Highlights of Spanish Astrophysics VII, Proceedings of the X Scientific Meeting of the Spanish Astronomical Society held on July 9 - 13, 2012, in Valencia, Spain. J. C. Guirado, L. M. Lara, V. Quilis, and J. Gorgas (eds.)

Photo-z study for the PAU@WHT survey

Pol Martí Sanahuja[1]

[1] IFAE Barcelona

Abstract

The PAU@WHT survey will study the properties of Dark Energy (DE) using the observations of Redshift Space Distortions (RSD) and Weak Gravitational Lensing (WL) from galaxy cross-correlations as main cosmological probes. The instrument used for this purpose, PAUCam, will be installed at the prime focus of the WHT in La Palma. Unlike other photometric instruments, PAUCam will use a filter set composed of 42 narrow bands and the 6 standard ugrizY wide bands. It can cover 2 deg^2 per night in all filters, delivering low-resolution (R50) spectra for 30000 galaxies, 5000 stars, 1000 quasars, and 10 clusters per night. Photo-z simulations applied on mock catalogues, with template-based codes, tell us that very precise redshifts $\sigma(z) \sim 0.35\%$ can be achieved for galaxies with $i_{AB} < 22.5$. A typical photo-z precision $\sigma(z) \sim 3.5\%$ is also achieved for galaxies with $22.5 < i_{AB} < 23.7$. Moreover, these results are valid not only for LRGs but all types of galaxies. Such redshift accuracy combined with a large galaxy density can provide a highly competitive determination of the DE parameters, even covering only a moderate area.

1 Introduction

In 1929 Edwin Hubble showed that Universe is expanding. The gravitational attraction of its matter and its energy content should slow down the expansion rate. However, in 1998, the study of the recession velocity of distant supernova showed that the expansion is in fact accelerating. According to the general relativity, this can only be possible if most of the content follows the equation of state $P = \omega\rho$, with $\omega < -0$, which means negative pressure. Nowadays, there is no empirical evidence of the existence of such kind of entity, so cosmologists have started to call it Dark Energy (DE). There are more proves supporting the presence of DE such as the Cosmic Microwave Background (CMB) or the Large Scale Structure (LSS). All this proves together suggest that Universe is made of $\sim 75\%$ DE, $\sim 20\%$ Dark Matter (DM) and $\sim 5\%$ Baryonic Matter (or visible matter). This agreement between different cosmological proves is called the Cosmic Concordance. The first step to understand the nature of the DE is to characterize its state parameter ω with enough precision to rule out different hypothesis such as the Modified Gravity, the Cosmological Constant Λ ($\omega = -1$), the Quintessence ($\omega \sim -1$), etc.

LSS analysis have demonstrated to be a powerful method to constrain cosmological parameters; Baryon Acoustic Oscillations (BAO), Cluster Counts, Redshift Space Distortions (RSD) or Weak Gravitational Lensing (WL), are some examples. All of them are based on the fact that the growth of structure is conditioned by the expansion rate of Universe. So, if we measure the formation of structures at different cosmic times (different redshifts z), we will be able to reconstruct the history of the expansion rate and therefore to know the DE necessary properties to produce it.

The LSS studies require huge amounts of data from distant galaxies at different redshifts. That is why large redshift surveys are being carried out currently. In [3] it is shown that RSD and WL measurements in combination with Magnification (MAG) made in the same area of the sky can improve DE constrains by a factor of ~ 100. Two samples are needed: one on the foreground with spectroscopic redshift precision of $\sigma(z) \sim 0.3\%$ for the RSD measurements and another one in the background with photometric redshift (photo-z) precision of $\sigma(z) \sim 3\%$ for the MAG.

PAU@WHT (www.pausurvey.org) is going to be a semi-spectroscopic survey that will cover 200 deg^2 of the sky in ~ 100 nights to provide two samples like these. PAU will mount a 1 deg^2 camera in the WHT telescope (La Palma) with an innovative system of 42 Narrow Bands (NB) of bandwidth ~ 100 Å and 6 Broad Bands (BB) ugrizY covering the optical wavelength range. It is expected to deliver photo-z of $\sigma(z) \sim 3\%$ with galaxies up to $i_{AB} < 22.5$ and a order of magnitude worse within $22.5 < i_{AB} < 23.7$. In this document, we want to show that this precision can be achieved, at least, at the simulations level.

2 The PAU photo-z

In order to simulate the photo-z measurements for the PAU@WHT survey we are going to generate a mock galaxy catalog following [4]. That is, sampling a luminosity function that has been calibrated with some real data from, for example, GOODS, UDF, COSMOS or VVDS surveys. Then, with a collection of different spectral templates of galaxies we can generate the relative magnitudes m_i in all the 48 PAU bands according to their spectral type t, their true z and their absolute magnitude M_0 in some band. We have used the Extended Coleman library from [2] and Charlot & Bruzual 1996 models, which contains 66 interpolated templates representing elliptical, spiral, irregular and starburst galaxies. We add Gaussian noise to all these magnitudes through the signal-to-noise $S/N = \sqrt{N_{\rm gal} + N_{\rm Sky} + RN^2}$, where $N_{\rm gal}$ is the number of counts in photons from the galaxy, $N_{\rm Sky}$ the number of counts from the sky brightness and RN the read-out noise of the camera. Finally, we split the catalog in two disjoint samples: the Bright Sample (BS) with ~ 70000 galaxies and magnitude $i_{AB} < 22.5$, which is the 5σ limiting magnitude of the NB, and the Faint Sample (FS) with ~ 110000 galaxies and magnitude $22.5 < i_{AB} < 23.7$, which is the 5σ limiting magnitude of the BB.

Once the mock catalog is ready, we can run a photo-z algorithm on it. There are two differentiated methods: the template based method and the training method. The first one uses spectral templates to fit the observed data and the second one fits some general function by learning from a spectroscopic subsample. We choose to use the template based method which, in principle, does not depend on any training sample. Moreover, we already have the spectral templates used in the catalog generation. Specifically, we choose the Bayesian Photometric Redshift (BPZ) code described in [1]. This is a template based method that takes advantage of the Bayesian probability to refine the photo-z performance $\Delta z \equiv z(\text{phot}) - z(\text{true})$ by reducing considerably the fraction of catastrophic redshifts (photo-z that differ a lot from the true value $|\Delta z| > 1$). In this framework, the probability to find a galaxy at redshift z when the observed magnitudes are m_i is given by the probability density distribution (PDF):

$$p(z \mid m_i, m_0) \propto \sum_t \pi(z, t \mid m_0) L(m_i \mid z, t). \tag{1}$$

The likelihood $L(m_i \mid z, t)$ is the exponential of the Chi square of the predicted magnitudes and the real ones. The prior $\pi(z, t \mid m_0)$ is the expected true redshift and type distribution at some reference magnitude m_0, which is modeled through a function of 11 parameters:

$$\pi(z, t \mid m_0) \propto f_t e^{-k_t(m_0 - 20.)} \cdot z^{\alpha_t} \exp\left\{ -\left[\frac{z}{z_{mt}(m_0)} \right]^{\alpha_t} \right\}. \tag{2}$$

The values of the parameters are obtained by fitting the function to the true reference magnitude, the true z and the true type t from a subsample of the mock catalog of 2000 galaxies. In the real world, we could think of having a spectroscopic training sample or a simulation that provide this information. The resulting values are showed in the Table 1:

Table 1: Parameters of the prior, $\pi(z, t \mid m_0)$

Spectral Type t	α	z_0	k_m	k	f
Ell	2.593	0.420	0.116	0.219	0.500
Sp	1.866	0.356	0.101	0.000	0.419
Irr/SB	1.387	0.188	0.146	–	–

In Fig. 1 we show the scatter plot $z(\text{phot})$ versus $z(\text{true})$ for the FS with and without using the prior. The prior helps to reduce the presence of catastrophics by a factor of more than 2.

Another way to improve the photo-z performance is to apply a photo-z quality cut. Imagine that the PDF is spread over all the redshift range with multiple peaks, then we could think that the photo-z value is more likely to be wrong than if all the PDF is concentrated around a single peak. We define the odds value that quantifies this as the integral of the PDF within an interval δz centered at $z(\text{phot})$:

$$\text{odds} \equiv \int_{z(\text{phot}) - \delta z}^{z(\text{phot}) + \delta z} p(z \mid m_i, m_0) dz. \tag{3}$$

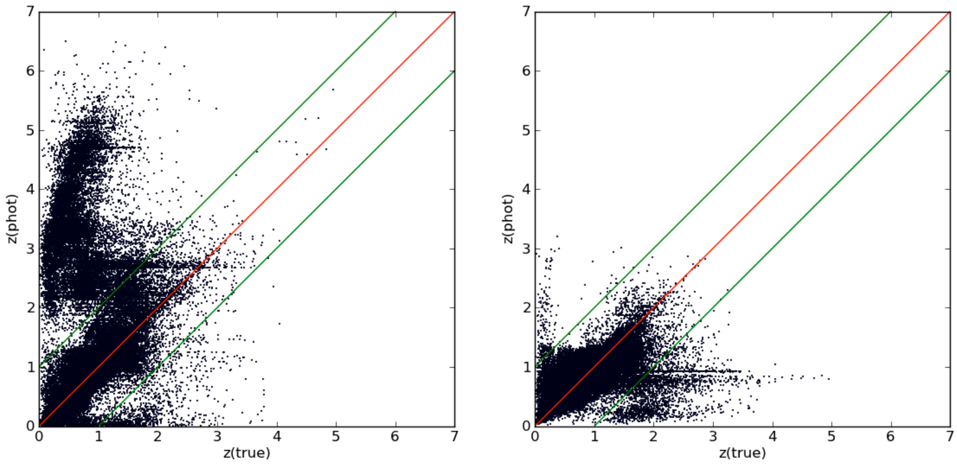

Figure 1: *Left panel*: z(phot) versus z(true) without using the prior in the FS. *Right panel*: Same using the prior. Prior helps to reduce catastrophics, defined as $\mid z(\text{phot}) - z(\text{true}) \mid > 1$ (green lines), by a factor of more than 2.

In Fig. 2 we show the scatter plot Δz versus odds for the BS (left) and the FS (right). Clearly, the dispersion increases as odds get lower. The green line shows the cumulated RMS for decreasing odds, so the curve increases when more galaxies with low odds are considered. Therefore, we can increase the photo-z precision by removing galaxies with bad odds. We cut galaxies with odds < 0.8 in the BS and with odds < 0.5 in the FS. This halves the catalog but allows us to achieve a photo-z precision of $\sigma_{68}(z) \sim 0.3\%$ and $\sigma_{68}(z) \sim 3\%$ respectively, which is exactly the PAU requirement.

In Fig. 3 we show the scatter plot z(phot) versus z(true) for the BS (left) and for the FS (right) after the odds cut.

In Fig. 4 we show some statistical analysis of $\Delta z/(1+z)$ through z(true). The layout is the same as in the previous plots. The top plots correspond to the bias. The black line is the mean and the blue line is the median. Bias is practically negligible in the BS, while in the FS it is manifested at low $z < 0.5$ and high $z > 1.5$ redshift. The middle plots show the photo-z precision. The black line is the RMS and the blue is the σ_{68}. We can see, in the BS, that σ_{68} is below the PAU photo-z requirement (the flat red line) while the RMS is partially above but very close. Note that at some redshifts the RMS values are particularly high and their error bar large. This is because the RMS is more sensitive to isolated outliers. In the FS, the typical photo-z precision is an order of magnitude worse, but it is still very competitive. At low and high redshift, we see again a bad behavior due to the catastrophics that prior has not been able to fix. The bottom plots correspond to the 3σ outliers fraction. This is the fraction of galaxies whose Δz is above 3 sigmas. We can use either RMS (black) or σ_{68} (blue) as reference. Realize that the outliers fraction from the σ_{68} is larger than the one from RMS. This makes sense since σ_{68} is generally below the RMS and therefore, the 3σ interval gets narrower and less permissive. At worst, we could say that the outliers fraction is $\leq 3\%$ in the BS and $\leq 10\%$ in the FS.

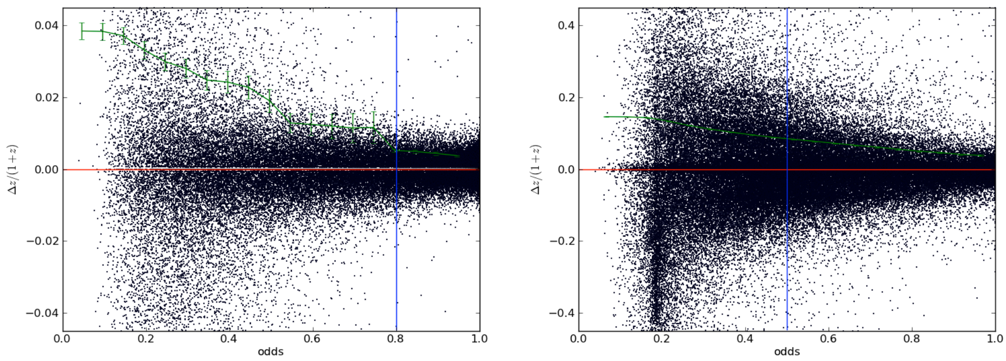

Figure 2: *Left panel*: Scatter plot of Δz versus odds in the BS. *Right panel*: The same in the FS. The dispersion is higher at low odds. This is reflected by the green curve which shows the cumulated RMS from high to low odds. An odds cut of odds > 0.8 in the BS and odds > 0.5 in the FS reduces the catalog to the half but lets achieve the photo-z precision of $\sigma_{68}(z) \sim 0.3\%$ in the BS and $\sigma_{68}(z) \sim 3\%$.

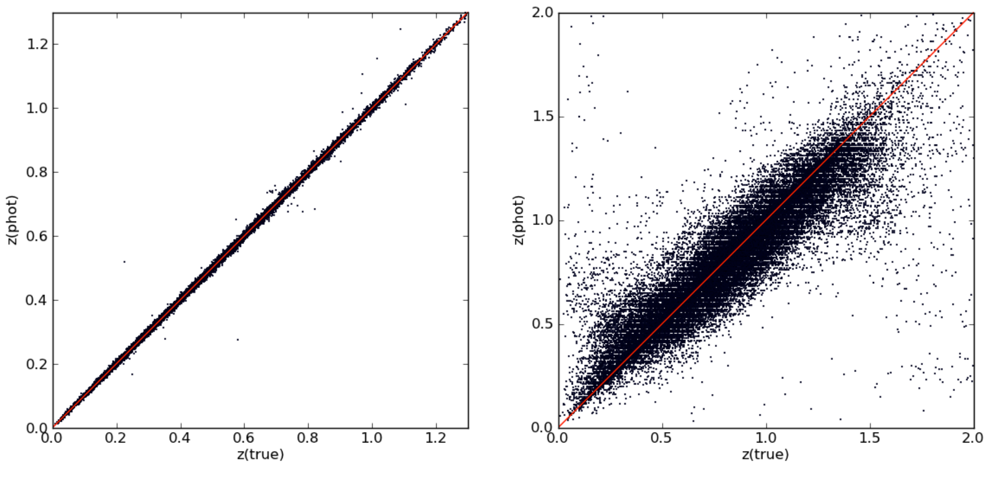

Figure 3: *Left panel*: Scatter plot of z(phot) versus z(true) in the BS. *Right panel*: The same in the FS.

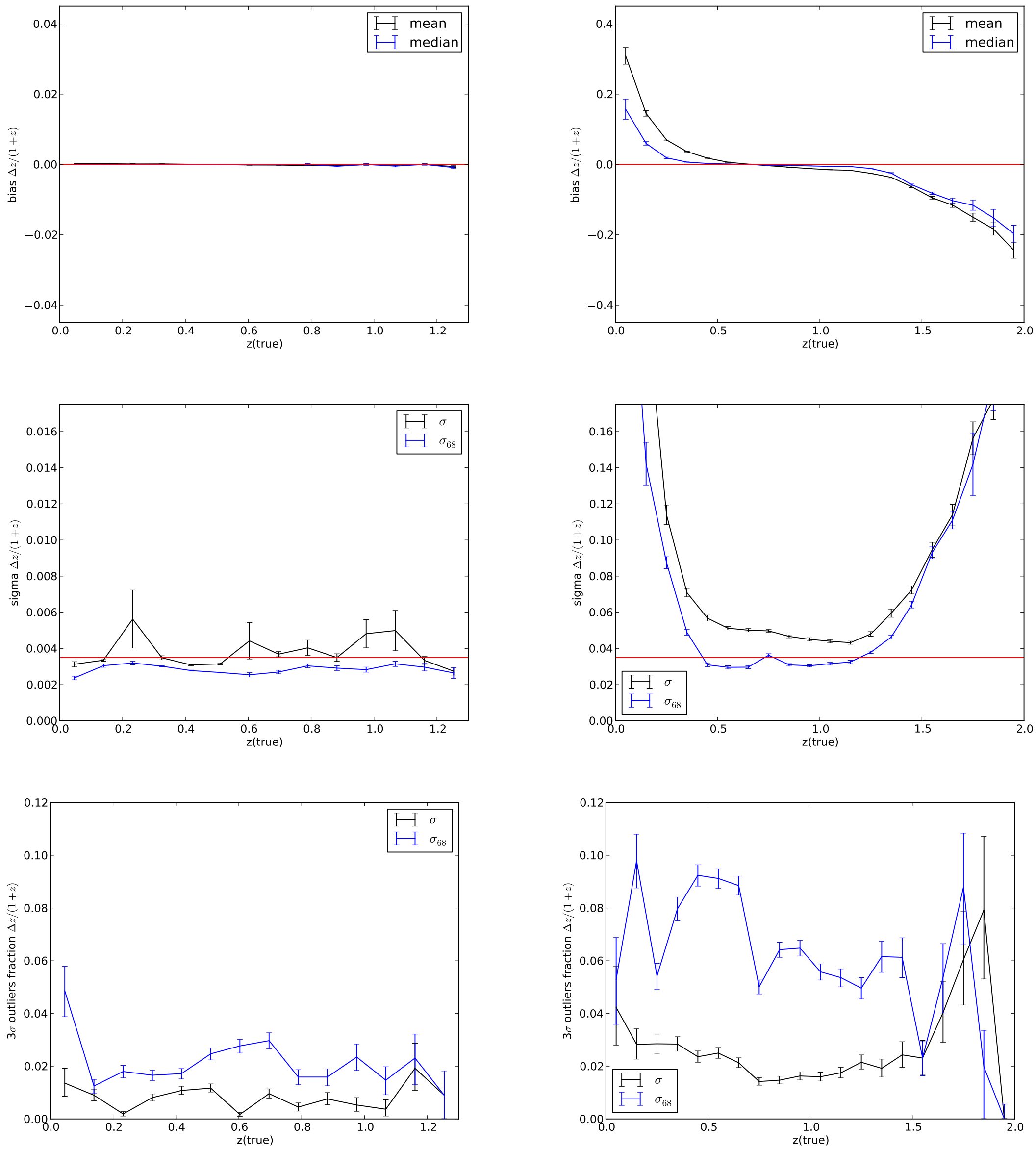

Figure 4: *Top*: Bias of $\Delta z/(1+z)$ through z(true) of the BS (left) and the FS (right). The blue line is the median and the black line is the mean. The red line shows non bias. *Middle*: The photo-z precision. The blue line is the σ_{68} and the black line is the RMS. The red line shows the precision requirement. Realize that the scale of the y axis is just an order of magnitude higher in the FS than in the BS. *Bottom*: The 3σ outliers fraction respect to σ_{68} (blue line) or the RMS (black line).

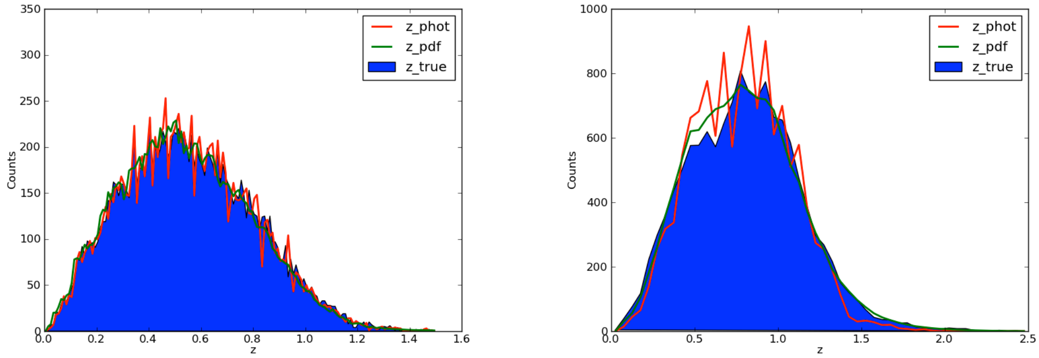

Figure 5: The redshift distribution through z(true) in the BS (*left*) and FS (*right*). The solid blue area is the true z distribution, the red line is the photo-z distribution and the green line is also the photo-z distribution but using the whole PDF information of each galaxy.

Finally, we have considered the possibility of using the whole pdf information of each galaxy to recover the true redshift distribution instead of using the single photo-z value. The results are shown in Fig. 5. On the right, the BS and on the left the FS. The solid blue area is the true redshift distribution, the red curve is the photo-z histogram and the green curve is the same but using the full PDF information. Note that the green curve is the one that fits better and in a smoother way the true distribution. This is because single galaxies contribute in all the redshift bins of the histogram. Of course, this contribution is higher in bins that are close to the actual photo-z, but the small contributions to distant bins tend to average counts and avoid spurious spikes.

References

[1] Benitez, N. 1998, ApJ, 536, 571

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

[3] Gaztañaga, E., Eriksen, M., Crocce, M., et al. 2012, MNRAS, 422, 2904

[4] Jouvel, S., Kneib, J. -P., Ilbert, O., et al. 2009, A&A, 504, 359