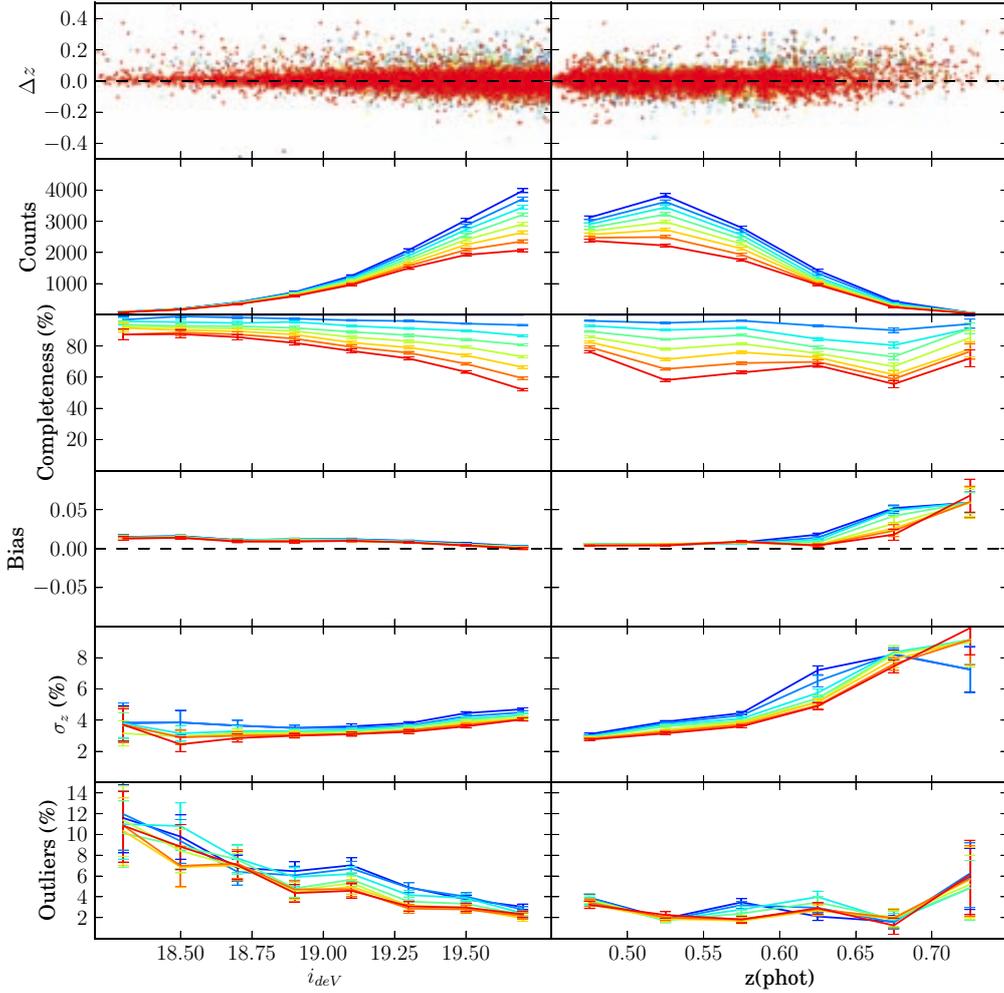
**Figure 4.5:** Top plot: the 2SLAQ photo- $z$  precision  $\sigma_z$  for different photo- $z$  quality cuts resulting in the completeness shown on the  $x$  axis. The error bars are computed using bootstrap (Efron, 1979). The color scale labels the different photo- $z$  quality cuts, here and also in Figs. 4.6 and 4.8. The nominal precision, without any *odds* cut, is 0.042. It can be improved by a factor 1.5 when the most aggressive cut, which leaves only 5% of the galaxies, is applied. However, on the bottom plot, we see that the most efficient cut (defined in (4.12)) is at 65% completeness, where by removing 35% of the galaxies we achieve 50% of the improvement, with  $\sigma_z \sim 0.035$ . Only cuts with completeness  $\geq 65\%$  are considered in the following.

factor of 1.5. Obviously, the best accuracy is obtained when the completeness is close to 0%, but this is very inefficient. Defining the efficiency of the cut as:

$$\text{Cut Efficiency } (x) = x \left[ \frac{\vec{\sigma}_z(100\%) - \vec{\sigma}_z(x)}{\vec{\sigma}_z(100\%) - \vec{\sigma}_z(0\%)} \right] \quad (4.12)$$

where  $x$  is the completeness of the catalog after the cut, we find that the most efficient photo- $z$  quality cut is at 65% of completeness, where  $\vec{\sigma}_z \sim 0.035$ , as shown in the bottom plot in Fig. 4.5. Since we cannot compute  $\vec{\sigma}_z(0\%)$  for lack of galaxies, we use instead in (4.12) the value at 5% completeness. From now on, we will refer to photo- $z$  quality cuts as all those in Fig. 4.5 that lead to a completeness between 100% and 65%, and which are labeled in different colors.

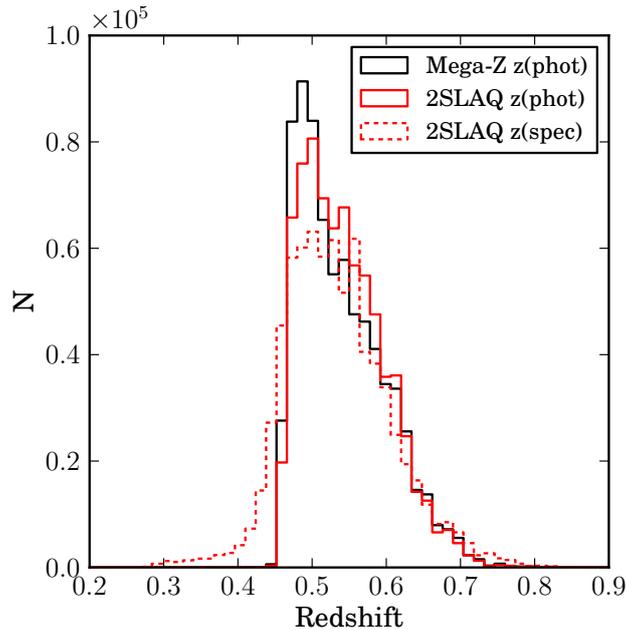
An exhaustive analysis of the photo- $z$  results is shown in Fig. 4.6. It consists of a series of plots where different statistical properties of  $\Delta z$ , in the rows, are shown as a function of two different variables, in the columns:  $i_{deV}$  magnitude on the left, and the photo- $z$  estimation,  $z(\text{phot})$  on the right. In the first row, we can see the  $\Delta z$  scatter from which the rest of plots are derived. As in Fig. 4.5, the color progression of the curves from blue to red corresponds to the different photo- $z$  quality cuts with completeness going from 100% to 65%, the most efficient cut, in steps of 5%. The number of galaxies, in the second row, grows for increasing magnitudes, but drops for increasing redshifts. The completeness, in the third row, drops at high magnitudes,



**Figure 4.6:** Statistics showing the 2SLAQ photo- $z$  performance. In the first row we show the scatter of  $\Delta z \equiv z(\text{phot}) - z(\text{spec})$  with respect to the  $i_{deV}$  magnitude (left) and  $z(\text{phot})$  (right). The scatter has been binned along these two variables and some statistical estimators have been computed in each bin. In descending order of rows, we show the galaxy population (in counts), the completeness, the bias (median), the photo- $z$  precision  $\sigma_z$  and the  $3\sigma$  outlier fraction. The color degradation from red to blue is the same as in Fig. 4.5 and labels different photo- $z$  quality cuts.

especially when quality cuts become harder. It is quite constant along redshift. The bias (median), in the fourth row, shows a general offset of  $\sim 0.02$ , so that photo- $z$ s are in general  $\sim 4\%$  larger than the actual redshift. The photo- $z$  precision  $\sigma_z$ , in the fifth row, degrades slightly for fainter galaxies, while it does by a factor of almost 3 for high- $z$  galaxies. As in Fig. 4.5, the harder the photo- $z$  quality cut, the better precision we get. Finally, the last estimator is the outlier fraction, defined as the fraction of galaxies with  $|\Delta z|$  above three times  $\sigma_z$ . It decreases from 10% to 3% for increasing magnitudes, while it keeps constant around 3% along redshift. The photo- $z$  quality cuts help reduce it at some magnitudes and redshifts.

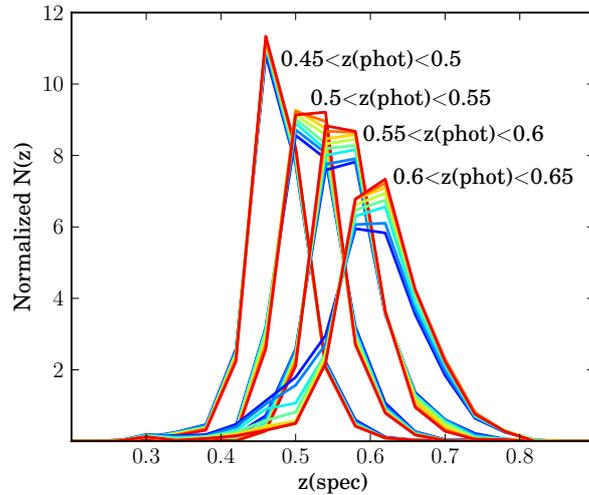
In Fig. 4.7, we have compared the 2SLAQ photo- $z$  distribution with the spectroscopic redshift distribution. Both distributions are clearly different at low red-



**Figure 4.7:** The photo- $z$  distribution of Mega-Z (black), and the spectroscopic (red dashed) and photometric (red solid) redshift distributions of 2SLAQ. The 2SLAQ distributions have been normalized to the Mega-Z number of galaxies for comparison.

shift. While the photo- $z$  distribution rises very sharply from  $z \sim 0.45$ , reaching the maximum immediately, the spectroscopic distribution rises much more gradually from  $z \sim 0.25$ . This is because the color cut in (4.5) acts as a photo- $z$  cut at  $z(\text{phot}) \gtrsim 0.45$ . Figure 4.7 also contains the photo- $z$  distribution of the Mega-Z galaxies. It closely resembles the photo- $z$  distribution in 2SLAQ, but they are not as similar as the magnitude distributions in Fig. 4.2.

Finally, we split the 2SLAQ and Mega-Z catalogs into four photo- $z$  bins of equal width 0.05. The width has been chosen to roughly match the photo- $z$  precision of the catalogs. We want to know the actual redshift distributions inside each of these photo- $z$  bins in order to make the predictions for clustering in the next section. For this purpose, we use the spectroscopic information in 2SLAQ. In Fig. 4.8 we show the spectroscopic redshift distributions of these four photo- $z$  bins in 2SLAQ at the different photo- $z$  quality cuts of Fig. 4.5, using the same color labeling. All the distributions have been normalized in order to compare them. As expected, the distributions become wider in the higher photo- $z$  bins. This is in agreement with the increasing  $\vec{z}$  with redshift seen in Fig. 4.6. On the other hand, photo- $z$  quality cuts tend to reduce the width of the distributions. For instance, the left tail of the last bin, at  $0.6 < z(\text{phot}) < 0.65$ , is drastically reduced when the most efficient photo- $z$  quality cut is applied.



**Figure 4.8:** The spectroscopic redshift distributions in the 2SLAQ catalog for the four photo-z bins in which we will measure galaxy clustering in section 4.3. They are all normalized to the same area under the curves. Different colors label different photo-z quality cuts, as in Figs. 4.5 and 4.6.

### 4.3 Galaxy clustering and the effect of the photo-z quality cuts

We will now compute the angular galaxy correlations in the four photo-z bins of Fig. 4.8, before and after applying the different photo-z quality cuts of Fig. 4.5. We want to see and characterize the impact of these cuts on clustering.

For this purpose, we use the Hierarchical Equal Area isoLatitude Pixelization (Healpix) framework (Gorski et al., 2005), developed for CMB data analysis. It provides pixelations of the sphere with pixels of equal area, with their centers forming a ring of equal latitude. The resolution of the grid is expressed by the parameter  $N_{side}$ , which defines the number of divisions along the side of a base-resolution pixel that is needed to reach a desired high-resolution partition. The total number of pixels in the sphere is given by  $N_{pix} = 12N_{side}^2$ . We will use  $N_{side} = 256$  for our maps, which divides the sphere into 786432 pixels.

Unlike some CMB maps, the galaxy maps do not usually cover the whole sky, so we need to define a mask for the observed area. Moreover, some surveyed regions, for some technical reasons, sometimes become deprived of galaxies. This may cause systematic distortions on the measured galaxy clustering if they are not removed from the mask. We construct the mask in several steps:

- First, the geometry of the Mega-Z sample is inferred by populating a low resolution Healpix map of  $N_{side} = 64$  with all the Mega-Z galaxies. All pixels with less than 65 galaxies per pixel are rejected. The threshold is chosen in order to cut out a long, low amplitude tail in the distribution of number of objects per pixel. This low resolution mask has the outline of the SDSS footprint, and also throws away underpopulated sky areas, which can be seen in Fig. 4.9 as small black patches inside the Mega-Z area. These regions include

data with poor quality or completeness. Some of the patches will be removed inappropriately, since they may be underpopulated due to normal fluctuations in the number counts of galaxies. We have tested that changes in the cut value used induce differences on the measured correlations that are small compared to the size of the correlation errors.

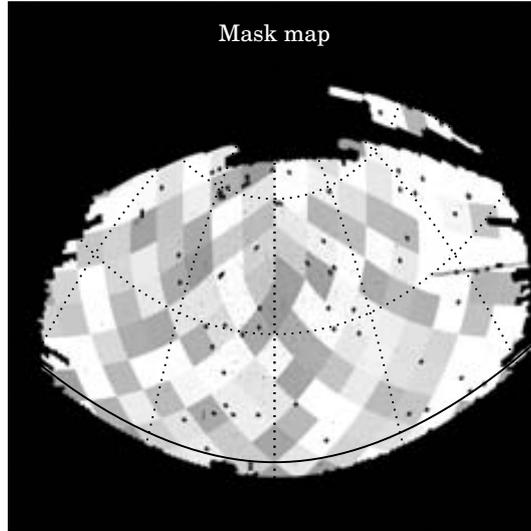
- Second, there are some areas in the SDSS footprint that have poor quality data, but are smaller than a Healpix pixel with  $N_{side} = 64$ . To eliminate these, we download 50 million stars from the SDSS database<sup>5</sup> with magnitude down to 19.6. The star catalog is much more spatially dense than the Mega-Z catalog; therefore, we can construct a mask of the SDSS footprint using the stars in the same manner, but with better resolution, than with the Mega-Z galaxies. We construct a map of the stars with  $N_{side}=512$ , and declare pixels as bad if they have less than 7 stars per pixel. This throws away bad regions such as the long thin horizontal stripe in the right side of Fig. 4.9. It also throws away some pixels at high galactic latitude (black dots in the center) that may lack stars due to normal fluctuations in star counts. However, the density of stars should be unrelated to the positions of the Mega-Z galaxies, and therefore throwing away a small area with the lowest stellar density should not bias measurements of the galaxy correlation function.
- Finally, we reduce the resolution of the mask to  $N_{side} = 256$ , which is the resolution that we will use in our galaxy maps.

More details and justification for computing the mask as described can be found in Cabré & Gaztanaga (2009).

Once we have the mask, we create the galaxy maps shown at the top of Fig. 4.10 and the *odds* maps shown at the top of Fig. 4.11. The galaxy maps are created by counting the number of galaxies that fall in each pixel of the mask, while the *odds* maps are created by averaging the *odds* of these galaxies in the pixel. When no galaxies fall in a pixel, we still need an *odds* value in that pixel, so we take the average value in the neighboring pixels within a circle of  $1^\circ$  radius. At the working resolution, this is a total of 19 pixels, enough so that at least one contains some galaxies, even in the last bin  $0.6 < z < 0.65$ , where the average number of galaxies per pixel is  $\sim 0.36$  when the 65% completeness cut is applied. We also create galaxy maps after applying each photo- $z$  quality cut defined in Fig. 4.5. On the second row of plots in Fig. 4.10 we show the galaxy maps after applying the most efficient cut with 65% completeness.

The first *odds* map at  $0.45 < z < 0.5$  is clearly redder than the other three, since it contains galaxies with higher photo- $z$  quality. On the contrary, bluer regions mark regions with bad photo- $z$  quality. Note that they are not uniformly distributed on the map. They form a pattern of horizontal strips that cross the entire Mega-Z footprint. This is most noticeable in the bins  $0.45 < z < 0.5$  and  $0.5 < z < 0.55$ . Therefore, when we remove low-*odds* galaxies, we are not taking them uniformly off the map. We can already see this in the galaxy maps of Fig. 4.10 when the *odds* cut is applied. If we focus on the bin  $0.5 < z < 0.55$ , we see that the map shows

<sup>5</sup><http://casjobs.sdss.org/>.



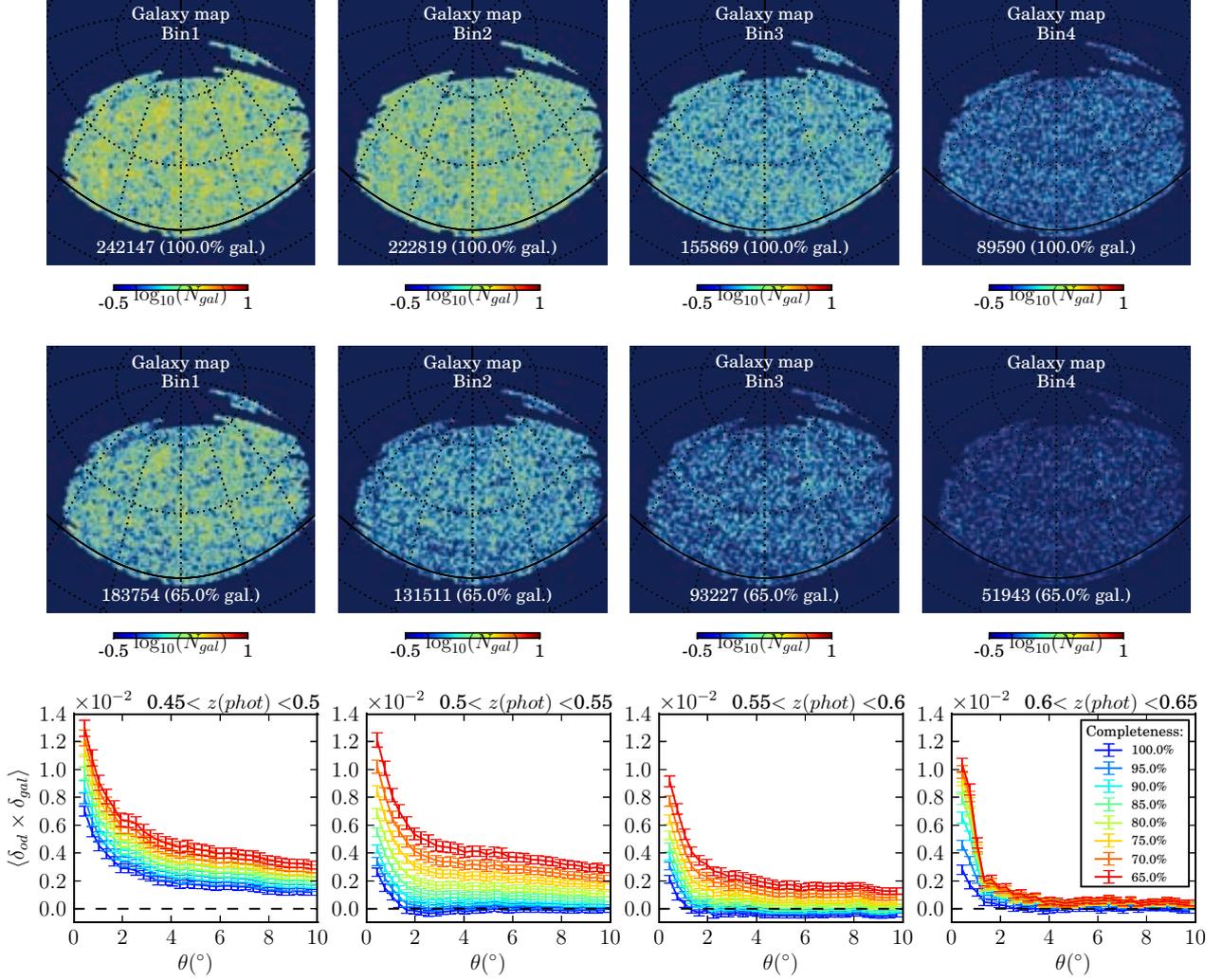
**Figure 4.9:** The Mega-Z DR7 mask in Healpix of  $N_{side}=512$ . It is obtained in two steps. First, using a low resolution Healpix map of  $N_{side} = 64$ , we reject all those pixels with less than 65 galaxies per pixel to get rid of the underpopulated areas (black patches) and obtain the overall geometry of the Mega-Z footprint. Second, we repeat the process with a star map to reject poor data quality regions of even smaller size (black dots and the long thin horizontal stripe on the right side). Different gray levels display the 174 jackknife zones used in (4.14) to compute the covariance of the angular correlations between different scales. They are low resolution pixels of a Healpix map with  $N_{side} = 8$ .

a strip pattern very similar to the bluer zones of its corresponding *odds* map. In Fig. 4 of Crocce et al. (2011) the authors show a map of the mean error on the  $r$  magnitude per pixel of a catalog similar to Mega-Z. Besides the regions with clearly bad photometry due to galactic extinction, regions with low photometric quality in the center of the footprint form patterns very similar to those on our odds maps, with horizontal strips crossing the whole Mega-Z footprint. They approximately coincide with the drift scan paths of the SDSS instrument, and, due to observations done at different nights with different photometric quality of the atmosphere, they have resulted in regions of poor photo-z quality on the sky.

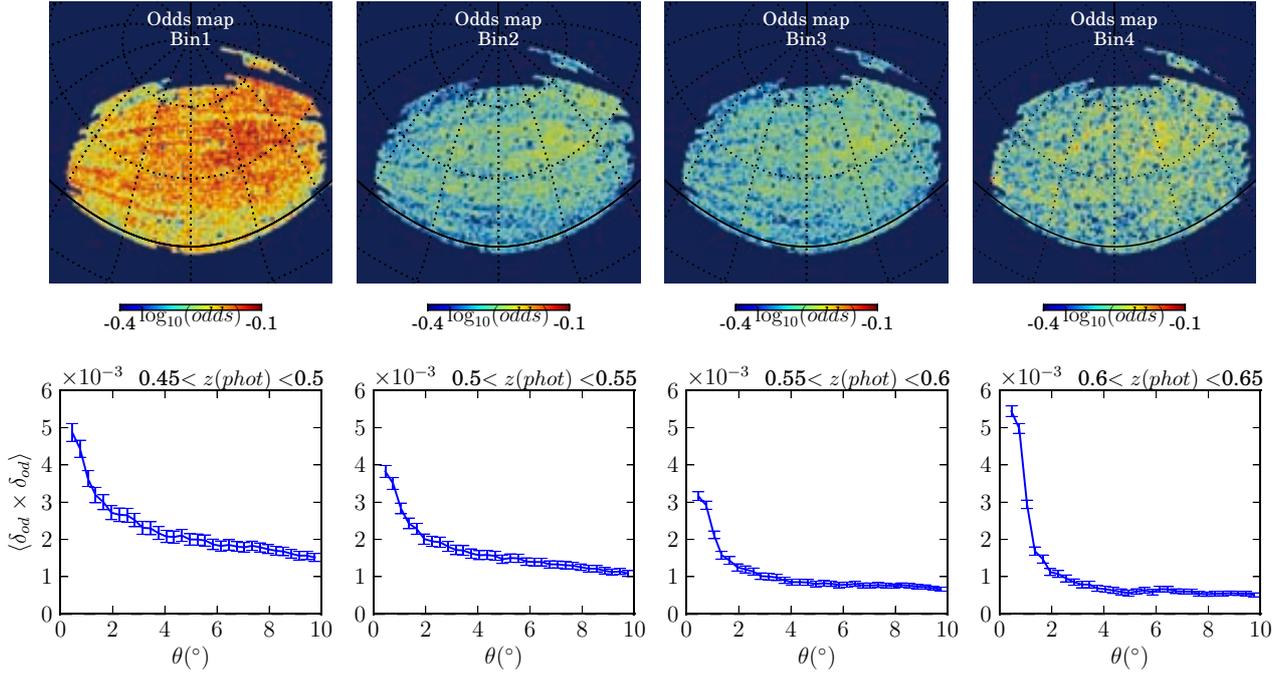
Next we want to compute the angular correlations on all these maps. The angular correlations between two Healpix maps,  $a$  and  $b$ , are given by:

$${}_{ab}(\theta) \equiv \frac{1}{N_\theta} \sum_{i,j}^{N_\theta} \delta_{a,i} \delta_{b,j} \quad (4.13)$$

where  $\delta_{a,i} = a_i - \bar{a}$  is the fluctuation of the map  $a$  at the pixel  $i$  with respect to the mean  $\bar{a}$ , and  $N_\theta$  is the total number of combinations of pixels  $i$  and  $j$  separated by an angular distance between  $\theta$  and  $\theta + \Delta\theta$ . Since the typical pixel resolution when  $N_{side} = 256$  is  $\sim 0.2^\circ$ , we choose  $\Delta\theta$  to be  $0.3^\circ$ . We also want to know the covariance of  ${}_{ab}(\theta)$ . We can compute it using the *jackknife* technique, also used in a similar study in Crocce et al. (2011) and explained in detail in Cabre et al. (2007). This technique consists of splitting the survey area within the mask into



**Figure 4.10:** On the first row, the galaxy maps of the Mega-Z catalog for the four photo- $z$  bins of Fig. 4.8. On the second row, the same after applying a photo- $z$  quality cut with 65% completeness. These are Healpix maps of  $N_{side} = 256$ . The number of galaxies is given in each map. On the bottom row, the angular cross correlations of the galaxy maps with the *odds* maps of Fig. 4.11, at the different photo- $z$  quality cuts of Fig. 4.5. Initially, both maps are not cross-correlated at scales  $> 2^\circ$  (except in the first bin,  $0.45 < z(phot) < 0.5$ ), but the *odds* cut introduce progressively larger cross-correlations between them.



**Figure 4.11:** On the first row, the *odds* maps of the Mega-Z catalog for the four photo-z bins of Fig. 4.8. These are Healpix maps of  $N_{side} = 256$  where the *odds* values per pixel are computed as the mean *odds* of all the galaxies in each pixel. Redder regions are regions with higher photo-z quality. They are not homogeneously distributed over the mask. On the bottom row, the angular auto correlations of the *odds* maps. They are auto correlated at all scales  $< 10^\circ$ , however the strength of these correlations is lower at higher  $z(\text{phot})$ .

$N_k$  sub-areas. We have used pixels of a low resolution Healpix map of  $N_{side} = 8$  as the different jackknife sub-areas. We end up with a total of 174 pixels lying on the mask, represented by different gray levels in Fig. 4.9. The correlations will be computed  $N_k$  times, each time removing each one of the sub-areas. This will result in the jackknife correlations  $\omega_k(\theta)$ . Then, the covariance between the correlation functions measured at angles  $\theta_n$  and  $\theta_m$  will be:

$$\text{Cov}_\omega(\theta_n, \theta_m) = \frac{(N_k - 1)}{(N_k)} \sum_{k=1}^{N_k} [\omega_k(\theta_n) - \bar{\omega}(\theta_n)] [\omega_k(\theta_m) - \bar{\omega}(\theta_m)] , \quad (4.14)$$

where  $\bar{\omega}(\theta) = \sum_k \omega_k(\theta) / N_k$  is the mean of all the jackknife correlations. Therefore, the diagonal errors of the measured correlation at  $\theta$  will be  $\sigma_\omega(\theta) = \sqrt{\text{Cov}_\omega(\theta, \theta)}$ .

On the bottom row of Fig. 4.11 we show the resulting auto-correlations of the *odds* maps. They are not zero at any scale, which confirms that the photo-z quality is not uniformly distributed on the sky. The higher the redshift, the lower the auto correlation. However, the highest auto correlation is reached at scales  $< 1^\circ$  in the fourth bin,  $0.6 < z < 0.65$ .

On the bottom row of Fig. 4.10, we show the resulting galaxy-*odds* cross-correlations, the cross-correlations between the maps on the top of this figure and those on Fig. 4.11. Different curves of different colors label correlations after the different

photo- $z$  quality cuts in Fig. 4.5. The redder curve corresponds to the most efficient cut with 65% completeness. Apart from the first bin  $0.45 < z < 0.5$ , the general tendency is that, initially, there is little correlation between the odds and the galaxies at scales  $> 2^\circ$ , but once the quality cuts are applied, the cross correlations start growing. The harder the cut, the higher the correlations. However, this growth is less noticeable at higher  $z$ . The reason is that, the spatial features of the odds maps become imprinted into the galaxy maps as the *odds* cuts are applied. So, if the auto correlations of the odds maps are small, the strength of this imprinting will also be small. This agrees with the behaviour of the odds auto-correlations in Fig. 4.11. The lower  $z$  bin is unusual in the sense that the galaxy-odds correlations are not initially 0 at any scale. This means that regions with an under or over density of galaxies coincide with regions with the best or worst photo- $z$  quality. This will be discussed in more detail in the following sections. Even so, the correlations still grow when the cuts are applied.

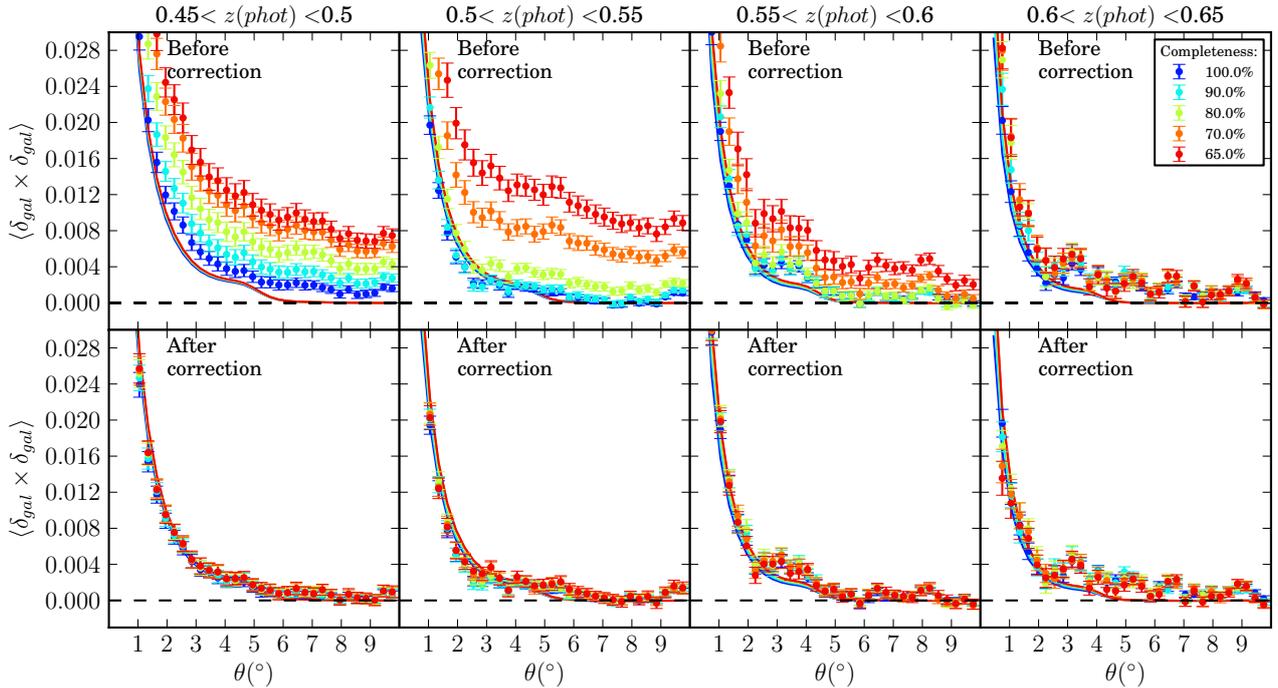
Finally, we compute the angular galaxy auto- and cross-correlation between the four photo- $z$  bins at different photo- $z$  quality cuts. We also compare our measured correlations with predictions. We compute the predictions for the angular correlation  $_{ab}(\theta)$  as described in Crocce et al. (2011):

$$_{ab}^{(theo)}(\theta) = \int dz_1 N_a(z_1) \int dz_2 N_b(z_2) \sigma^s(z_1, z_2, \theta) \quad (4.15)$$

where  $N_i(z)$  are the selection functions, which in our case are the curves in Fig. 4.8,  $\sigma^s(z_1, z_2, \theta)$  is the redshift space correlation of the pairs of galaxies at redshift  $z_1$  and  $z_2$  subtending an angle  $\theta$  with the observer. We use the non-linear power spectrum (Smith et al., 2003) with  $\Lambda$ CDM with  $\Omega_M = 0.25$ ,  $\Omega_\Lambda = 0.75$  and  $H_0 = 70$  (km/s)/Mpc, the linear Kaiser (1984) model of redshift space distortions for the correlation function, and a linear bias model with evolution:  $b(z) = 1.5 + 0.6(z - 0.1)$  (Cabré & Gaztanaga, 2009). We have not included the effect of magnification due to gravitational lensing, which turns out to be negligibly small for this sample. Note that the inclusion of photo- $z$  errors in the predictions is only through the  $N_i(z)$ .

The results for the auto-correlations are on the top row of Fig. 4.12, and those for the cross-correlations are on the top row of Fig. 4.13. Solid lines represent the predicted correlations obtained with (4.15) and points with error bars the measurements using (4.13) and (4.14), respectively. Different colors label different photo- $z$  quality cuts.

Focusing on the auto-correlations, we see that the results before any cut (blue), are slightly above the predicted curves (roughly 1 or 2 $\sigma$ ) in the first three photo- $z$  bins, depending on the angle, and up to  $\sim 3\sigma$  in the last bin. The extra clustering in bin  $0.6 < z(\text{phot}) < 0.65$  is an issue already known from other studies such as Blake, Collister & Lahav (2008), Thomas, Abdalla & Lahav (2011a), and, in Crocce et al. (2011), it is justified by the fact that the number of galaxies in this bin is low enough for systematics to introduce significant distortions on the correlations. The photo- $z$  quality cuts also introduce extra clustering, which is larger the harder the cut, as was seen in the galaxy-odds correlations in Fig. 4.10. Moreover, we see that it is also related to how much the odds are auto-correlated, since the increase of the correlations with the cut is lower at higher photo- $z$  bins, which have smaller



**Figure 4.12:** The angular galaxy auto correlations in the four photo-z bins of Fig. 4.8 with different photo-z quality cuts labeled with colors. The upper plots show the results before applying the *odds* correction in (4.23), while the lower plots show the results after applying it. Points with error bars correspond to measurements and curves to predictions obtained using the  $N_i(z)$  selection functions in Fig. 4.8 in Eq. (4.15).

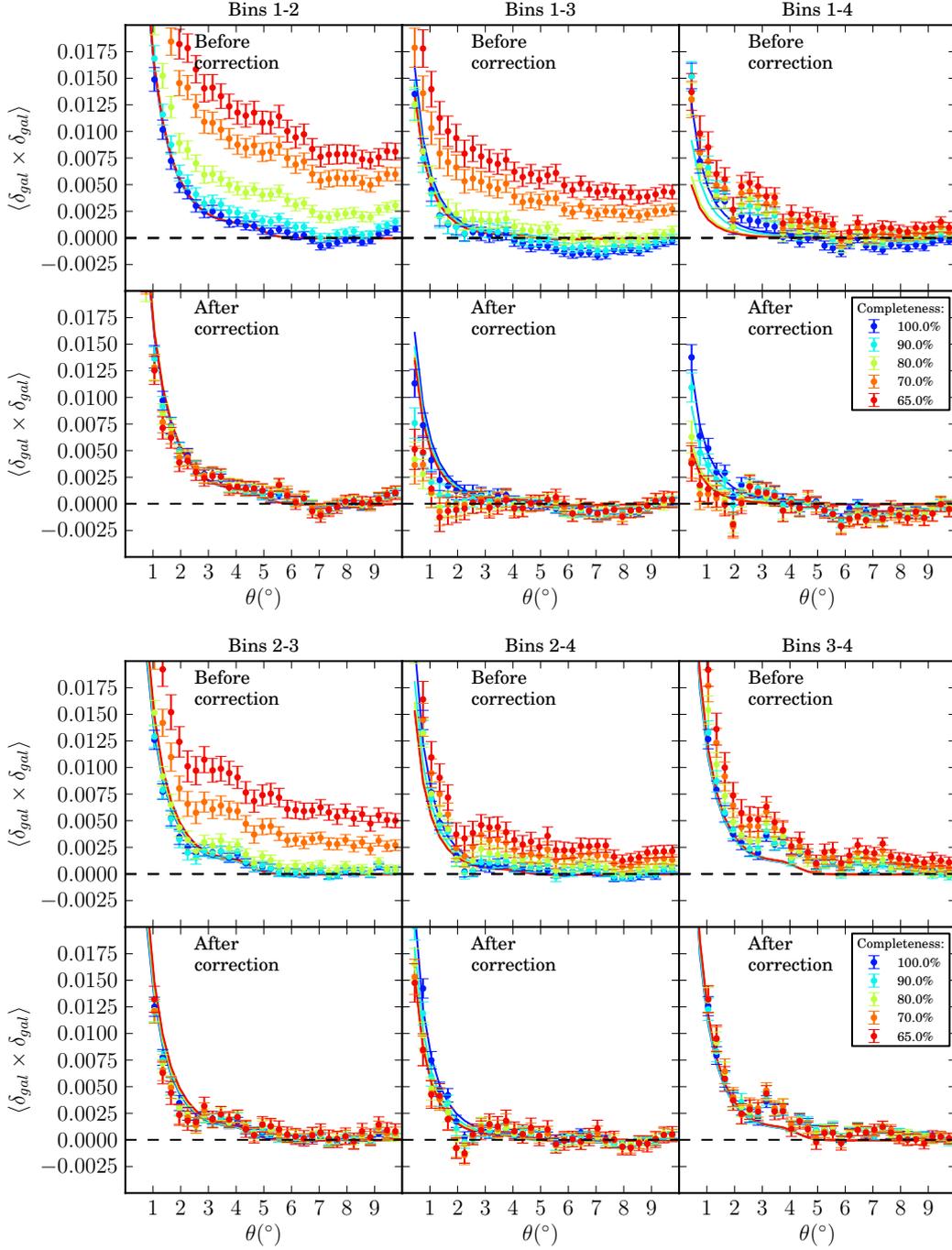
*odds* auto-correlations (Fig. 4.11). However, the correlations in the first bin do not increase as much as in the second bin, even having the most auto-correlated *odds* map. This might be related to the fact that galaxies in this bin are already correlated with the *odds* value before any cut, as we can see in the bottom-left plot of Fig. 4.10, and, in those cases, the extra clustering may not be as additive as for a completely uncorrelated clustering.

The cross-correlations show a similar behaviour, with the photo-z quality cuts introducing extra clustering. This effect is less significant in the cross-correlation of bins 1-4, 2-4 and 3-4 than in those of bins 1-3, 2-3 and 1-2. The reason is that the fourth bin is the one whose *odds* map has the lowest auto correlation.

## 4.4 Correcting the effect of photo-z quality cuts on galaxy clustering

In the previous section we have seen that the photo-z quality cuts introduce extra clustering in the angular galaxy correlations, since they remove galaxies non-homogeneously from the sky. In this section, we want to find a way to correct for it.

We will use the framework presented in Ho et al. (2012) and Ross et al. (2011), which the authors use for the treatment of systematic effects that influence the



**Figure 4.13:** All the possible combinations of the angular galaxy cross correlations between the different photo- $z$  bins of Fig. 4.8 with different photo- $z$  quality cuts labeled with different colors. The upper plots show the results before applying the *odds* correction in (4.22), while the lower plots show the results after applying it. Points with error bars correspond to measurements and curves to predictions obtained using the  $N(z)$  selection functions in Fig. 4.8 in Eq. (4.15).

#### 4.4 Correcting the effect of photo-z quality cuts on galaxy clustering 91

SDSS-III galaxy clustering, such as the stellar contamination, the sky brightness or the image quality of the instrument. Following Ho et al. (2012) and Ross et al. (2011), the density fluctuation  $\delta_i$  of the systematic effect  $i$  modifies the true galaxy density fluctuation  $\delta_g^t$  through a linear contribution modulated by  $\epsilon_i$ , so that the observed galaxy density fluctuation becomes:

$$\delta_g = \delta_g^t + \sum_i \epsilon_i \delta_i \quad (4.16)$$

where the contribution of the systematic effect must be small compared with  $\delta_g^t$ .

Therefore, assuming that there is no intrinsic cross-correlation between the true galaxy fluctuations and the systematic effect,  $\delta_g^t \delta_i = 0$ , the cross-correlation between the observed galaxy density fluctuation and the systematic effect is:

$$\delta_g \delta_i = (\delta_g^t + \sum_j \epsilon_j \delta_j) \delta_i = \epsilon_i \delta_i \delta_i + \sum_{j \neq i} \epsilon_j \delta_j \delta_i \quad (4.17)$$

If we consider that the only systematic effect acting is the *odds* distribution, the previous equations reduce to:

$$\delta_g = \delta_g^t + \epsilon_{od} \delta_{od} \quad (4.18)$$

$$\delta_g^t \delta_{od} = 0 \quad (4.19)$$

$$\delta_g \delta_{od} = \epsilon_{od} \delta_{od} \delta_{od} \quad (4.20)$$

Then, the angular galaxy cross-correlation between two different galaxy maps, 1 and 2, at, for instance, different redshifts, is:

$$\begin{aligned} \delta_{g1} \delta_{g2} &= (\delta_{g1}^t + \epsilon_{od1} \delta_{od1}) (\delta_{g2}^t + \epsilon_{od2} \delta_{od2}) \\ &= \delta_{g1}^t \delta_{g2}^t + \epsilon_{od1} \epsilon_{od2} \delta_{od1} \delta_{od2} \\ &= \delta_{g1}^t \delta_{g2}^t + \frac{\delta_{g1} \delta_{od1}}{\delta_{od1} \delta_{od1}} \frac{\delta_{g2} \delta_{od2}}{\delta_{od2} \delta_{od2}} \delta_{od1} \delta_{od2} \end{aligned} \quad (4.21)$$

where in the first equality we have used (4.18), in the second (4.19) and in the third (4.20). What we measure is the left-hand side of the equation, and, therefore, we need to subtract the second term on the right-hand side to obtain the true values of the correlations.

Therefore, the *odds* correction for the angular cross-correlations of two different galaxy maps is:

$${}^t_{g1 g2}(\theta) = {}_{g1 g2}(\theta) - \frac{{}_{g1 od1}(\theta)}{{}_{od1 od1}(\theta)} \frac{{}_{g2 od2}(\theta)}{{}_{od2 od2}(\theta)} {}_{od1 od2}(\theta) \quad (4.22)$$

where  ${}^t_{g1 g2}(\theta) \equiv \delta_{g1}^t \delta_{g2}^t$  is the true galaxy cross-correlation,  ${}_{g1 g2}(\theta) \equiv \delta_{g1} \delta_{g2}$  is the observed one,  ${}_{g1 od1}(\theta) \equiv \delta_{g1} \delta_{od1}$  is the cross-correlation of the galaxies with the *odds* in map 1 (the same for map 2),  ${}_{od1 od1}(\theta) \equiv \delta_{od1} \delta_{od1}$  is the auto-correlation of the *odds* in map 1 (the same for map 2), and  ${}_{od1 od2}(\theta) \equiv \delta_{od1} \delta_{od2}$  is the cross-correlation between the *odds* maps 1 and 2.

If we were only interested in this correction for auto-correlations Eq. (4.22) would reduce to:

$${}^t_{g g}(\theta) = {}_{g g}(\theta) - \frac{{}_2_{g od}(\theta)}{{}_{od od}(\theta)} \quad (4.23)$$

where  ${}^t_{g g}(\theta) \equiv \delta_g^t \delta_g^t \theta$  is the true galaxy auto-correlation,  ${}_{g g}(\theta) \equiv \delta_g \delta_g \theta$  is the observed one,  ${}_{g od}(\theta) \equiv \delta_g \delta_{od} \theta$  is the galaxy-odds cross-correlation, and  ${}_{od od}(\theta) \equiv \delta_{od} \delta_{od} \theta$  is the odds auto-correlation.

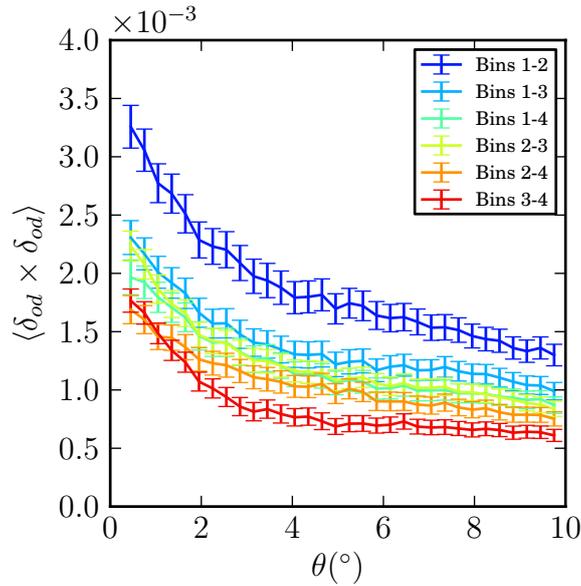
The structure of Eqs. (4.22) and (4.23) is quite intuitive: we have the correlations of the galaxies, shown on the top row of Figs. 4.12 and 4.13, minus the cross-correlations of them with the photo- $z$  quality, shown on the bottom row of Fig. 4.10, properly normalized by the auto-correlations of the *odds*. Both  ${}_{g g}(\theta)$  and  ${}_{g od}(\theta)$  grow when the quality cuts are applied. Therefore, the increase in the auto-correlation will be compensated by the increase in the cross-correlation.

The origin of the galaxy-odds cross-correlation is probably manifold. On the one hand, the *odds* map can be seen as a proxy for other systematic effects, such as sky brightness, seeing, airmass, etc, and correcting for it, even without any photo- $z$  quality cut, could therefore partially correct for these other systematic effects. For this reason, we will also apply these corrections when no cut is applied. On the other hand, in a galaxy catalog containing both early- and late-type galaxies (unlike Mega-Z, which contains almost exclusively early-type galaxies), the *odds* corrections could also reflect the fact that early-type galaxies cluster more than late-type galaxies, and, at the same time, their photo- $z$  quality tends to be better, thereby creating a cross-correlation between galaxy clustering and *odds*.

In Eq. (4.22) the cross-correlations between different *odds* maps,  $\delta_{od1} \delta_{od2}$ , are needed. In our case, those are the maps on the top row of Fig. 4.11. We compute all the cross-correlations and display them on Fig. 4.14. We see that all the *odds* maps are cross-correlated with each other, at least up to angles  $< 10^\circ$ . However, the higher the  $z$ , the lower the correlation. For example, the bins 1-2 are the most cross-correlated, while the bins 3-4 are the least. All the other combinations have similar values.

Finally, we apply the corrections to the measured angular galaxy clustering, before and after applying the photo- $z$  quality cuts. The covariance  $\text{Cov}(\theta_n, \theta_m)$  of the correlation functions after applying the *odds* correction is obtained using Eq. (4.14) after applying the correction to each one of the jackknife correlation function  ${}_{k}(\theta_n)$ . The results are shown on the bottom rows of Fig. 4.12 for auto-correlations and Fig. 4.13 for cross-correlations. First, we see that the corrections work very well. After applying them, all measurements agree, regardless of the photo- $z$  quality cut used in the analysis. Second, we see that the corrections not only work to remove the extra clustering introduced by the quality cuts, they also correct for any intrinsic extra clustering that might be present before any cut. For example, on the left-most plot of Fig. 4.12, the auto-correlations were 1 or 2 $\times$  above the prediction before any cut (blue). After applying the correction, most of the data points agree with their corresponding prediction.

The predictions in Fig. 4.12, and most of those in Fig. 4.13 do not change much after the quality cuts, since they depend through (4.15) on the  $N_i(z)$  functions in Fig. 4.8, which themselves change very little. In fact, the size of the measured error



**Figure 4.14:** All the possible combinations of the angular *odds* cross-correlations between the different maps on the top row of Fig 4.11. They are needed for the *odds* correction in (4.22). We see that all odds maps are cross-correlated with each other, at least up to angles  $< 10^{\circ}$ . However, the higher the  $z$ , the smaller the correlation.

bars are not small enough to distinguish between different predictions. So much so that, although we have been able to correct for the extra clustering introduced by the quality cuts, the cuts themselves do not help in any relevant way in the clustering analysis, and, therefore, in this case there may be no obvious advantage in applying them. Furthermore, the relative errors in the corrected correlation functions can be substantially larger than those in the uncorrected ones: from a few percent larger to almost twice as large, depending on the angular scale, the photo- $z$  bin and the value of the *odds* cut.

Things are different for cross-correlations. The strength of the signal of the cross-correlation between two different photo- $z$  bins is mainly given by the amount of overlap in their  $N_i(z)$ , or, in other words, the fraction of galaxies that are at very similar true redshifts but, due to their photo- $z$  uncertainty, end up in separate photo- $z$  bins. In Fig. 4.8, we saw that the low tail of bin  $0.6 < z < 0.65$ , reduces considerably when the photo- $z$  quality cuts are applied. Consequently, the overlap between this bin and the rest will also reduce, particularly the overlap with the farthest bin,  $0.45 < z < 0.5$ . This should result in differences in the predicted cross-correlations large enough to be distinguished by our measurements. If we look at the cross-correlation of the bins 1-4 on the top-right plot of Fig. 4.13, we see that, at angles  $< 3^{\circ}$  where cross-correlations are not zero, the predicted curves differ more than the size of the error bars in the measurements. Even then, the corrections again put the measurements on top of their corresponding predicted curves. Note that, as for the auto-correlations, this is also true even when no *odds* cut is applied. This may have consequences for methods of photo- $z$  calibration based on the study of the cross-correlations between photometric galaxy samples (Benjamin et al., 2010).

## 4.5 Effects on the extraction of the BAO scale

Having studied in the previous sections the effects of photo- $z$  quality cuts on the measured galaxy auto- and cross-correlations and the way to correct for them, next we want to see how the extraction of the Baryon Acoustic Oscillations (BAO) scale from the measured galaxy auto-correlations in Fig. 4.12 is affected by the photo- $z$  quality cuts and the subsequent correction.

The BAO scale has been proven to be a successful standard ruler to constrain cosmological parameters (Eisenstein et al., 2005, Gaztanaga, Miquel & Sánchez, 2009, Hutsi, 2006, Okumura et al., 2008, Padmanabhan et al., 2007, Percival et al., 2007). It was originated when primordial overdensities on matter caused acoustic (pressure) waves in the photon-baryon fluid that traveled freely across space until photons and baryons decoupled at the drag epoch (when the baryons were released from the Compton drag of the photons (Hu & Sugiyama, 1996))  $z_d = 1059.25 \pm 0.58$  (Ade & Others, 2013), and those waves stopped traveling. At that time, they had traveled  $r_s(z_d) = 147.49 \pm 0.59$  Mpc (Ade & Others, 2013) away from the primordial overdensities, where  $r_s(z_d)$  is the sound horizon scale at the drag epoch. Later, when structure started forming, these waves seeded the formation of galaxies resulting in a small excess in the two-point angular galaxy clustering at the correspondent angular scale:

$$\theta_{BAO}(z) = r_s(z_d) r(z) \quad (4.24)$$

where

$$r(z) = \int_0^z \frac{dz}{H_0 \sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}} \quad (4.25)$$

is the comoving distance at redshift  $z$  that only depends on the Hubble constant  $H_0 = 67.3 \pm 1.2$  (km/s)/Mpc and the fraction of matter  $\Omega_M = 0.315^{+0.016}_{-0.018}$  (Ade & Others, 2013) in a flat  $\Lambda$ CDM cosmological model. In this case  $\Omega_\Lambda = 1 - \Omega_M = 0.685^{+0.018}_{-0.016}$ .

A first detection and measurement of the BAO scale  $\theta_{BAO}$  in angular clustering was presented in Carnero et al. (2012). They made use of a method described in Sánchez et al. (2011) that consists of fitting the empirical function

$$\langle \theta \rangle = A + B\theta^\gamma + C e^{-(\theta - \theta_{FIT})^2 / 2\sigma^2} \quad (4.26)$$

to the observed angular correlation function in a range of angular separations that encloses the BAO peak, where  $A$ ,  $B$ ,  $\gamma$ ,  $C$ ,  $\theta_{FIT}$ ,  $\sigma$  are the parameters of the fit.  $A$  takes into account any possible global offset,  $B$  and  $\gamma$  describe the typical decreasing profile of  $\langle \theta \rangle$ , with  $\gamma$  negative, and the rest is a Gaussian that characterizes the BAO peak. A non-zero value of  $C$  will tell us that the BAO peak has been detected,  $\theta_{FIT}$  is the location of that peak and  $\sigma$  its width. Since angular correlations are measured in redshift bins of finite width due to the intrinsic photo- $z$  dispersion, projection effects can cause a mismatch between  $\theta_{FIT}$  and  $\theta_{BAO}$ . Sánchez et al. (2011) shows that this can be successfully corrected as follows:

$$\theta_{BAO}^{(obs)} = \alpha(z, \Delta z) \theta_{FIT} \quad (4.27)$$

where  $\alpha$  is a factor that only depends on the mean redshift  $z$  and the width  $\Delta z$  of the bin. Figure 3 in Sánchez et al. (2011) shows these dependences. In general

terms, the wider the redshift bin, the larger the shift of  $\theta_{FIT}$  towards smaller angles. For example, the shift is  $\sim 5\%$  of  $\theta_{BAO}$  when  $\Delta z \sim 0.05$ . For higher widths, only bins at  $z \gtrsim 0.5$  continue shifting, while the others saturate.

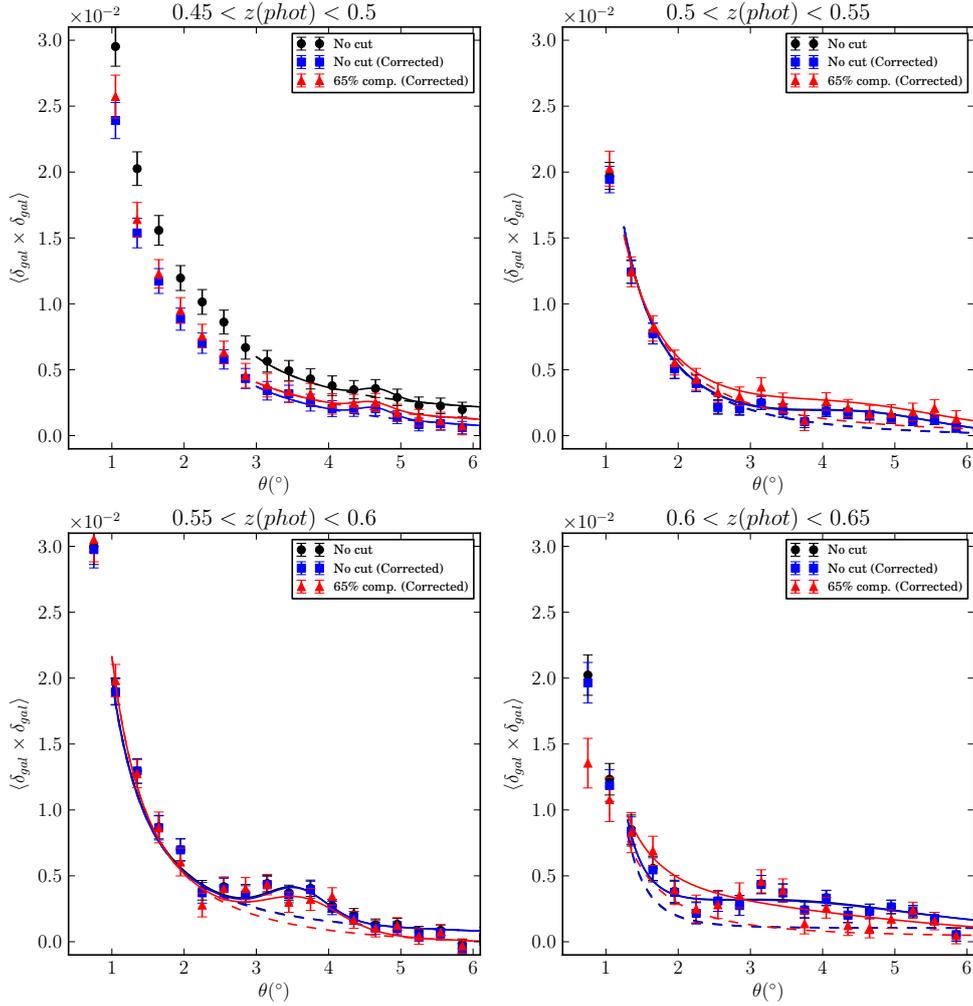
In section 4.2, we chose a width of 0.05 for our photo- $z$  bins in Fig. 4.8. This was in photo- $z$  space. In real (spectroscopic) redshift space, this translates into a FWHM of  $\sim 0.08$ , except in the last bin  $0.6 < z < 0.65$  where, as a consequence of the poorer photo- $z$  performance, the width becomes  $\sim 0.12$ . We find that for those bins the shift in  $\theta_{FIT}$  amounts to  $\sim 7\%$ , except in the last bin, where it is  $\sim 8\%$ .

Finally, we fit (4.26) to the galaxy auto-correlation measurements of Fig. 4.12. The results are shown on Fig. 4.15. Fits are performed in three different cases: when no photo- $z$  quality cut and odds correction are applied (black), when no photo- $z$  quality cut is applied but the correction is (blue), and when the photo- $z$  quality cut of 65% completeness and the correction are applied (red). Error bars correspond to observations and are the same as in Fig. 4.12. Solid and dashed lines correspond to the best fit, but in the dashed the BAO feature has been removed by setting  $C = 0$ . The range of the fit is slightly changed in each photo- $z$  bin to make sure that it completely encloses the BAO peak.

The results of the fits are summarized in Table 4.2. Looking at the parameter  $C$ , we see that we find evidence for the BAO peak (a non-zero value of  $C$ ) in the first three bins, although with different significances in different bins. Typically, applying the photo- $z$  quality cut and its correction (last column in Table 4.2) results in a decrease of the significance of the BAO peak, since 35% of the galaxies are removed from the sample. The first and second column both correspond to the case in which no photo- $z$  quality cut is applied. However, for the results in the second column the *odds* correction has been nevertheless applied using the formulas in (4.23). This will correct for any intrinsic correlation between *odds* value and galaxy position that may create a spurious correlation in the galaxy map. Since higher *odds* may be due to larger signal-to-noise, the *odds* value may be seen as a proxy for airmass, seeing, extinction, etc. Therefore, it may make sense to correct for the *odds* effect even when applying no explicit *odds* cut in order to try to mitigate these effects. A comparison of the results in the second and third columns in Table 4.2 shows that correcting for *odds* when no cut is applied gives very similar results to simply not applying the correction. This is consistent with the findings in Ho et al. (2012) and Ross et al. (2011), where they fail to see significant systematic effects in the galaxy auto-correlations due to differences in seeing, survey depth, extinction, etc.

After applying the quality cut and its correction (last column), we recover values of  $\theta_{BAO}^{obs}$  that are fully consistent with those found without applying an *odds* cut (first and second columns), demonstrating that our correction is consistent. In the last photo- $z$  bin ( $0.60 < z_{phot} < 0.65$ ), however, the photo- $z$  quality cut wipes out the BAO peak. We attribute this to the fact, already mentioned in section 4.3, that in this bin there are fewer galaxies than in the others, and even fewer after removing 35% of them with the quality cut, and moreover, the photo- $z$  performance is worse in this bin.

It is worth mentioning that for some photo- $z$  bins the theoretically-expected position of the BAO peak at the mean redshift in that photo- $z$  bin,  $\theta_{BAO}^{theo}$ , changes slightly when the photo- $z$  quality cut is applied. This is because  $\theta_{BAO}^{theo}$  depends on



**Figure 4.15:** Results of the fits of (4.26) to the galaxy auto-correlations of Fig. 4.12 to extract the BAO scale. In black when no photo- $z$  quality cut is applied, in blue the same but correcting for *odds* and, in red when the 65% completeness cut is applied and corrected. Error bars correspond to observations. Solid and dashed lines correspond both to the best fit, but in the dashed the BAO peak has been removed by setting  $C = 0$  in (4.26).

$z$  (see Eq. (4.24)) and, since the selection functions  $N(z)$  change when quality cuts are applied, the mean redshift may also shift a little, shifting also  $\theta_{BAO}$ .

We can see that in some bins there are some discrepancies between the extracted values of  $\theta_{BAO}^{obs}$  and the expected  $\theta_{BAO}^{theo}$ . This is most likely due to systematic effects that lie beyond the scope of this chapter. The errors quoted in Table 4.2 only contain the statistical uncertainties from the fit. Actually, the results of the fits are rather fragile, showing a significant dependence on details of the fits, such as the exact  $\theta$  range chosen. Fits performed using the whole data covariance matrices (estimated with jack-knife) result in values for the  $\omega^2$  per degree of freedom of order 2, signaling a poor fit quality. The same fits performed using only the diagonal elements of the covariance matrices result in very similar central values for  $\theta_{FIT}$ , but with larger errors and values of the  $\omega^2$  per degree of freedom around 1. The

**Table 4.2:** Results of the BAO fits for the four photo-z bins, under three different conditions: no *odds* cut and no correction, no *odds* cut but correction for the *odds* effect, and finally after applying an *odds* cut that retains 65% of the galaxies and the corresponding correction. The fit parameters are defined in Eq. (4.26). The rows labeled  $\theta_{BAO}^{theo}$  contain the position of the BAO peak at the mean redshift in each bin derived from the latest Planck results (Ade & Others, 2013). The corresponding errors are derived by propagating the errors in the cosmological parameters in (4.24) and (4.25). Since the quality cut slightly changes the true redshift distribution inside each photo-z bin, the expected BAO scale in the bin may also change slightly.

	No odds cut	No odds cut + Correction	Odds cut (65% eff.) + Correction
$0.45 < z < 0.5$			
C	$(0.7 \pm 0.3) \cdot 10^{-3}$	$(0.7 \pm 0.3) \cdot 10^{-3}$	$(0.6 \pm 0.4) \cdot 10^{-3}$
$\theta_{FIT}$	$4.68 \pm 0.09$	$4.65 \pm 0.09$	$4.62 \pm 0.12$
$\theta_{BAO}^{obs}$	$5.03 \pm 0.09$	$5.00 \pm 0.08$	$4.97 \pm 0.12$
$\theta_{BAO}^{theo}$	$4.51 \pm 0.08$	$4.51 \pm 0.08$	$4.50 \pm 0.08$
$0.5 < z < 0.55$			
C	$(1.2 \pm 0.4) \cdot 10^{-3}$	$(1.2 \pm 0.4) \cdot 10^{-3}$	$(1.5 \pm 0.6) \cdot 10^{-3}$
$\theta_{FIT}$	$4.58 \pm 0.33$	$4.59 \pm 0.32$	$4.40 \pm 0.60$
$\theta_{BAO}^{obs}$	$4.92 \pm 0.27$	$4.94 \pm 0.27$	$4.73 \pm 0.41$
$\theta_{BAO}^{theo}$	$4.16 \pm 0.08$	$4.16 \pm 0.08$	$4.16 \pm 0.08$
$0.55 < z < 0.6$			
C	$(2.2 \pm 0.4) \cdot 10^{-3}$	$(2.2 \pm 0.4) \cdot 10^{-3}$	$(2.2 \pm 0.6) \cdot 10^{-3}$
$\theta_{FIT}$	$3.56 \pm 0.06$	$3.57 \pm 0.06$	$3.63 \pm 0.13$
$\theta_{BAO}^{obs}$	$3.83 \pm 0.06$	$3.84 \pm 0.06$	$3.90 \pm 0.10$
$\theta_{BAO}^{theo}$	$3.88 \pm 0.07$	$3.88 \pm 0.07$	$3.88 \pm 0.07$
$0.6 < z < 0.65$			
C	$(2.1 \pm 0.5) \cdot 10^{-3}$	$(2.0 \pm 0.5) \cdot 10^{-3}$	$(2.0 \pm 4.4) \cdot 10^{-3}$
$\theta_{FIT}$	$3.34 \pm 0.72$	$3.30 \pm 0.77$	$1.90 \pm 7.35$
$\theta_{BAO}^{obs}$	$3.63 \pm 0.77$	$3.59 \pm 0.81$	$2.07 \pm 4.29$
$\theta_{BAO}^{theo}$	$3.75 \pm 0.07$	$3.75 \pm 0.07$	$3.64 \pm 0.07$

main point of this section, however, can be qualitatively understood from looking only at the data points in Fig. 4.15, regardless of the fits: the position of the BAO peak does not change when going from data without *odds* cut and without correction to data without *odds* cut but with correction, and then to data with *odds* cut and with correction.

Finally, we have also tried to extract the BAO scale from the uncorrected auto-correlation functions after applying the most stringent *odds* cut. While in some bins, the fit has trouble converging, in those in which it does, the resulting BAO scale is only biased by a few percent, proving again the known fact that the BAO scale is very robust against systematic errors, even those, like this one, that grossly bias the overall shape and normalization of the correlation function. For example, in the bin with  $0.55 < z < 0.60$ , the BAO peak is found at  $\theta_{BAO}^{obs} = (3.91 \pm 0.09)$  deg

when applying the *odds* cut and no correction, to be compared with  $\theta_{BAO}^{obs} = (3.90 \pm 0.10)$  deg, obtained applying the correction.

The main result of this section, however, is not another measurement of the BAO scale with the SDSS LRG sample, but rather the proof that after applying a tight photo- $z$  quality cut that eliminates 35% of the galaxies and severely distorts the shape of the galaxy-galaxy auto-correlation, the correction technique outlined in the previous section delivers a corrected auto-correlation function from which the BAO feature can be extracted without introducing any additional bias.

## 4.6 Discussion and Conclusions

In the previous sections we have seen how photo- $z$  quality cuts, if left uncorrected, can severely bias the measured galaxy angular auto- and cross-correlations. The effect, as seen in Figs. 4.12 and 4.13, consists mostly of a large increase in the correlation across the whole range in angular separation, although slightly more prominent at larger separations. This is not unlike the effect reported in other clustering studies based on very similar samples, such as those in Thomas, Abdalla & Lahav (2011a) and in Crocce et al. (2011). In those papers an excess of clustering has been observed in the photometric redshift bin  $0.6 < z < 0.65$ . In at least one of the papers (Crocce et al., 2011) a photo- $z$  quality cut is performed, eliminating about 16% of the galaxies, but no attempt is made to correct for the possible effect of this cut on the measured correlations. While a quick look at Fig. 4.12 reveals that in our case the effect of the photo- $z$  quality cut is more prominent at lower redshifts, the issue may deserve more thorough study, which lies beyond the scope of this chapter.

It is intriguing to see that, at least in the  $0.45 < z < 0.50$  photo- $z$  bin, there is a correlation between galaxy density and *odds* value even before any cut on the value of the *odds* (bottom-left plot in Fig. 4.10). This correlation then leads to extra galaxy auto- and cross-correlations whenever that first photo- $z$  bin is involved (top-left plot in Fig. 4.12 and top-right plot in Fig. 4.13). The correction method we propose eliminates this extra correlation very effectively (see the corresponding plots in Figs. 4.12 and 4.13), but the question remains: what is it that we are actually eliminating? Or: where does this galaxy-*odds* correlation come from?

One possibility is that it comes from systematic effects in the survey that are otherwise uncorrected: differences in seeing conditions, airmass, extinction, etc. between different areas of the survey will lead to correlated differences in galaxy density and in the value of the *odds* parameter. In general, these systematic effects would result in an additive extra correlation. In this case, correcting for this spurious correlation will mitigate the effects of those systematic issues.

However, in general, another possibility is that the galaxy-*odds* correlation is due to the fact that different galaxy types have different clustering amplitudes (different biases) and, at the same time, also have different mean photo- $z$  precisions, hence different mean *odds*. In this case, the *odds* correction could be removing genuine galaxy-galaxy correlations. Alternatively, one could interpret that the correction would be changing the average galaxy type (and hence bias) of the sample. Hence, this would result in a multiplicative extra correlation.

In our case, since the sample is largely composed of LRGs, the galaxy-*odds*

correlation we observe in the lower photo- $z$  bin is likely due to the uncorrected systematic effects mentioned above, and, therefore, it makes sense to apply the *odds* correction even without an explicit *odds* cut. We observe that, indeed, the extra correlation we observe seems to be additive in nature (i.e. roughly constant as a function of angular scale). Figures 4.12 and 4.13 show that, even with no *odds* cut, the agreement with the predictions improves once the *odds* correction has been applied.

In summary, using the Mega-Z DR7 galaxy sample and the BPZ photometric redshift code, we have shown that applying moderate galaxy photo- $z$  quality cuts may lead to large biases in the measured galaxy auto- and cross-correlations. However, a correction method derived within the framework presented in Ho et al. (2012), Ross et al. (2011) manages to recover the original correlation functions and, in particular, does not bias the extraction of the BAO peak. It remains to be seen whether this correction might eliminate some or all of the excess correlation observed by several groups in galaxy samples essentially identical to the one we use.



# Chapter 5

## Summary and Outlook

Chapters 3 and 4 conform the scientific work of this thesis and both have become in two scientific publications; Martí et al. (2014b) and Martí et al. (2014a) respectively.

In chapter 3 we have shown through simulated data that a photometric system composed of 40 narrow bands of 100 Å width, as in the PAU@WHT survey (Benítez et al., 2009) camera (PAUCam), can deliver photometric redshifts with a precision of  $\sigma_z \sim 0.0035(1+z)$  for the  $\sim 50\%$  of galaxies with the best photo- $z$  quality up to magnitude  $i_{AB} = 22.5$ . We have found that these galaxies are mainly ellipticals and irregulars. Contrary to what had been believed, prominent spectral emission lines ([OII] and [OIII]) in irregular galaxies make the resulting photo- $z$  performance being even better than that obtained by tracking the 4000Å break in elliptical galaxies. With such a photometric resolution, the photo- $z$  method becomes closer to the spectroscopic, where the position in wavelength of a single spectral line may be used to determine the redshift of the whole spectrum.

We have also seen that broad-band filters (*ugrizY*), together with the narrow bands, help go deeper in magnitude. For  $22.5 < i_{AB} < 23.7$ , the photo- $z$  precision degrades by about an order of magnitude.

Finally, we have studied how small deviations on the original narrow-band filters, in terms of width, wavelength coverage, etc., affect the photo- $z$  performances. We have found that results are very similar to the default, so we deduce that the original set is close to optimal.

PAUCam will see its first light in 2014, so real data will be available very soon leaving mock catalogs in the background. The photometric mock catalog used here is preliminary and relatively unsophisticated. The aim was just to have a first-order version of the PAU@WHT photo- $z$  performance. However, in the near future and in order to calibrate systematic effects, it would be very useful to have a more realistic mock catalog, including dust extinction on the galaxy photometry (reddening), adding more spectral type diversity (e.g. starburst galaxies, quasars or even stars), including evolution on the spectra, etc. Special emphasis should be given to increase the resolution of spectral templates, since template libraries for photo- $z$  determination (e.g. Coleman, Wu & Weedman, 1980, Kinney et al., 1996) are intended to be used with broad-band filters ( $\gtrsim 100$  nm), so that thin spectral features are omitted without affecting much the results. However, we have shown that spectral lines make a substantial difference when using narrow-bands ( $\sim 10$  nm). It would be interesting

to study what spectral lines ( $Ly_\alpha$ ,  $H_\alpha$ ,  $H_\beta$ , [OII], [OIII], [OIII], etc.), with different ratios produce the better results. Furthermore, data could be simulated at the pixel level, so that the fluxes would be measured more realistically as well, including all possible systematics derived from the photometric extraction.

Beyond this work, and as was said at the beginning of chapter 3, we have also tried other methods, besides BPZ, for the PAU@WHT photo- $z$  determination: LePhare, which is also a template-fitting based method and some training-based methods (ANNz, ArborZ, random forest algorithms, nearest neighbors, etc). While template-fitting methods give very similar results, consistent with BPZ, the training methods give uneven results. Some of them are unable to cope with the combinatorial growth of the complexity of using a large number of filters (inputs), so they run into difficulties, while others are able to reproduce or even improve the results presented here.

In the fall of 2012, a new galaxy survey called the Dark Energy Survey (DES) saw its first light. When finished, it will have mapped about 5000 sq. deg. of the southern sky up to a depth  $i_{AB} \sim 24$ , providing measurements of the positions in the sky, photometric redshifts and shapes of about 300 million galaxies up to redshift 1.2. During winter 2012/13, a Science Verification (SV) period of observations provided science-quality images for about 150 sq. deg. In Sánchez et al. (in preparation), the photo- $z$  performance of DES is studied using these data together with spectroscopic information of about 15000 galaxies from other surveys. If PAU@WHT surveys an area that overlaps the DES footprint, its quasi-spectroscopic information could be used as well to calibrate the DES photo- $z$ s.

In chapter 4 we have shown that applying photo- $z$  quality cuts on galaxy catalogs, similar to the ones applied in chapter 3 to achieve the required PAU@WHT photo- $z$  precision, can grossly bias the measured galaxy correlations within and across photometric redshift bins, since typically galaxies are not removed uniformly over the sky. Extending the work of Ho et al. (2012) and Ross et al. (2011), we have developed a method to correct for this using the data themselves. We have applied it to the Mega-Z catalog (Collister et al., 2007), containing  $\sim 1$  million luminous red galaxies in the redshift range  $0.45 < z < 0.65$  and after splitting the sample into four  $\Delta z = 0.05$  photo- $z$  bins using the BPZ algorithm, we have seen how the corrections bring the measured galaxy auto- and cross-correlations into agreement with expectations.

In contrast to chapter 3, the Mega-Z catalog only contains elliptical galaxies. Therefore, it would also be interesting to study how these cuts affect a galaxy catalog containing different spectral types. As shown in Figs. 3.12 and 3.13, different galaxy types provide different photo- $z$  quality. Since they also cluster with different amplitudes, the photo- $z$  quality cuts may bias the measured clustering. DES will provide photometry for different galaxy types with redshift up to  $z \lesssim 1.2$  and  $i_{AB} \lesssim 24$ , which will make it very suitable to test if our correction also work in this case. Moreover, there are realistic DES mock catalogs that do not suffer from the systematic effects causing most of the extra clustering studied here, so they would help us disentangle the contributions due to observational systematics from those due to differences in the intrinsic clustering.

We would like to note that although the correction has only been applied to the

---

BAO-scale extraction, this may not be the best case to exemplify its usefulness. As shown in Fig. 4.8, photo- $z$  quality cuts are especially useful in order to reduce the long tails in the real redshift distribution within a given photo- $z$  bin. This reduces considerably the overlap between distant photo- $z$  bins, which is particularly useful in the measurement of weak-lensing magnification. This is because the contribution to the angular correlation of the terms due to the autocorrelation of overlapping galaxies at different photo- $z$  bins would be reduced with respect to the terms due to lensing, as can be deduced from Eq. 3.17.

Summarizing, in this work we have shown that, although photometric redshifts are typically considered as low quality determinations of the real redshift, they can also perform highly reliably and precisely redshift determinations, when a narrow-band filter set and photo- $z$  quality cuts are used. However, one must be careful when applying those cuts because the measured galaxy clustering can be highly biased due to either the inhomogeneous photometric quality over the sky or the different photo- $z$  quality associated with different galaxy types. We also provide a simple method to correct for this effect. We conclude that, this work opens a new path to precise photometric redshift determinations, in current and upcoming galaxy redshift surveys.



# Bibliography

- Abazajian K. et al., 2004, *Astron. J.*, 128, 502
- Abazajian K. N. et al., 2009, *Astrophys. J. Suppl. Ser.*, 182, 543
- Ade P., Others, 2013, arXiv:1303.5076
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Gi-  
allongo E., 1999, *Mon. Not. R. Astron. Soc.*, 310, 540
- Assef R. J. et al., 2008, *Astrophys. J.*, 676, 286
- Astier P., Pain R., 2012, *Comptes Rendus Phys.*, 13, 521
- Bahcall N. A., Soneira R. M., Burgett W. S., 1986, *Astrophys. J.*, 311, 15
- Baum W. A., 1962, in *IAU Symp.*, p. 390
- Benitez N., 2000, *Astrophys. J.*, 536, 571
- Benítez N. et al., 2009, *Astrophys. J.*, 691, 241
- Benjamin J., Van Waerbeke L., Ménard B., Kilbinger M., 2010, *Mon. Not. R. Astron.  
Soc.*, 408, 1168
- Bertin E., Arnouts S., 1996, *Astron. Astrophys. Suppl. Ser.*, 117, 393
- Blake C., Collister A., Lahav O., 2008, *Mon. Not. R. Astron. Soc.*, 385, 1257
- Bolzonella M., Miralles J., Pell R., 2000, *Astron. Astrophys.*, 492, 476
- Bonnett C., 2013, arXiv e-prints, 13, 1
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *Astrophys. J.*, 686, 1503
- Bruzual A. G., Charlot S., 1993, *Astrophys. J.*, 405, 538
- Cabre A., Fosalba P., Gaztanaga E., Manera M., 2007, *Mon. Not. R. Astron. Soc.*,  
381, 1347
- Cabré A., Gaztanaga E., 2009, *Mon. Not. R. Astron. Soc.*, 393, 1183
- Cannon R. et al., 2006, *Mon. Not. R. Astron. Soc.*, 372, 425

- Carliles S., Budavári T., Heinis S., Priebe C., Szalay A. S., 2010, *Astrophys. J.*, 712, 511
- Carnero A., Sánchez E., Croce M., Cabré A., Gaztanaga E., 2012, *Mon. Not. R. Astron. Soc.*, 419, 1689
- Carrasco Kind M., Brunner R. J., 2013, *Mon. Not. R. Astron. Soc.*, 432, 1483
- Castander F. J. et al., 2012, in *Soc. Photo-Optical Instrum. Eng. Conf. Ser.*, McLean I. S., Ramsay S. K., Takami H., eds., p. 84466D
- Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, *Astron. J.*, 132, 926
- Coleman G. D., Wu C.-C., Weedman D. W., 1980, *Astrophys. J. Suppl. Ser.*, 43, 393
- Colless M. et al., 2001, *Mon. Not. R. Astron. Soc.*, 328, 1039
- Collister A. et al., 2007, *Mon. Not. R. Astron. Soc.*, 375, 68
- Collister A. A., Lahav O., 2004, *Publ. Astron. Soc. Pacific*, 116, 345
- Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995, *Astron. J.*, 110, 2655
- Croce M., Gaztanaga E., Cabré A., Carnero A., Sánchez E., 2011, *Mon. Not. R. Astron. Soc.*, 417, 2577
- Dahlen T., Mobasher B., Somerville R. S., Moustakas L. A., Dickinson M., Others, 2005, *Astrophys. J.*, 631, 126
- Dawson K. S. et al., 2013, *Astron. J.*, 145, 10
- de Vaucouleurs G., 1948, *Ann. d Astrophysique*, 11, 247
- Dodelson S., 2003, *Modern Cosmology*. Academic Press
- Drinkwater M. J. et al., 2010, *Mon. Not. R. Astron. Soc.*, 401, 1429
- Efron B., 1979, *Ann. Stat.*, 7, 1
- Eisenstein D. J. et al., 2005, *Astrophys. J.*, 633, 560
- Fry J. N., Gaztanaga E., 1993, *Astrophys. J.*, 413, 447
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *Astron. J.*, 111, 1748
- Gaztanaga E., Eriksen M., Croce M., Castander F. J., Fosalba P., Martí P., Miquel R., Cabré A., 2012, *Mon. Not. R. Astron. Soc.*, 422, 2904
- Gaztanaga E., Miquel R., Sánchez E., 2009, *Phys. Rev. Lett.*, 103, 091302

- Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, *Astrophys. J.*, 715, 823
- Giavalisco M. et al., 2004, *Astrophys. J.*, 600, L93
- Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *Astrophys. J.*, 622, 759
- Gunn J. E. et al., 1998, *Astron. J.*, 116, 3040
- Gunn J. E. et al., 2006, *Astron. J.*, 131, 2332
- Hamilton A. J. S., 1992, *Astrophys. J.*, 385, L5
- Hamuy M. et al., 1993, *Astron. J.*, 106, 2392
- Hildebrandt H. et al., 2010, *Astron. Astrophys.*, 523, A31
- Ho S. et al., 2012, *Astrophys. J.*, 761, 14
- Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 1996, arXiv:astro-ph/0210394v1, 1
- Hu W., Sugiyama N., 1996, *Astrophys. J.*, 471, 542
- Hubble E., 1929, *Proc. Natl. Acad. Sci.*, 15, 168
- Hutsi G., 2006, *Astron. Astrophys.*, 449, 891
- Ilbert O. et al., 2006, *Astron. Astrophys.*, 457, 841
- Johnson H. L., Morgan W. W., 1953, *Astrophys. J.*, 117, 313
- Jouvel S. et al., 2009, *Astron. Astrophys.*, 504, 359
- Kaiser N., 1984, *Astrophys. J.*, 284, L9
- Kaiser N., Tonry J. L., Luppino G. A., 2000, *Publ. Astron. Soc. Pacific*, 112, 768
- Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, *Astrophys. J.*, 467, 38
- Komatsu E., Others, 2009, *Astrophys.J.Suppl.*, 180, 330
- Koo D. C., 1985, *Astron. J.*, 90, 418
- Kotulla R., Fritze U., 2009, *Mon. Not. R. Astron. Soc. Lett.*, 393, L55
- Le Fèvre O. et al., 2005, *Astron. Astrophys.*, 439, 845
- Li I. H., Yee H. K. C., 2008, *Astron. J.*, 135, 809
- Linder E., 2005, *Phys. Rev. D*, 72, 043529
- Loh E. D., Spillar E. J., 1986, *Astrophys. J.*, 303, 154

- Martí P., Miquel R., Bauer A., Gaztanaga E., 2014a, *Mon. Not. R. Astron. Soc.*, 437, 3490
- Martí P., Miquel R., Castander F. J., Gaztanaga E., Eriksen M., Sánchez C., 2014b, *arXiv e-prints*, 1402.3220, 19
- Moles M. et al., 2008, *Astron. J.*, 136, 1325
- Oke J. B., Gunn J. E., 1983, *Astrophys. J.*, 266, 713
- Oke J. B., Schild R. E., 1970, *Astrophys. J.*, 161, 1015
- Okumura T., Matsubara T., Eisenstein D. J., Kayo I., Hikage C., Szalay A. S., Schneider D. P., 2008, *Astrophys. J.*, 676, 889
- Padmanabhan N. et al., 2005, *Mon. Not. R. Astron. Soc.*, 359, 237
- Padmanabhan N. et al., 2007, *Mon. Not. R. Astron. Soc.*, 378, 852
- Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C., Szalay A. S., 2007, *Mon. Not. R. Astron. Soc.*, 381, 1053
- Perlmutter S. et al., 1999, *Astrophys. J.*, 517, 565
- Perlmutter S., Schmidt B. P., 2003, *Lect. Notes Phys.*, 598, 195
- Puschell J. J., Owen F. N., Laing R. A., 1982, *Astrophys. J.*, 257, L57
- Riess A. G. et al., 1998, *Astron. J.*, 116, 1009
- Ross A. J. et al., 2011, *Mon. Not. R. Astron. Soc.*, 417, 1350
- Sánchez E. et al., 2011, *Mon. Not. R. Astron. Soc.*, 411, 277
- Schechter P., 1976, *Astrophys. J.*, 203, 297
- Smith R. E. et al., 2003, *Mon. Not. R. Astron. Soc.*, 341, 1311
- Sowards-Emmerd D., Smith J. A., McKay T. A., Sheldon E., Tucker D. L., Castander F. J., 2000, *Astron. J.*, 119, 2598
- Tegmark M., Zaldarriaga M., 2002, *Phys. Rev. D*, 66, 103508
- Thakar A. R., Szalay A., Fekete G., Gray J., 2008, *Comput. Sci. Eng.*, 10, 30
- Thomas S. A., Abdalla F. B., Lahav O., 2011a, *Phys. Rev. Lett.*, 106, 241301
- Thomas S. A., Abdalla F. B., Lahav O., 2011b, *Mon. Not. R. Astron. Soc.*, 412, 1669
- Tyson J., Wittman D., Hennawi J., Spergel D., 2003, *Nucl. Phys. B - Proc. Suppl.*, 124, 21
- Vanzella E. et al., 2004, *Astron. Astrophys.*, 423, 761

Williams R. E. et al., 1996, *Astron. J.*, 112, 1335

Wolf C., 2009, *Mon. Not. R. Astron. Soc.*, 397, 520

Yanny B. et al., 2009, *Astron. J.*, 137, 4377

York D. G. et al., 2000, *Astron. J.*, 120, 1579