

UNIVERSITAT AUTÓNOMA DE BARCELONA

DOCTORAL THESIS

**Accurate photometric redshifts with
narrow bands filters, photo-z quality
cuts and their impact on the galaxy
clustering**

Author:

Pol MARTÍ

Supervisor:

Dr. Ramon MIQUEL

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy*

in the

Research Group Name
Institut de Física d'Altes Energies (IFAE)

November 2013

“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”

Dave Barry

UNIVERSITAT AUTÓNOMA DE BARCELONA

Abstract

Faculty Name

Institut de Física d'Altes Energies (IFAE)

Doctor of Philosophy

**Accurate photometric redshifts with narrow bands filters, photo-z quality
cuts and their impact on the galaxy clustering**

by Pol MARTÍ

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Chapter 1

Introduction

Chapter 2

Cosmological Framework

2.1 The geometry of the universe

The cosmological principle asserts that the distribution of matter-energy in the universe at large scales is homogeneous and isotropic and, therefore, there is no preferred direction in the sky.

According to the General Theory of Relativity, the unique metric compatible with that principle is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which space-time line element can be written as:

$$ds^2 = -cdt^2 + a^2(t)dx^2 \quad (2.1)$$

The left-hand side $-ct$ of the sum is the temporal contribution while the right-hand side is the spatial, which is the product of a scale factor $a(t)$ and a comoving space line element dx . In the case of a perfect flat space, which seems to be in accord with the observations, and using spherical coordinates:

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

2.2 Cosmological Redshift

When the light that comes from a galaxy travels across the space of an expanding universe suffers an effect called Cosmological Redshift. As the equivalent redshift due to the Doppler effect of a source in motion where the wavelength of the light is dilated, in

the cosmological redshift the wavelength is also dilated but due to the dilatation of the space itself.

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

$$1 + z = \frac{a}{a_0} \quad (2.4)$$

2.3 The evolution of the universe

Assuming that the matter-energy content of the universe can be described as a perfect fluid, which follows the state equation $P = \omega\rho$, the gravitational field equation gives us the temporal evolution $a(t)$ of the FLRW metric in differential form.

Friedman equation

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \sum_{\omega} \left(\frac{a_0}{a}\right)^3 (1 + w) \quad (2.5)$$

Hubble constant $H_0 \sim 70(km/s)/Mpc$

$$H_{\Lambda}(z) = H_0 \sqrt{\Omega_M(1+z)^3 + \Omega_{\Lambda}} \quad (2.6)$$

$$\Omega_M + \Omega_{\Lambda} = 1$$

$$\text{Dark Energy relative density } \Omega_{\Lambda} \sim 0.73$$

$$\text{Mater relative density } \Omega_M \sim 0.26$$

2.4 Luminosity and Angular distances

Comoving distance

$$r \equiv \frac{c}{H_0} \int_0^z \frac{dz}{H(z)} \quad (2.7)$$

Luminosity distance

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

Angular distance

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

2.5 Photometry

2.5.1 Flux and Apparent magnitude

Apparent magnitude

$$m - m_0 \equiv -2.5 \log \frac{F}{F_0} \quad (2.10)$$

$$F \equiv \int_0^\infty f(\nu) R(\nu) \frac{d\nu}{\nu} = \int_0^\infty f(\lambda) R(\lambda) \lambda \frac{d\lambda}{c} \quad [31]$$

$$\nu\lambda = c \implies \frac{d\nu}{\nu} = -\frac{d\lambda}{\lambda}$$

$$f(\nu)d\nu = f(\lambda)d\lambda \implies \nu f(\nu) = \lambda f(\lambda)$$

$$[F] = [f(\nu)] = \left[\frac{\text{Flux}}{\text{time}^{-1}} \right], [f(\lambda)] = \left[\frac{\text{Flux}}{\text{longitude}} \right]$$

2.5.2 The AB Systems

The AB system [43]

$$m_0 = 0$$

$$F_0 \text{ such that } f_0^{AB}(\nu) = 3631Jy$$

$$1Jy = 10^{-23} erg \cdot cm^{-2} \cdot s^{-1} \cdot Hz^{-1} = 1.51 \cdot 10^7 \cdot photons \cdot m^{-2} \cdot s^{-1} \cdot dlog^{-1}\lambda$$

$$m_{AB} \equiv -2.5 \log \left(\frac{F}{F_0} \right) \iff F = F_0 10^{-0.4m_{AB}}$$

$$m_{AB}(\nu) = -2.5 \log f_\nu(\nu) - 48.6 \quad [?]$$

2.5.3 Absolute magnitude and K-correction

Absolute magnitude

$$M = m - DM - K \quad (2.11)$$

Distance Modulus

$$DM \equiv 5 \log \left[\frac{D_L}{10pc} \right] \quad (2.12)$$

K-correction

$$K \equiv -2.5 \log \left[\frac{1}{1+z} \frac{\int_0^\infty f(\lambda/1+z) R(\lambda) \lambda d\lambda}{\int_0^\infty f(\lambda) R(\lambda) \lambda d\lambda} \right] \quad (2.13)$$

2.5.4 Luminosity function

Schechter function

$$n(x) \equiv \frac{dN}{dx} = \phi^* x^a e^{-x} \quad (2.14)$$

where $x \equiv \frac{M}{M^*}$

2.5.5 Measuring Magnitudes

Model magnitudes: identical apertures

Apper magnitudes

Auto, psf

fibre

De Vaucouleurs

Petrosian

2.6 Photometric Redshifts

2.6.1 Template Fitting Methods

$$\chi^2(z, t) = \sum_i \frac{(m_i - m_i^t(z))^2}{\sigma_m^2} \quad (2.15)$$

2.6.2 Bayesian statistics

BPZ [3]

Probability density distribution

$$p(z | m_i, m_{ref}) \propto \sum_t \Pi(z, t | m_{ref}) L(m_i | z, t) \quad (2.16)$$

Likelihood $L(m_i | z, t) \propto \exp \chi^2(z, t)$

Prior

$$\Pi(z, t \mid m) \propto f_t e^{-k_t(m - m_0)} \cdot z^{\alpha_t} \exp \left\{ - \left[\frac{z}{z_{mt}(m)} \right]^{\alpha_t} \right\} \quad (2.17)$$

$$z_{mt}(m) = z_{0t} + k_{mt}(m - m_0)$$

2.6.3 Quality cuts

$$odds \equiv \int_{z(phot)-\delta z}^{z(phot)+\delta z} p(z \mid m_i, m_{ref}) dz \quad (2.18)$$

2.7 Baryonic Acoustic Oscilations

$$r_{BAO} \sim \frac{c_s}{H_0} \int_{1100}^{\infty} \frac{dz}{H_{\Lambda}(z)} \sim 150 Mpc \quad (2.19)$$

where $c_s \sim c \sqrt{\frac{1}{3}}$

Chapter 3

High photo-z precision with a narrow-band filter set: The PAU@WHT Survey

3.1 Introduction

Galaxy surveys are a fundamental tool in order to understand the large-scale structure of the universe as well as its geometry, content, history, evolution and destiny. Spectroscopic surveys (2dF, Colless et al. [16]; VVDS, Le Fèvre et al. [40]; WiggleZ, Drinkwater et al. [22]; BOSS, Dawson et al. [21]) 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. [55]; PanSTARRS, Kaiser, Tonry & Luppino [37]; LSST, Tyson et al. [54]) 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. 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, and, hence, overcoming the limitations of spectroscopic surveys [4].

In this paper we present the study of the optimization of the filters for PAU. We look for the filter set that minimizes the overall photometric redshift (photo-z) precision in a

sample consisting of all galaxies of all types with $i_{AB} < 22.5$. Based of detailed simulation studies [26], 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, and, actually, these galaxies play a crucial role in the PAU science case [26]. We will study the performance that the different filter sets achieve in a fainter galaxy sample with $22.5 < i_{AB} \lesssim 24.$, **but we will not impose any requirement on it.**

There are two main sets of techniques for measuring photometric redshifts (or *photo-zs*): template methods (e.g. Hyperz, Bolzonella, Miralles & Pell [7]; BPZ, Benitez [3] & Coe et al. [14]; LePhare, Ilbert et al. [34]; EAZY, Brammer, van Dokkum & Coppi [8]), in which the measured broadband galaxy spectral energy distribution (SED) is compared to a set of redshifted templates until a best match is found, thereby determining both the galaxy type and its redshift; training methods (e.g. ANNz, Collister & Lahav [18]; ArborZ, Gerdes et al. [28]; TPZ, Carrasco Kind & Brunner [13]), in which a set of galaxies for which the redshift is already known is used to train a machine-learning algorithm (an artificial neural network, for example), which is then applied over the galaxy set of interest. Each technique has its own advantages and disadvantages, whose discussion lies beyond the scope of this paper.

Throughout the paper we will be using the Bayesian Photo-Z (BPZ) template-based code from Benitez [3], after adapting it to our needs. We have tried to use several photo-z codes based on training methods, but we have encountered difficulties having to do with the large, $\mathcal{O}(50)$, number of filters. This is described in detail in Sanchez, Abdalla & Miquel [49].

The outline of the paper is as follows. In section 3.2 we present the default PAU filter set. Section 3.3 discusses the mock galaxy samples that we use in our study, the noise generation, and the split into a brighter and a fainter galaxy samples. In section 4.3 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. In section 3.5 we try several modifications to the filter set (wider/narrower filters, bluer/redder filters, etc.), see how they perform on the brighter and fainter galaxy samples and find the optimal set. Finally, in section 4.7, we discuss the results and offer some conclusions.

3.2 Default filter set-up

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

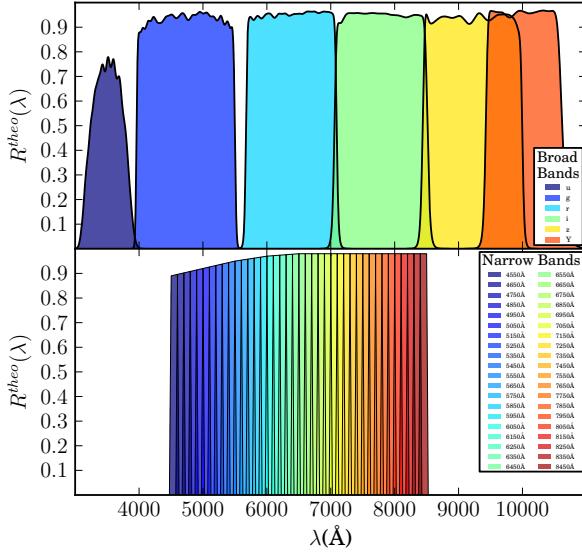


FIGURE 3.1: The nominal (or theoretical) response $R^{theo}(\lambda)$ of the *ugrizY* PAU Broad Bands (top) and the 40 Narrow Bands (bottom). The *u* band is the same as in the USNO 40-in telescope at Flagstaff Station (Arizona), while the *grizY* are the same as in the DECam mounted in the Blanco Telescope (CTIAO, Chile). Narrow bands have 100Å width and have two additional lateral wings of 25Å width that overlap 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 bands.

3.2.1 Nominal response

The PAUCam will mount two sets of filters: the Broad-Band (BB) filters, composed of 6 bands *ugrizY*¹, 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 of 100Å width ranging from 4500Å to 8500Å. 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 25Å width. This induces an overlap of $\sim 20\%$ between contiguous bands. Additionally, we set the overall NB response to match that from the ALHAMBRA survey instrument [42], since it is the one galactic survey with similar technical specifications on its bands known to this day.

¹The *u* band is the same as the used in the USNO 40-in telescope at Flagstaff Station (Arizona) and can be obtained from <http://www.sdss.org/dr7/algorithms/standardstars/Filters/response.html>, while the rest are the same as in the DECam mounted in the Blanco Telescope (CTIAO, Chile) and can be obtained from <http://des-docdb.fnal.gov:8080/cgi-bin>ShowDocument?docid=4295> (private link).

3.2.2 Effective response

The filter responses $R^{theo}(\lambda)$ in Fig. 3.1 are nominal: this is the response that we would measure if light went only through the filter. Actually, 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, CCDs detector 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 $T_i(\lambda)$ of these effects i . In our case, these curves are shown in the top plot of Fig. 3.2. The QE curve (blue), correspond to the measured QE of CCDs provided by *Hammamatsu*, the measured efficiency curve of the mirror's reflexion (green) corresponds to that from the *William Herschel Telescope (WHT)* primary mirror, and the atmospheric transmission curve (red) is taken from the *Apache Point Observatory (APO)* at New Mexico. The resulting effective response $R(\lambda)$ is derived with the expression:

$$R(\lambda) = R^{theo}(\lambda) \prod_i T_i(\lambda) \quad (3.1)$$

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 $T_i(\lambda)$ 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. Note that atmospheric absorption valleys, 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, note the case of the narrow band with central wavelength at $\sim 7550\text{\AA}$ (orange), whose profile is drastically changed.

3.2.3 5σ -limiting magnitudes

Next, we compute the 5σ -limiting magnitudes, $m_{AB}(5\sigma)$, for all the PAU bands in the AB photometric system² [43]. This is the apparent magnitude whose Signal-to-Noise ratio, given by

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

²According to Hogg et al. [31], 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 $m_{AB} \equiv -2.5 \log_{10} [\int f(\nu)R(\nu)\frac{d\nu}{\nu} / \int (3631\text{Jy})R(\nu)\frac{d\nu}{\nu}]$, where $1\text{Jy} = 10^{-23}\text{erg} \cdot s^{-1} \cdot cm^{-2} \cdot Hz^{-1} = 1.51 \cdot 10^7 \text{photons} \cdot m^{-2} \cdot s^{-1} \cdot \frac{\lambda}{d\lambda}$.

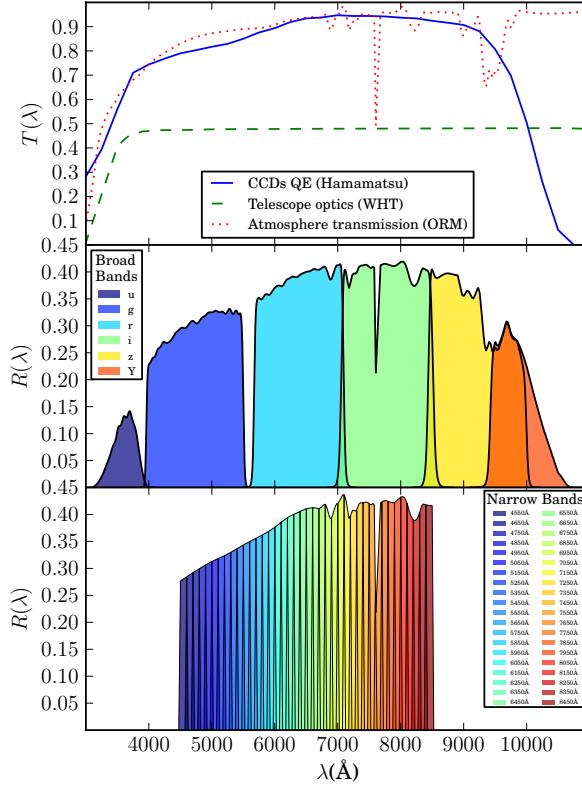


FIGURE 3.2: Top: the QE curve (blue) of the PAUCam CCDs, the efficiency curve of the reflexion (green) in the WHT primary mirror and the atmospheric transmission (red) at the APO (Apache Point Observatory), that affect the final response of the PAU bands. The two plots below, are the same as in Fig.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

is equal to 5, where

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

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

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 ref???, 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 $5 \text{ trays} \times 8 \text{ NB} = 40$

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).

ϕ	Telescope's mirror diameter	4.2 m
α	Focal Plane Scale	0.265 arcsec/pix
RN	Read-out Noise	5 electrons/pix
A	Galaxy Aperture	2 arcsec 2
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 on 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

NB, the entire number of 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 in its particular tray with its particular exposure time. 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 $m_{AB}(5\sigma)$ 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 $f_{sky}(\lambda)$ 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 t_R for the NB increase in steps, due to this arrangement in groups per tray. However, this increment is not enough to compensate the sky brightness as we can see with the descending $m_{AB}(5\sigma)$ with wavelength. On the other hand, 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 . Also note that there are large drops in $m_{AB}(\sigma)$ for the NB with central wavelength 5550Å, 6250Å and 6350Å. This is due to emission lines in the sky spectrum $f_{sky}(\lambda)$ at these wavelengths.

3.3 The mock catalog

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

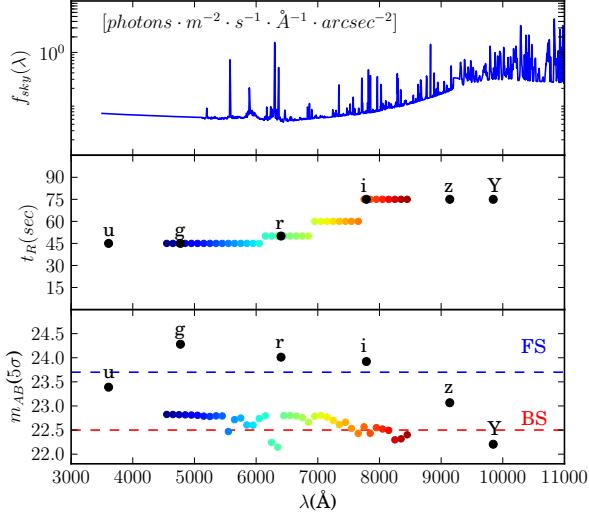


FIGURE 3.3: Top: a model of the spectral density flux of the sky brightness $f_{\text{sky}}(\lambda)$ at LaPalma (day 7 on the lunar cycle) assuming an airmass of 1.0, used in (3.4). Middle: the exposure times t_R for each PAU band used in (3.3) and (3.4). Bottom: the resulting limiting magnitudes $m_{AB}(5\sigma)$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

type t .

3.3.1 Noiseless magnitudes

We use a similar method as described in Jouvel et al. [35], which consists on sampling the cumulative Luminosity Function (LF):

$$N(z, t) = \int_{-\infty}^{M_{lim}(z, t)} \phi x^\alpha e^{-x} dx, \quad (3.5)$$

where $x \equiv 10^{-0.4(M - M^*)}$, 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, $\{M^*, \phi, \alpha\}$ are the parameters of the LF, which also depend on z and t . We assume that their redshift dependency is:

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

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 [20], where three different spectral types, 1=Elliptical(Ell), 2=Spiral(Sp), 3=Irregular(Irr)/StarBurst(SB), are distinguished.

TABLE 3.3: Parameters $\{a, b, c\}$ that, through (3.6), give the values and evolution in redshift of the LF parameters $\{\log_{10} \phi, M^*, \alpha\}$ for a given spectral type t . These values are based on the LFs in Dahlen et al. [20] where three galaxy types are distinguished: 1=Ell, 2=Sp and 3=Irr/SB.

t	$\log_{10} \phi$			M^*			α		
	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

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 extracted from Hogg et al. [31]:

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

where $D_L(z)$ is the Luminosity Distance of the galaxy at redshift z in Mpc, which in a Λ CDM universe is expressed as:

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

with cosmological parameters: $H_0 = 75$ (km/s)/Mpc, $\Omega_M = 0.25$ and $\Omega_\Lambda = 0.75$, and $K(z, t)$ is the K-correction [31]:

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

where $R_0(\lambda)$ is the response of the reference band, $f_t(\lambda)$ is the Spectral Energy Density (SED) of the galaxy with spectral type t in the rest frame, and $f_t((1+z)\lambda)$ the same SED at redshift z . As a representation of these SEDs, we use the CWW extended template library from the LePhare³ photo-z code. It contains 66 templates ranging through Ell→Sp→Irr/SB and shown on Fig. 3.4. We split them in three groups: Ell = (0-17), Sp = (17-55) and Irr/SB = (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.

When we already have the $\{z, t\}$ values for all galaxies, we also assign them an absolute magnitude M randomly within the range $[-\infty, M_{lim}(z, t)]$ following the LF probability distribution in (3.5). Then, the apparent magnitude m_0 , in our reference band r_{SDSS} ,

³The 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.

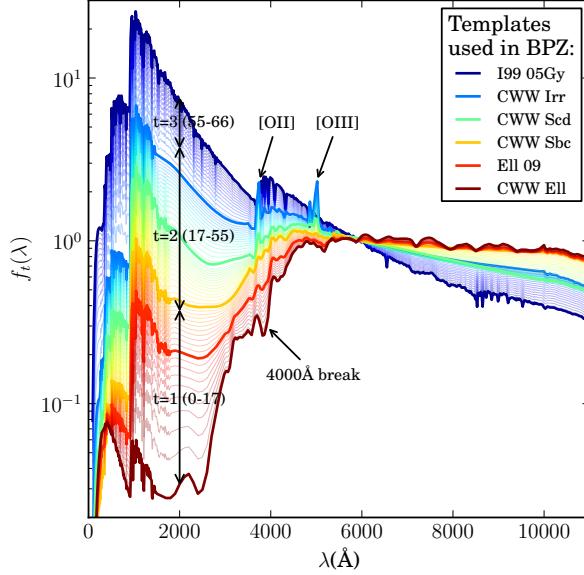


FIGURE 3.4: The 66 spectral templates in the rest frame extracted from the extended CWW library. They are used, in (3.7) and (3.10), to generate the photometry of the PAU mock catalog. They evolve from Ellipticals (0-17, in red) to Spirals (17-55, in yellow, green and cyan), and finally to Irregulars/Starburst (55-66, in blue). Black curves separate these three main spectral types of our LFs. Wider and deeper curves highlight the five templates used in BPZ to compute the photo-zs.

is computed from (3.7). The other magnitudes m_j at any band j are obtained through:

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

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

3.3.2 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 σ_{m_j} the expected magnitude error which is related to the Signal-to-Noise in (3.2) as follows:

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

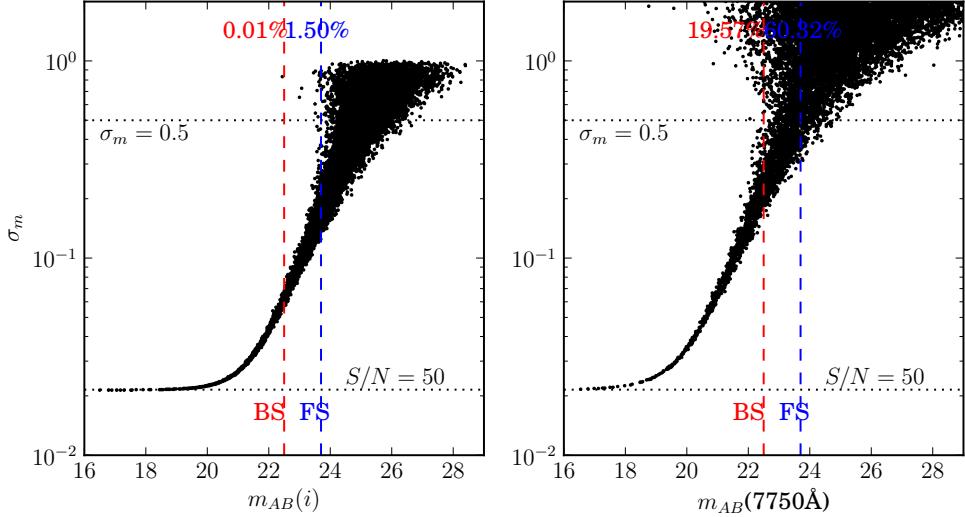


FIGURE 3.5: Scatter plots of σ_m vs. m for the i BB (left) and the 7750\AA NB (right). For the sake of clarity only 10000 galaxies randomly selected are plotted. The magnitude limits of the BS (red) and FS (blue) are also plot in vertical-dashed lines. The bottom dotted line in both plots shows the calibration error ($S/N = 50$) added in quadrature to σ_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.

Additionally, we add an extra component of noise of size ~ 0.022 , corresponding to a $S/N = 50$, in quadrature to σ_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 σ_m vs. m scatter plots, in the i BB and the 7750\AA NB, are shown in Fig. 3.5, where for the sake of clarity we only plot 10000 galaxies randomly selected. The 7750\AA band is chosen because its central wavelength is very similar to that of the i band. Note how on both bands, σ_m starts being flat at ~ 0.022 (the calibration error), and then, at fainter magnitudes, when the sky brightness and the CCDs read-out noise become important, it grows and the scatter becomes wider.

3.3.3 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 [26]). The BS should contain galaxies bright enough to have a decent Signal-to-Noise in all bands, including the narrow, and, therefore, reach the necessary photo-z accuracy to measure RSD and the position of the lenses for MAG. We see in Fig. 3.3 that the 5σ -limiting magnitudes for the NB are close to 22.5, so we define the BS as all those galaxies with $i_{AB} \equiv m_{AB}(i) < 22.5$ (magnitude

in the i band). The FS will contain the rest of the galaxies within $22.5 < i_{AB} \lesssim 24$. The upper limit has been chosen to roughly match with the 5σ -limiting magnitudes of the gri broad bands (see Fig. 3.3). In fact, we choose it to be exactly 23.7 since, as we will see in next section, is up to this magnitude where the photo-z precision requirement is still fulfilled after applying a photo-z quality cut that removes half of the catalog, the lower completeness that we can afford.

Considering 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 are not observed in the i band in the BS, while $\sim 1.5\%$ are not in the FS. Similarly, $\sim 19.57\%$ are not observed in the 7750Å band in the BS, while $\sim 60.32\%$ are not 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-zs. However, part of these FS galaxies will be the magnified sources for MAG and they can afford a poorer photo-z quality according to [26].

In Fig. 3.6 we show the resulting distributions of the magnitude $m_{AB}(i)$ (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 ~ 25.0 , so that the BS and FS are on the brighter tail of the distribution and account for only $\sim 8.4\%$ 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 redshift limit is $z \sim 5$, the BS limit is $z \sim 1.5$ and the FS limit is $z \sim 3$. Finally, we see that both BS and FS have a similar proportion of Spiral galaxies ($t=3$), $\sim 55\%$; however the BS contains more Elliptical galaxies (33%) than the FS (19%), and consequently, the FS contains more Irregular and Starburst galaxies.

3.4 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 outliers fraction. We also apply some photo-z quality 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 Gaztañaga et al. [26].

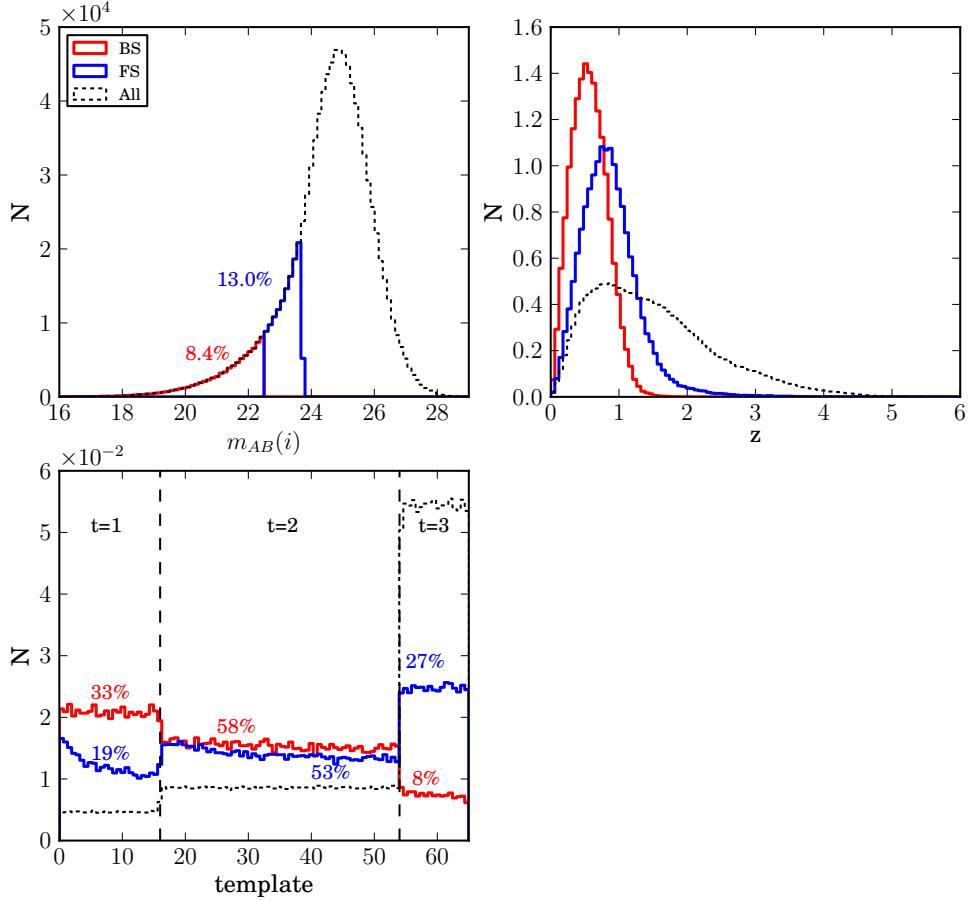


FIGURE 3.6: The observed magnitude in the i band (top-left), true redshift (top-right) and spectral type distributions for the whole catalog (black), the BS $i_{AB} < 22.5$ (red) and the FS $22.5 < i_{AB} < 23.7$ (blue). We also show the proportion of galaxies respect to the whole catalog of both samples,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

Photo-zs z_{ph} are obtained using the Bayesian Photometric Redshifts⁴ (BPZ) template-fitting code described in Benítez [3]. It uses Bayesian statistics to produce a posterior probability density function $p(z|m_j)$ that a galaxy is at redshift z when its magnitudes in the different bands j are m_j :

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

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 shows that $p(z|m_j)$ must be properly normalized. Finally, the

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

photometric redshift z_{ph} of the galaxy will be taken as the position of the maximum of $p(z|m_j)$.

3.4.1 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_Ell.sed`, `Ell_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 the photo-z performance so much.

3.4.2 Prior

A crucial point of BPZ is the prior probability $\Pi(z, t | m)$ that helps improve the photo-z performance. Benitez [3] proposes the following empirical function:

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

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 $\{f, k, \alpha, z_0, k_m\}$ 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 $1 \rightarrow (0-17)$, $2 \rightarrow (17-55)$ and $2 \rightarrow (55-66)$. From now on, we will use t for either the 66 templates or these 3 galaxy type groups. Finally, we fit (4.10) to the training sample and recover the prior parameters. We show the resulting values in

t	f	k	α	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 galaxies randomly selected from the BS and FS together. The three galaxy types $t = 1, 2, 3$ are the same that were defined in Section 3.3 for the Luminosity Functions (LF). f_3 and k_3 do not appear because $\Pi(t = 3 | m_0)$ is deduced by normalization.

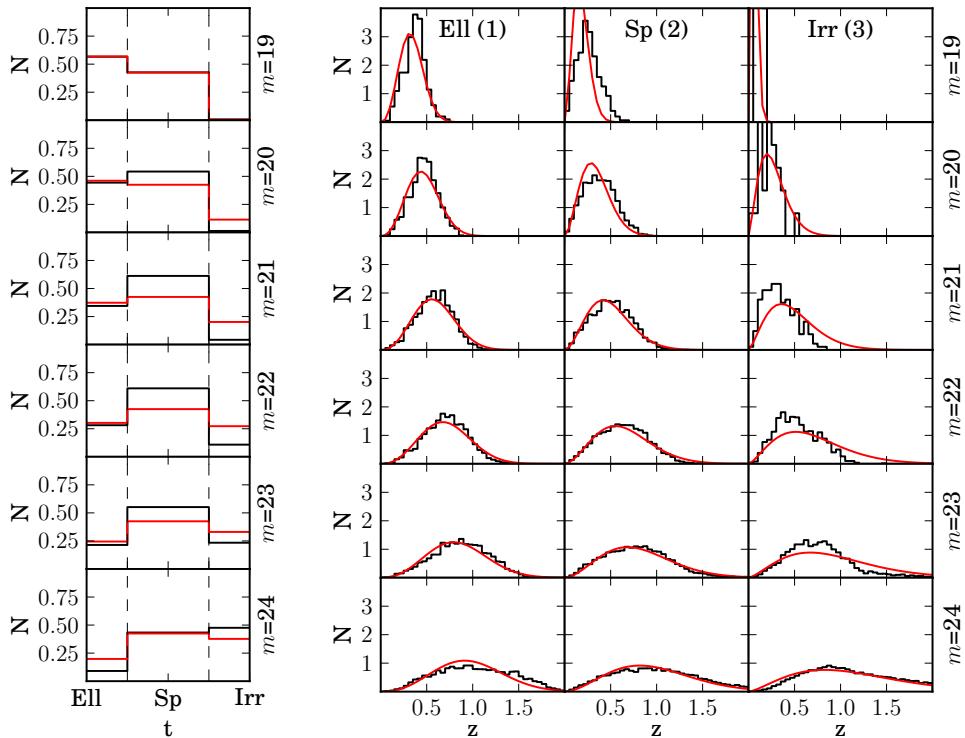


FIGURE 3.7: Left: Comparison between the fitted $\Pi(t|m)$ prior (red) of Eq. (4.10) and the actual distribution (black). Right: The same for $\Pi(z|t,m)$. We only differentiate between 3 galaxy types t (as we did for the LFs in section 3.3): 1=Elliptical, 2=Spiral and 3=Irregular. Rows correspond to 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.

Table 4.1. 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 $m(i_{AB}) = 19$, the reference magnitude, to $m(i_{AB}) = 24$. 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 magnitude,

are defined positive. 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 panes on the right plot of Fig. 3.7), but for Irr and magnitudes close to the 19 (reference) 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 only $\sim 0.2\%$ 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 σ_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 σ_z almost does not change in the BS, while it improves by a factor ~ 1.8 in the FS by going from $\sigma_z \sim 16\%$ to $\sim 8.86\%$ when adding the prior.

3.4.3 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

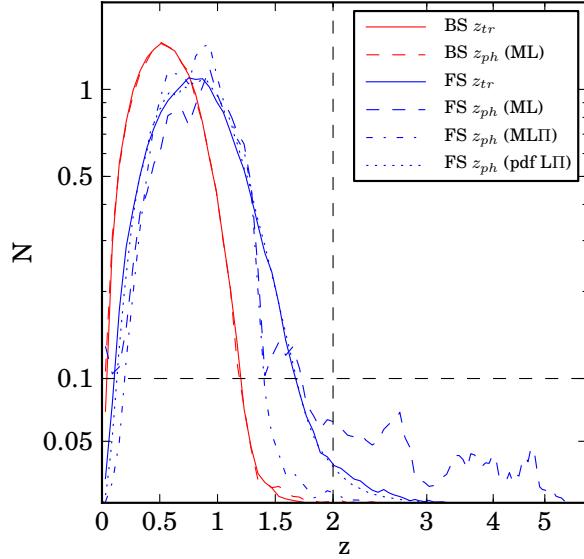


FIGURE 3.8: Comparison between the z_{tr} (solid) and z_{ph} (discontinuous) distributions for the BS (red) and FS (blue). For the sake of clarity, we have set the x- and y-axis scales to be linear below the black dashed lines and logarithmic above them. All distributions have been normalized to area. 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).

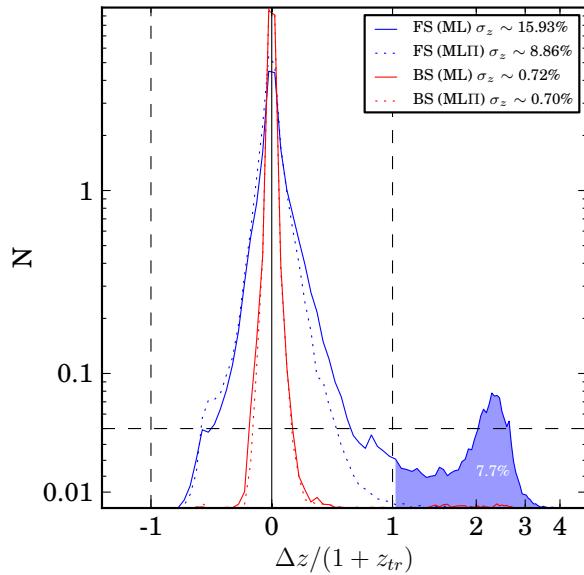


FIGURE 3.9: $\Delta z / (1 + z_{tr})$ distributions, where $\Delta z \equiv z_{ph} - z_{tr}$, for the BS (blue) and FS (red) with the z_{ph} obtained by maximizing only the likelihood $L(m_j|z, t)$ (ML) (solid) or when also including the prior (MLII) (dotted). Photo-z precision σ_z values, as half of the symmetric interval that encloses the 68% of the distribution area around the maximum, 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, 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 right-hand tail (blue region) of catastrophic outliers ($|\Delta z| / (1 + z_{tr}) > 1$) that accounts for a $\sim 7.7\%$ of the sample. This reduces σ_z by a factor ~ 1.8 .

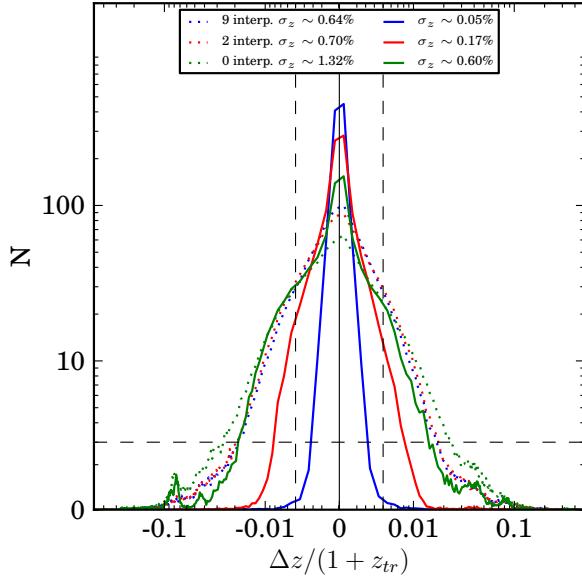


FIGURE 3.10: The $\Delta z/(1 + z_{tr})$ distributions for the BS using different number of interpolated templates in BPZ, when input magnitudes are noiseless (solid) and when they are noisy (dotted). Photo-z precision σ_z values for each case are shown on the legend. The x-axis scale is linear between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

applying Eq. 3.11), while dotted lines include the noise. The σ_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 σ_z , which gets worse by a factor of ~ 3 at each step, for noisy magnitudes these differences are much smaller. In fact, going from 9 to 2 interpolated templates the differences are negligibly small, and from 2 to 0, less than a factor of 2.

3.4.4 The Odds parameter

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

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

where δz 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 σ_z in these samples for the PAU Survey (see the σ_z plots in Fig. 3.11). A bad choice of δz

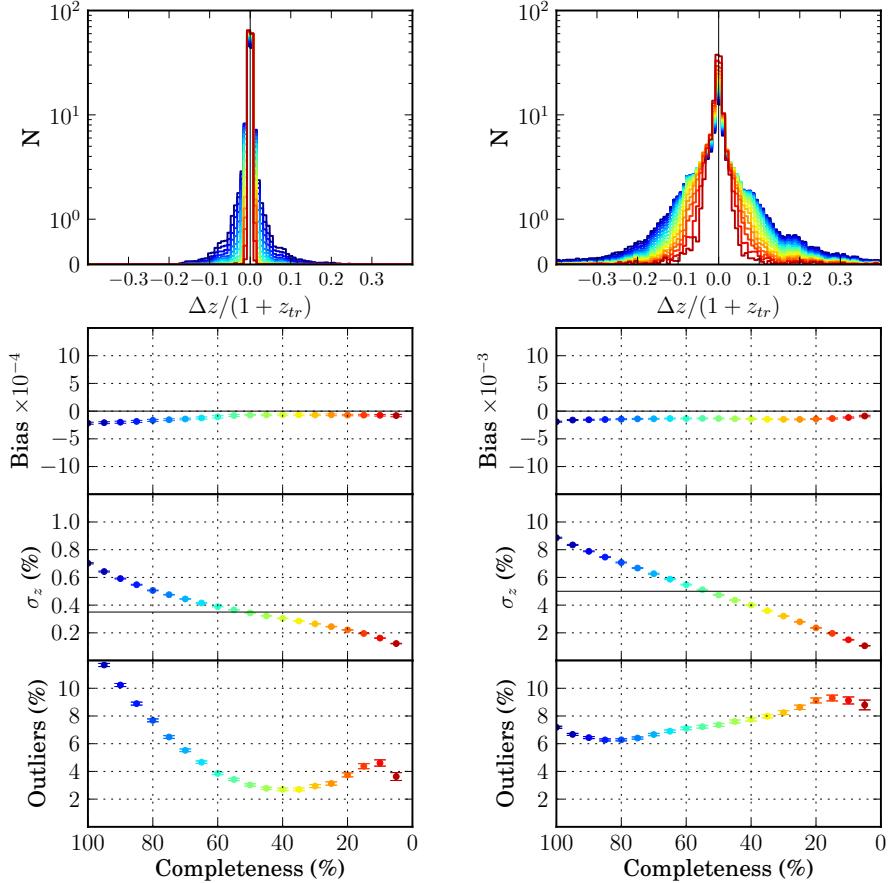


FIGURE 3.11: Top: $\Delta z/(1 + z_{tr})$ distributions for the BS (left) and the FS (right) at different photo-z quality cuts whose completeness, shown in a color degradation, ranges from 100% (bluest) to 5% (reddest) in 5% steps. On the bottom and by rows: the bias (median), the photo-z precision (σ_z) and the 3σ -outliers 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 σ_z panes changes by an order of magnitude between the two samples. The black-solid horizontal lines on the σ_z panes show the photo-z precision requirement defined in Gaztañaga et al. [26]:

$$\sigma_z < 0.35\% \text{ (BS)} \text{ and } < 0.5\% \text{ (FS).}$$

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 $|\Delta z|$.

3.4.5 Performance vs. Odds

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 δz . Therefore, we have translated these odds cuts into the fraction of the galaxy sample remaining after a certain cut has been applied, in such a way that the bluest curve

corresponds to 100% completeness while the reddest to 5% 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 (σ_z) and the 3σ -outliers fraction, evolve with each odds cut of completeness given in the x-axis. The 3σ -outliers 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 [23] for the bias and σ_z , and, by computing the σ_{68} of a binomial distribution with mean $n_{outlier}/N$ for the outliers fraction. As we expected, σ_z decreases as the odds cuts get harder. 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. [26], 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 σ_z in both samples towards negative Δz values. It practically vanishes when high odds cuts are applied. The 3σ -outliers 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 $\sim 10\%$ completeness. Therefore, we deduce that the gain on σ_z with the odds cut occurs basically through the cleaning of outliers. However, in the FS, even if σ_z decreases with the odds cuts, the outliers fraction increases from $\sim 7\%$ to $\sim 9\%$ at $\sim 15\%$ completeness.

On the last three rows of plots of 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. On 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}) in the x-axis.

On 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's read-out starts to be comparable to the noise coming from the signal. This is reflected in Fig. 3.5 as a turning point on the slope of the σ_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 σ_z from $\sim 2.2\%$ to $\sim 0.35\%$ and the outliers 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 high magnitude are spread 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 mean of z_{tr} is shifted from

~ 0.6 to ~ 0.4 . The worst bias, σ_z and outliers fraction are obtained for Spiral galaxies ($t \sim 10\text{-}30$). The odds cuts mitigate these results, but even after applying them, spiral galaxies still have the worst bias and σ_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 σ_z gets flatter over all z_{tr} as the odds cuts are harder.

Regarding the z_{ph} precision requirement $\sigma_z < 0.35\%$ (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 σ_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. [4], it was assumed that elliptical galaxies (or rather Luminous Red Galaxies) were the best to obtain the highest 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 σ_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 were a single spectral feature is tracked, Irregulars show two clear emission lines at $\sim 3737\text{\AA}$ [OII] and $\sim 5000\text{\AA}$ [OIII] (Fig. 3.4, see Kennicutt, Robert C. [38] for an explanation of their origin) that break any possible degeneracy in color space.

On the FS (Fig. 3.13), we see that the scatters $\Delta z/(1 + z_{tr})$ are much more spread, as could be 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) is still kept. We recognize behaviors similar to those in the BS in most of aspects of the z_{ph} performance, although they are substantially exaggerated. 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 ($t \sim 10\text{-}30$) 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, σ_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 σ_z increase is smaller the harder the odds cuts. This is because the gain in photo-z precision is at expense of keeping brighter galaxies each time. The mean i_{AB} magnitude goes from ~ 23.2 to ~ 22.8 with the odds cuts, and the shift in the z_{tr} mean is from ~ 0.86 to ~ 0.81 . Unlike in the BS, we see that the best σ_z is obtained at the extremal spectral types: $t \sim 0$ (Elliptical) and $t \sim 66$ (Irregular). However, once the hardest odds cuts are

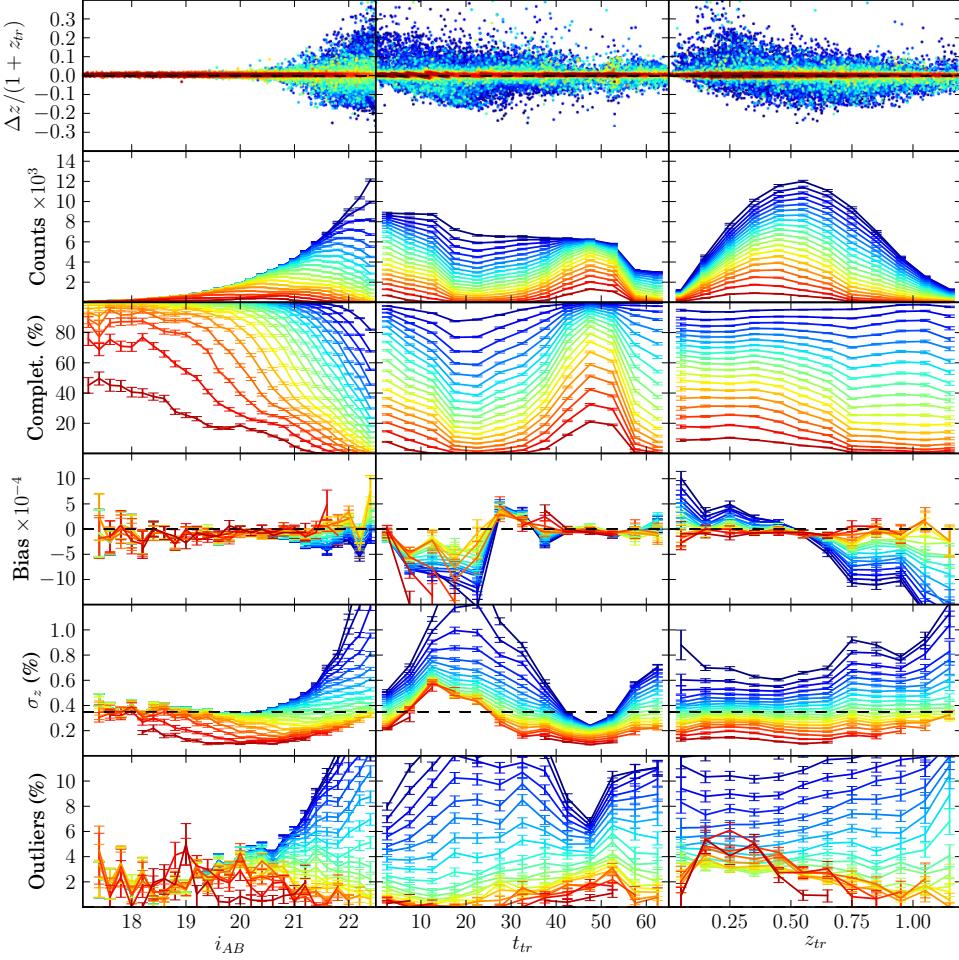


FIGURE 3.12: Statistics showing the PAU-BS photo-z performance in function of the observed i_{AB} magnitude (left), true galaxy type t_{tr} (center) and true redshift z_{tr} (right). In the first row we show the scatter $\Delta z/(1+z_{tr})$. Then, in a descending order of rows, we show the number of galaxies, the completeness, the bias (median), the photo-z precision (σ_z) and the 3σ -outliers 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 σ_z panes show the photo-z precision requirement defined in Gaztañaga et al. [26].

applied, irregular galaxies with $t \sim 50$ are again the ones with the best σ_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 σ_z take values much larger at these redshifts. The photo-z precision requirement, $\sigma_z < 5\%$, when the odds cut of 50% completeness is applied is fulfilled up to magnitude ~ 23.1 , for all galaxy types except Spirals, and at the z_{tr} interval from ~ 0.4 to ~ 1.3 . As we already saw on Fig. 3.11, the 3σ -outliers fraction grows with the odds cuts. In general, its values are higher where σ_z is lower, since the outliers criterion becomes more stringent.

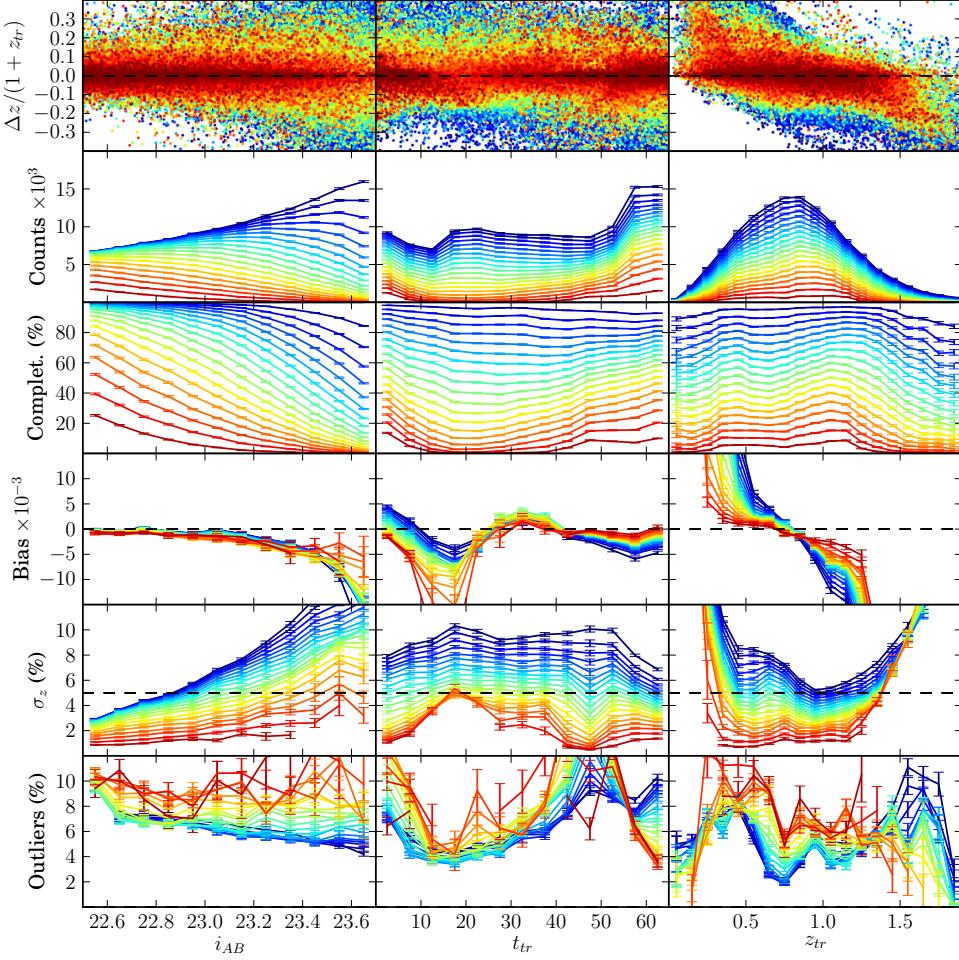


FIGURE 3.13: Statistics showing the PAU-FS photo-z performance in the same layout as in Fig. 3.12.

3.4.6 Narrow bands vs. Broad bands

We want to quantify the improvement that NB bring to the photo-z performance. For this purpose, we run BPZ on the BS and FS using NB and BB separately. Then, in Fig. 3.14 and Table 3.5, we compare results between them and also with the original ones when BB and NB are used together (BB+NB). Fig. 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). By eye, we see that resulting distributions when using only BB (dashed) show overall shapes close to Gaussian with perhaps larger tails on both sides. However, when NB are also included (solid) the peaks of the distributions become clearly sharper. This is more noticeable in the BS than in the FS, probably because the non-observed condition $\sigma_m > 0.5$ defined in Section 3.3 make that most of NB will be not used in the photo-z determination in the FS. Table 3.5 shows bias (median), σ_z (σ_{68}) and 3σ -outliers fraction of each distribution. σ_z in the BS is reduced ~ 4.8 times going from $\sim 3.34\%$ to $\sim 0.7\%$ when NB are included, while the

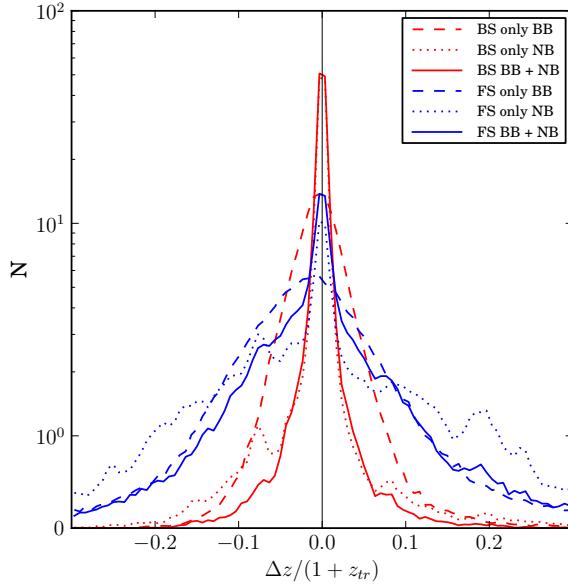


FIGURE 3.14: Normalized $\Delta z/(1+z_{tr})$ distributions for the BS (red) and the FS (blue) using only BB (dashed), only NB (dotted) or both together BB+NB (solid). Bias, σ_z and outliers fraction of each distribution are shown on Table 3.5.

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

TABLE 3.5: Bias (median), σ_z (σ_{68}) and 3σ -outliers fraction (rows) when using only BB, only NB or both together BB+NB (columns). Left column corresponds to the BS and right to the FS.

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 outliers fraction increases, but as was mentioned this is due to the fact that improvements on σ_z penalizes the outliers fraction. On the other hand, we see that using NB alone slightly degrades all metrics in the BS and the FS, except the outliers fraction in the FS which is improved for the same reason conversely. In fact, σ_z in the FS gets almost twice worst than when only using BB or BB+NB. It seems that, in the FS, NB themselves only help on the improvement of the Bias, while if they are used together with the BB, the improvement also extends to σ_z .

TABLE 3.6: Global photo-z performance results for each filter set shown in Fig. 3.15. On the left column we show results for the BS and in the right for the FS. Photo-z performance is parametrized through the three metrics: Bias (median), σ_z (σ_{68}) and the 3σ -outliers fraction. A photo-z quality cut of completeness 50 per cent is applied in all the cases.

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

3.5 Optimization of the PAU filter set

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

3.5.1 NB filter set variations

We propose five variations of the original NB filter set whose response is shown in Fig. 3.15. All the variations conserve the number of filters. In the order that appear in Fig. 3.15, the proposed filter sets are:

- **Default:** This is the default filter set already shown in Fig. 3.2.

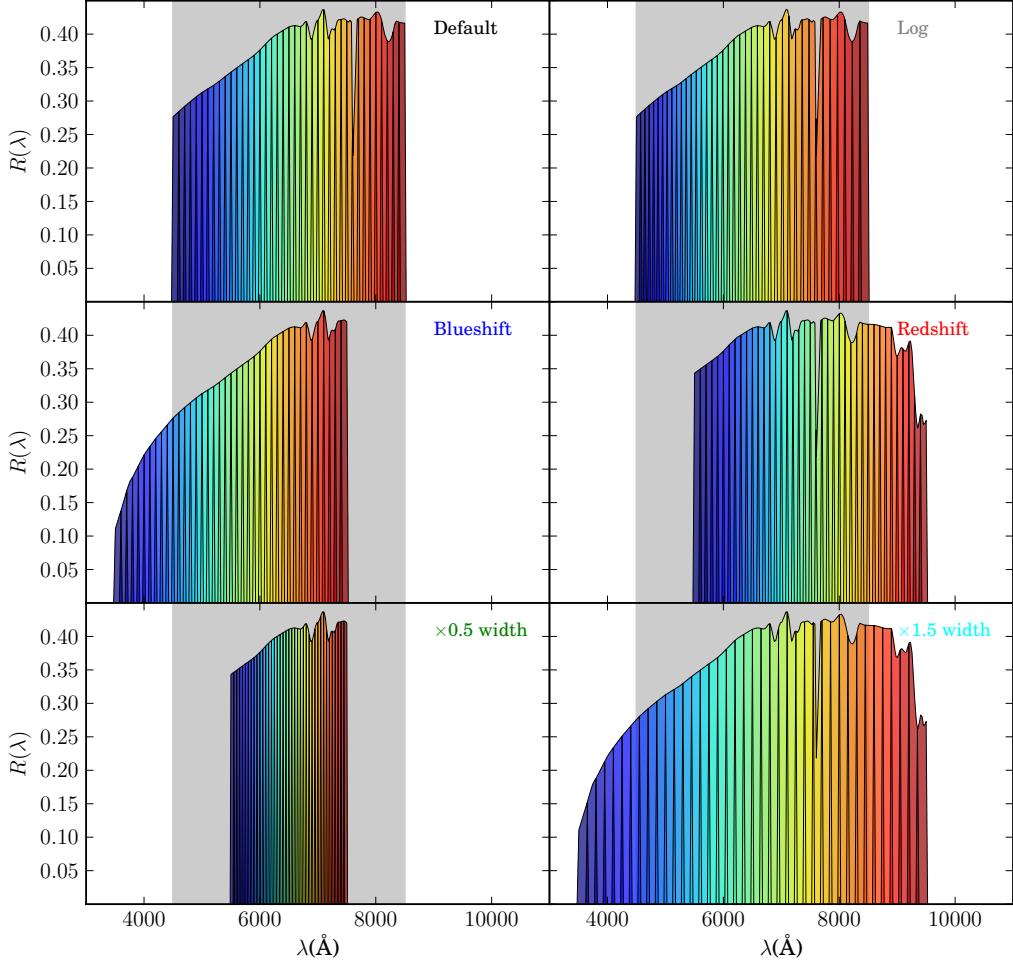


FIGURE 3.15: On the top-left, the original PAU NB filter set (same as Fig. 3.2), the rest are the five variations proposed to be compared in terms of photo-z performance. Grayed areas show the covered wavelength range by the **default** filter set. In descending order from left to right we have: the **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 1000Å towards bluer wavelengths, the **Redshift** filter set which is the same as before but shifted towards redder wavelength, the **$\times 0.5$ width** filter set whose band widths are half the **Default** ones and the **$\times 1.5$ width** filter set whose bands are 1.5 times wider. The overall wavelength range of these last two is chosen to be centered respect to the **Default** one.

- **Log:** In this filter set, band widths increase in wavelength logarithmically, so they fulfill $\lambda_0/\Delta\lambda = ctn.$, where $\Delta\lambda$ is the width of the band and λ_0 is the wavelength of the band where the response starts to be non-zero (without taking into account the lateral wings). We impose the wavelength range covered by the overall bands to be the same as the **Default** filter set. Given that the total number of bands is kept to be 40, we obtain that the bluest filter has a width of 72Å while the reddest 134Å. The point is that, when spectra are redshifted, their spectral features are moved to redder wavelengths, but also their widths are stretched as $\Delta\lambda' = (1+z)\Delta\lambda$. If the photo-z determination depends strongly on the tracking of any spectral feature,

such as the 4000Å break in the elliptical galaxies, a filter set like `Log` will continue enclosing the same part of the spectral 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 early-type galaxies. The down side of this filter set is that transmission curves $T_i(\lambda)$ shown in Fig. 3.2 turn to be very inefficient at the ultraviolet zone where these bands reach and therefore, their effective response is considerably reduced (see middle-left of Fig. 3.15). That is the same issue that makes the *u* band to have such low response.
- **Redshift:** This is the same variation as previous but shifting bands 1000Å towards redder wavelengths. In contrast, we expect to get better photo-z performance at high redshift and later-type galaxies. Moreover, this filter set does not suffer from the problem of the bad transmission curves $T_i(\lambda)$ at the ultraviolet, so its band responses are much more uniform and efficient over the covered range (see middle-right of Fig. 3.15).
- **$\times 0.5$ width:** This is a filter set whose band widths are the half of the `default` ones. Lateral wings are also reduced to the 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 slightly, 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 covered wavelength range is also reduce to the half. We choose it to be centered respect to the `Default`, so that it ranges from 5500Å to 7500Å. This makes the wavelength coverage take the worst of the `Blueshift` and the `Redshift` filter sets, the lack of coverage on both the bluer and the redder parts of the spectra. This could lead to degradation of photo-zs at very low and high redshift, although broad bands will probably attenuate this effect.
- **$\times 1.5$ width:** This is a filter set whose band widths are 1.5 times wider to the default ones. We expect a slight degradation of the photo-z precision for galaxies with good Signal-to-Noise ratio on their photometry, such as in the Bright Sample, however the rise of the limiting magnitudes will probably help on the Faint Sample. Moreover, the fact that the number of bands is kept, makes the covered wavelength range also increases 1.5 times. We choose the new range to be centered respect to the `Default` set, so that it ranges from 3500Å to 9500Å. In contrast to the $\times 0.5$ width filter set, this takes the best of the `Blueshift` and the `Redshift` filter sets in terms of wavelength coverage, so we expect to see a more uniform photo-z performance over the whole redshift range.

We generate an independent mock catalog for each filter set as described in Section 3.3 using the same exposure times per NB as in the `Default` filter set. We are aware that, since exposure times have been specifically chosen to work optimal with this configuration, the `Default` filter set plays in advantage. However, the only aim of this study is to see how, in spite of this, the photo-z performance changes. Once the 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 `Default` filter set. There is no need to calibrate a different prior for each filter set since it was initially calibrated on the broad band i which is also shared among all these filter sets.

3.5.2 Global photo-z performance results

Global photo-z performance results for each filter set are shown on Table 3.6 throughout the same metrics used in Section 4.3: Bias (median), σ_z (σ_{68}) and the 3σ -outliers fraction. We find that the `width` set gives the best Bias (it completely vanishes), and σ_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 `width` set which gives the best σ_z ($\sim 18\%$ better). On the contrary, the `width` set gives the worst Bias (a factor 4.6 worse) and σ_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 `width` which gives the worst σ_z ($\sim 46\%$ worse). Regarding the outliers fraction, its direct comparison is trickier since it strongly depends on the value of σ_z , so we do not want to do much emphasis on it. Even so, we see that in the BS the `Log` set gives its best value while in the FS the `width` set gives its worst.

General conclusions are that the `Log` set gives almost the same photo-z performance as the `Default` set with a slightly increase of 3% in σ_z in the BS. Therefore, at the end, we see that the logarithmically increase 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 photo-z performance results. On the contrary, wider bands in the FS are the ones which give better photo-z precision, but this is because their Signal-to-Noise ratio in that sample, even being quite limited, is higher enough that more bands fulfill the non-observed cut $\sigma_m < 0.5$ introduced in Section 3.3. Finally, we also see that shifting bands towards bluer or redder wavelengths has a direct impact on the Bias which is mostly noticeable in the Faint Sample. The redder their are shifted, the less Bias we observe. This is probably due to the fact that we are measuring redshifted spectral so the most part of the signal

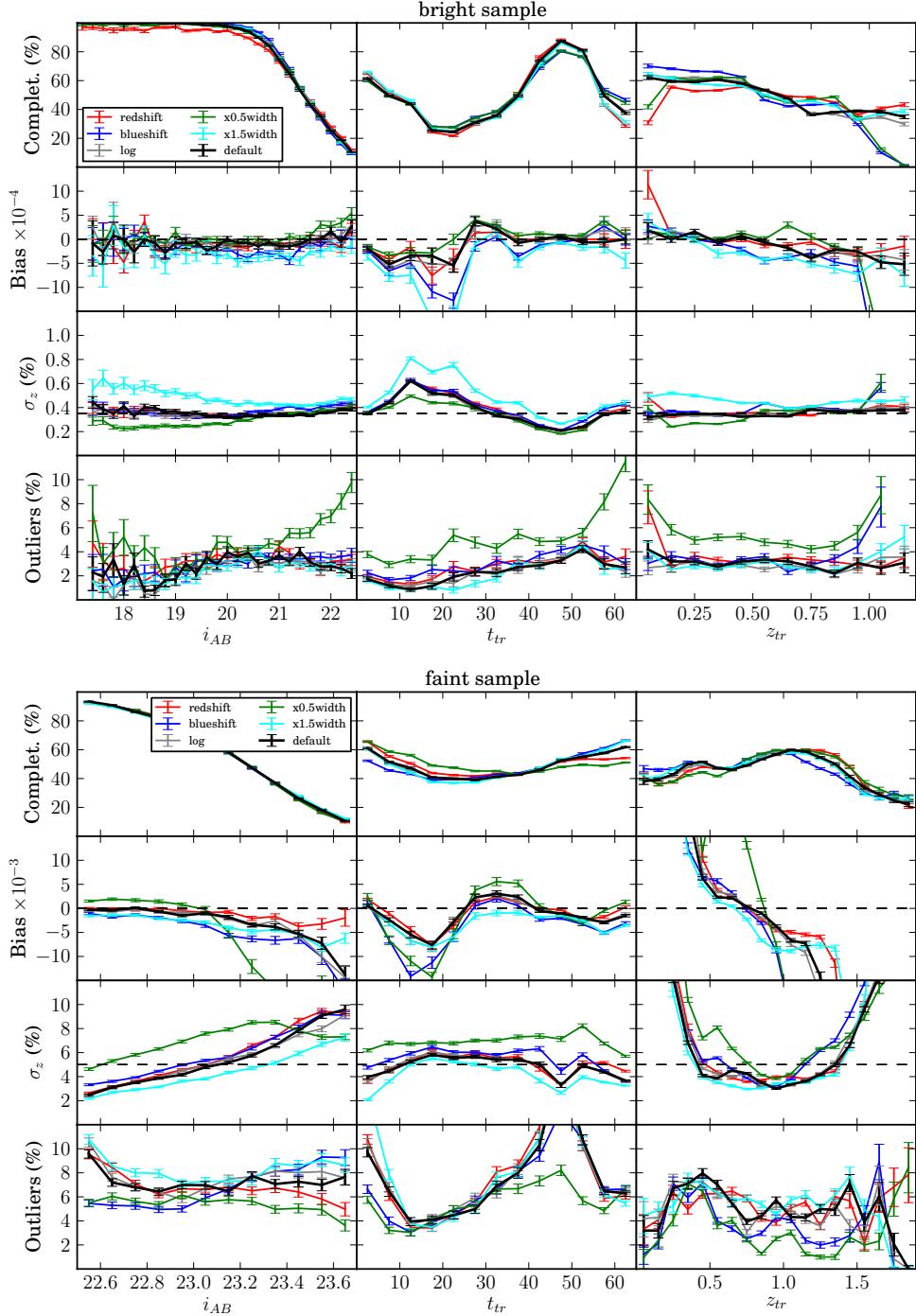


FIGURE 3.16: Plots with very similar layout as in Figs. 3.12 and 3.13 showing the different photo-z performance metrics (rows): Completeness after applying the 50% of global completeness photo-z quality cut, Bias (median), σ_z (σ_{68}) and 3 σ -outliers fraction, in function of i_{AB} , t_{tr} and z_{tr} (columns), in the BS (top) and FS (bottom) when using the different filter sets shown in Fig. 3.15.

will come from red, and that transmission curves $T_i(\lambda)$ are substantially less efficient at the blue part of the spectrum (see Fig. 3.2).

3.5.3 Results in function of i_{AB} , t_{tr} and z_{tr}

In Fig. 3.16 we show very similar plots as in Figs. 3.12 and 3.13 of the photo-z performance metrics in 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. 3.15. 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 much deviate 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 of each filter set respectively as we have already mentioned before. Some kind of the same happens when band widths and consequently the covered wavelength range are reduced to the half (green curve). Since the resulting filter set gets the lack of coverage on both high and low wavelengths, the resulting photo-z performance is worse at both low and high z_{tr} . However, the photo-z precision σ_z is slightly better at intermediate redshifts $0.15 < z_{tr} < 0.5$, spiral galaxies $10 < t_{tr} < 30$ and low magnitudes $i_{AB} < 20.5$. Increase the band width a factor 1.5 (cyan curve) does not suppose an improvement in any case. Bias and σ_z degrade all over the range of the three variables, i_{AB} , t_{tr} and z_{tr} . This in full agreement with the already shown results in Table 3.6 where this filter set was giving the worst photo-z performance. Also in agreement with the table, 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 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 **Redshift** and the $\times 0.5$ width filter sets which work again as the opposite. The **Blueshift** filter set also seems to cause a big degradation on the Bias for faint, spiral and at high redshift galaxies, and also a considerable worst σ_z over

all the ranges than the `default`. In return, the `Redshift` filter set shows better bias at high magnitudes and redshifts. On the other hand, we observe that the `×0.5 width` filter set shows much more accused behaviors than the other filter sets 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 the `Default` filter set. In accordance with Table 3.6, we observe that the worst photo-z precision σ_z is found for the `×0.5 width`, while the best is for the `×1.5 width` filter set, at about the whole three variable ranges. This is exactly the opposite behavior seen at the BS. Curiously, we find that σ_z at $i_{AB} > 23.4$ for the `×0.5 width` filter set gets even better than `Default` and very close to `×1.5 width`. However, in general we can say that narrower bands are useful in the BS, but not on the FS.

3.6 Discussion and Conclusions

* Default filter ser OK

* Results for bright and faint

* Relevance for cosmological analysis

Chapter 4

Photo-z Quality Cuts and their Effect on the Measured Galaxy Clustering

4.1 Introduction

During the last decades, galaxy surveys (SDSS, York et al. [55]; PanSTARRS, Kaiser, Tonry & Luppino [37]; 2dF, Colless et al. [16]; LSST, Tyson et al. [54]; VVDS, Le Fèvre et al. [40]; WiggleZ, Drinkwater et al. [22]; BOSS, Dawson et al. [21]) have become crucial tools for our understanding of the geometry, content and destiny of the universe. Spectroscopic surveys provide 3D images of the galaxy distribution in the near universe, but many suffer from limited depth, incompleteness and selection effects. Imaging surveys 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 galaxy position along the line of sight, which is obtained measuring the galaxy redshift through photometric techniques using a set of broadband filters (but see Benítez et al. [4] for alternative ideas).

There are two main sets of techniques for measuring photometric redshifts (photo-zs): template methods (e.g. Hyperz, Bolzonella, Miralles & Pell [7]; BPZ, Benitez [3] & Coe et al. [14]; LePhare, Ilbert et al. [34]; EAZY, Brammer, van Dokkum & Coppi [8]), in which the measured broadband galaxy spectral energy distribution (SED) is compared to a set of redshifted templates until a best match is found, thereby determining both the galaxy type and its redshift; and training methods (e.g. ANNz, Collister & Lahav [18]; ArborZ, Gerdes et al. [28]; TPZ, Carrasco Kind & Brunner [13]), in which a set of galaxies for which the redshift is already known is used to train a machine-learning algorithm (an artificial neural network, for example), which is then applied over the

galaxy set of interest. Each technique has its own advantages and disadvantages, whose discussion lies beyond the scope of this paper.

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 [3]. Applying cuts on the value of this quality parameter, one can clean up the sample from galaxies with an unreliable photo-z determination [3], or even select a smaller sample of galaxies with significantly higher photo-z precision [41].

However, we will show in this paper 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. [30] to deal with, among others, the effect of stars contaminating a galaxy sample.

The outline of the paper is as follows. Section 4.2 discusses the galaxy samples that we use in our study: the Mega-Z photometric galaxy sample [17] and its companion, the 2SLAQ spectroscopic sample [11], 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 4.4 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 4.5 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 4.6 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 4.7 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.

4.2 Data Samples

The Mega-Z LRG DR7¹ catalog [17] includes ~ 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 $\sim 7750 \text{ deg}^2$ of the sky that is displayed in Fig. 4.1. This

¹An 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>.

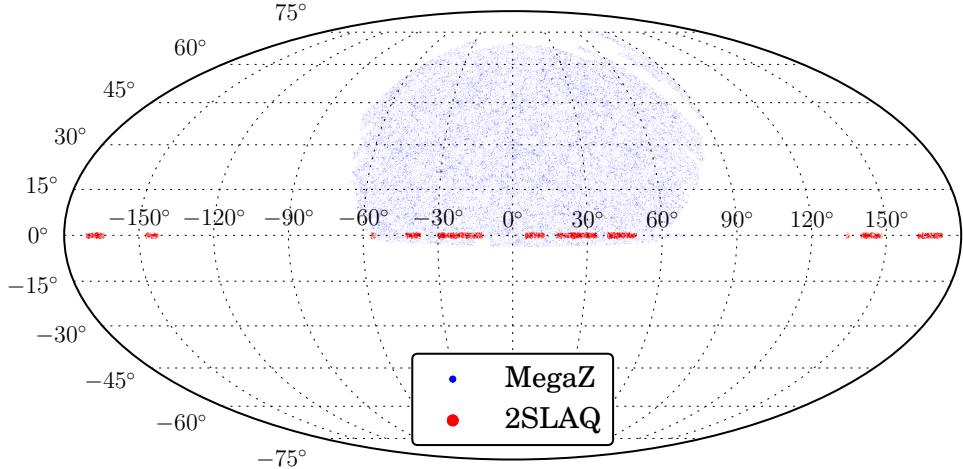


FIGURE 4.1: Mega-Z and 2SLAQ maps in Mollweide projection plotted in blue and red respectively. For the sake of clarity only a hundred thousand galaxies randomly selected from Mega-Z and five thousand from 2SLAQ have been plotted. The Mega-Z sample covers a total of 7750 deg^2 , while 2SLAQ covers 180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

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 Quasar² (2SLAQ) catalog [11] was obtained using the same selection criteria as the Mega-Z catalog and includes ~ 13100 LRGs with spectroscopic redshifts. Its sky coverage of only 180 deg^2 can be seen in Fig. 4.1 in red. The galaxies are located on a strip of 2° along the celestial equator, the area subtended by the 2dF spectrograph. Only non-repeated objects ($ind = 1$) with high spectroscopic redshift confidence level ($hqz \geq 3$) 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 4.3). 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 magnitude $i_{deV} = 20$, 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

²The whole 2SLAQ data can be downloaded from http://www.2slaq.info/query/2slaq_LRG_webcat_hdr.txt.

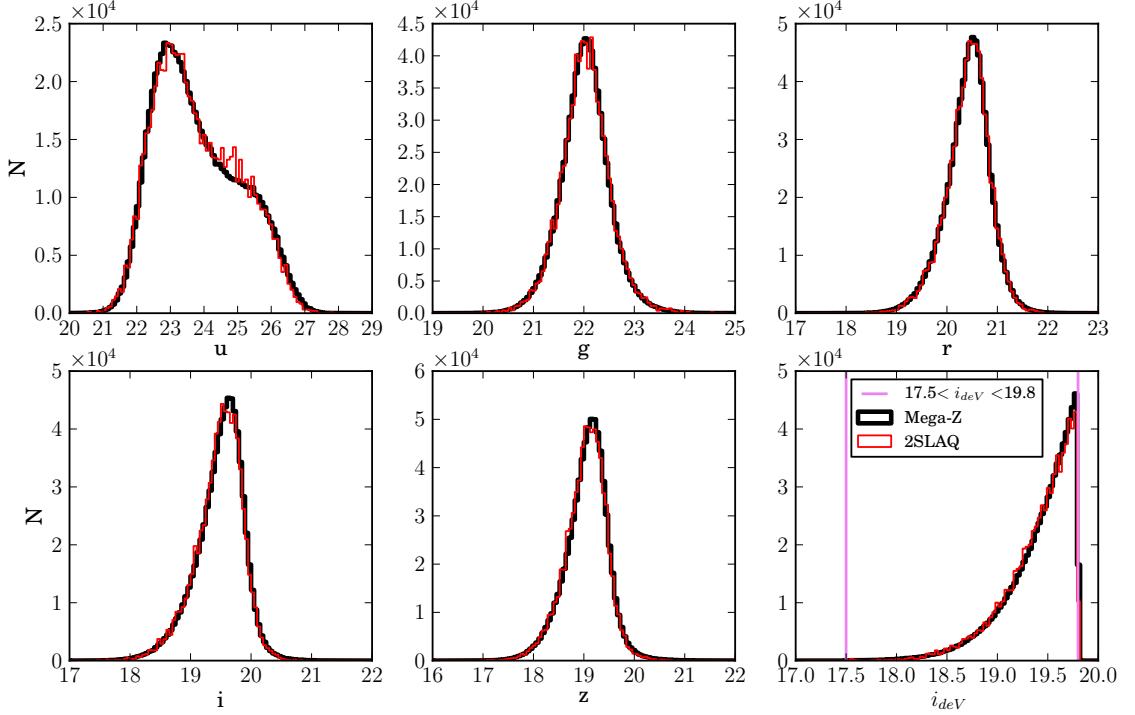


FIGURE 4.2: From top-left to bottom-right, the model magnitude distributions in the $ugriz$ bands for 2SLAQ in red and Mega-Z in black, normalized to each other. The last plot corresponds to the i band de Vaucouleurs magnitude. All magnitudes are corrected for galactic extinction. The purple lines in 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 ~ 25 .

galaxies leaving a total of ~ 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:

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

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

$$c_{\parallel} \equiv 0.7(g - r) + 1.2(r - i - 0.18) > 1.6 \quad (4.4)$$

$$d_{\perp} \equiv (r - i) - (g - r)/8.0 > 0.55. \quad (4.5)$$

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. [17] the d_{\perp} cut is set to 0.5 for the Mega-Z catalog, but, once again, the 2SLAQ completeness within $0.5 < d_{\perp} < 0.55$ 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 $d_{\perp} > 0.55$.

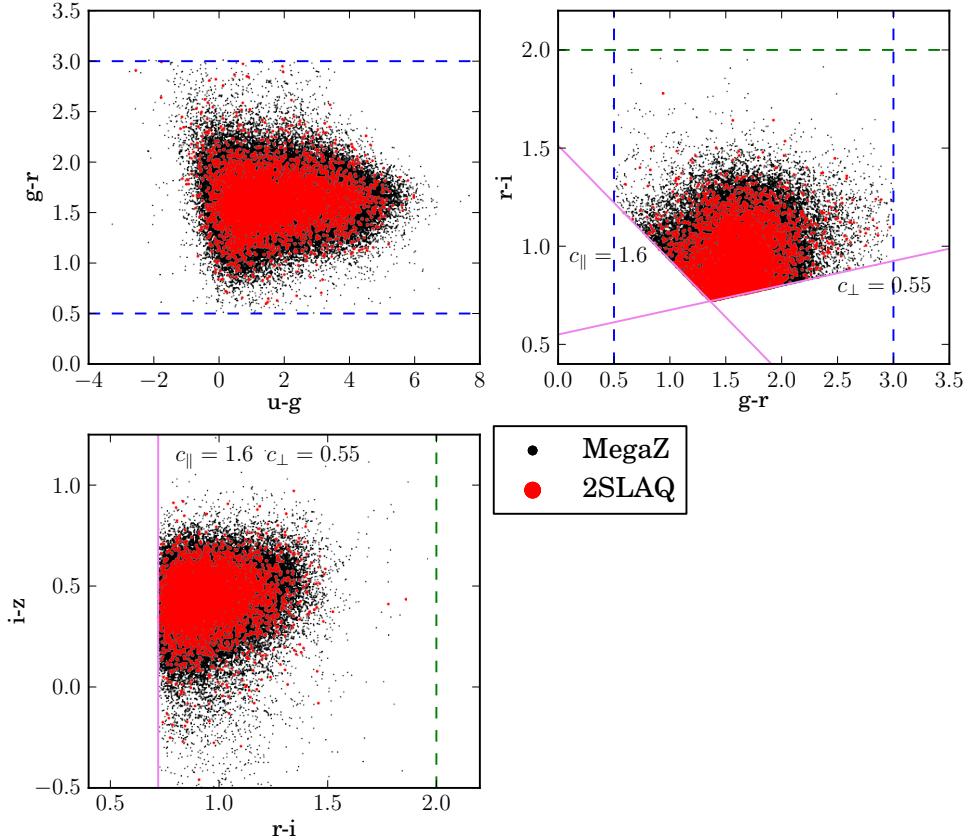


FIGURE 4.3: Color-color diagrams for both the Mega-Z catalog in black and the 2SLAQ catalog in red. For the sake of clarity, we have only plot a hundred-thousand galaxies for Mega-Z and five thousand for 2SLAQ. We can see that 2SLAQ covers the same color area as Mega-Z, so that we can conclude that it 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.

These magnitude and color cuts leave a total of 749152 objects in the Mega-Z 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. [17], 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 [18] 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 δ_{sg} that objects be galaxies. Removing all objects with probability below 0.2 reduces the stellar contamination from 5% to 2% [17]. In 2SLAQ, we can remove stars by simply getting rid of all those objects with redshift less than 0.01.

4.3 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 Redshifts³ (BPZ) template-fitting code described in Benitez [3] 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 :

$$p(z|m_i) \propto \sum_t L(m_i|z, t) \Pi(z, t | m_i), \quad (4.9)$$

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

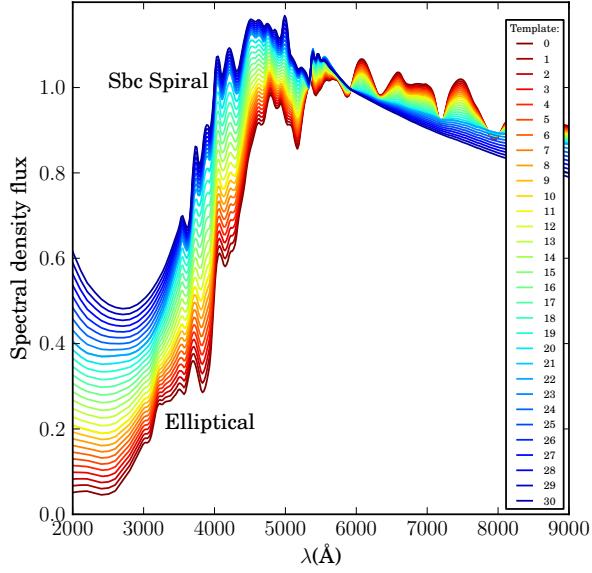


FIGURE 4.4: The spectral galaxy templates used in the determination of the 2SLAQ and Mega-Z photo-zs. There are a total of 31 templates that range from elliptical galaxies, the vast majority in both samples, to Sbc spiral galaxies.

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`⁴, another template-based photo-z code described in Arnouts et al. [2], Ilbert et al. [34]. Both libraries are based on Coleman, Wu & Weedman [15], Kinney et al. [39], 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: `Ell_01`, `Ell_09`, `Ell_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.

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. [25], whose throughputs are obtained from <http://home.fnal.gov/~annis/astrophys/filters.new>.

⁴The 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.

t	f	k	α	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.

A crucial point of BPZ is the prior probability $\Pi(z, t | m)$ that helps improve the photo-z performance. Benitez [3] proposes the following empirical function:

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

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 $m_0 = 18.5$. 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 σ_z as half of the symmetric interval that encloses the 68% of the Δz distribution area around the maximum, we also find that its value for the non-catastrophic determinations improves by a factor 1.7, down to $\sigma_z \sim 0.042$, in agreement with Collister et al. [17], Padmanabhan et al. [45], Thomas, Abdalla & Lahav [53].

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 $\sigma_z \sim 0.041$.

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 $z(phot)$, or something more complex, but the aim is the same. In BPZ, this indicator is called *odds*, and, it is defined as

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

where δz 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 $z(phot) \pm \delta z$. 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 δz 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 Δz , in the sense that higher *odds* should correspond to lower $|\Delta z|$. In Fig. 4.5, we show σ_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 δz . 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 $\sigma_z \sim 0.042$ to ~ 0.028 , a 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:

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

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 $\sigma_z \sim 0.035$, as shown in the bottom plot in Fig. 4.5. Since we cannot compute $\sigma_z(0\%)$ 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.

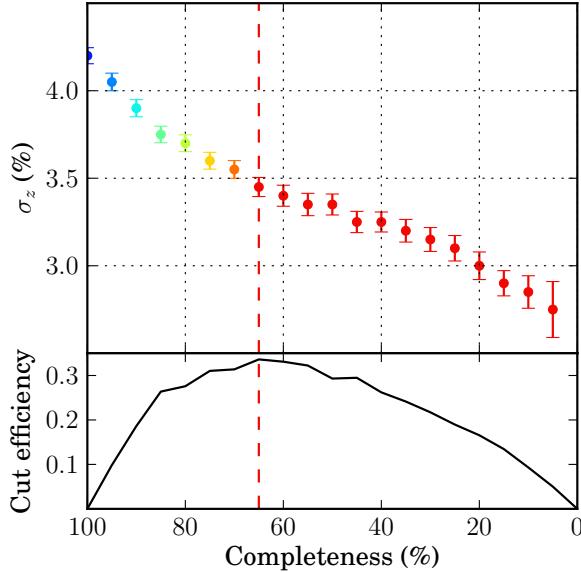


FIGURE 4.5: Top plot: the 2SLAQ photo-z precision σ_z for different photo-z quality cuts resulting in the completeness shown on the x axis. The error bars are computed using bootstrap [23]. The color scale labels the different photo-z quality cuts, here and also in Figs. 4.6 and 4.8. The nominal precision, 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.

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 Δz , 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, $z(phot)$ on the right. In the first row, we can see the Δz 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, especially when quality cuts become harder. It is quite constant along redshift. The bias (median), in the forth row, shows a general offset of ~ 0.02 , so that photo-zs are in general $\sim 4\%$ larger than the actual redshift. The photo-z precision σ_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 σ_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.

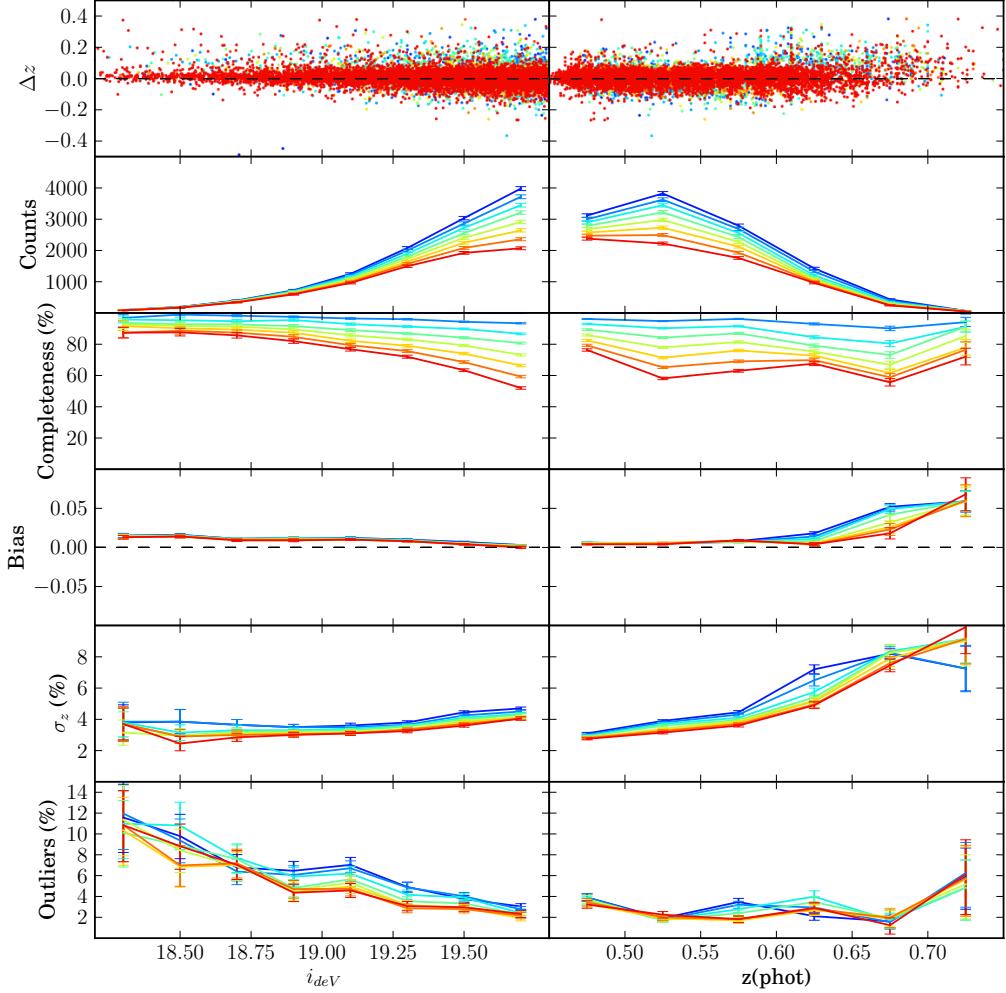


FIGURE 4.6: Statistics showing the 2SLAQ photo-z performance. In the first row we show the scatter of $\Delta z \equiv z(\text{phot}) - z(\text{spec})$ with respect to the i_{deV} magnitude (left) and $z(\text{phot})$ (right). The scatter has been binned along these two variables and some statistical 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 σ_z and the 3σ outlier fraction. The color degradation from red to blue is the same as in Fig. 4.5 and labels different photo-z quality cuts.

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

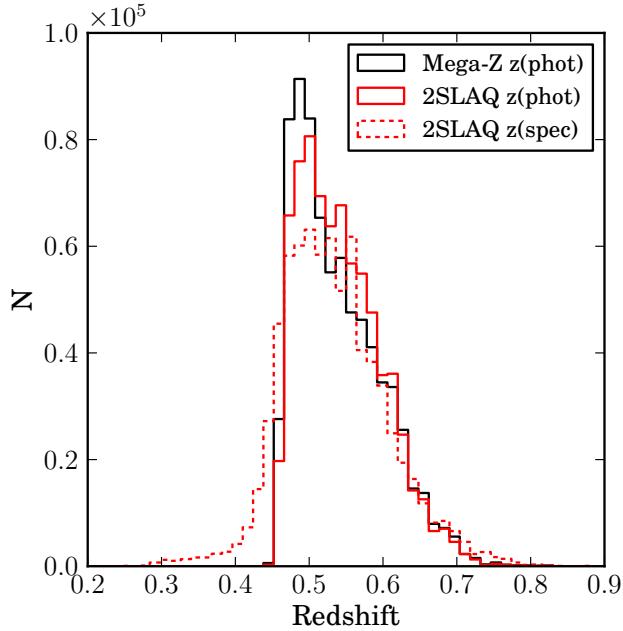


FIGURE 4.7: The photo-z distribution of Mega-Z (black), and the spectroscopic (red dashed) and photometric (red solid) redshift distributions of 2SLAQ. The 2SLAQ distributions have been normalized to the Mega-Z number of galaxies for comparison.

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 σ_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.

4.4 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 [29], developed for CMB data analysis. It provides pixelations of the sphere

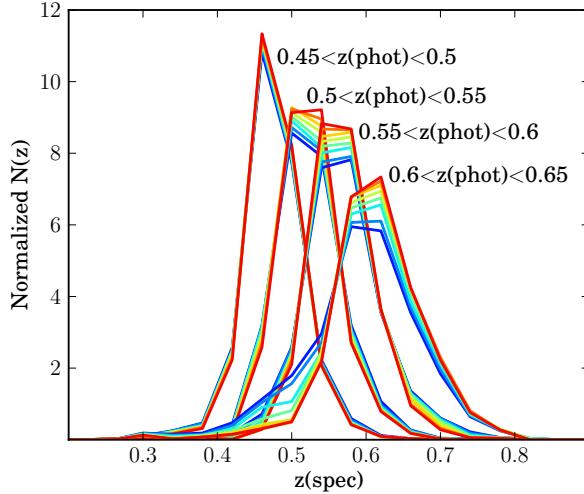


FIGURE 4.8: The spectroscopic redshift distributions in the 2SLAQ catalog for the four photo-z bins in which we will measure galaxy clustering in section 4.4. They are all normalized to the same area under the curves. Different colors label different photo-z quality cuts, as in Figs. 4.5 and 4.6.

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 database⁵ 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 & Gaztañaga [10].

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° 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 ~ 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

⁵<http://casjobs.sdss.org/>.

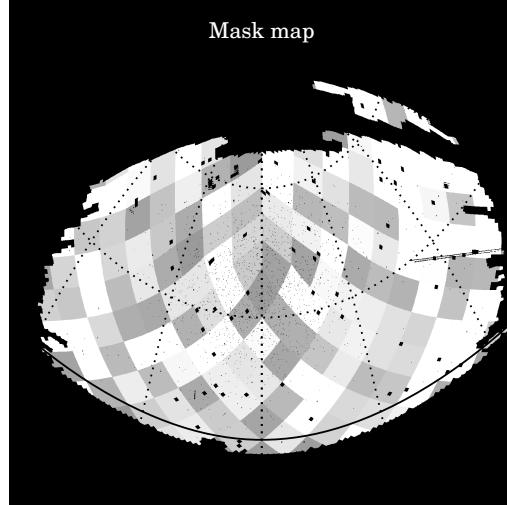


FIGURE 4.9: The Mega-Z DR7 mask in Healpix of $N_{side}=512$. It is obtained in two steps. First, using a low resolution Healpix map of $N_{side} = 64$, we reject all those pixels with less than 65 galaxies per pixel to get rid of the underpopulated areas (black patches) and obtain the overall geometry 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 $N_{side} = 8$.

on the bin $0.5 < z < 0.55$, we see that the map shows a strip pattern very similar to the bluer zones of its corresponding *odds* map. In Fig. 4 of Crocce et al. [19] 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:

$$\omega_{ab}(\theta) \equiv \langle \delta_a \delta_b \rangle(\theta) = \frac{1}{N_\theta} \sum_{i,j}^{N_\theta} \delta_{a,i} \delta_{b,j}, \quad (4.13)$$

where $\delta_{a,i} = a_i/\bar{a} - 1$ is the fluctuation of the map a at the pixel i with respect to the mean \bar{a} , and N_θ is the total number of combinations of pixels i and j separated by an angular distance between θ 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° . We also want to know the covariance

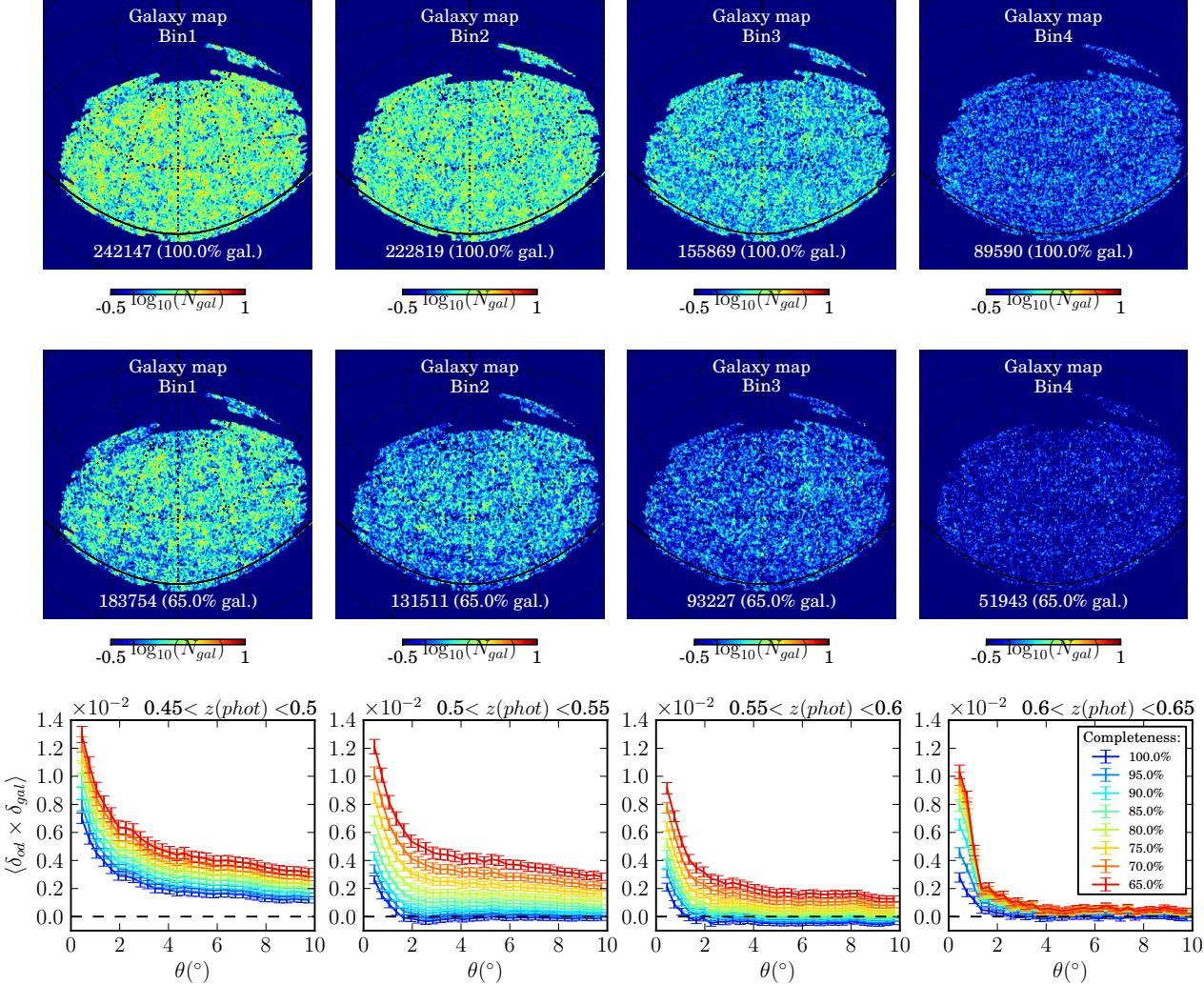


FIGURE 4.10: On the first row, the galaxy maps of the Mega-Z catalog for the four photo-z bins of Fig. 4.8. On the second row, the same after applying a photo-z quality cut with 65% completeness. These are Healpix maps of $N_{side} = 256$. The number of galaxies is given in each map. On the bottom row, 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.

of $\omega_{ab}(\theta)$. We can compute it using the *jackknife* technique, also used in a similar study in Crocce et al. [19] and explained in detail in Cabre et al. [9]. This technique consists of splitting the survey area within the mask into 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

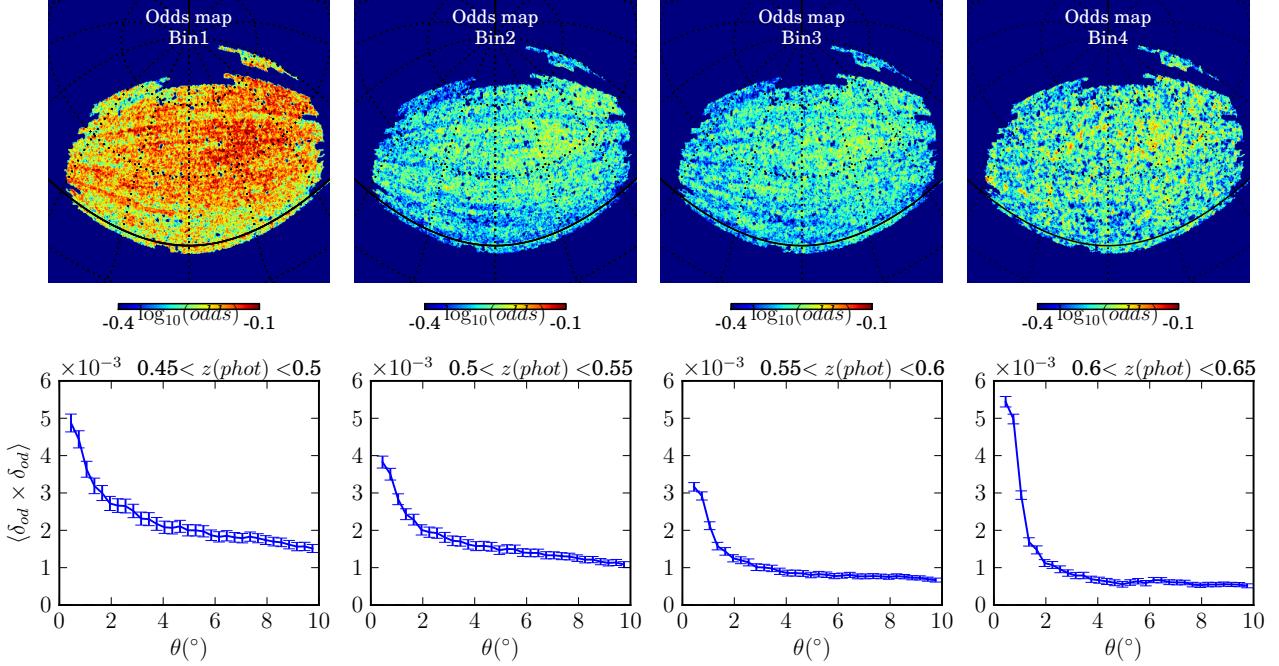


FIGURE 4.11: On the first row, the *odds* maps of the Mega-Z catalog for the four photo-z bins of Fig. 4.8. These are Healpix maps of $N_{\text{side}} = 256$ where the *odds* values per pixel are computed as the mean *odds* of all the galaxies in each pixel. Redder regions are regions with higher photo-z 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 $z(\text{phot})$.

between the correlation functions measured at angles θ_n and θ_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 $\bar{\omega}(\theta) = \sum_k^{N_k} \omega_k(\theta)/N_k$ is the mean of all the jackknife correlations. Therefore, the diagonal errors of the measured correlation at θ 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 $\omega_{ab}(\theta)$ as described in [19]:

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

where $N_i(z)$ are the selection functions, which in our case are the curves in Fig. 4.8, $\xi^s(z_1, z_2, \theta)$ is the redshift space correlation of the pairs of galaxies at redshift z_1 and z_2 subtending an angle θ with the observer. We use the non-linear power spectrum [51] with ΛCDM with $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$ and $H_0 = 70$ (km/s)/Mpc, the linear Kaiser [36] model of redshift space distortions for the correlation function, and a linear bias model with evolution: $b(z) = 1.5 + 0.6(z - 0.1)$ [10]. 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 2σ) 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 [6], Thomas, Abdalla & Lahav [52], and, in Crocce et al. [19], 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

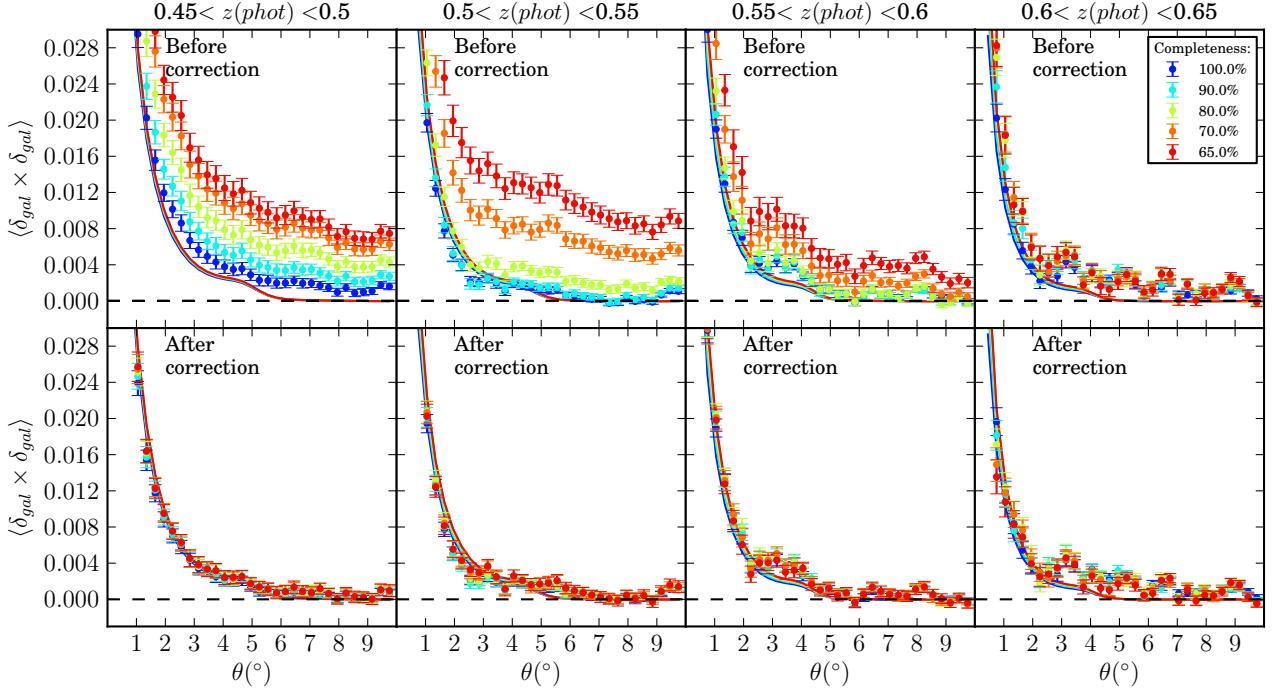


FIGURE 4.12: The angular galaxy auto correlations in the four photo-z bins of Fig. 4.8 with different photo-z quality cuts labeled with colors. The upper plots show the results before applying the *odds* correction in (4.23), while the lower plots show the results after applying it. Points with error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

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 *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.

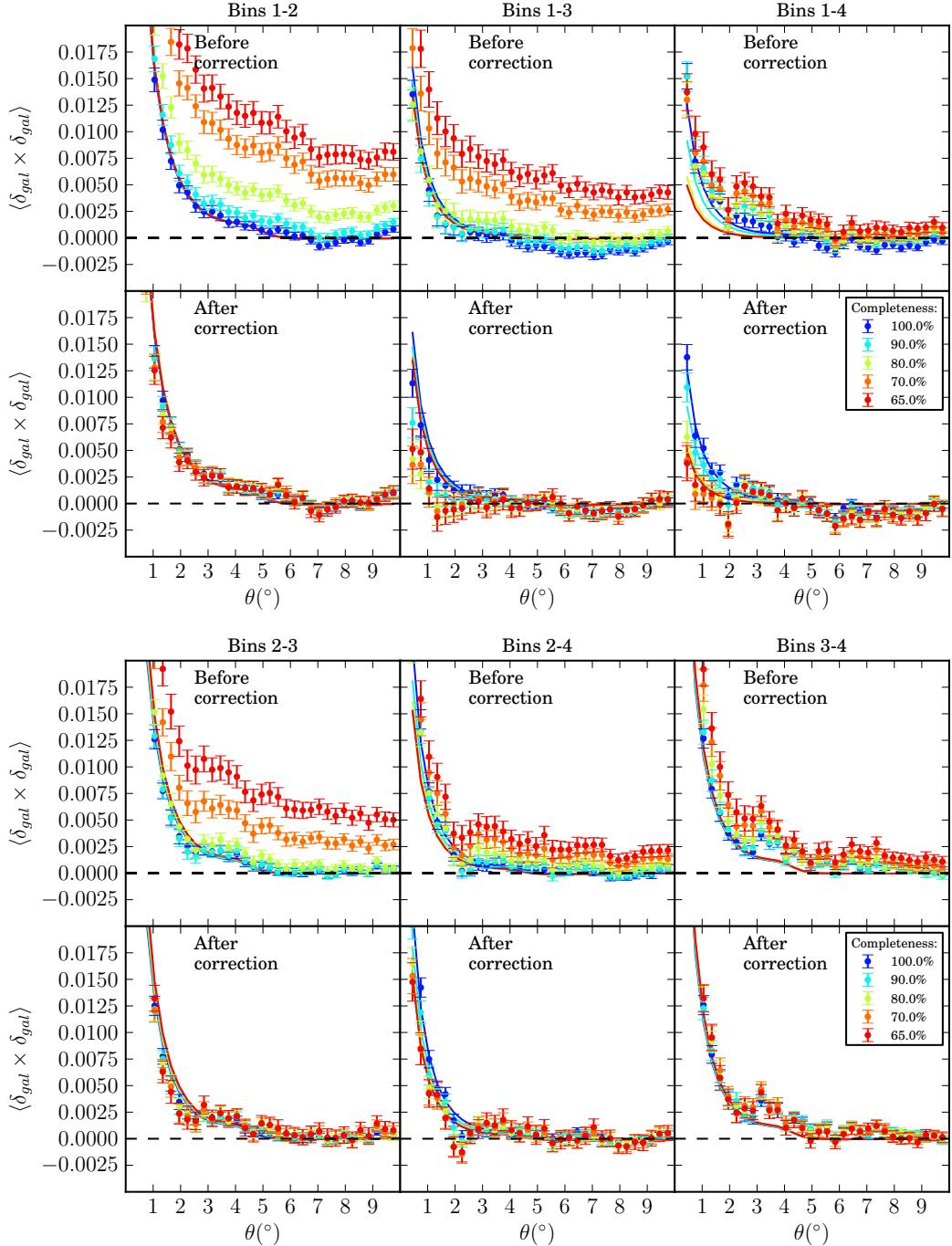


FIGURE 4.13: All the possible combinations of the angular galaxy cross correlations between the different photo-z bins of Fig. 4.8 with different photo-z quality cuts labeled with different colors. The upper plots show the results before applying the *odds* correction in (4.22), while the lower plots 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).

4.5 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. [30] and Ross et al. [48], which the authors use for the treatment of systematic effects that influence the SDSS-III galaxy clustering, such as the stellar contamination, the sky brightness or the image quality of the instrument. Following Ho et al. [30] and Ross et al. [48], the density fluctuation δ_i of the systematic effect i modifies the true galaxy density fluctuation δ_g^t through a linear contribution modulated by ϵ_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 δ_g^t .

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

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

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

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

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

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

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

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

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:

$$\omega_{g1,g2}^t(\theta) = \omega_{g1,g2}(\theta) - \frac{\omega_{g1,od1}(\theta)}{\omega_{od1,od1}(\theta)} \frac{\omega_{g2,od2}(\theta)}{\omega_{od2,od2}(\theta)} \omega_{od1,od2}(\theta), \quad (4.22)$$

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

$$\omega_{g,g}^t(\theta) = \omega_{g,g}(\theta) - \frac{\omega_{g,od}^2(\theta)}{\omega_{od,od}(\theta)}, \quad (4.23)$$

where $\omega_{g,g}^t(\theta) \equiv \langle \delta_g^t \delta_g^t \rangle_\theta$ is the true galaxy auto-correlation, $\omega_{g,g}(\theta) \equiv \langle \delta_g \delta_g \rangle_\theta$ is the observed one, $\omega_{g,od}(\theta) \equiv \langle \delta_g \delta_{od} \rangle_\theta$ is the galaxy-*odds* cross-correlation, and $\omega_{od,od}(\theta) \equiv \langle \delta_{od} \delta_{od} \rangle_\theta$ 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 $\omega_{g,g}(\theta)$ and $\omega_{g,od}(\theta)$ 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, $\langle \delta_{od1} \delta_{od2} \rangle$, are needed. In our case, those are the maps on the top row of Fig. 4.11. We compute all the

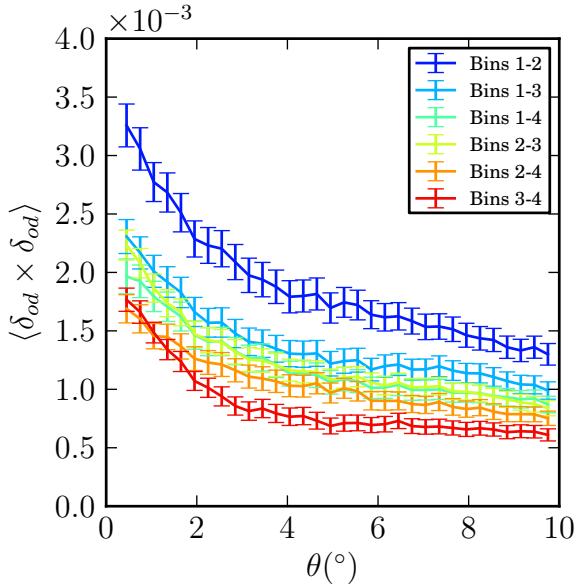


FIGURE 4.14: All the possible combinations of the angular *odds* cross-correlations between the different maps on the top row of Fig 4.11. They are needed for the *odds* correction in (4.22). We see that all odds maps are cross-correlated with each other, at least up to angles $< 10^\circ$. However,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

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 $\text{Cov}_\omega(\theta_n, \theta_m)$ 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 $\omega_k(\theta_n)$. 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 2σ 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 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 [5].

4.6 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 [24, 27, 33, 44, 46, 47]. 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 [32]) $z_d = 1059.25 \pm 0.58$ [1], and those waves stopped traveling. At that time, they had traveled $r_s(z_d) = 147.49 \pm 0.59$ Mpc [1] 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 $\Omega_M = 0.315^{+0.016}_{-0.018}$ [1] in a flat Λ 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 θ_{BAO} in angular clustering was presented in Carnero et al. [12]. They made use of a method described in Sánchez et al. [50] that consists of fitting the empirical function

$$\omega(\theta) = A + B\theta^\gamma + Ce^{-(\theta-\theta_{FIT})^2/2\sigma^2} \quad (4.26)$$

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 γ describe the typical decreasing profile of $\omega(\theta)$, with γ 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, θ_{FIT} is the location of that peak and σ 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 θ_{FIT} and θ_{BAO} . Sánchez et al. [50] shows that this can be successfully corrected as follows:

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

where α is a factor that only depends on the mean redshift z and the width Δz of the bin. Figure 3 in Sánchez et al. [50] shows these dependences. In general terms, the wider the redshift bin, the larger the shift of θ_{FIT} towards smaller angles. For example, the shift is $\sim 5\%$ of θ_{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 4.3, 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 ~ 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 ~ 0.12 . We find that for those bins the shift in θ_{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. [30] and Ross et al. [48], 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 θ_{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 4.4, 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, θ_{BAO}^{theo} , changes slightly when

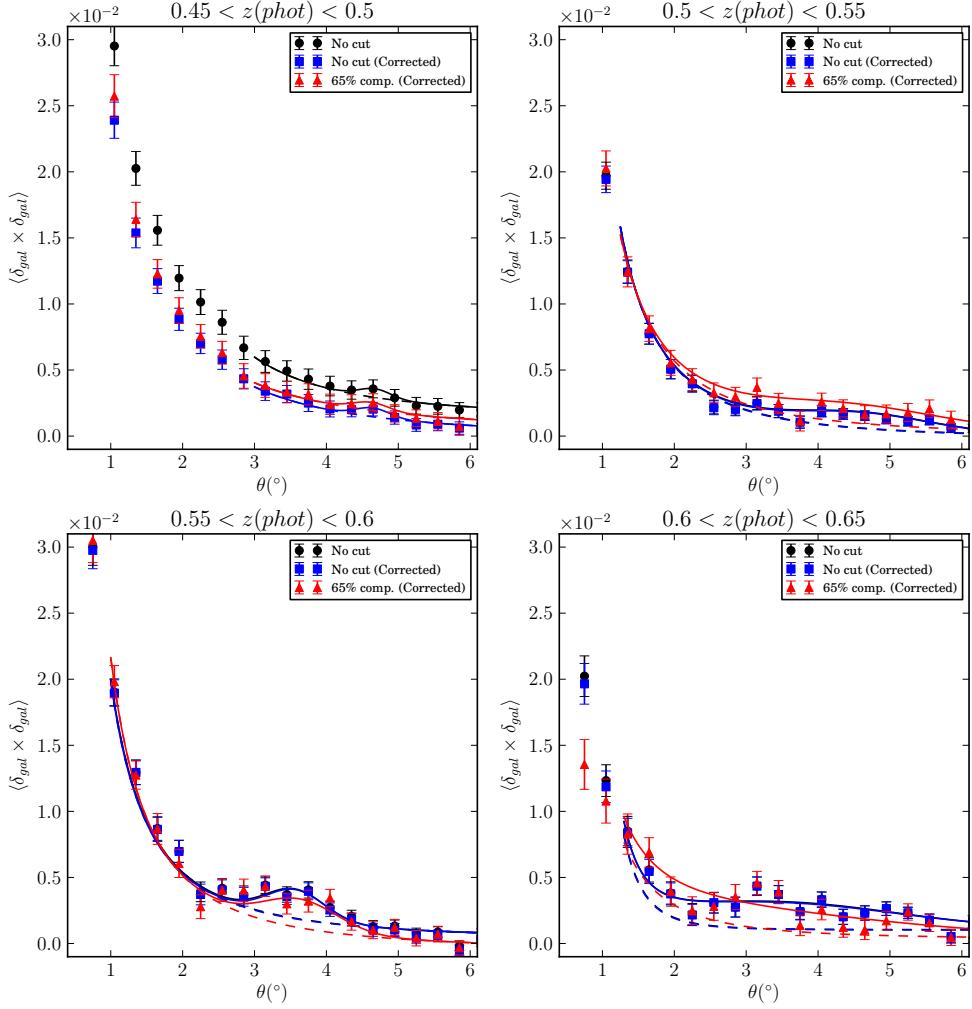


FIGURE 4.15: Results of the fits of (4.26) to the galaxy auto-correlations of Fig. 4.12 to extract the BAO scale. In black when no photo-z quality cut is applied, in blue the same but correcting for *odds* and, in red when the 65% completeness cut is applied and corrected. Error bars correspond to observations. 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).

the photo-z quality cut is applied. This is because θ_{BAO}^{theo} depends on 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 θ_{BAO} .

We can see that in some bins there are some discrepancies between the extracted values of θ_{BAO}^{obs} and the expected θ_{BAO}^{theo} . This is most likely due to systematic effects that lie beyond the scope of this paper. 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 θ range chosen. Fits performed using the whole data covariance matrices (estimated with jack-knife) result in values for the χ^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

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 θ_{BAO}^{theo} contain the position of the BAO peak at the mean redshift in each bin derived from the latest Planck results [1]. 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	$(0.7 \pm 0.3) \cdot 10^{-3}$	$(0.7 \pm 0.3) \cdot 10^{-3}$	$(0.6 \pm 0.4) \cdot 10^{-3}$
θ_{FIT}	4.68 ± 0.09	4.65 ± 0.09	4.62 ± 0.12
θ_{BAO}^{obs}	5.03 ± 0.09	5.00 ± 0.08	4.97 ± 0.12
θ_{BAO}^{theo}	4.51 ± 0.08	4.51 ± 0.08	4.50 ± 0.08
0.5 < z < 0.55			
C	$(1.2 \pm 0.4) \cdot 10^{-3}$	$(1.2 \pm 0.4) \cdot 10^{-3}$	$(1.5 \pm 0.6) \cdot 10^{-3}$
θ_{FIT}	4.58 ± 0.33	4.59 ± 0.32	4.40 ± 0.60
θ_{BAO}^{obs}	4.92 ± 0.27	4.94 ± 0.27	4.73 ± 0.41
θ_{BAO}^{theo}	4.16 ± 0.08	4.16 ± 0.08	4.16 ± 0.08
0.55 < z < 0.6			
C	$(2.2 \pm 0.4) \cdot 10^{-3}$	$(2.2 \pm 0.4) \cdot 10^{-3}$	$(2.2 \pm 0.6) \cdot 10^{-3}$
θ_{FIT}	3.56 ± 0.06	3.57 ± 0.06	3.63 ± 0.13
θ_{BAO}^{obs}	3.83 ± 0.06	3.84 ± 0.06	3.90 ± 0.10
θ_{BAO}^{theo}	3.88 ± 0.07	3.88 ± 0.07	3.88 ± 0.07
0.6 < z < 0.65			
C	$(2.1 \pm 0.5) \cdot 10^{-3}$	$(2.0 \pm 0.5) \cdot 10^{-3}$	$(2.0 \pm 4.4) \cdot 10^{-3}$
θ_{FIT}	3.34 ± 0.72	3.30 ± 0.77	1.90 ± 7.35
θ_{BAO}^{obs}	3.63 ± 0.77	3.59 ± 0.81	2.07 ± 4.29
θ_{BAO}^{theo}	3.75 ± 0.07	3.75 ± 0.07	3.64 ± 0.07

result in very similar central values for θ_{FIT} , but with larger errors and values of the χ^2 per degree of freedom around 1. The 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.

4.7 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 [52] and in Crocce et al. [19]. 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 [19] 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 paper.

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. [30], Ross et al. [48] 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.

Appendix A

Appendix Title Here

Write your Appendix content here.

Bibliography

- [1] Ade P., Others, 2013, arXiv:1303.5076
- [2] Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *Mon. Not. R. Astron. Soc.*, 310, 540
- [3] Benitez N., 2000, *Astrophys. J.*, 536, 571
- [4] Benítez N. et al., 2009, *Astrophys. J.*, 691, 241
- [5] Benjamin J., Van Waerbeke L., Ménard B., Kilbinger M., 2010, *Mon. Not. R. Astron. Soc.*, 408, 1168
- [6] Blake C., Collister A., Lahav O., 2008, *Mon. Not. R. Astron. Soc.*, 385, 1257
- [7] Bolzonella M., Miralles J., Pell R., 2000, *Astron. Astrophys.*, 492, 476
- [8] Brammer G. B., van Dokkum P. G., Coppi P., 2008, *Astrophys. J.*, 686, 1503
- [9] Cabre A., Fosalba P., Gaztanaga E., Manera M., 2007, *Mon. Not. R. Astron. Soc.*, 381, 1347
- [10] Cabré A., Gaztañaga E., 2009, *Mon. Not. R. Astron. Soc.*, 393, 1183
- [11] Cannon R. et al., 2006, *Mon. Not. R. Astron. Soc.*, 372, 425
- [12] Carnero A., Sánchez E., Crocce M., Cabré A., Gaztañaga E., 2012, *Mon. Not. R. Astron. Soc.*, 419, 1689
- [13] Carrasco Kind M., Brunner R. J., 2013, *Mon. Not. R. Astron. Soc.*, 432, 1483
- [14] Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, *Astron. J.*, 132, 926
- [15] Coleman G. D., Wu C.-C., Weedman D. W., 1980, *Astrophys. J. Suppl. Ser.*, 43, 393
- [16] Colless M. et al., 2001, *Mon. Not. R. Astron. Soc.*, 328, 1039

- [17] Collister A. et al., 2007, *Mon. Not. R. Astron. Soc.*, 375, 68
- [18] Collister A. A., Lahav O., 2004, *Publ. Astron. Soc. Pacific*, 116, 345
- [19] Crocce M., Gaztañaga E., Cabré A., Carnero A., Sánchez E., 2011, *Mon. Not. R. Astron. Soc.*, 417, 2577
- [20] Dahlen T., Mobasher B., Somerville R. S., Moustakas L. A., Dickinson M., Others, 2005, *Astrophys.J.*, 631, 126
- [21] Dawson K. S. et al., 2013, *Astron. J.*, 145, 10
- [22] Drinkwater M. J. et al., 2010, *Mon. Not. R. Astron. Soc.*, 401, 1429
- [23] Efron B., 1979, *Ann. Stat.*, 7, 1
- [24] Eisenstein D. J. et al., 2005, *Astrophys. J.*, 633, 560
- [25] Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *Astron. J.*, 111, 1748
- [26] Gaztañaga E., Eriksen M., Crocce M., Castander F. J., Fosalba P., Martí Sanahuja P., Miquel R., Cabré A., 2012, *Mon. Not. R. Astron. Soc.*, 422, 2904
- [27] Gaztañaga E., Miquel R., Sánchez E., 2009, *Phys. Rev. Lett.*, 103, 091302
- [28] 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
- [29] Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *Astrophys. J.*, 622, 759
- [30] Ho S. et al., 2012, *Astrophys. J.*, 761, 14
- [31] Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 1996, arXiv:astro-ph/0210394v1, 1
- [32] Hu W., Sugiyama N., 1996, *Astrophys. J.*, 471, 542
- [33] Hütsi G., 2006, *Astron. Astrophys.*, 449, 891
- [34] Ilbert O. et al., 2006, *Astron. Astrophys.*, 457, 841
- [35] Jouvel S. et al., 2009, *Astron. Astrophys.*, 504, 359
- [36] Kaiser N., 1984, *Astrophys. J.*, 284, L9
- [37] Kaiser N., Tonry J. L., Luppino G. A., 2000, *Publ. Astron. Soc. Pacific*, 112, 768
- [38] Kennicutt, Robert C. J., 1992, *Astrophys. J. Suppl. Ser.*, 79, 255

- [39] Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, *Astrophys. J.*, 467, 38
- [40] Le Fèvre O. et al., 2005, *Astron. Astrophys.*, 439, 845
- [41] Martí P., Miquel R., Castander F. J., Others, 2014, Prep.
- [42] Moles M. et al., 2008, *Astron. J.*, 136, 1325
- [43] Oke J. B., Schild R. E., 1970, *Astrophys. J.*, 161, 1015
- [44] Okumura T., Matsubara T., Eisenstein D. J., Kayo I., Hikage C., Szalay A. S., Schneider D. P., 2008, *Astrophys. J.*, 676, 889
- [45] Padmanabhan N. et al., 2005, *Mon. Not. R. Astron. Soc.*, 359, 237
- [46] Padmanabhan N. et al., 2007, *Mon. Not. R. Astron. Soc.*, 378, 852
- [47] 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
- [48] Ross A. J. et al., 2011, *Mon. Not. R. Astron. Soc.*, 417, 1350
- [49] Sanchez C., Abdalla F., Miquel R., 2014, Prep.
- [50] Sánchez E. et al., 2011, *Mon. Not. R. Astron. Soc.*, 411, 277
- [51] Smith R. E. et al., 2003, *Mon. Not. R. Astron. Soc.*, 341, 1311
- [52] Thomas S. A., Abdalla F. B., Lahav O., 2011, *Phys. Rev. Lett.*, 106, 241301
- [53] Thomas S. A., Abdalla F. B., Lahav O., 2011, *Mon. Not. R. Astron. Soc.*, 412, 1669
- [54] Tyson J., Wittman D., Hennawi J., Spergel D., 2003, *Nucl. Phys. B - Proc. Suppl.*, 124, 21
- [55] York D. G. et al., 2000, *Astron. J.*, 120, 1579