UAB
Universitat Autònoma de Barcelona

A study of photometric redshifts and intrinsic alignments with PAUS and Euclid

David Navarro Gironés

A study of photometric redshifts and intrinsic alignments with PAUS and Euclid

David Navarro Gironés

Supervisor: Dr. Héctor Martin Crocce
Dr. Enrique Gaztañaga Balbas

Departament de Física
Universitat Autònoma de Barcelona

This dissertation is submitted for the degree of
Doctor of Philosophy in Physics

September 2024

Al meu avi Joan

Acknowledgements

Este largo viaje que ha sido el doctorado parece ser que termina con este manuscrito. Durante estos 4 años, mucha gente me ha acompañado y he tenido la oportunidad de conocer y trabajar con personas maravillosas, a las cuales me gustaría agradecer.

En primer lugar, quiero dar las gracias a mis supervisores, Martin y Enrique, por vuestra inestimable ayuda durante estos años. He aprendido mucho de vosotros, no solo sobre cosmología, que también, sino sobre cómo ser un buen investigador. De cada una de las reuniones que hemos tenido, he salido con excelentes sugerencias y buenas ideas, que seguro que se reflejan en los resultados de esta tesis.

También me gustaría agradecer el excelente ambiente del grupo de cosmología del ICE. Tanto a investigadores senior, postdocs y estudiantes de doctorado, siempre he podido aprender algo de vosotros. Me gustaría agradecer especialmente a Benjamin, con quien he compartido despacho y supervisores durante estos cuatro años. He aprendido mucho contigo en este tiempo y me gusta pensar que tengo un nuevo amigo en ti. Gracias especiales también a Sylvain, Michel-Andrès y Gabriele, quienes también me han dado excelentes consejos y con los que he pasado muy buenos momentos. Por supuesto, gracias también a todos los compañeros con los que he compartido despacho estos años. A todos, os echaré en falta.

Gracias inmensas también a la gente del PIC, especialmente a Jorge y Pau, que proporcionan una infraestructura estupenda, con la que he realizado la mayoría de los trabajos en esta tesis. También quiero agradecer a Eli su amistad y ayuda en este tiempo, tus consejos para sacar adelante esta tesis han sido muy útiles.

A las personas que forman el experimento de PAUS. Este ha sido el primer experimento en el que he trabajado como investigador y le tengo un especial cariño. Todos los miembros de PAUS hacen posible que haya un ambiente de trabajo genial y espero poder seguir colaborando con vosotros mucho tiempo.

Me gustaría agradecer al grupo de IST-NL de Euclid. Con ellos me introduje en este experimento tan emocionante y siempre tuvieron un momento para ayudarme, sobre todo en los primeros momentos. Quiero dar las gracias especialmente a Guada, que me dio una clase magistral de qué es Euclid y de cómo funciona. També vull donar les gràcies a l'Isaac, amb

qui vaig tenir l'oportunitat de fer una estada de recerca a Toulouse, sempre aprenc molt amb tu.

I would also like to thank the group from Bochum, where I had the opportunity to stay for a couple of months. Special thanks to Anna, with whom I had the pleasure to work with during my stay, and to Hendrik, who invited me to his group.

I would like to thank the panel of my thesis, Elisa Chisari, Alex Alarcon and Benjamin Joachimi, for accepting the role of being the reviewers and examiners of this PhD thesis. I also want to thank Henk Hoekstra, for hiring me as a postdoc in the Observatory of Leiden and believing I can be a good asset for his research group.

Me gustaría también hacer un agradecimiento un tanto inusual. En este caso, a la educación pública de este país, que me ha permitido, desde la educación infantil a los estudios de doctorado, formarme en una educación de calidad. Evidentemente, siempre hay ámbitos que mejorar, pero creo que es de recibo poner en valor la educación pública que tenemos.

Por supuesto, me gustaría agradecer a mis amigos, en especial a Laura y Anna Laura, por siempre apoyarme y darme ánimos con el doctorado y, simplemente, por siempre estar cuando se les necesita. També m'agradaria donar les gràcies als meus amics de la universitat, molts dels quals també estan fent el doctorat, per compartir el viatge amb mi.

A mi familia, especialmente a mis padres y mi hermana. Gracias infinitas por todo.

Por último, gracias a mi pareja, Javier, por siempre apoyarme en este viaje. Nos conocimos al principio de este doctorado, por lo que hemos compartido todo este viaje juntos, y siempre has estado ahí para apoyarme. Mil gracias.

Abstract

The field of cosmology is undergoing an era of unprecedented new observational data, with galaxy surveys collecting information from billions of galaxies. This will enable the scientific community to perform precise measurements of the large-scale structure of the universe, enhancing our understanding about the cosmological model that best describes it. This thesis focuses on photometric redshift (photo-z) estimation and the study of intrinsic alignments (IA) with the Physics of the Accelerating Universe Survey (PAUS) and Euclid.

First, we employ PAUS to estimate the distances to galaxies, via photo-z, for $\sim$1.8 million objects in an area of $\sim$50 deg^2 and down to $i_{\rm AB} < 23$. We employ a template-fitting code to estimate photo-z and implement a novel zero-point calibration, which does not require spectroscopic redshifts. We define an innovative weighting scheme, where the photo-z estimates from PAUS and broad-band (BB) data are combined. This allows us to achieve photo-z accuracies one order of magnitude better than BB estimates for bright objects, while reducing the number of outliers and the bias. We analyse the photo-z accuracy as a function of the galaxy colour, finding similar performances for red and blue objects. Additionally, we provide calibrated probability density distributions of the photo-z for each object.

Next, using the previously computed photo-z, we measure the galaxy clustering (GC) and the IA of galaxies in the PAUS fields, for 400000 objects in the redshift range $0.1 < z < 1.0$. We analyse the dependence on colour, redshift and luminosity and constrain the GC and IA parameters for each scenario. The results show significant detection of IA for red galaxies, with an increase in the signal at high redshifts and luminosities. In contrast, blue galaxies show null or weak alignments. In general, our results are consistent with previous literature, extending the IA amplitude-luminosity relation towards fainter objects.

Finally, we implement the Tidal Alignment and Tidal Torquing (TATT) IA model in Euclid's likelihood pipeline. Then, we perform forecasts on how IA affects cosmological parameter inference, demonstrating that the mismodelling of the IA effect might lead to biases in the cosmological parameters. Therefore, we highlight the need of adding flexible IA models for accurate cosmological inference in upcoming galaxy surveys.

Resum

El camp de la cosmologia està experimentant una era sense precedents pel que fa a noves dades observacionals, amb els anomenats *galaxy surveys*. Això permetrà a la comunitat científica portar a terme mesures precises de l'estructura a gran escala de l'univers, millorant el nostre coneixement sobre el model cosmològic que millor el descriu. Aquesta tesi se centra en el càlcul de *redshifts* fotomètrics (photo-z) i en l'estudi dels *intrinsic alignments* (IA) de les galàxies amb el Physics of the Accelerating Universe Survey (PAUS) i Euclid.

En primer lloc, utilitzem PAUS per estimar les distàncies a les galàxies, a partir dels photo-z, per $\sim$1.8 milions d'objectes en una àrea de $\sim$50 graus quadrats i una magnitud de $i_{\rm AB} < 23$. Amb un codi de *template-fitting*, calculem els photo-z i implementem una nova calibració dels punts-zero, que no requereix *redshifts* espectroscòpics. Definim una nova estimació dels photo-z, combinant dades de PAUS i de banda ampla, aconseguint qualitats un ordre de magnitud millor que les estimacions amb banda ampla, pel cas d'objectes brillants, i reduint també el nombre de valors atípics i el biaix. Analitzem la qualitat dels photo-z en funció del color de les galàxies, trobant valors similars per objectes vermells i blaus. A més a més, proporcionem distribucions de probabilitat calibrades dels photo-z de cada objecte.

Seguidament, utilitzant els photo-z calculats prèviament, mesurem el galaxy clustering (GC) i els IA de les galàxies als camps de PAUS, per 400000 objectes en el rang de *redshift* de $0.1 < z < 1.0$. Analitzem la dependència en color, *redshift* i lluminositat i prediem els paràmetres de GC i IA per cada cas. Els resultats mostren una detecció significativa per galàxies vermelles, amb un increment del senyal per *redshifts* i lluminositats elevades. D'altra banda, les galàxies blaves mostren alineaments nuls o molt dèbils. En general, els nostres resultats són consistents amb la literatura anterior, estenent la relació d'amplitud dels IA i la lluminositat envers objectes més dèbils.

Finalment, implementem el model de IA Tidal Alignment and Tidal Torquing (TATT) a la *pipeline* de probabilitat de Euclid. Realitzem previsions de com els IA afecten la inferència de paràmetres cosmològics, demostrant que un modelatge erroni dels IA pot incórrer en biaixos en els paràmetres cosmològics. Per tant, remarquem la necessitat d'afegir models de IA flexibles per tal d'obtenir paràmetres cosmològics precisos en els futurs *galaxy surveys*.

Resumen

El campo de la cosmología está experimentando una era sin precedentes en cuanto a nuevos datos observacionales, con los llamados *galaxy surveys*. Esto permitirá a la comunidad científica llevar a cabo medidas precisas de la estructura a gran escala del universo, mejorando nuestro conocimiento respecto al modelo cosmológico que mejor lo describe. Esta tesis se centra en la estimación de *redshifts* fotométricos (photo-z) y el estudio de los *intrinsic alignments* (IA) con el Physics of the Accelerating Universe Survey (PAUS) y Euclid.

Primero, estimamos las distancias a las galaxias con PAUS, mediante photo-z, para $\sim$1.8 millones de objetos en una área de 50 grados cuadrados hasta una magnitud de $i_{\rm AB} < 23$. Mediante un código de *template-fitting*, estimamos los photo-z e implementamos una nueva calibración de los puntos-cero, que no requiere de *redshifts* espectroscópicos. Combinamos datos de PAUS y banda ancha, logrando photo-z con una precisión un orden de magnitud mejor que las estimaciones de banda ancha, para objetos brillantes, y reduciendo el número de valores atípicos y el sesgo. Analizamos la calidad de los photo-z en función del color de las galaxias, encontrando valores similares para objetos rojos y azules. Además, proporcionamos distribuciones de probabilidad calibradas de los photo-z para cada objeto.

Seguidamente, utilizando los photo-z previamente calculados, medimos el galaxy clustering (GC) y el IA de las galaxias de los campos de PAUS, para 400000 objetos en el rango de *redshift* de $0.1 < z < 1.0$. Analizamos la dependencia en color, *redshift* y luminosidad y predecimos los parámetros de GC e IA para cada caso. Los resultados muestran una detección significativa de IA para galaxias rojas, con un incremento de la señal para *redshifts* y luminosidades elevadas. Por otro lado, las galaxias azules muestran alineamientos nulos o débiles. En general, nuestros resultados son consistentes con literatura previa, extendiendo la relación de la amplitud de los IA con la luminosidad hacia objetos más débiles.

Finalmente, implementamos el modelo de IA de Tidal Alignment and Tidal Torquing (TATT) en el *pipeline* de probabilidad de Euclid. Demostramos que el modelado incorrecto de este efecto puede dar lugar a sesgos en la inferencia de parámetros cosmológicos. Por lo tanto, remarcamos la necesidad de añadir modelos de IA flexibles para una inferencia de parámetros precisa en los futuros *galaxy surveys*.

Table of contents

List of figures

List of tables

Acronyms

2PCF two-point correlation function.

2dFGRS Two-degree Field Galaxy Redshift Survey.

AEGIS All-Wavelength Extended Groth Strip International Survey.

AP Alcock-Paczyński.

APO Apache Point Observatory.

BB broad-band.

CCD Charge-Coupled Device.

CDFS Chandra Deep Field South.

CDPU Control and Data Processing Unit.

CFHTLenS Canada-France-Hawaii Telescope Lensing Survey.

CFHTLS Canada-France-Hawaii Telescope Legacy Survey.

CIGALE Code Investigating GALaxy Emission.

CLOE Cosmology Likelihood for Observables in Euclid.

CMB Cosmic Microwave Background.

COBE Cosmic Background Explorer.

CosmoSIS COSMOlogical Survey Inference System.

CP Cosmological Principle.

DE Dark Energy.

DES Dark Energy Survey.

DESI Dark Energy Spectroscopic Instrument.

DM Dark Matter.

DUNE Dark Universe Explorer.

DV data vector.

EAF Euclid Auxiliary Fields.

EDF Euclid Deep Fields.

EDS Euclid Deep Survey.

EFT effective field theory.

ESA European Space Agency.

ESO European Southern Observatory.

EWS Euclid Wide Survey.

FLRW Friedmann-Lemaître-Robertson-Walker.

FoM Figure of Merit.

FoV Field of View.

FWHM Full Width at Half Maximum.

GAMA Galaxy And Mass Assembly.

GC Galaxy Clustering.

GOODS Great Observatories Origins Deep Survey.

HOD Halo Occupation Distribution.

HSC SSP Hyper Suprime-Cam Subaru Strategic Program.

HST Hubble Space Telescope.

IA Intrinsic Alignments.

IMF initial mass function.

IR infrared.

JK jackknife.

JWST James Webb Space Telescope.

KiDS Kilo-Degree Survey.

LA Linear Alignment.

LCO Las Campanas Observatory.

LSS large-scale structure.

LSST Legacy Survey of Space and Time.

MCMC Monte-Carlo Markov Chain.

NB narrow-band.

NIRCam Near Infrared Camera.

NISP Near Infrared Spectrometer and Photometer.

NLA Non-Linear Alignment.

PAUdm PAU data management.

PAUS Physics of the Accelerating Universe Survey.

PIC Port d'Informació Científica.

PM particle-mesh.

PSF Point Spread Function.

R.A Right ascension.

RSD redshift-space distortions.

SDSS Sloan Digital Sky Survey.

SED spectral energy distribution.

SHAM subhalo abundance matching.

SNR signal-to-noise ratio.

SOM Self Organising Map.

SPACE Spectroscopic All-Sky Cosmic Explorer.

sSFR specific star formation rate.

SVD singular value decomposition.

SXDS Subaru/XMM-Newton Deep Survey.

TATT Tidal Alignment Tidal Torquing.

UV ultraviolet.

VIKING VISTA Kilo-degree INfrared Galaxy.

VIPERS VIMOS Public Extragalactic Redshift Survey.

VIS Visible Imaging Instrument.

VLT Very Large Telescope.

VVDS VIMOS VLT Deep Survey.

WHT William Herschel Telescope.

WMAP Wilkinson Microwave Anisotropy Probe.

Introduction

Why are we here? Where do we come from? What is the Universe history and fate? All these are questions that humans have been asking themselves since the dawn of time. All these are questions that have allowed the progress of Cosmology. But, what is Cosmology? According to the Oxford English Dictionary, Cosmology is "The science or theory of the universe as an ordered whole, and of the general laws which govern it". From this definition, it is not surprising that Cosmology has been studied for many centuries, leading to significant progress over time.

Even though there are records from ancient civilizations, such as the Mesopotamian, the Babylonian or the Chinese, amongst others, that contributed to the development of mathematics and astronomy, and also provided theories on the origin of the cosmos (although these were mostly of mythological nature), it was the Greek civilization that systematically studied the cosmos and started to propose cosmological theories away from mythology.

The most enduring theory from the ancient Greeks is that of Aristotle, presented in his treatise *De Caelo et Mundo*. Building on previous works from Plato and Eudoxus, Aristotle proposed a cosmological model in which celestial objects moved in concentric spheres, with Earth at the centre of the universe and the outermost sphere containing the *fixed stars*. More than 400 years after, it was Claudius Ptolemy who modified Aristotle's theory by introducing the epicycles to account for the retrograde motion of the planets. This model would prevail in the Western world until the 16th century.

With the Renaissance era, new cosmological theories emerged that advocated to locate the Sun in the middle of the Universe, going from a geocentric to a heliocentric model. This was the case of Nicolaus Copernicus' model, in which all planets moved around the Sun in circular orbits. This model could effectively explain the retrograde movements of the planets with orbits closer to the Sun than Earth's. A century later, Johannes Kepler, embracing the Copernican model, introduced the idea of elliptical orbits, instead of circular ones, which could explain better the observations made by his colleague Tycho Brahe. Galileo Galilei's observations of the night sky with a telescope also supported the Copernican model, specially

his observations of moons orbiting around Jupiter and the phases of Venus. Finally, under his theory of universal gravitation, Isaac Newton synthesized the previous works from Kepler and Galileo and proposed that gravity governed the motion of celestial objects.

Modern Cosmology was born almost 250 years later, with the theory of General Relativity introduced in 1915 by Albert Einstein, which described gravity in terms of a relation of spacetime with mass and energy. This theory allowed the development of modern Cosmology in the 20th century and culminated, together with observational evidences, with the ΛCDM model.

The ΛCDM model is currently the most accepted cosmological model. It is described by a universe dominated by cold Dark Matter (DM), which makes up for a $\sim 27\%$ of the matter-energy budget of the universe, and Dark Energy (DE), which makes up for another $\sim 68\%$. In contrast, baryonic matter, such as stars and galaxies, only accounts for a $\sim 5\%$ of this content. On the one hand, cold DM is composed by gravitationally interacting matter, which does not interact (or does it very weakly) through electromagnetic or nuclear forces. It is considered to be cold, in the sense that its velocity is much smaller than the velocity of light, allowing it to cluster effectively in the early universe. On the other hand, DE is a form of energy that causes the universe to accelerate, counteracting the gravitational force from matter. DE is explained with a cosmological constant in the ΛCDM model. Both the nature of DM and DE are unknown and constitute major key challenges in cosmology.

In order to disentangle the origin of DM and DE, cosmologists have designed and built the so-called *galaxy surveys*. These surveys map the distribution of galaxies in the sky, both in the angular and radial directions, with the objective of analysing the large-scale structure (LSS) of galaxies, from which one can study the geometry, the content and the evolution of the universe. From the first generation of galaxy surveys, which observed thousands of galaxies, to the most recent ones, capable of observing hundreds of millions of objects, the information has not ceased to increase, allowing us to observe deeper and fainter objects in the sky.

The first goal of this thesis consists on the measurement of galaxy distances, through photometric redshifts, using the PAUS galaxy survey. PAUS has been specifically designed to allow for the accurate measurement of redshifts, using information from 40 narrow-band filters. In this thesis, we estimate the redshift for the PAUS footprint, which amounts to a total of $\sim$50 deg^2 and 1.8 million objects. These photometric redshift estimates are essential for many studies requiring galaxy distances estimates, not only in PAUS but also in other surveys.

One of the studies requiring galaxy distances consists on the measurement of the so-called *Intrinsic Alignments (IA)*. This effect accounts for the correlation of intrinsic galaxy shapes and constitutes one of the main systematic in weak lensing studies, since it can mimic its signal, thus interfering in the cosmological parameter inference. The second goal of this thesis consists on the measurement of IA in PAUS, studying this effect by splitting the galaxy samples in colour, redshift and luminosity.

The last goal of this thesis is the implementation of an IA model into the theory and likelihood pipeline of Euclid, which is a last generation galaxy survey that will cover 14000 $\deg^2$ and observe 1.5 billion galaxies. With this implementation, we add flexibility in the likelihood pipeline by introducing an IA model that is expected to describe this effect at non-linear scales. Together with the implementation, we study the differences between the two most popular IA models and we analyse the contribution of IA to the total signal coming from the weak lensing and the galaxy-galaxy lensing observables. Finally, we focus on the impact that mismodelling the IA effect may have on cosmological parameter inference.

The structure of this thesis is as follows: In Chapter 1, we review the cosmological framework on which the content of this thesis is based. In Chapter 2, we describe the imaging surveys from which the data for this work were obtained. Next, in Chapter 3, we present the photometric redshift study performed within PAUS. Then, in Chapter 4, we use the photometric redshifts from Chapter 3 to measure the IA of galaxies in PAUS. In Chapter 5, we detail the implementation of an IA model in the Euclid likelihood pipeline and analyse the effect of mismodelling IA. Finally, in Chapter 6, we present the conclusions of this thesis.

Chapter 1

Cosmological framework

1.1 The Cosmological Principle

The Cosmological Principle (CP) is a fundamental assumption that asserts that the universe is homogeneous and isotropic on sufficiently large scales. Here, homogeneity describes the fact of being identical everywhere in space, while isotropy describes the fact of being identical in all directions. Certainly, this is only true if considering the average observations on scales larger than the typical ones traced by the LSS of galaxies. Isotropy does not necessarily imply homogeneity, unless the Copernican Principle is assumed. The Copernican Principle states that the observer is not in a privileged place of the universe. For example, if the Copernican Principle was not satisfied and the observer was located in a privileged position from where the universe is isotropic, homogeneity could be wrongly assumed. Thus, the combination of the observed isotropy and the Copernican Principle lead us to the CP.

The isotropy and homogeneity of our universe allows simplifying its global structure, leading to relatively simple cosmological models. At the time the CP was proposed, little was known about the true distribution of the structure of the universe, so simplified principles were required to allow for the development of theoretical Cosmology. That is the reason why scientists like Einstein, at first, and others, later, worked within the assumption of the CP to develop their theories without any clear observational evidence that justified it.

1.2 An expanding universe

In the decade of the 1920s, some physicists such as Alexander Friedmann and Georges Lemaître theorized, following Einstein's equations of General Relativity, that the universe

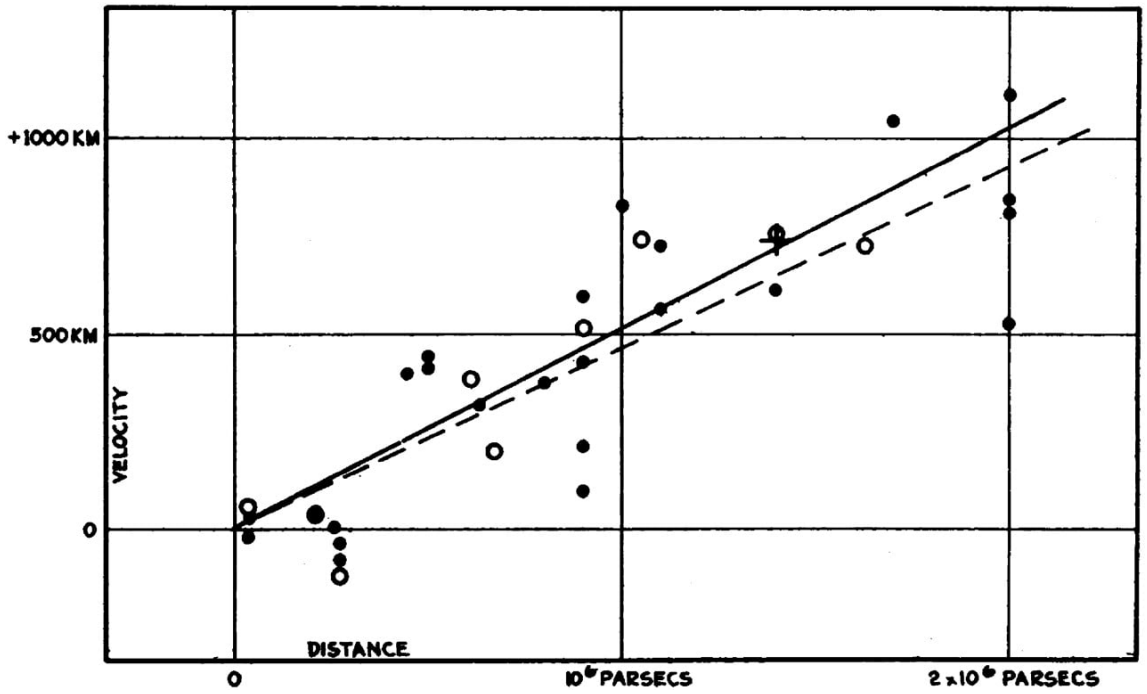

Fig. 1.1 Radial velocity versus distance for extra-galactic nebulae studied by Hubble. Figure extracted from [108]. Note the error in the velocity units, which should be km/s.

could be expanding. This was later confirmed in 1929 by Edwin Hubble [108], who measured the distances of extra-galactic nebulae from Cepheid variables, to then combine these measurements with the redshifts measured by Vesto Slipher [199] on the same objects. Based on these studies, he proved that those objects had a recession velocity with respect to us. Additionally, Hubble could establish a relation between distance (d) and redshift (or velocity, v), which indicated a linear trend, with the slope being the Hubble constant (H_0), such that:

$$v = H_0 d \tag{1.1}$$

Fig. 1.1 shows the original plot by Hubble of the velocities of extra-galactic nebulae, measured through the redshift, versus the distance, measured through Cepheid variables. The fitted lines indicate a linear relation between both magnitudes, which can be described following eq. 1.1.

The Hubble law is an indication that we live in an expanding universe (for a review on other probes proving the expansion of the universe, see [111, 9, 79]). This implies that the distance between us and distant objects was shorter in the past than what it is now. The rate at which the universe is expanding can be described with the scale factor a, which increases with time. The scale factor is related to the Hubble parameter by:

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)}, \tag{1.2}$$

where $\dot{a}(t)$ is the time derivative of a. The scale factor today (t_0) is set to $a(t_0) = 1$ and the Hubble parameter today is the Hubble constant H_0 ($H(t_0)$).

If not specified otherwise, throughout the rest of this Chapter 1, the units for the speed of light are set to $c = 1$.

1.2.1 FLRW metric

Regarding the geometry of our universe, it can be defined as open, closed or Euclidean. To understand the difference between each case, imagine two particles travelling freely in parallel. These particles will diverge, converge or stay parallel for an open, closed or Euclidean universe, respectively. Current observations indicate that our universe has a Euclidean geometry, also referred to as *flat universe*.

The Friedmann-Lemaître-Robertson-Walker (FLRW) metric defines a expanding universe that satisfies the CP (Section 1.1). From the metric, we can convert observer-dependent coordinates into invariants. In particular, the distance squared is an invariant quantity.

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu, \tag{1.3}$$

where $g_{\mu\nu}$ is the metric and μ and ν range from 0 to 3, with the first coordinate describing the time and the other three describing the space. In the case of a FLRW metric, eq. 1.3 becomes:

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2 \right), \tag{1.4}$$

where k describes the curvature and has values of $k = 0, -1, 1$ for a flat, open and closed universe, respectively. The term $dr^2/1-kr^2$ represents the radial part of the spatial coordinate and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ is the solid angle, representing the angular part in spherical coordinates.

It is also useful to define the metric expressed in eq. 1.4 in comoving distance (χ, see Section 1.2.3), such that:

$$ds^2 = -dt^2 + a(t)^2 \left(d\chi^2 + S_k^2(\chi) d\Omega^2 \right), \tag{1.5}$$

where

$$S_k(\chi) \equiv \begin{cases} \sin(\chi) & k > 0 \\ \chi & k = 0 \\ \sinh(\chi) & k < 0 \end{cases} \tag{1.6}$$

The evolution of a with time depends on the constituents of the universe and can be described through the Friedmann equations.

1.2.2 Einstein's and Friedmann's equations

The Einstein's equations in General Relativity relates the geometry of the universe with its energy, such that:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.7}$$

where $G_{\mu\nu}$ is the Einstein tensor and is defined as:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \tag{1.8}$$

with $R_{\mu\nu}$ being the Ricci tensor, a quantity that depends on the metric and tells us how its geometry changes locally with respect to that of a Euclidean space. R is the Ricci scalar and is the contraction of the Ricci tensor, $R \equiv g^{\mu\nu} R_{\mu\nu}$. Λ and G are the cosmological and Newton's constants, respectively, and $T_{\mu\nu}$ is the energy-momentum tensor, which in the case of a perfect fluid can be expressed as:

$$T_{\mu\nu} = -P g_{\mu\nu} + (P + \rho) u_\mu u_\nu, \tag{1.9}$$

where $u_\mu = (1,0,0,0)$, P is the pressure and ρ is the energy density of all the constituents of the universe: baryonic matter, DM, radiation and DE.

The beauty of eq. 1.7 relies on the fact that it relates the geometry of the universe (left-hand side) with its energy content (right-hand side) in a very simple equation.

Making use of Einstein's equation and the FLRW metric, we can derive the Friedmann equations, which tell us how the scale factor evolves with time. The first Friedmann equation is obtained by solving the time-time component of the Einstein equation, such that:

$$H^2(a) = \frac{8\pi G}{3} \rho - \frac{k}{a^2}, \tag{1.10}$$

where H is the Hubble parameter defined in eq. 1.2. This equation relates the rate of expansion of the universe (left-hand side) with its energy content and curvature (right-hand side).

The second Friedmann equation is obtained from the space-space components of eq. 1.7 and is:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{1.11}$$

This equation describes the acceleration of the expansion of the universe (left-hand side) with its energy content (right-hand side).

A useful expression can be derived by combining both Friedmann equations, allowing to study how the energy density changes for each constituent of the universe:

$$\dot{\rho} + 3H(\rho + P) = 0 \tag{1.12}$$

In the case of non-relativistic matter, $P_m = 0$, so that $\rho_m \propto a^{-3}$. For radiation, $P_r = \rho_r/3$, so $\rho_r \propto a^{-4}$. Finally, in the case of a cosmological constant describing the DE, we expect $\dot{P}_\Lambda = 0$, so that $\rho_\Lambda = -P_\Lambda$.

By defining the equation of state parameter w_s, we obtain the following values:

$$w_s \equiv \frac{P_s}{\rho_s} = \begin{cases} 0 & \text{for matter} \\ 1/3 & \text{for radiation} \\ -1 & \text{for } \Lambda \end{cases} \tag{1.13}$$

where s corresponds to any constituent of the universe. By integrating eq. 1.12, we can express the evolution of the energy density per constituent as a function of the scale factor and the w_s parameter:

$$\rho_s(a) = \rho_s(t_0) a^{-3\cdot(1+w_s)}, \tag{1.14}$$

where w_s is assumed to be time-independent and $\rho_s(t_0)$ is the current energy density of the constituent s.

It is also conventional to define the density parameters today:

$$\Omega_{s,0} \equiv \frac{\rho_s(t_0)}{\rho_{\text{cr}}}, \tag{1.15}$$

where ρ_{cr} is the critical density, defined as:

$$\rho_{\text{cr}} \equiv \frac{3H_0^2}{8\pi G} \tag{1.16}$$

With these expressions, the first Friedmann equation, eq. 1.10, can be redefined by dividing it by the Hubble constant:

$$\frac{H^2(t)}{H^2(t_0)} = \sum_{s=r,m,\Lambda} \Omega_{s,0} \left[a(t)\right]^{-3(1+w_s)} + \Omega_k [a(t)]^{-2}, \tag{1.17}$$

where the sum is done over radiation, matter (baryonic and dark) and the cosmological constant. $\Omega_k \equiv 1 - \Omega_0 \equiv 1 - \sum_s \Omega_{s,0}$ is the term that accounts for the curvature and is $\Omega_k = 0$ for a flat universe. This equation allows understanding how the different constituents of the universe contribute to the evolution of the scale factor $a(t)$.

From now on, if not indicated otherwise, we will set $k = 0$ to express the upcoming equations for the case of a flat Universe, since this is the most supported hypothesis based on observational evidences.

1.2.3 Distances

There are different ways of defining distances in an expanding universe. We can define the *physical distance*, which corresponds to the distance we would measure between two objects at a given time. It can also be thought as the actual distance one would have to travel between those two objects. This distance grows with time due to the expansion of the universe.

We can also define the *comoving distance*, which is not affected by the expansion of the universe. The differential comoving distance travelled by light in a time interval dt is $dx = dt/a$. By integrating this quantity from time t to the time today t_0, we obtain the comoving distance:

$$\chi(t) = \int_t^{t_0} \frac{dt'}{a(t')} = \int_{a(t)}^{1} \frac{da'}{a'^2 H(a')} = \int_0^z \frac{dz'}{H(z')}, \tag{1.18}$$

where we can see that it can be expressed either as a function of time, scale factor or redshift.

An interesting case of eq. 1.18 occurs when $t = 0$, which represents the comoving distance light has travelled since the beginning of time:

$$\eta(t) \equiv \int_0^t \frac{dt'}{a(t')} \tag{1.19}$$

This distance is known as the comoving horizon, and regions of the universe separated by comoving distances larger than η are not causally connected.

The *angular diameter distance* (d_A) is another way of defining distances in the universe. It is defined as the ratio of the physical size of an object (l) and the angle (θ) with respect to the Earth. If θ is small, the angular diameter distance is:

$$d_A = \frac{l}{\theta} \tag{1.20}$$

In an Euclidean universe, eq. 1.20 can be related to the comoving distance following eq. 1.21:

$$d_A^{\text{Euc}} = a\chi \tag{1.21}$$

Finally, we can define the *luminosity distance* (d_L) by studying the relation between the flux (F) of an object of known luminosity (L), such that:

$$F = \frac{L}{4\pi d_L^2}. \tag{1.22}$$

Due to the expansion of the universe, one can rewrite eq. 1.22 in terms of the scale factor and the comoving distance as:

$$F = \frac{La^2}{4\pi\chi^2} \tag{1.23}$$

From eq. 1.21 and 1.23, the luminosity distance in a Euclidean universe is:

$$d_L^{\text{Euc}} = \frac{\chi}{a} \tag{1.24}$$

Thus, the angular diameter and the luminosity distances can be related by $d_L = d_A/a^2$, a relation that is also satisfied for the case of a curved universe.

1.3 CMB

In 1964, while working on the Holmdel Horn Antenna, a radio antenna used for satellite communications and radio astronomy, Arno Penzias and Robert Wilson unexpectedly discovered a signal coming from the Cosmic Microwave Background (CMB). Prior to the discovery, the radio antenna had been used for satellite communications at a wavelength of $\lambda = 7.35$cm. When they pointed it towards the sky, they measured an isotropic signal which corresponded to a background noise of about 3.5K, for which they did not have an explanation. Penzias and Wilson contacted the group led by Robert Dicke in Princeton, which had deduced that the universe should be filled with microwave radiation if it had started from a hot dense state,

backing the theory of a Big Bang event. After this, two publications followed, describing the measurement of the signal by Penzias and Wilson and its implication of a relic of a hot, dense and opaque early state of the universe by the group in Princeton.

Years after the discovery of Arno Penzias and Robert Wilson, the CMB was accurately measured for the first time by the Cosmic Background Explorer (COBE) [19], a satellite launched in 1989 into a 900km orbit above Earth's surface. Even though COBE could detect fluctuations in the temperature in 1 in 10^4 parts, it found that the spectrum of the CMB was really close to an ideal blackbody with a temperature of T=2.725K. The angular resolution to study the CMB was further increased with more advanced instruments, in particular, the Wilkinson Microwave Anisotropy Probe (WMAP) [14] and the Planck ([172]) satellites, both located at the L2 point of the Sun-Earth system. These instruments were able to map precisely the temperature fluctuations of the CMB, as seen in Fig. 1.2, which shows the Planck measurements.

The CMB has its origin in the early universe, when photons decoupled from electrons. Prior to that, baryonic matter formed an ionized plasma, from which photons scattered constantly through a process named Thomson scattering. As the universe expanded, the mean energy per photon decreased and electrons and protons were able to form atoms, in a process called recombination. Just after recombination, photons decoupled from electrons and streamed freely through space. Related to this moment, one can define the last scattering time, which is the time when a CMB photon had its last scattering from an electron. This corresponds to a redshift of $z \sim 1090$. One can also define the last scattering surface as the surface associated to the last scattering time.

From the angular size of CMB temperature fluctuations, we can learn about the inhomogeneities of the universe and their physical sizes. Using the definition of the angular diameter distance (eq. 1.20), we can relate the angular ($\delta\theta$) and physical (l) sizes of the temperature fluctuations:

$$d_A = \frac{l}{\delta\theta} \simeq \frac{d_{\rm hor}(t_0)}{z_{\rm ls}}, \tag{1.25}$$

where the second expression is an approximation for the angular diameter distance at $z \to \infty$, which is the case for the redshift at the last scattering surface ($z_{\rm ls} \sim 1190 \gg 1$). Here, $d_{\rm hor}(t_0)$ is the current horizon distance, which is $d_{\rm hor}(t_0) \simeq 14000$Mpc. Thus, the physical and the angular size of fluctuations can be related as:

$$l = 12.8{\rm Mpc}\left(\frac{\delta\theta}{1{\rm rad}}\right), \tag{1.26}$$

where the value 12.8Mpc corresponds to the ratio 14000Mpc/1190.

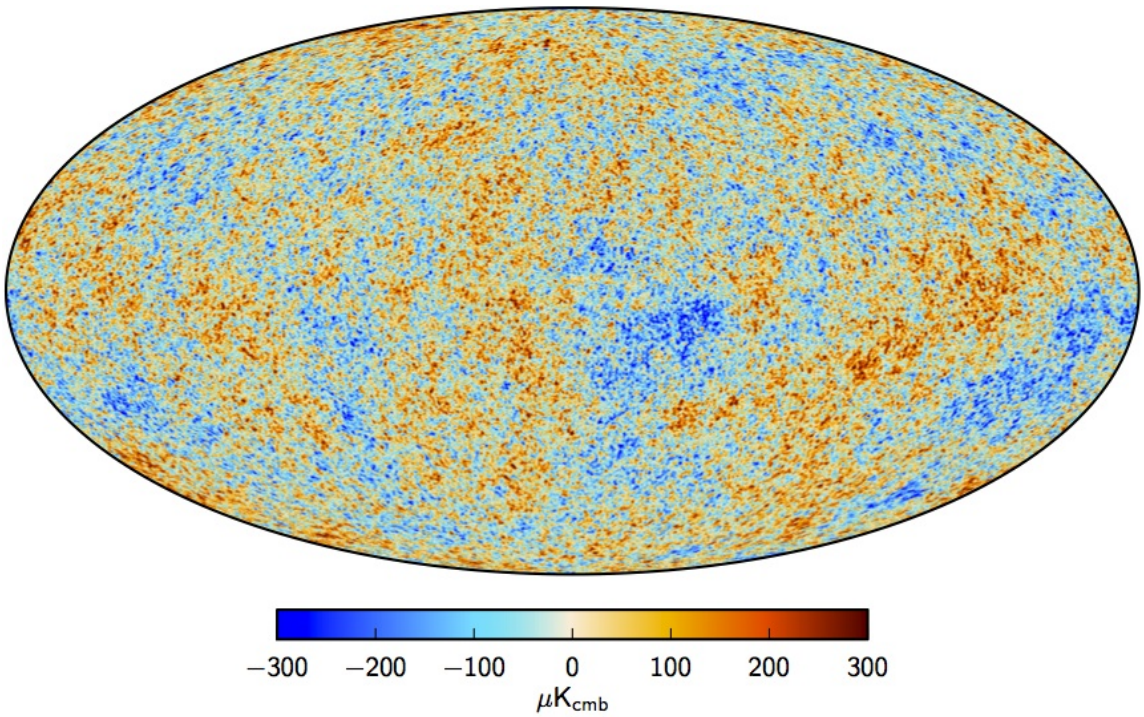

Fig. 1.2 Temperature variations in the CMB signal measured by Planck. Figure extracted from [171]

The CMB temperature fluctuations are usually studied by expanding them in spherical harmonics, such that:

$$\frac{\delta T}{T}(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}Y_{lm}(\theta,\phi), \tag{1.27}$$

where θ and ϕ define the angular positions in the sky, Y_{lm} are the spherical harmonic functions and a_{lm} are their amplitudes. Note that the l employed in this last equation differs from the one defined in eq. 1.25.

From this, the correlation function of the CMB temperature fluctuations can be written as:

$$C(\theta) = \left\langle \frac{\delta T}{T}(\hat{n})\frac{\delta T}{T}(\hat{n}') \right\rangle_{\hat{n}\cdot\hat{n}'=\cos\theta} = \frac{1}{4\pi}\sum_{l=0}^{\infty}(2l+1)C_l P_l(\cos\theta), \tag{1.28}$$

where the first equality is the average product of eq. 1.27 at directions $\hat{n}$ and $\hat{n}'$, separated by an angle θ. In the second equality, P_l are Legendre polynomials and C_l are the multipole moments.

The temperature fluctuations are usually decomposed in terms of l as:

$$\Delta_T = \left(\frac{l(l+1)}{2\pi}C_l\right)^{1/2}\langle T\rangle, \tag{1.29}$$

which describes the contribution per logarithmic bin l to the CMB temperature fluctuations.

Fig. 1.3 shows the temperature fluctuations following eq. 1.29. Considering that the angular scale and the multipole l are related by $\theta \sim 180^\circ/l$, it is noticeable that we can divide the behaviour of the fluctuations in small (large l) and large (small l) angular scales. This division is roughly delimited by the angular size of the horizon distance, $\theta_{\rm hor} \sim 1.1^\circ$ ($l_{\rm hor} \sim 160$).

On the one hand, the fluctuations seen at large scales are dominated by the DM potential wells at the time of the last scattering. Photons located in local minimum (maximum) would lose (gain) energy, which is proportional to a decrease (increase) in temperature. This process is known as the Sachs-Wolfe effect. As a consequence, the temperature fluctuations at large angular scales map the fluctuations of the DM potential wells. The fact that ΔT is constant over a broad range of large scales indicate that the fluctuations in the potential wells were also similar.

On the other hand, fluctuations in small angular scales result from the interaction of photons and baryons with the gravitational force of the DM. If the photon-baryon fluid is in a potential well, it starts compressing towards the centre of the well, gaining energy and temperature. However, due to the increase in density of the fluid, the pressure will eventually cause its expansion, thus losing energy and temperature. This process continues, and the oscillations experimented by the fluid are known as acoustic oscillations. As a conclusion, the first peak seen in the CMB temperature fluctuations in Fig. 1.3 represents the DM potential wells, in which the photon-baryon plasma experimented maximum compression at the last scattering surface. The position and amplitude of the first peak is a cosmological tool that allows to constrain the curvature and the baryon density of the universe, respectively.

1.4 Structure formation

While the CMB is a powerful cosmological probe that allows us to test the early-universe, the LSS traced by galaxies enables the study of the late-universe, which is the focus of this thesis. In particular, the LSS of the universe is one key cosmological probe that allows to study the acceleration of the universe, the growth of structure, the properties of initial conditions of inflation, DM and neutrinos, amongst others. Therefore, there are currently a great deal of surveys studying the LSS, such as Euclid, the Legacy Survey of Space and Time (LSST) or the Dark Energy Spectroscopic Instrument (DESI).

As seen in Section 1.1, the universe is isotropic and homogeneous on large scales. However, in order to explain the structure we see in the universe today, some small inho-

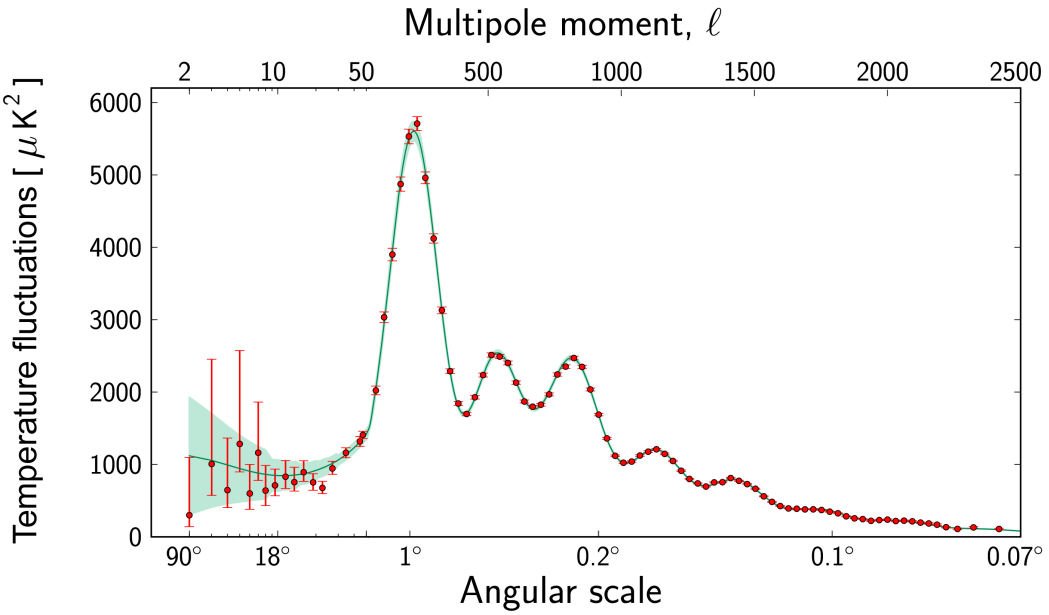

Fig. 1.3 Power spectrum of the temperature fluctuations of the CMB. Figure extracted from the Planck team.

mogeneities had to be present. These inhomogeneities produced gravitational instabilities, such that regions with slightly higher densities were subject to larger gravitational forces, attracting more surrounding matter and collapsing into structures. We can describe these inhomogeneities by expressing the matter density field (ρ_m) as:

$$\rho_m(\mathbf{x},t) = \bar{\rho}_m\left[1+\delta_m(\mathbf{x},t)\right], \tag{1.30}$$

where ρ_m depends on the vector position ($\mathbf{x}$) and the time (t). $\bar{\rho}_m$ represents the mean background matter density, while $\delta_m(\mathbf{x},t)$ is the so-called *density fluctuation*. In the case where $\delta_m(\mathbf{x},t) = 0$, we recover the homogeneous matter distribution, while $\delta(\mathbf{x},t) > 0$ indicate overdensities and $\delta(\mathbf{x},t) < 0$ indicate underdensities.

LSS can be studied by combining the Einstein equations, that describe gravity, and the Boltzmann equations, which describe macroscopically the evolution of an ensemble of particles in the phase-space (position+momentum) through the distribution function $f_m(\mathbf{x},\mathbf{p},t)$, where $\mathbf{p}$ describes the momentum. DM and baryons are the main components in the universe that allow us to study structure formation. Considering that DM has no interactions (or very small) and neglecting the interaction of baryons, we can treat them

as a unique fluid, for both ρ_m and $f_m(\mathbf{x},\mathbf{p},t)$, the evolution of which is described to a good approximation by the collisionless Boltzmann equation:

$$\frac{df_m}{dt} = 0 = \frac{\partial f_m}{\partial t} + \frac{\partial f_m}{\partial x^i}\frac{dx^i}{dt} + \frac{\partial f_m}{\partial p^i}\frac{dp^i}{dt}. \tag{1.31}$$

Assuming non-relativistic matter on sub-horizon scales, the geodesic equations $\frac{dx^i}{dt}$ and $\frac{dp^i}{dt}$ can be expressed as in the Newtonian case plus the scale factor a, such that:

$$\frac{dx^i}{dt} = \frac{p^i}{am}, \tag{1.32}$$

$$\frac{dp^i}{dt} = -Hp^i - \frac{m}{a}\frac{\partial \psi}{\partial x^i}, \tag{1.33}$$

with ψ being the gravitational potential.

By substituting eq. 1.32 and 1.33 into eq. 1.31 and by using the Poisson equation (eq. 1.34), we are left with these set of equations, usually referred to as Vlasov-Poisson system, describing the evolution of matter in phase-space under a gravitational potential:

$$\nabla^2\psi = \frac{3}{2}\Omega_m(\eta)\,(aH)^2\,\delta_m \tag{1.34}$$

$$\frac{df_m}{dt} = 0 = \frac{\partial f_m}{\partial t} + \frac{\partial f_m}{\partial x^i}\frac{p^i}{am} - \frac{\partial f_m}{\partial p^i}\left[Hp^i + \frac{m}{a}\partial_i\psi\right] \tag{1.35}$$

This is a set of equations of 7 dimensions (3 for $\mathbf{x}$, 3 for $\mathbf{p}$ and 1 for t), where we need to integrate the Boltzmann equation to obtain f_m and δ, which is then replaced in the Poisson equation. Therefore, this is very difficult to solve, and a common solution is to take a perturbative approach, by taking moments (integrate over the momentum variables) of the Vlasov-Poisson equations.

1.4.1 Perturbation theory

The idea behind perturbation theory, which is usually valid for mildly non-linear perturbations $\delta_m \sim 1$, is that it is possible to locally expand the density and the velocity fields, such that:

$$\delta_m(\mathbf{x},\eta) = \sum_{n=1}^{\infty} \delta_m^{(n)}(\mathbf{x},\eta), \tag{1.36}$$

$$\theta_m(\mathbf{x},\eta) = \sum_{n=1}^{\infty} \theta_m^{(n)}(\mathbf{x},\eta), \tag{1.37}$$

where $\theta_m = \nabla \cdot u_m$ is the divergence of the velocity field and we have neglected the vorticity in the velocity field. In this framework, $\delta_m^{(n)}$ and $\theta^{(n)}$ are the nth order solutions, after taking moments, of the Vlasov-Poisson equations, with $\delta_m^{(1)}$ and $\theta_m^{(1)}$ the linear solutions, $\delta_m^{(2)}$ and $\theta_m^{(2)}$ the quadratic solutions, etc.

By taking moments of the Vlasov-Poisson equations, and after neglecting higher order moments and the velocity dispersion, we are only left with space and time dimensions. The set of equations we obtain are the fluid equations, which are the continuity equation (eq. 1.38), the Euler equation (eq. 1.39) and the Boltzmann equations (eq. 1.34):

$$\frac{\partial \delta_m}{\partial \eta} + \frac{\partial}{\partial x^j}\left[(1+\delta_m)\, u_m^j\right] = 0 \tag{1.38}$$

$$\frac{\partial u_m^i}{\partial \eta} + u_m^j \frac{\partial}{\partial x^j} u_m^i + aH u_m^i + \frac{\partial \psi}{\partial x^i} = 0 \tag{1.39}$$

Note that these expressions are derived with respect to the conformal time η, which is defined as $dt = a(\eta)d\eta$. The velocity term u_m corresponds to the peculiar velocity, related to the comoving velocity by $v_m(\mathbf{x},\eta) = aH\mathbf{x} + u_m(\mathbf{x},\eta)$. Thus, the structure formation can be analysed by studying the perturbations of the matter density (ρ_m), the velocity (u_m) and the gravitational potential (ψ) fields.

The simplest case is to solve the fluid equations for the linear solutions. We can do that by taking the divergence of the Euler equation (eq. 1.39) and separating between linear and non-linear terms:

$$\frac{\partial \delta_m}{\partial \eta} + \theta_m = -\delta_m \theta_m - u_m^j \frac{\partial}{\partial x^j}\delta_m \tag{1.40}$$

$$\frac{\partial \theta_m}{\partial \eta} + aH\theta_m + \nabla^2\psi = -u_m^j \frac{\partial}{\partial x^j}\theta_m - \left(\partial_i u_m^j\right)\left(\partial_j u_m^i\right), \tag{1.41}$$

If we neglect the right-hand side terms of eq. 1.40 and eq. 1.41, which correspond to the non-linear terms in the variables in which we are interested, we can combine the fluid equations and obtain a second order ordinary differential equation for δ_m:

$$\frac{\partial^2 \delta_m}{\partial \eta^2}(\mathbf{x},\eta) + aH\frac{\partial \delta_m}{\partial \eta(\mathbf{x},\eta)} = \frac{3}{2}\Omega_m(\eta)\left(aH\right)^2 \delta_m(\mathbf{x},\eta) \tag{1.42}$$

The fluctuations of the density field will grow independently at all $\mathbf{x}$ if we redefine:

$$\delta_m(\mathbf{x},\eta) = D(\eta)\delta_0(\mathbf{x}), \tag{1.43}$$

where D is the growth factor and δ_0 is the density field at initial conditions. Substituting eq. 1.43 into eq. 1.42, we arrive at:

$$\frac{\partial^2 D}{\partial \eta^2}(\eta) + aH\frac{\partial D}{\partial \eta(\eta)} = \frac{3}{2}\Omega_m(\eta)\,(aH)^2 D(\eta) \tag{1.44}$$

This equation has two independent solutions, with a fast and a slow growing mode, $D^{(+)}$ and $D^{(-)}$, so that we can express the matter density and the velocity divergence fields as:

$$\delta_m(\mathbf{x},\eta) = D^{(+)}(\eta)A(\mathbf{x}) + D^{(-)}(\eta)B(\mathbf{x}) \tag{1.45}$$

$$\theta_m(\mathbf{x},\eta) = -aH\left[A(\mathbf{x})f(\eta) + B(\mathbf{x})g(\eta)\right], \tag{1.46}$$

with

$$f(\eta) = \frac{d\ln D^{(+)}}{d\ln a} \tag{1.47}$$

and

$$g(\eta) = \frac{d\ln D^{(-)}}{d\ln a} \tag{1.48}$$

where $A(\mathbf{x})$ and $B(\mathbf{x})$ correspond to functions described by the initial density and divergence fields. In the case of matter domination ($\Omega_m = 1$), the solutions for the different terms are:

$$D^{(+)} = a \qquad D^{(-)} = a^{-3/2} \qquad f = 1 \qquad g = -3/2 \tag{1.49}$$

with the terms $D^{(+)}$ and f being the dominant ones. The growth factor, which reflects the density fluctuations, grows linearly with the scale factor.

Another interesting case to solve is when DE is considered, so that $\Omega_m + \Omega_\Lambda = 1$. The solutions for this case are:

$$D^{(+)} = \frac{5\Omega_m a^3 H(a)}{2}\int_0^a \frac{da}{a^3 H^3(a)} \qquad D^{(-)} = H(a) \qquad f \sim \Omega_m^{5/9} \tag{1.50}$$

This tells us that, when we consider DE, the structure formation is suppressed, since the growth factor is smaller.

1.4.2 N-body simulations

The fluid equations described in the previous Section 1.4.1 do not describe accurately the evolution of structure in the regime of non-linear scales. In particular, the fluid equations are not able to describe the so-called *shell crossing*. In the context of gravitational collapse of matter, this phenomenon occurs when different shells cross each other. Given that DM is collisionless, shell crossing is unimpeded. A simple way to understand this phenomenon is to imagine a gravitationally collapsing sphere, with higher density than its environment. This sphere has different layers, defined by the radius with respect to its centre. Shells with smaller radii collapse before than shells with larger ones. Thus, the shells that have already collapsed have velocities closer to 0, while the outer shells that have not collapsed yet will have larger velocities. All this leads to positions having different velocities, so that the phase-space sheet is multivalued. In the fluid treatment defined in the previous section, a fluid can only have a single velocity value at each position and time, hence being unable to describe shell crossing. In that case, perturbation theory is not valid anymore and we must resort to other options, such as N-body simulations.

These type of simulations also make use of the Poisson and collisionless Boltzmann equations, eq. 1.34 and eq. 1.35, respectively. Nevertheless, the goal of N-body simulations is not to solve the 7D computations of the previous equations. Instead, the approach is to consider that matter lives in the phase-space sheet, so that it is possible to discretize it and evolve it numerically. To do so, we focus on the geodesic equations, introduced in eq. 1.32 and eq. 1.33, and follow the evolution of phase-space elements, which are usually referred to as particles in the N-body terminology. A common strategy is to define the so-called *superconformal momentum* $\mathbf{p}_c = a\mathbf{p}$ and to introduce it in the geodesic equations, so that they become:

$$\frac{dx^i}{dt} = \frac{p_c^i}{a^2 m}, \tag{1.51}$$

$$\frac{dp_c^i}{dt} = -m\frac{\partial \psi}{\partial x^i}. \tag{1.52}$$

The advantage of this change of variable is that $\mathbf{p}_c$ is conserved if there are no perturbations in ψ.

The objective is to study the evolution of $\mathbf{x}$ and $\mathbf{p}_c$ for each particle. A possible approach is to use the so-called *leapfrog scheme*, where the position and momentum are computed at staggered times. This process can be described in these four steps:

1. Compute the gravitational potential ψ and take its gradient ($\nabla\psi$).

2. Change the particle momentum through a *kick*, such that

$$\mathbf{p}_c^{(i)}(t+\Delta t/2) = \mathbf{p}_c^{(i)}(t-\Delta t/2) - m\nabla\psi(\mathbf{x}^{(i)},t)\Delta t \tag{1.53}$$

3. Move the position of each particle given the momentum computed in the previous step

$$\mathbf{x}^{(i)}(t+\Delta t) = \mathbf{x}^{(i)}(t) + \frac{\mathbf{p}_c^{(i)}(t+\Delta t/2)}{ma^2(t+\Delta t/2)}\Delta t \tag{1.54}$$

4. Repeat

Here, the superscript i specifies each particle and Δt is the time step, where $\mathbf{x}$ and $\mathbf{p}_c$ are computed with a time interval difference of $\Delta t/2$.

The initial conditions for the N-body simulations can be obtained using the Zel'dovich approximation, which can be considered as a 1st order Lagrangian perturbation theory solution. While, in the previous Section 1.4.1, the theory was derived using Eulerian perturbation theory, where the fluid element corresponds to a fixed volume in comoving space, in the Lagrangian case the fluid element follows the particle trajectory, which can be described as:

$$\mathbf{x}(\eta) = \mathbf{q} + S(\mathbf{q},\eta), \tag{1.55}$$

where $\mathbf{q}$ and $\mathbf{x}$ are the initial and final positions of the particle, respectively, and S is the displacement field, which defines the motion of each element of the fluid. In Lagrangian perturbation theory, one solves for the nth solution of the displacement field, $S^{(n)}$, in analogy with eq. 1.36 and 1.37. The Zel'dovich approximation is the first order solution of the displacement field, such that eq. 1.55 can be written as:

$$\mathbf{x}(\eta) = \mathbf{q} + D(\eta)s(\mathbf{q}), \tag{1.56}$$

where $s(\mathbf{q})$ is the initial displacement field and D is the linear growth factor (eq. 1.43). This derivation uses the fact that the displacement field is related to the initial density field,

$\delta_m(\mathbf{x} = \mathbf{q})$, and the growth factor by:

$$\nabla_{\mathbf{q}} \cdot S(\mathbf{q}, \eta) = -D(\eta)\delta_m(\mathbf{q}), \tag{1.57}$$

This approximation is valid up to the shell-crossing phenomenon described before, so that it is an effective way to set the initial conditions for N-body simulations.

A critical aspect of N-body simulations is the computation of $\nabla\psi$, that is, the force experienced by particles. One approach would be to compute the force experienced by a given particle as the sum of the forces exerted by all the other particles:

$$\nabla\psi(\mathbf{x_i}) = Gm \sum_{j \neq i} \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_j - \mathbf{x}_i|^3} \tag{1.58}$$

However, the computational cost of this scales as N^2 for N particles, which is ineffective to do. Thus, other approaches with computational costs that scale as $N \log N$ have been developed, such as particle-mesh (PM) and tree algorithms. On the one hand, in the case of PM algorithms, particles are assigned onto grids of a certain resolution, obtaining a density field for each grid. Then, the Poisson equation is solved for each grid and the gradient of the potential is interpolated to the position of each particle. The grids defined in PM algorithms can be of fixed size, leading to fixed resolutions below which the structure evolution can not be resolved, or of adaptive size, where the size of the grid decreases (thus increasing the resolution) in regions where the number of particles are larger. On the other hand, tree algorithms do not use grids but group particles together to compute the effect of other distant particles. This way, a tree is generated of *meta* particles that carry the mass of several other particles, from which we can compute $\nabla\psi$.

The final output of an N-body simulation is the position and velocity of particles that form the simulation at different times, usually referred to as *snapshots*. These define the positions in the phase-space sheet after gravitational evolution, which can be used to compute the density field and measure the matter power spectrum.

Dark matter halos

A main limitation of N-body simulations is that they only include gravity, excluding gas and star formation, thus being unable to simulate galaxies. However, one can look for gravitationally bound particles in these simulations, which are called *dark matter halos*. These are important because it is possible to explain the structure and galaxy formation and evolution from them. In particular, we can understand the location of galaxies, since they are

expected to reside in DM halos. Consequently, if we know how galaxies populate DM halos, the distribution of galaxies can be understood using only N-body simulations.

The first step is to identify DM halos. The so-called *halo finders* are employed for this task. Although there are different algorithms doing this and differing in the details, the main idea is to work with the catalogue of particle positions and find local maxima in the density field. Later, neighbouring particles are evaluated to check if they are gravitationally bound to that density maxima, by comparing their kinetic energy to the gravitational potential produced by the local density maxima.

Another key step is to determine the mass of halos. Although it might seem reasonable to just sum the mass of all bound particles, this is not an appropriate solution since it is hard to connect this to observations, given that we do not observe DM, but tracers of it, such as galaxies. Instead, a common approach is to compute the mass M_Δ of a sphere enclosing a fixed density, expressed in terms of Δ times the cosmic mean density ($\bar{\rho}_m$):

$$\frac{M(<R_\Delta)}{4\pi R_\Delta^3/3} = \Delta \cdot \bar{\rho}_m(t_0), \tag{1.59}$$

where $M_\Delta = M(<R_\Delta)$ and R_Δ are the mass and the radius of the halo, respectively. A typical extended value for Δ in the literature is $\Delta = 200$. An interesting case is when $\Delta = 1$, where the radius obtained is referred to as Lagrangian radius, which physical interpretation is the comoving size of a sphere containing the particles in the initial conditions that will eventually form a bound structure of mass M.

Once the mass of the halos is determined, one can study the halo abundance, which is the most important statistical property of DM halos. It is usually studied using the halo mass function, which is defined as the mean number density of halos in logarithmic mass bins, $dn/d\ln M$ vs. $d\ln M$. Qualitatively, the halo mass function can be described by a power-law for small masses with an exponential cut-off at high masses. This means that there are many more low-mass halos, corresponding to smaller galaxies or subhalos, while higher mass halos, corresponding to galaxy clusters, are more scarce. Typical analytical derivations describing the halo mass function have been presented by Press-Schechter [177] and Sheth-Tormen [194].

Finally, we need to populate halos with galaxies. There are two main methods used in N-body simulations to perform this task: Halo Occupation Distribution (HOD) and abundance matching techniques. On the one hand, HOD works under the assumption that galaxies reside in DM halos as a function of their mass, such that one can establish a probability distribution $P(N_g|M_h)$, with N_g the number of galaxies and M_h the mass of the halo. The

number of galaxies to be assigned to each halo usually depends on the mass of the halo and on the type of galaxies, typically differentiating between central and satellite galaxies. Within this technique, it is possible to specify the minimum mass that a halo must have in order to host a galaxy and the relation between the mass of a halo and the number of satellite galaxies. The function $P(N_g|M_h)$ has several free parameters which are varied in order to match observations of similar populations. On the other hand, the abundance matching technique is simpler than HOD algorithms. It assigns galaxies to DM halos based on both their properties. A typical approach is to populate halos with galaxies based on their mass, such that massive galaxies populate massive halos. The number of free parameters in abundance matching is reduced with respect to HOD techniques. An extension of abundance matching is subhalo abundance matching (SHAM), where the DM halos are divided between central massive halos and less massive subhalos, allowing for a better assignment of satellite galaxies.

1.4.3 Two-point correlation function, power spectrum and angular correlations

The two-point correlation function (2PCF) is a widely used tool in cosmology that allows to statistically describe the LSS of the universe. The 2PCF is defined as the average, over an integrated volume V, of the product of the scalar field f($\mathbf{x}$) with itself separated by a distance vector $\mathbf{r}$. Without loss of generalization, we will use the overdensity field, δ_m, defined in eq. 1.30, as the scalar field to define the 2PCF, since we are interested in the distribution of the LSS of the universe:

$$\xi_m(\mathbf{r}) = \frac{1}{V}\int_V d\mathbf{x}\delta_m(\mathbf{x})\delta_m(\mathbf{x}+\mathbf{r}) = \langle \delta_m(\mathbf{x})\delta_m(\mathbf{x}+\mathbf{r})\rangle\,, \tag{1.60}$$

If we assume that the universe is isotropic, as the CP indicates, for a large enough value of V, the 2PCF will also be isotropic, so that eq. 1.60 only depends on r, the modulus of $\mathbf{r}$.

An interesting case for the 2PCF is when the scalar field f($\mathbf{x}$) describes the distribution of galaxies in the universe. In that case, f($\mathbf{x}$) is a discrete function with the galaxy counts as a function of $\mathbf{x}$ and the correlation function measures the excess probability, with respect to a Poisson distribution, of finding a galaxy at a separation r from another galaxy. Thus, positive (negative) values of $\xi(\mathbf{r})$ indicate that the distribution of galaxies at a separation r is higher (lower) than that expected from a random distribution. We will show this in more detail in Section 1.5.

The 2PCF is defined in configuration space. However, it is also common to define the analogous expressions in Fourier space, the power spectrum. First, the Fourier transform of a scalar field $f(\mathbf{x})$ in the $\mathbb{R}^3$ space is defined as:

$$\mathscr{F}[f(\mathbf{x})] = \int_{\mathbb{R}^3} d^3\mathbf{x} f(\mathbf{x}) e^{-i\mathbf{kx}} = \tilde{f}(\mathbf{k}), \tag{1.61}$$

where $\mathbf{k}$ is the wavenumber in Fourier space.

The inverse Fourier transform is defined as:

$$\mathscr{F}^{-1}[\tilde{f}(\mathbf{k})] = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} d^3\mathbf{k} \tilde{f}(\mathbf{k}) e^{i\mathbf{kx}} = f(\mathbf{x}), \tag{1.62}$$

where the $(2\pi)^{-3}$ is a normalisation constant.

The convolution product of two scalar fields, f_1 and f_2, in the $\mathbb{R}^3$ space can be expressed as:

$$(f_1 * f_2)(\mathbf{x}) = \int d^3\mathbf{x}' f_1(\mathbf{x} - \mathbf{x}') f_2(\mathbf{x}') \tag{1.63}$$

The *convolution theorem* in Fourier space states that the Fourier transform of the convolution of two functions equals the pointwise product of the Fourier transforms of each function, such that:

$$\mathscr{F}[(f_1 * f_2)(\mathbf{x})] = (2\pi)^3 \mathscr{F}[f_1(\mathbf{x})] \mathscr{F}[f_2(\mathbf{x})] \tag{1.64}$$

Thus, we can rewrite eq. 1.63 in Fourier space as:

$$(f_1 * f_2)(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k} e^{i\mathbf{kx}} (\tilde{f}_1 \cdot \tilde{f}_2)(\mathbf{k}) \tag{1.65}$$

From all this, we can relate the 2PCF and the power spectrum, since the former is the inverse Fourier transform of the latter:

$$\xi_m(r) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} d^3\mathbf{k} P_{\delta\delta}(\mathbf{k}) e^{i\mathbf{kx}} \tag{1.66}$$

Therefore, the matter power spectrum is defined as the average of the Fourier transform of δ_m, in particular, $|\tilde{\delta}_m(\mathrm{k})|^2$, over the solid angle ($\Omega$):

$$P_{\delta\delta}(k) = \frac{1}{4\pi} \int d^2\Omega |\tilde{\delta}_m(\mathrm{k})|^2, \tag{1.67}$$

which is equivalent to the average over the whole ensemble given by the following definition:

$$\left\langle \tilde{\delta_m}(\mathbf{k})\tilde{\delta_m}^*(\mathbf{k}')\right\rangle_\Omega = (2\pi)^3 P_{\delta\delta}(k)\delta_D^{(3)}(\mathbf{k}-\mathbf{k}'), \tag{1.68}$$

where $\delta_D^{(3)}(\cdot)$ is the Dirac delta function.

The information included in both the 2PCF and the power spectrum is three-dimensional. Nevertheless, in some case, due to inaccurate redshifts yielding inaccurate measurements of the distance, one might need to project the 2PCF into two dimensions, leading to the two-point angular correlation function, $w(\theta)$, or to project the power spectrum, leading to the two-dimensional (or angular) power spectrum, the $C(l)$. Note that, as in the case of three dimensions, the $C(l)$ is the Fourier transform of $w(\theta)$.

Using the Limber approximation [147], which is valid for small angular scales, the angular power spectrum can be written as:

$$C(l) = \frac{1}{l}\int_0^\infty dk P(k) W^2\left(\frac{l}{k}\right), \tag{1.69}$$

where W is a selection (or window) function that describes the radial distribution of objects and the multipole l is defined as $l = k\chi$.

1.5 Galaxy clustering

As seen in Section 1.4, the distribution of matter in our universe can be studied from the matter power spectrum. The majority of matter is in the form of DM, while galaxies and dust account for a smaller portion of the matter content. Nevertheless, many of the current spectroscopic and imaging surveys observe galaxies, so a relation between the matter and the galaxy distributions need to be established, using galaxies as *tracers* of matter. Galaxy Clustering (GC) is the branch of Cosmology that studies the relation between matter and galaxies and all the effects intervening in it.

In order to measure the positions of galaxies and obtain the galaxy density field, both spectroscopic and imaging surveys can be used. In the case of spectroscopic surveys, the three-dimensional distribution of galaxies is obtained, using spectroscopic redshifts as radial estimators. This can be an expensive task, as we will see in Section 1.7. That is why, in the case of imaging surveys, we resort to estimating the angular distribution of galaxies, dividing the redshift space, estimated with photometric redshifts, in a certain number of bins, usually referred to as *tomographic bins*.

The relation between the matter and the galaxy distributions is encoded in the *galaxy bias* term(s). The simplest model one can use is given by a constant term [167]:

$$\delta_g = b_1 \delta_m, \tag{1.70}$$

where b_1 is usually referred to as *linear galaxy bias*. Although this model is very simplistic, it is known to work well at large scales. Nevertheless, at small scales, the non-linearities of the LSS need more sophisticated models. In this sense, the galaxy bias can be expressed as:

$$\delta_g(\mathbf{x},t) = \sum_O b_O O(\mathbf{x},t), \tag{1.71}$$

where O are operators, or statistical fields, that capture the properties of the galaxies' environment on which their density can depend and b_O are constant numbers, which are not predicted by theory and are treated as nuisance parameters. In the case of the linear galaxy bias, the expansion is truncated at first order, so that $b_O = b_1$ and $O(\mathbf{x},t) = \delta_m(\mathbf{x},t)$. The power spectrum for GC in the linear bias approximation is defined as:

$$P_{gg}(k) = b_1^2 P_{\delta\delta}^{\rm lin}(k). \tag{1.72}$$

with $P_{\delta\delta}^{\rm lin}$ the linear matter power spectrum.

The analogous quantity of $P_{gg}(k)$ can be defined in configuration space by measuring the 2PCF of galaxies, $\xi_{gg}(\mathbf{r})$, via:

$$\left\langle N_{g,1}(\mathbf{x}) N_{g,2}(\mathbf{x}+\mathbf{r}) \right\rangle = \bar{n}^2 \left[1 + \xi_{gg}(\mathbf{r})\right] \delta V_1 \delta V_2, \tag{1.73}$$

where $\bar{n}$ is the mean number of galaxies per unit volume and $N_{g,1}$ and $N_{g,2}$ are the number of galaxies in two regions separated by $\mathbf{r}$ and with volumes δV_1 and δV_2, respectively.

1.5.1 Redshift-space distortions and Alcock-Paczyński parameters

As we will see in Section 1.7, the redshifts observed by both the spectroscopic and imaging surveys are not only affected by the expansion of the universe, but also by a Doppler shift, which is caused by the peculiar velocities of the objects with respect to the Hubble flow. This means that, even if the measurement of the redshift is very accurate, it can not always be translated into a comoving distance, since the peculiar velocities need to be taken into account. Besides, these peculiar velocities are not random, but are correlated with the matter

distribution. These effects are usually treated under the name of *redshift-space distortions (RSD)*.

Another effect that needs to be taken into account is the fact that, in order to translate redshifts into distances, one needs to assume a fiducial cosmological model, which may not agree with that of the real universe, adding some distortions to the expected observed GC. This is referred to as the Alcock-Paczyński (AP) effect.

Taking both of these phenomena into account, one can relate the observed three-dimensional position ($\mathbf{x}_{\rm obs}$) with the true position ($\mathbf{x}$) as:

$$\mathbf{x}_{\rm obs} = \mathbf{x} + \left(\delta\chi(z) + \frac{1}{aH} u_{\parallel}(\mathbf{x}) \right) \hat{n}, \tag{1.74}$$

where $\delta\chi(z)$ is the shift in the comoving distance incurred when assuming a fiducial cosmological model different from the truth, $u_{\parallel}$ is the peculiar velocity of the galaxy and the unit vector $\hat{n}$ is defined as:

$$\hat{n} = \frac{\mathbf{x}_{\rm obs}}{|\mathbf{x}_{\rm obs}|} \tag{1.75}$$

The difference between the observed and the true positions expressed in eq. 1.74 is translated into the observed and true matter overdensities. By considering that the number of galaxies in a region of the universe is conserved in both the observed and the true coordinate positions:

$$n_{g,\rm obs}(\mathbf{x}_{\rm obs}) d^3 x_{\rm obs} = n_g(\mathbf{x}) d^3 x \tag{1.76}$$

and expanding to first order in perturbations, the following expression arises:

$$1 + \delta_{g,\rm obs}(\mathbf{x}_{\rm obs}) = \bar{J} \left[1 + \delta_g(\mathbf{x}) - \frac{1}{aH} \frac{\partial}{\partial x} u_{\parallel}(\mathbf{x}) \right], \tag{1.77}$$

where

$$\bar{J} = 1 - 2\frac{\delta\chi(\bar{z})}{\bar{\chi}} + H^{-1}(\bar{z}) \delta H(\bar{z}), \tag{1.78}$$

and $\delta H(z) = H(z) - H_{\rm fid}(z)$ accounts for the shift between the Hubble rates of the true and the fiducial rates, respectively. To simplify the computation, it has been assumed that we are observing galaxies in a narrow redshift slice around $\bar{z}$, so that $\bar{J}$ can be treated as a prefactor in eq. 1.77.

The RSD and AP effect mentioned above can be seen in eq. 1.77. On the one hand, $\bar{J}$ carries the effect of assuming a wrong fiducial cosmology through δH and $\delta\chi$. On the other

hand, the last term on the right-hand side of eq. 1.77 yields the distortion due to peculiar velocities.

If we focus only on the RSD effect, by using eq. 1.70 and eq. 1.77, setting $\bar{J} = 1$ (effectively ignoring the AP effect) and assuming a flat-sky approximation, the matter overdensity due to RSD can be written in Fourier space with the following expression:

$$\delta_{g,\mathrm{RSD}}(\mathbf{k}) = \left[b_1 + f\mu_k^2\right] \delta_m(\mathbf{k}), \tag{1.79}$$

where f is the linear growth rate defined in eq. 1.47 and μ_k is the cosine of the angle between the line of sight and the wavevector $\hat{\mathbf{k}}$. An important remark one can extract from eq. 1.79 is that it is possible to extract cosmological information from RSD through the linear growth rate.

The power spectrum of GC accounting for RSD can be derived from eq. 1.79 to be:

$$P_{gg,\mathrm{RSD}}(k,\mu_k,\bar{z}) = P^{\mathrm{lin}}_{\delta\delta}(k,\bar{z}) \left[b_1 + f\mu_k^2\right]^2 \tag{1.80}$$

We can also include the AP effect in eq. 1.80 by considering how the true and observed positions are affected by a wrong fiducial cosmological model. This effect can be expressed as:

$$\mathbf{x}[\mathbf{x}_{\mathrm{obs}}] = \left([1-\alpha_\perp]x_{\mathrm{obs}^1}, [1-\alpha_\perp]x_{\mathrm{obs}^2}, [1-\alpha_\parallel]x_{\mathrm{obs}^3}\right), \tag{1.81}$$

where

$$\alpha_\perp = \left.\frac{\delta\chi}{\chi_{\mathrm{fid}}}\right|_{\bar{z}} \tag{1.82}$$

and

$$\alpha_\parallel = \left.\frac{\delta H}{H_{\mathrm{fid}}}\right|_{\bar{z}} \tag{1.83}$$

are the perpendicular and parallel distortions due to a wrong distance-redshift relation, respectively. The power spectrum accounting for both RSD and AP effects can be expressed following eq. 1.80 but now evaluated at the true wavevector derived from eq. 1.81:

$$P_{gg,\mathrm{obs}}(\mathbf{k}_{\mathrm{obs}},\bar{z}) = \left(P^{\mathrm{lin}}_{\delta\delta}(k,\bar{z}) \left[b_1 + f\mu_k^2\right]^2\right)|_{\mathbf{k}=\left([1+\alpha_\perp]k_{\mathrm{obs}^1}, [1+\alpha_\perp]k_{\mathrm{obs}^2}, [1+\alpha_\parallel]k_{\mathrm{obs}^3}\right)} \tag{1.84}$$

1.5.2 Angular correlations

The three-dimensional spectrum derived in eq. 1.84 requires the precise measurement of redshifts. In the case of many imaging galaxy surveys, this information is not available. Instead, these type of surveys have photometric redshifts, which act as a proxy to the spectroscopic ones in the case of broad-band imaging surveys. The projected overdensity Δ_g is the quantity that these surveys are able to measure and it corresponds to the number of galaxies in different regions of the sky, which can be interpreted as the projection over the radial direction of the three-dimensional galaxy density $\delta_{g,\mathrm{obs}}$, weighted by the radial distribution of galaxies $W(\chi)$:

$$\Delta_g(\hat{\mathbf{n}}) = \int_0^\infty d\chi W(\chi)\delta_{g,\mathrm{obs}}(\mathbf{x} = \hat{\mathbf{n}}\chi, \eta = \eta_0 - \chi), \tag{1.85}$$

where

$$W(\chi) = \frac{1}{N_g}\frac{dN_g}{d\chi} \tag{1.86}$$

and N_g is the total number of galaxies. Note that the argument $\eta = \eta_0 - \chi$ in eq. 1.85 reflects the fact that $\delta_{g,\mathrm{obs}}$ is evaluated at different conformal times η, since distant galaxies are observed at earlier times than closer ones.

The GC angular power spectrum can be computed by averaging Δ_g, expanded in spherical harmonics ($\Delta_g(\hat{\mathbf{n}}) \to \Delta_{g,lm}$):

$$\left\langle \Delta_{g,lm}\Delta^*_{g,l'm'} \right\rangle = \delta_{ll'}\delta mm' C_{gg}(l). \tag{1.87}$$

Using the Limber approximation, the GC angular power spectrum can be expressed as:

$$C_{gg}(l) = \int \frac{d\chi}{\chi^2} W^2(\chi) P_{gg,\mathrm{obs}}\left(k = \frac{l+1/2}{\chi}, \mu_k = 0, \eta(\chi)\right) \tag{1.88}$$

As a final remark, the angular correlation function $w_g(\theta)$ is related to $C_{gg}(l)$ by:

$$w_g(\theta) = \int_0^\infty \frac{dl}{2\pi} l C_{gg}(l) J_0(l\theta), \tag{1.89}$$

where J_0 is the 0th order Bessel function of the first kind.

1.6 Weak gravitational lensing

One of the predictions of general relativity is the deflection of light produced by gravity. In fact, it was the first observational test of Einstein's theory, when the group led by Frank Dyson observed the deflection of light by the Sun's gravitational field in the total eclipse of May 29, 1919 [63].

Based on Fermat's principle, light paths follow trajectories, which are called geodesics, such that the light travel time is stationary. In a FLRW flat metric (Section 1.2.1), this corresponds to straight trajectories. Nevertheless, in the presence of gravitational forces which curve the space-time metric, this leads to a curvature in the light trajectory.

The phenomenon that the gravity associated to matter inhomogeneities distort the light paths of distant sources, as they travel towards us, is known as *gravitational lensing*. The degree to which the path lights are distorted lead to different scenarios. On the one hand, in the case that light sources are close, in angular position, to massive objects (which act as lenses), the light may be bent in multiple directions, generating multiple images of the same object. This effect is called *strong gravitational lensing* and distorts the shape of the objects into tangential arcs centred around the massive object, while at the same time magnifies its size. On the other hand, if the sources and the lenses are apart in position and/or the lens is not a very massive object, the distortion of the light and its magnification will be much lower, so that it does not generate multiple images of the source but just distorts the shape of the object slightly. This phenomenon is called *weak gravitational lensing*.

Weak gravitational lensing is a powerful cosmological tool that is able to map the mass distribution of the universe, not only from galaxies but from DM as well. The usual objects for which we study the shape distortions in weak gravitational lensing studies are galaxies. Given the slight distortions in the shapes of galaxies, one needs to perform statistical analyses with thousands or millions of galaxies to estimate the mass distribution. Typically, the process in weak lensing studies is to define a set of background galaxies, referred to as *sources*, for which one measures the shape distortions produced by the intervening foreground mass, which may be DM halos or galaxy clusters and are usually referred to as *lenses*. Together with the estimation of the mass distribution of intervening objects, weak lensing analyses also allow measuring the non-linear matter power spectrum. In comparison with galaxy clustering (Section 1.5), this cosmological probe has the advantage that one is able to probe the mass distribution directly, without having to make assumptions about the relation between the matter and the galaxy distributions, that is, the galaxy bias.

1.6.1 Deflection of light paths

The relation between the true source light position, θ_S, and the observed one, that is, after being distorted by a gravitational field, θ, can be described as:

$$\theta_S^i = \theta^i + \Delta\theta^i, \tag{1.90}$$

where

$$\Delta\theta_i(\theta) = 2\int_0^\chi d\chi' \frac{\partial\phi}{\partial x^i}(\mathbf{x}(\theta,\chi'))\left(1-\frac{\chi'}{\chi}\right) \tag{1.91}$$

represents the shift in the light position due to an intervening gravitational potential and

$$\mathbf{x}(\theta,\chi')) = \left(\chi'\theta^1, \chi'\theta^2, \chi'\right) \tag{1.92}$$

is the relation between θ and the transverse (first two coordinates in eq. 1.92) and radial (last coordinate) positions. ϕ is a perturbation in the metric due to the spatial curvature generated by the gravitational potential.

From eq. 1.92, the relation $\partial/\partial x^i = \chi'^{-1}\partial/\partial\theta^i$ can be derived, which allows redefining eq. 1.91 as:

$$\Delta\theta_i(\theta) = \frac{\partial}{\partial\theta^i}\phi_L(\theta), \tag{1.93}$$

where ϕ_L is called the *lensing potential*:

$$\phi_L(\theta) = 2\int_0^\chi \frac{d\chi'}{\chi}\phi(\mathbf{x}(\theta,\chi'))\left(1-\frac{\chi'}{\chi}\right). \tag{1.94}$$

Finally, another useful quantity is the *distortion tensor*, which is defined as the derivative of the deflection angle $\Delta\theta^i$:

$$\psi_{ij} \equiv \frac{\partial\Delta\theta^i}{\partial\theta^j} = 2\int_0^\chi d\chi' \frac{\partial^2\phi}{\partial x^i\partial x^j}(\mathbf{x}(\theta,\chi'))\left(1-\frac{\chi'}{\chi}\right), \tag{1.95}$$

where we have used again the relation between the partial derivatives of x and θ. Both eq. 1.94 and 1.95 are very useful when computing the relation between the gravitational lensing and the matter distribution.

1.6.2 Measuring galaxy shapes

Since weak gravitational lensing studies mainly use galaxy shapes to obtain estimates of the mass distribution, we will briefly review how the galaxy shapes are measured and how these are affected by gravitational lensing.

One can define galaxy shapes from the second moments of their image:

$$q_{ij} = \left\langle \theta_i \theta_j \right\rangle_{I_{\rm obs}} = \frac{1}{F} \int d^2\theta I_{\rm obs}(\theta) \theta_i \theta_j, \tag{1.96}$$

where $I_{\rm obs}$ is the observed specific intensity, which is defined as the incident energy ($E_{\rm obs}$) to the detector per solid angle, per unit area, per time and per unit frequency:

$$I_{\rm obs} = \frac{dE_{\rm obs}}{d\Omega dA_{\perp} dt d\nu}, \tag{1.97}$$

and the normalization factor is the total flux of the image:

$$F = \int d^2\theta I_{\rm obs}(\theta) \tag{1.98}$$

Note that both eq. 1.96 and 1.98 assume that the image is centred at the origin, such that $(\theta_x, \theta_y) = (0, 0)$.

We can write eq. 1.96 in terms of the ellipticity components:

$$q_{ij} = \frac{1}{2} q \begin{pmatrix} 1+\varepsilon_1 & \varepsilon_2 \\ \varepsilon_2 & 1-\varepsilon_1 \end{pmatrix} \tag{1.99}$$

where q is the trace of q_{ij} and ε_1 and ε_2 are the ellipticity terms, corresponding to the E- and B-modes, respectively. From this, we can extract the following relations:

$$\varepsilon_1 = \frac{q_{11} - q_{22}}{q} \tag{1.100}$$

$$\varepsilon_2 = \frac{2q_{12}}{q} \tag{1.101}$$

How do we link the expression in eq. 1.100 and eq. 1.101 to the effect of gravitational lensing? It is common to define the matrix A, which captures the change in the angle θ_S with respect to θ, that is, the change between the source and the image:

$$A_{ij} = \frac{\partial \theta_S^i}{\partial \theta^j} = \begin{pmatrix} 1-k-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-k+\gamma_1 \end{pmatrix} \tag{1.102}$$

where k is the *convergence* and captures the isotropic focusing, affecting the change in flux and size of the image, and $\gamma = \gamma_1 + i\gamma_2$ is the *gravitational shear*.

Note that, using eq. 1.90 and eq. 1.95, one can redefine A_{ij} as:

$$A_{ij} = \frac{\partial \theta_S^i}{\partial \theta^j} = \frac{\partial \left(\theta^i + \Delta\theta^i\right)}{\partial \theta^j} = \delta_{ij} + \psi_{ij}, \tag{1.103}$$

where k and γ are then fully defined by the integrated gravitational potential ϕ.

In the limit of small light deflections, galaxy ellipticities are related to the shear by

$$\varepsilon_i = \varepsilon_i^{\text{true}} + 2\gamma_i, \tag{1.104}$$

where $\varepsilon_i^{\text{true}}$ is the ellipticity of the galaxy before being distorted by the gravitational shear. Now, if we assume that the average of all galaxy ellipticities vanish ($\langle \varepsilon_i^{\text{true}} \rangle = 0$), we find that the average of the observed ellipticities is proportional to the average of the gravitational shear:

$$\langle \varepsilon_i \rangle = 2 \langle \gamma_i \rangle \tag{1.105}$$

which indicates that the observed ellipticity is an unbiased estimator of the gravitational shear. We will show in Section 1.8 that the assumption $\langle \varepsilon_i^{\text{true}} \rangle = 0$ is not necessarily true, so that one should take into account these *intrinsic* ellipticities.

1.6.3 Shear power spectrum and shear correlation function

We can compute the average weak lensing shear by integrating the second derivatives of the gravitational potential ϕ over the line of sight

$$\langle \gamma \rangle = \int_0^{\chi_H} d\chi p(\chi) \int_\chi^{\chi_H} d\chi' \frac{(\chi' - \chi)\chi'}{\chi} \left[(\partial_y^2 - \partial_x^2 + 2i\partial_x\partial_y) \right] \phi(x, y, \chi), \tag{1.106}$$

where χ_H is the comoving horizon distance ($\chi_H \sim c/H_0$) and $p(\chi)$ is the probability distribution of the source galaxies as a function of the comoving distance.

The average weak lensing shear is expected to vanish for large areas of the sky, due to the isotropy of the universe. Thus, the common approach is to use two-point statistics to measure the weak gravitational lensing. One can adopt two different two-point statistics: correlations between pairs of galaxy shapes and correlations between galaxy shapes and positions. In this

section, we focus on the former, which is also usually referred to as *cosmic shear*, leaving the latter for Section 1.6.4. The angular power spectrum of weak gravitational shear is formally defined as:

$$\left\langle \tilde{\gamma}^{(i)}(\mathbf{l})\tilde{\gamma}^{(j)}(\mathbf{l}')\right\rangle = (2\pi)^2\delta_D^2(\mathbf{l}+\mathbf{l}')C_{\gamma\gamma}^{(ij)}(l), \tag{1.107}$$

where δ_D^2 is the 2-dimensional Dirac delta function. Thus, the angular power spectrum of weak gravitational shear can be derived from eq. 1.106 using the Limber approximation to be:

$$C_{\gamma\gamma}^{(ij)}(l) = \int_0^{\chi_H} d\chi \frac{q^i(\chi)q^j(\chi)}{\chi^2} P_{\delta\delta}\left(\frac{l}{\chi},\chi\right), \tag{1.108}$$

where q^i corresponds to the lensing efficiency, defined as:

$$q^i(\chi) = \frac{3H_0^2\Omega_m}{2c^2}\frac{\chi}{a(\chi)}\int_\chi^{\chi_H} d\chi' p^{(i)}(\chi')\frac{\chi'-\chi}{\chi'} \tag{1.109}$$

Some interesting remarks can be extracted from eq. 1.108. First, the weak gravitational shear angular power spectrum is proportional to the matter power spectrum, without the need of any additional galaxy bias terms, which reduces the number of nuisance parameters. Second, eq. 1.108 can be used to constrain Ω_m, since it enters the lensing efficiency expression. Third, this expression allows us to study the expansion history of the universe through the comoving distance χ to the source galaxies, which enters the amplitude of the angular power spectrum, by means of the distance-redshift relation. Finally, eq. 1.108 accounts for the correlation of the E-modes of the shear, while the correlation of the B-modes is expected to be null, thus consisting on an excellent test to detect systematics in the observations.

Even though cosmic shear is a promising tool in cosmology, some challenges arise when it comes to observations. To begin with, the gravitational shear changes the shapes of the galaxies at a percent level, thus the signal-to-noise ratio (SNR) is rather small, so a large sample of galaxies is needed to obtain a significant measurement of cosmic shear. Next, a large population of galaxies will entail a large fraction of faint galaxies, leading to a complication in the measurement of their shapes. Additionally, the term $p(\chi)$ requires a good estimation of the distances of the source galaxies, which are usually computed using photometric redshifts. Nevertheless, the accuracy of those is not always good enough, so that information from external measurements is sometimes needed. Finally, the distortion of galaxy shapes is produced not only by gravitational shear, but only by the intrinsic ellipticity of galaxies, as indicated in eq. 1.104.

The shear correlation function is an alternative to the weak gravitational shear angular power spectrum, which is usually more difficult to measure due to the complex angular selection functions that need to be taken into account. The shear correlation function is the Fourier transform of the shear power spectrum and can be defined as:

$$\xi_{\pm}(\theta) = \langle \gamma_+ \gamma_+ \rangle (\theta) \pm \langle \gamma_x \gamma_x \rangle (\theta), \tag{1.110}$$

where θ indicates the angular separation, γ_+ is usually defined as the tangential shear, that is, the component of the shear parallel or perpendicular to the line connecting a pair of galaxies, and γ_x corresponds to the cross-component of the shear, shifted 45º with respect to γ_+.

The shear correlation function is related to the shear angular power spectrum via:

$$\xi_{\pm}(\theta) = \int \frac{ldl}{2\pi} J_{0,4}(l\theta) C_{\gamma\gamma}(l), \tag{1.111}$$

where $J_{0,4}$ are the 0th and 4th order Bessel functions of the first kind.

1.6.4 Shear cross-correlations

In this section, we focus on the other relevant two-point statistic in weak lensing analyses: the correlation between galaxy shapes and positions, also referred to as *shear cross-correlations* or *galaxy-galaxy lensing correlations*. In this scenario, two galaxy populations are defined: the lens galaxies, that define the positions of the galaxies, and the source galaxies, that are used to measure the shear. In order to obtain a not null galaxy shear correlation, the source galaxies need to be at higher redshift than the lens galaxies, given that this correlation measures the lensing produced by the mass associated to the lens galaxies. The angular galaxy-galaxy lensing power spectrum is similar to the one expressed in eq. 1.108, but substituting the lensing efficiency by the distribution of the lens galaxies and the matter power spectrum by the galaxy-matter power spectrum, such that:

$$C_{g\gamma}^{(ij)}(l) = \int_0^{\chi_H} d\chi \frac{p^i(\chi) q^j(\chi)}{\chi^2} P_{g\delta}\left(\frac{l}{\chi}, \chi\right), \tag{1.112}$$

It is important to note that, on linear scales, $P_{g\delta} \sim b_1$, as opposed to $P_{gg} \sim b_1^2$, as we saw in Section 1.5. This allows to break degeneracies in b_1 by means of combining measurements of shear cross correlations and galaxy correlation functions. More important for current imaging surveys is the combination of shear, shear cross and galaxy correlation functions

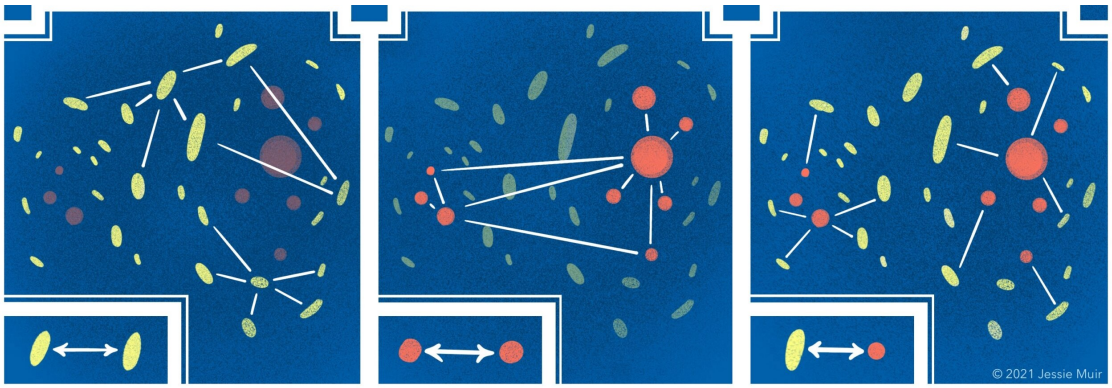

Fig. 1.4 Picture of the three 2-point statistics employed in 3x2pt analyses. From left to right: shear-shear, galaxy-galaxy and galaxy-galaxy lensing correlations. Yellow objects indicate source galaxies and red objects indicate lens galaxies. Credit to Jessie Muir and the Dark Energy Survey (DES) collaboration.

to tighten the constraints and break degeneracies on cosmological and nuisance parameters. The combination of these 3 correlation functions is usually referred to as 3x2pt. Fig. 1.4 depicts a picture of the three 2-point statistics employed in these type of analyses. Yellow objects indicate the source galaxies, while red objects indicate lens galaxies. From left to right, the shear-shear, galaxy-galaxy and galaxy-galaxy lensing correlations are depicted.

Finally, the galaxy-galaxy lensing cross correlation in configuration space is defined as:

$$\xi_{g+}(\theta) = \langle \Delta_g(0)\gamma_+(\theta)\rangle = -\int \frac{ldl}{2\pi} J_2(l\theta) C_{g\gamma}(l), \tag{1.113}$$

where Δ_g is defined in eq. 1.85. The cross-correlation with γ_x is expected to be null if there are not B-modes, so ξ_{gx} can also be used to ensure the absence of systematic effects, as in the case of the correlation between the B-modes of γ.

1.7 Photometric redshifts

The increase in wavelength experimented by photons when travelling throughout the universe is called redshift. It is an essential observable used by astrophysicists and cosmologists, and it is defined as the normalised wavelength difference of a photon at the time of emission ($\lambda_{\rm em}$) and at the time of observation ($\lambda_{\rm obs}$):

$$z = \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}} \tag{1.114}$$

This stretch of the photon's wavelength can be caused by three main causes: the cosmological redshift, the Doppler effect and the gravitational redshift. The cosmological redshift is caused by the acceleration of the universe, which we reviewed in Section 1.2, and is the dominant factor of redshift for distant objects. The acceleration of the universe causes distant objects to recede with respect to us, leading to a stretch in redshift. Thus, this cosmological redshift is directly related to the distance of the objects to us. Given the fact that the comoving wavelength between the time of observation and the time of emission do not change, one can find the following relation between the scale factor and the cosmological redshift:

$$1+z=\frac{1}{a(t)} \tag{1.115}$$

The Doppler effect happens due to the relative motion of radiative sources. For an object that is moving at a velocity u with respect to us, the redshift generated by the Doppler effect is given by:

$$1+z=\sqrt{\frac{1+u/c}{1-u/c}} \tag{1.116}$$

This Doppler effect is the cause of RSD (see Section 1.5.1), which difficults the relation between redshift and distance one can infer from the cosmological redshift.

Finally, the gravitational redshift is caused by the fact that photons encounter gravitational potentials during their journey. This causes a loss in energy, which translates in an increase in wavelength. The opposite effect, that is, the gain of energy in photons due to gravitational potentials, is known as gravitational blueshift.

The methodology behind redshift estimation lies in the *spectral energy distribution (SED)* of an object, which represents the energy emitted by an object as a function of the wavelength, and is composed by a continuum and by absorption and emission lines. As the universe expands, the wavelength associated to the SED stretches towards larger values by a factor $1+z$ (eq. 1.115). Thus, the redshift of an object can be estimated by detecting identifiable features on the SED and computing how much the wavelength has stretched. In particular, the emission and absorption lines are key features that allow to estimate the redshift. Besides, the shape of the continuum can be affected by two distinguishable features: the Balmer break and the Lyman break. On the one hand, the Balmer break is characterised by the absorption of photons more energetic than 3646 Å, which corresponds to the Balmer limit. On the other hand, the Lyman break correspond to the absorption of photons in the Lyman limit, at 912 Å.

One can infer redshift estimates using spectroscopy or photometry. The former case yields the so-called spectroscopic redshifts (spec-z), while the latter yields photometric

redshifts (photo-z). In the case of spec-z, the SED of an object is determined with great resolution with a spectrograph, which disperses the light of an object into its component wavelengths. This allows excellent accuracy in the estimation of the redshifts. However, this technique is expensive, since one can only measure a limited number of objects at once and a preselection of objects has to be performed beforehand, yielding a lower completeness. The case of the photo-z estimation is different, since the flux of the objects present in the field of view of the camera are measured at once, allowing to reconstruct its SED when measured through filters of different wavelengths. Nevertheless, this reconstruction depends on the filters used, with broader filters performing worse, since it is harder to differentiate key features in the SED, especially absorption and emission lines. For this reason, the accuracy of the photo-z is worse than the one of spec-z. In fact, a key step in the estimation of photo-z is its validation, which is usually estimated using samples of spectroscopic redshifts that need to be as representative of the photometric sample as possible, in order to validate them in all their parameter space.

The mapping between the observed colours, defined as the difference between the flux observed in two bands, and the redshift is key in the photo-z estimation method. The left-hand side plot in Fig. 1.5 shows the SED of different types of objects, such as elliptical (Ell, in red), starburst (SB, dark blue) and spiral galaxies with small bulge (Sc, grey), together with two kinds of AGN, luminous quasars (QSO, light blue) and low luminosity AGN (Sy1.8, orange). The differences in the shape and the amplitude of the SEDs helps to classify the objects of interest and to determine their redshift. In the same plot, we can also see the Lyman and Balmer breaks as vertical lines, together with the Balmer break at $z = 1.1$ and the transmission curves of the i and z filters around it (as green and pink contours, respectively). One of the methods to estimate photo-z is related to the decrease in the continuum flux from the red to the blue part of the SED wavelength, caused by the Balmer and Lyman breaks. This can be detected by comparing a difference in fluxes in adjacent filters, which would indicate the presence of a break. Thus, when designing an imaging survey, it is important to select a filter set that includes SED key features at the redshifts of interest. On the right-hand side of Fig. 1.5, we can see the observed i-z colour as a function of redshift for the different templates presented before. In particular, the maximum observed at $z = 1.1$ corresponds to the Balmer break between the i-z colour. Nevertheless, it is noticeable that there are some degeneracies present in the form of other maxima at specific photo-z values. These degeneracies might be resolved by including multiple colours in the survey, allowing to discard some of those maxima points. Therefore, an imaging survey will, in general, benefit from a broader wavelength coverage that allows to reduce the impact of photo-z degeneracies.

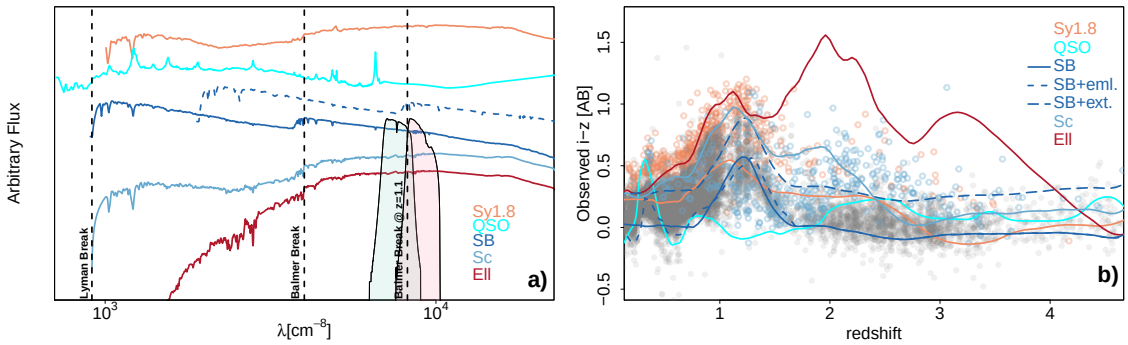

Fig. 1.5 Colour-redshift mapping: a) SED templates for different kind of objects: elliptical (Ell), starburst (SB) and spiral galaxies (Sc), QSO (light blue) and low luminosity AGN (Sy1.8, orange). Dashed vertical lines indicate the Lyman and the Balmer breaks. The position of the Balmer break at $z = 1.1$ is also shown, together with the positions of the i and z filters, shown in green and pink contours, respectively. b) Observed $i - z$ colour as a function of the redshift for the different SED templates presented in a) (solid lines). Grey dots represent the spectroscopic redshifts of galaxies, while orange and blue dots represents spectroscopic redshift for Sy1.8 and QSO, respectively. Plot extracted from [182]

Following the colour-redshift mapping technique, one can estimate photo-z based on template-fitting methods, machine learning techniques or a hybrid version of both cases.

The template-fitting method is physically motivated, since the observed fluxes are compared with SED templates, which can be theoretical templates generated from stellar population synthesis models ([74, 27, 151]) or observed templates ([43, 126, 173]). The photo-z accuracy will strongly depend on how the selected templates cover the colour-redshift space. The dust in the interstellar medium absorbs and scatters the stellar light, causing the SED continuum to appear redder. As a consequence, it is important to model this dust attenuation in template-fitting algorithms. This specially affect objects at $z > 1$, since optical data at those redshifts will correspond to rest-frame ultraviolet (UV) wavelengths, being the most affected by this phenomenon. Besides this effect, the SED templates also need to be corrected for the dust attenuation of the Milky Way and by the absorption that light can suffer from neutral gas in the inter-galactic medium. The last step when modelling the SED templates is to integrate them through the transmission curves of the filters used, together with any other possible systematics from the telescope or the camera.

As for machine learning techniques, they can be divided into supervised or unsupervised, depending on whether they rely on samples with spectroscopic redshifts or not. The latter only uses photometry and works by grouping the input data into groups of similar objects. A common unsupervised machine learning technique are the Self Organising Maps (SOMs)

([33]). In the case of supervised methods, they use photometry and spectroscopic redshift in order to be able to generate a relation between the multidimensional space of colours (or fluxes) and the redshift estimates. Once the training is completed, the flux of the objects for which one wishes to compute the redshifts are evaluated using this relation and the photo-z estimates are provided. It is of utmost importance that the training data is representative of the sample for which we want to estimate the redshifts, since otherwise extrapolations will be performed and the output might not be accurate. Neural Networks constitute one of the most popular supervised machine learning algorithms for photo-z estimation. They are built on several layers, each of them corresponding to a transformation. At the end of all the layers, the output is the photo-z estimate, which is estimated from a non-linear combination (in the form of a matrix) from all the layers of the input properties. The training data in these algorithms is used to optimise the matrix transformations, by minimizing the difference between the estimated and the spectroscopic redshifts of the training sample.

As a final note, the SED template-fitting algorithms and the machine learning techniques can both benefit from the use of priors, which constrain the possible redshift ranges of certain galaxy populations. However, these should be used cautiously in order not to introduce artificial constraints.

1.8 Intrinsic alignments

Galaxies are one of the main tools that astrophysicists and cosmologists use to study the universe. Their high abundance and the variety in their physical properties allow the study of a plethora of fields. Of particular interest to cosmologists is the fact that galaxies trace the DM distribution along the line of sight, as seen in Sections 1.4 and 1.5, allowing for the constraint of cosmological parameters and for the study of DM itself. Many physical properties can be measured from galaxies, e.g., the angular positions, the distance, the flux, the metallicity, the apparent size, the orientation and the ellipticity. In this section, we focus on the orientation and ellipticity of images and what we can extract from them.

One may think that, for an ensemble of galaxies, their orientations and ellipticities would be random with respect to one another, given that our universe is supposed to be homogeneous and isotropic (Section 1.1). However, observations show that this is not the case, since correlations between intrinsic galaxy shapes are not null. This phenomenon is known as *Intrinsic Alignments (IA)* of galaxies and is thought to originate from the interaction of galaxies with gravitational forces, usually DM, at the time of formation and during their evolution.

The importance of the study of IA is twofold. On the one hand, they are interesting from an astrophysics point of view, since they allow the study of galaxy formation and evolution by analysing galactic IA with the different constituents of the cosmic web, that is, clusters of galaxies, filaments, sheets and voids; as well as their study for different kinds of galaxies, such as central and satellite, early-type and late-type or red and blue and for different redshift scenarios. On the other hand, IA can be seen as a contaminant to the weak lensing signal (Section 1.6), since they can mimic the effects produced by it, that is, the correlation of galaxy shapes between them. Thus, it is necessary to account for IA in cosmological analyses in order not to get biased cosmological constraints.

In the last decade, there has been an increasing interest in the field of galaxy alignments, given the ongoing cosmological galaxy surveys, such as Euclid, LSST and DESI. These are going to constrain cosmological models at unprecedented precisions using, amongst other techniques, weak gravitational lensing. For that reason, the field has to advance in terms of observations and modelling of IA to maximally mitigate biases in cosmological parameters.

In the following, we will briefly describe the IA for different types of galaxies, central and satellites (Section 1.8.1) and early and late-type (Section 1.8.2). Next, we will review the most popular list of models proposed to explain this phenomenon (Section 1.8.3). Finally, we will summarise their impact on cosmological analyses (Section 1.8.4).

1.8.1 Alignments for central and satellite galaxies

Galaxies are expected to reside in DM halos (see Section 1.4.2). Depending on the mass of the halos, they can host one or more galaxies. In the latter case, we can distinguish between central and satellite galaxies. On the one hand, central galaxies are characterised by being the most massive and luminous object in the DM halo, while also being closer to its centre. On the other hand, satellite galaxies are usually fainter and smaller than central ones and are distributed around the centre of the halo. For that reason, the shapes of central galaxies have been easier to measure than those of satellite galaxies, given the technical problems of measuring shapes for faint galaxies, since the light that we observe from them is dimmer and is usually confused with the light from other neighbouring galaxies.

From the study of the alignment of central galaxies with their host halos, it has been established that they tend to align with them, as traced by the satellite galaxies of the halo. This was first detected by Sastry [185] and subsequently confirmed for individual galaxies in Brainerd [24] and in clusters of galaxies in Binggeli [16]. This has also been confirmed in simulations by Wang et al. [216].

In the case of satellite galaxies, it has been observationally established that they are randomly oriented with respect to their host halo, both in the case of groups and clusters of galaxies ([110, 187, 40, 195]). Nevertheless, this is not supported by N-body and hydrodynamic simulations, where there exists radial alignment of satellite galaxies with the centre of the halo, specially important within its virial radius in the case of N-body simulations ([132, 71, 130, 128, 168]) and which increases with redshift and mass of the halo in the case of hydrodynamic simulations ([129, 208]). This difference between alignments in observations and simulations may be explained by the fact that satellite galaxies are aligned via the tidal torquing mechanism [118] (see Section 1.8.3).

1.8.2 Alignments for early and late-type galaxies

In contrast with the alignment characterization in the previous Section 1.8.1, in which one considers the environment of galaxies, one can also study the alignment of galaxy samples that are usually employed in weak lensing analyses. These samples are usually selected to have numerous objects, large contiguous area and high number densities, so that the statistics used to constrain cosmological parameters can be quantified with high accuracy. A common approach to characterize the IA in these samples is to split them in early and late-type galaxies.

On the one hand, early-type galaxies have an elliptical shape and are formed by old and low mass stars. Their star formation rate is low and they usually have redder colours. Observations indicate that these type of galaxies align radially with their neighbouring LSS ([149, 100, 164, 119, 123, 122, 221]) and are usually well described by linear alignment models (see Section 1.8.3).

On the other hand, late-type galaxies have a spiral shape and are formed by a rotating disk and a central part, known as bulge, with a high concentration of stars. Their star formation rate is high and they tend to have bluer colours. The alignment of these kinds of galaxies is commonly linked to their angular momentum via the tidal torquing mechanism. Up to date, there is no clear detection of galaxy IA of the shapes of late-type galaxies, although there are some studies indicating upper limits on these measurements, such as Hirata et al. [100], Mandelbaum et al. [148] and Johnston et al. [123].

1.8.3 Modelling

In this section, we review the main models that are expected to describe galaxy IA. These are derived by the physical mechanism that is thought to source the alignments. We can mainly divide these mechanisms in linear alignment, quadratic alignment and halo alignments.

Linear Alignment (LA)

This model works under the assumption that the DM halo, in which the galaxy resides, is tidally distorted by the gravitational field exerted by the surrounding LSS. The stellar content of the galaxy follows this distortion at the time of its formation and/or during its evolution, being tidally aligned with its host DM halo. This mechanism is thought to explain IA for elliptical galaxies at linear scales.

Following Hirata and Seljak [101], the intrinsic shear of an object can be described as:

$$\gamma^I = -\frac{\bar{C}_1}{4\pi G}\left(\Delta_x^2 - \Delta_y^2, \Delta_x\Delta_y 2\right) S[\psi_P], \tag{1.117}$$

where $\bar{C}_1 = 5x10^{-14} M_\odot^{-1} h^{-2}\mathrm{Mpc}^3$ is a normalization constant whose value was set by Brown et al. [26] for low-redshift IA measurements in SuperCOSMOS [94], S acts as a smoothing filter for ψ_P, which is the Newtonian potential at the time of galaxy formation, and Δ is the comoving derivative.

It is not clear whether IA are set at the time of galaxy formation or evolve with galaxy evolution. As a consequence, the amplitude of the alignment (A_1) can be defined in various ways. If IA are set at the time of galaxy formation, A_1 can be described as:

$$A_1(z) = -\frac{\bar{C}_1 \rho_{m,0}\,(1+z)}{\bar{D}(z)}, \tag{1.118}$$

where $\rho_{m,0}$ is the matter density field today, $\bar{D}(z) = (1+z)\,D(z)$ and $D(z)$ is the growth factor.

Another option includes the evolution of the IA caused by galaxy evolution processes, in which case the alignment is set at z_{IA} and evolves with z:

$$A_1(z) = -\bar{C}_1 \rho_{m,0}\,(1+z_{\mathrm{IA}})\,\frac{D(z_{\mathrm{IA}})}{D(z)} \tag{1.119}$$

Finally, in the case where the amplitude is considered to be set at the redshift at which the observation is performed, the amplitude of the alignment is:

$$A_1(z) = -\bar{C}_1 \rho_{m,0} (1+z), \tag{1.120}$$

In any of the cases, the intrinsic-intrinsic and matter-intrinsic IA power spectra for the LA model are related to the linear matter power spectrum ($P_{\delta\delta}^{\rm lin}$) by:

$$P_{II}(k,z) = A_1^2(z) P_{\delta\delta}^{\rm lin}(k,z) \tag{1.121}$$

and

$$P_{\delta I}(k,z) = A_1(z) P_{\delta\delta}^{\rm lin}(k,z) \tag{1.122}$$

Non-Linear Alignment (NLA)

The Non-Linear Alignment (NLA) model is an extension of the LA model, where the linear matter power spectrum in eq. 1.121 and 1.122 is substituted by the non-linear matter power spectrum ($P_{\delta\delta}$). This is an empirical model used vastly in literature that allows to reproduce the IA of red galaxies on smaller scales than the LA model.

Tidal alignment tidal torquing (TATT)

The alignment of disc galaxies is expected to be governed by their angular momentum. First, the host halo is distorted via tidal shearing from the LSS. Later, the disc galaxy angular momentum axis is aligned with that of the host halo, generating quadratic alignments.

This model is called Tidal Alignment Tidal Torquing (TATT) and is an expansion of the NLA model to second order. This expansion is expressed as a function of the tidal field s and the matter overdensity δ_m:

$$\gamma_{ij}^I = C_1 s_{ij} + C_{1\delta} \delta_m s_{ij} + C_2 \sum_k s_{ik} skj + ..., \tag{1.123}$$

where both γ_{ij}^I and s_{ij} are 3x3 tensors. The first term of the expansion in eq. 1.123 corresponds to the LA model. The second term corresponds to the so-called *density-weighting*, given that alignment can only be observed at the positions where galaxies reside. The third term accounts for the *tidal torquing*.

The II and δI power spectra [1] in the case of TATT are [18]:

$$\begin{aligned} P_{II}(k,\mu_k) =& C_1^2 p(\hat{k})^2 P_{\delta\delta}(k) + 2C_1 C_{1\delta} p(\hat{k}) \left[A_{0|0E}(k,\mu_k) + C_{0|0E}(k,\mu_k)\right] \\ &+ C_{1\delta}^2 A_{0E|0E}(k,\mu_k) + C_2^2 A_{E2|E2}(k,\mu_k) \\ &+ 2C_1 C_2 p(\hat{k}) \left[A_{0|E2}(k,\mu_k) + B_{0|E2}(k,\mu_k)\right] + 2C_{1\delta} C_2 D_{0E|E2}(k,\mu_k), \end{aligned} \tag{1.124}$$

and

$$\begin{aligned} P_{\delta I}(k,\mu_k) =& C_1 p(\hat{k}) P_{\delta\delta}(k) + C_{1\delta} \left[A_{0|0E}(k,\mu_k) + C_{0|0E}(k,\mu_k)\right] \\ &+ C_2 \left[A_{0|E2}(k,\mu_k) + B_{0|E2}(k,\mu_k)\right], \end{aligned} \tag{1.125}$$

where $p(\hat{k})$ is a projection operator, $p(\hat{k}) = 1 - \hat{k}_z^2$, the terms $A_{0|0E}$, $C_{0|0E}$, $A_{0E|0E}$, $A_{E2|E2}$, $A_{0|E2}$, $B_{0|E2}$, $D_{0E|E2}$ are defined in Blazek et al. [18] and the amplitudes C_1, $C_{1\delta}$ and C_2 are characterised by:

$$C_1 = -A_1 \frac{\bar{C}_1 \rho_{\rm crit} \Omega_m}{D(z)}, \tag{1.126}$$

$$C_{1\delta} = -A_{1\delta} \frac{\bar{C}_1 \rho_{\rm crit} \Omega_m}{D(z)}, \tag{1.127}$$

$$C_2 = A_2 \frac{5\bar{C}_1 \rho_{\rm crit} \Omega_m}{D(z)^2}, \tag{1.128}$$

where C_1 is the amplitude that appears in the LA/NLA models, $C_{1\delta}$ is the amplitude of the density weighting parameter and is related to C_1 by $C_{1\delta} = b_1 C_1$, with b_1 the linear bias, and C_2 is the amplitude of the tidal torquing.

Halo model

The physical mechanism for the halo model for IA builds from the halo model for GC [47]. In this model for IA, proposed by Schneider et al. [186], the stellar content of the galaxies and their DM halos share the same dynamical properties, leading to their alignment. In the case of the satellite galaxies, they are radially aligned with respect to the centre of their halo, while for central galaxies, the alignment happens with central galaxies from other halos.

[1] They correspond to P_{EE} and $P_{\delta E}$, since there is also a not null P_{BB} contribution in the TATT model, although this is expected to be subdominant.

Thus, it is a common approach to express the power spectra of this model as a function of a one halo term, corresponding to small-scale correlations between satellite galaxies, and a two halo term, corresponding to the alignment of the central galaxies, such that:

$$P_{II}(k,z) = P_{II}^{1\mathrm{h}}(k,z) + P_{II}^{2\mathrm{h}}(k,z) \tag{1.129}$$

and

$$P_{\delta I}(k,z) = P_{\delta I}^{1\mathrm{h}}(k,z) + P_{\delta I}^{2\mathrm{h}}(k,z) \tag{1.130}$$

In the case of the two halo term contributions, if we neglect the contribution of satellite galaxies, we can split the populations in red and blue galaxies, which are expected to have different alignments, so that

$$P_{II}^{2h}(k,z) = \left(f_{\mathrm{cen}}^{\mathrm{red}}\right)^2 P_{II}^{2\mathrm{h,red}}(k,z) + \left(f_{\mathrm{cen}}^{\mathrm{blue}}\right)^2 P_{II}^{2\mathrm{h,blue}}(k,z), \tag{1.131}$$

$$P_{\delta I}^{2h}(k,z) = f_{\mathrm{cen}}^{\mathrm{red}} P_{\delta I}^{2\mathrm{h,red}}(k,z) + f_{\mathrm{cen}}^{\mathrm{blue}} P_{\delta I}^{2\mathrm{h,blue}}(k,z), \tag{1.132}$$

where $f_{\mathrm{cen}}^{\mathrm{red/blue}}$ are the fraction of red/blue central galaxies.

In the case of the one halo term, driven by satellite galaxies, the P_{II} and the $P_{\delta I}$ are expressed by

$$P_{II}^{1h}(k,z) = \int dM n(M) f_s^2(z) \frac{\langle N_S\,(N_S-1) \mid M\rangle}{\bar{n}_s^2(z)} \left|\hat{\gamma}_s^I(k \mid M)\right|^2, \tag{1.133}$$

$$P_{\delta I}^{1h}(k,z) = \int dM n(M) \frac{M}{\bar{\rho}_m} f_s^2(z) \frac{\langle N_S \mid M\rangle}{\bar{n}_s(z)} \left|\hat{\gamma}_s^I(k \mid M)\right| u(k,M), \tag{1.134}$$

where M is the mass of the halo, $n(M)$ is the halo mass function, f_s is the fraction of satellite galaxies, $\langle N_S \mid M\rangle$ is the halo occupation distribution of satellite galaxies, $\bar{n}_s$ is the mean number density of satellites, $\hat{\gamma}_s^I$ is the average of the density-weighted satellite intrinsic ellipticity and $u(k,M)$ is the normalised matter density profile. Thus, the strength of the alignment of satellite galaxies strongly depends on the mass of the halo and the distance to its centre.

The halo model may apply to elliptical galaxies for smaller scales than the linear alignment mechanism.

1.8.4 Cosmological applications

Intrinsic alignments act as a contaminant of the weak lensing and the galaxy-galaxy lensing signal, affecting the ellipticity-ellipticity and the position-ellipticity correlations, respectively.

In the first case, when correlating two galaxy ellipticities, terms from both the shear and the intrinsic ellipticities arise:

$$\left\langle \varepsilon_i \varepsilon_j \right\rangle = \left\langle \gamma_i \gamma_j \right\rangle + \left\langle \varepsilon_i^I \varepsilon_j^I \right\rangle + \left\langle \gamma_i \varepsilon_j^I \right\rangle + \left\langle \varepsilon_i^I \gamma_j \right\rangle , \tag{1.135}$$

where ε is the observed galaxy ellipticity, γ is the ellipticity from the shear and ε^I is the intrinsic ellipticity. The left-hand side of eq. 1.135 is the one we observe. The first right-hand side term corresponds to the cosmic shear contribution, commonly referred to as GG, the second term is the intrinsic-intrinsic contribution, known as II, and the third and fourth terms correspond to the shear-intrinsic and intrinsic-shear contributions, known as GI. Note that one of the GI terms is expected to be null, since the shear of a foreground galaxy is not correlated with the IA of a source galaxy. The subscripts i and j indicate the objects being correlated.

In the second case, the galaxy position (represented as the density field n) correlates with both the shear and the intrinsic ellipticity of another galaxy:

$$\left\langle n_i \varepsilon_j \right\rangle = \left\langle n_i \gamma_j \right\rangle + \left\langle n_i \varepsilon_j^I \right\rangle , \tag{1.136}$$

where the first term indicates the correlation between the galaxy position and the shear, known as gG, and the second term represents the correlation between the galaxy position and the intrinsic ellipticity, known as gI. Here, we have neglected the contribution coming from magnification.

If the intrinsic ellipticities were randomly oriented, only the GG term in eq. 1.135 and the gG term in eq. 1.136 would be not null. However, since this is not the case, one has to expect contamination from the galaxy IA to our observed ellipticity-ellipticity and position-ellipticity correlators. On the one hand, concerning eq. 1.135, given the fact that IA are a local effect, the II term will be negligible when correlating objects with a significant difference in redshift. Thus, one is able to ignore this II term if studying samples well separated in redshift. Unfortunately, this is not the case for the GI term, since this indicates the correlation of the gravitational shear, which is not a local effect, and the IA. On the other hand, regarding eq. 1.136, the gI term is also supposed to vanish for galaxies separated in redshift. Fig. 1.6 depicts all the terms discussed in eq. 1.135 and eq. 1.136.

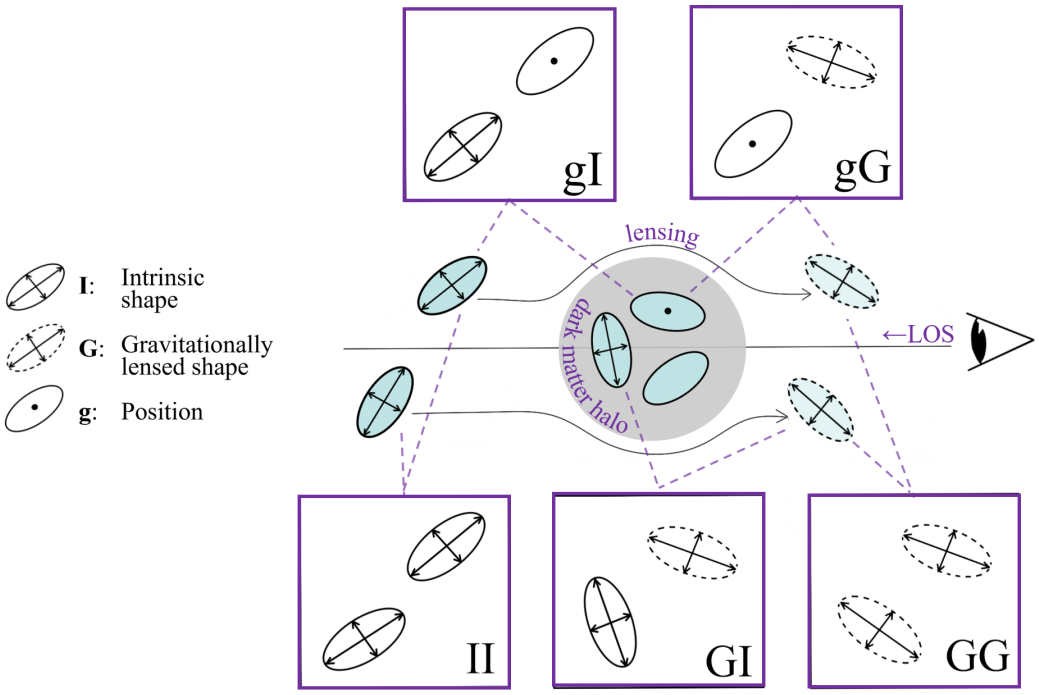

Fig. 1.6 Diagram of the different terms present in the shape-shape (eq. 1.135) and position-shape (eq. 1.136) correlations. This figure has been extracted from [137].

Cosmologists have to deal with the contamination produced by IA to weak lensing and galaxy-galaxy lensing when constraining cosmological parameters (see Section 1.9). A common approach to perform these constraints is to compute the likelihood (L) of a set of parameters (θ) for a given dataset (D), covariance (C) and model (M):

$$-2\ln L(\theta) = \sum [D - M(\theta)]\, C^{-1}\, [D - M(\theta)]^T . \tag{1.137}$$

The modelling of the data has to be well understood in order to avoid biasing our parameters of interest θ. Nevertheless, the modelling of galaxy IA is still incomplete. Moreover, the addition of extra parameters, usually called *nuisance parameters*, so as to better fit the data to the model, can increase the error bars associated to the parameters θ. Thus, IA of galaxies may lead to both biases in cosmological parameter constraints and the increase of their error bars.

There are different approaches to reduce the impact of IA on cosmological analyses. These are related with the terms in eq. 1.137, that is, the model, the data and the covariance. In the first case, a complete model of IA will lead to unbiased cosmological constraints (in the absence of other non IA systematic effects). Moreover, if the IA parameters are related with the cosmological parameters, it would be even possible to reduce the error bars in the

latter, since extra information from IA would be beneficial. In the second case, the data vector can be modified in order to remove the IA contributions, in a method which is known as *nulling*. Nevertheless, nulling usually involves losing part of the information in which we are interested, leading to larger error bars in the cosmological parameters. Finally, in the third case, the covariance matrix may be modified, increasing its errors and performing a *path integral* which marginalises over a non-parameter functional effect hard to model [127].

Finally, one can also combine different probes, such as weak lensing and galaxy-galaxy lensing, to reduce the impact of cosmological parameter biases due to IA. This method is known as *self-calibration* and takes advantage from the different sensitivity to IA by different estimators, internally calibrating the nuisance parameters [117, 183].

Fig. 1.7 shows an example of the contour plots one might obtain for the different cases we just described for parameters A and B, which are marginalised over a larger parameter space. The fiducial values are the ones represented by the black cross. The black contour accounts for the case when there are no systematics, recovering the unbiased parameters A and B. The red contour accounts for the case where there are correctly modelled systematics, with correct strong priors on the nuisance parameters, which are also dependent on the parameters of interest. The blue contour represents the case when the systematics are incorrectly modelled, so that biased parameters are obtained. Last, the green contour accounts for the presence of a systematic, which has been correctly modelled (but does not depend on the parameters of interest) or nulling or functional marginalisation have been applied.

1.9 Parameter inference

As seen in the previous Section 1.8.4, parameter inference is a crucial aspect in cosmological analyses, where some cosmological observations are fitted to a given theoretical model, which depends on a set of cosmological parameters. In this section, we review the basis of cosmological parameter inference, focusing on the Bayesian statistics framework and on how the parameter space is sampled.

1.9.1 Bayesian statistics

The Bayesian statistics framework is based on the Bayes' Theorem, which provides the probability distribution of a set of parameters (θ), given a theoretical model (M) and some observed data (D):

$$P(\theta|D,M) = \frac{L(D|\theta,M)\Pi(\theta|M)}{Z(D|M)}, \tag{1.138}$$

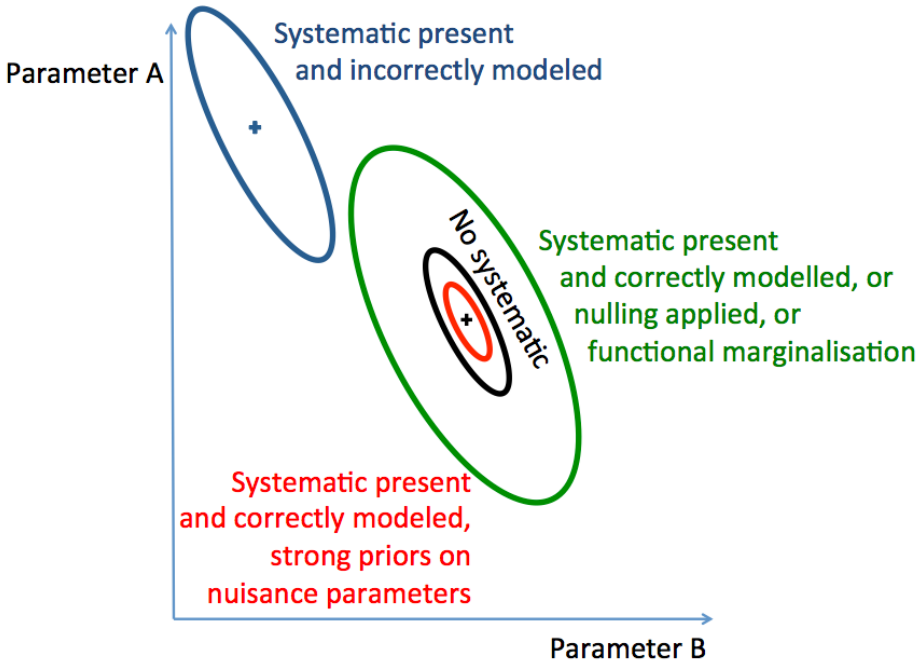

Fig. 1.7 Contour plots for the parameters A and B, marginalised over a larger parameter space. Different contour plots indicate the presence and treatment of systematic effects. Black: No systematics and recovery of unbiased parameters. Red: Correctly modelled systematics, which depend on the parameters of interests, allowing to reduce the error bars on the recovered parameters. Blue: Systematics incorrectly modelled and biased parameters A and B. Green: Systematics present but correctly modelled or nulling or functional marginalisation applied. This figure has been extracted from [118].

where $P(\theta|D,M)$ is called the posterior distribution, $L(D|\theta,M)$ is the likelihood function (which was already introduced in eq. 1.137), $\Pi(\theta|M)$ is the prior distribution and $Z(D|M)$ is called the evidence. Bayesian statistics work under the assumption that the set of parameters (θ), as long as they are treated as random variables, have probability distributions associated to them.

The likelihood function indicates the probability of the observations, given the set of parameters θ and the model M, so that it measures how compatible is the data with the model. The likelihood is usually expressed as a multivariate Gaussian distribution of n dimensions:

$$L(D|\theta,M) = \frac{1}{(2\pi)^n\sqrt{|C|}} e^{-1/2(D-M(\theta))C^{-1}(D-M(\theta))^T}, \tag{1.139}$$

where n is the number of data points of D, $|C|$ is the determinant of the covariance matrix and $M(\theta)$ is the model evaluated at some set of θ values, which is usually referred to as the theoretical data vector.

The prior distribution expresses the probability of the parameters θ given a model M. The priors may also take into account external information from other observed or simulated data, or limitations to theory, in order to set its constraints. Priors are usually employed to set constraints on the possible values of the parameters θ. Typical examples of priors are Gaussian priors, where a given parameter θ is constrained around a mean value with a given standard deviation, or flat priors, where a minimum and maximum values are allowed for a given parameter in the set θ.

Finally, the evidence yields the probability of measuring the data given the model M. It is also usually referred to as *marginal likelihood* and acts as a normalization of the posterior distribution.

1.9.2 Sampling the posterior distribution

If the case of study is to infer the parameters θ of the model M that best fits our data D, it is common to get rid of the evidence Z, since it only acts as a normalization factor. This way, the posterior distribution (eq. 1.139) can be expressed as:

$$P(\theta|D,M) \propto L(D|\theta,M)\Pi(\theta|M). \tag{1.140}$$

In order to identify the best-fitting values and obtain a posterior distribution sampled across the entire domain, the parameter space needs to be explored over a sufficiently wide range. To do this, it is necessary to employ numerical tools that randomly sample the

posterior distribution. The most popular option in cosmological analyses is to use Monte-Carlo Markov Chain (MCMC) algorithms. On the one hand, Monte Carlo methods generate random samples that allow to explore complex probability distributions. On the other hand, a Markov chain is a sequence of events where the probability of obtaining a random variable at a certain step only depends on the value of that random variable in the previous step, and not on the whole history of the sequence. Furthermore, the elements of a Markov chain are selected, after some specific number of steps (called *burn-in phase*), from the region close to the maximum of the likelihood.

A simple case of MCMC can be explained using an iterative process, where the set of parameters θ are evaluated. The algorithm begins with an initial guess of the set of parameters that may fit the data. At the next iteration, a new set of parameters is proposed and the fit to the data is again evaluated. The goal of this iterative process is to sample the posterior distribution and maximize the likelihood function. At each iteration, the new set of parameters can be accepted, creating a new set of values in the chain, or rejected, in which case the values of the chain are still the ones from the previous iteration. The acceptance or rejection depends on a selection criterion, which is usually related to the maximization of the likelihood function. This selection criterion changes as a function of the sampling algorithm employed, being the ones used in this thesis the Metropolis-Hastings and the nested sampling algorithms, which will be described below. The convergence of the chain is usually determined by the fact that its values do not change with time and the elements of the chain are selected from the region close to the maximum of likelihood, although it is hard to define a quantitative measure of convergence. In Section 4.4.6, we describe the convergence criterion used for the Metropolis-Hastings case. Finally, the points of the chain define the posterior distribution described in eq. 1.140.

Metropolis-Hastings

The Metropolis-Hastings algorithm [158, 96] is one of the most popular algorithms to sample the posterior distribution. The selection criterion for the sampling is based on the following ratio, called *acceptance ratio*:

$$a = \frac{p(\theta')}{p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}, \tag{1.141}$$

where p is the posterior distribution and $q(\theta'|\theta)$ is the *proposal distribution*, which represents the conditional probability of moving to a new set of parameter values θ' from the current set of values θ. The proposal distribution q is usually defined as a Gaussian distribution. At each iteration, a realization of θ' is drawn from $q(\theta'|\theta)$ and the acceptance

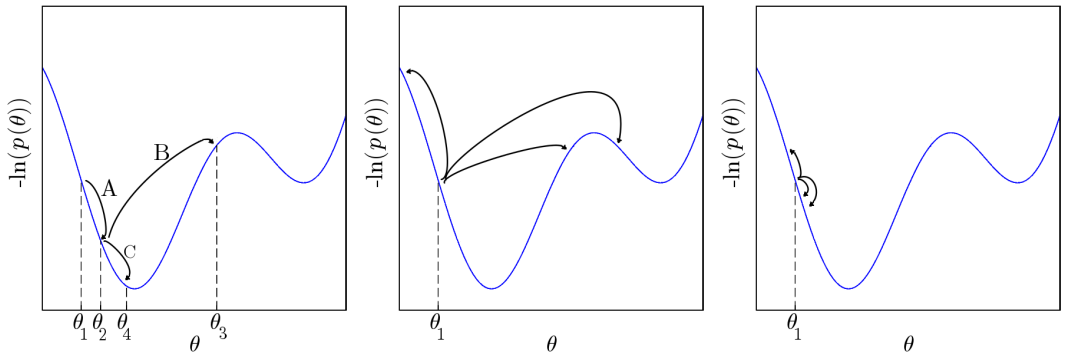

Fig. 1.8 Sampling in a Metropolis-Hastings algorithm. Left: Example of a chain that starts at θ_1. The proposal at θ_2 is accepted (step A). The proposal at θ_3 is rejected (step B), given that it is an unlikely step. The proposal at θ_4 is accepted. Central: Example of a chain with too broad moves in the parameter space. All the steps are rejected because they are unlikely, and the chain is stuck. Right: Example of a chain with too narrow moves. The parameter space is sampled very slowly. Fig. extracted from [142]

ratio in eq. 1.141 is computed. If $a > 1$, θ' is accepted and becomes a new member of the chain, otherwise it is accepted with a probability of a. If it is not accepted, the current θ is repeated in the chain. Note that this algorithm satisfies the two properties that a Markov chain follows: the independence on the number of steps close to the maximum of the likelihood (after the burn-in phase) and that each iteration only depends on the previous one.

Fig. 1.8 shows an example of how the Metropolis-Hastings algorithm works for different cases. The plot on the left shows an example that starts at the parameter value θ_1. The value θ_2 is proposed and accepted (step A), while the value θ_3 is proposed and rejected (step B), given that it involves a decrease in the log-posterior. Finally, the value θ_4 is accepted. The resulting chain is then $\{\theta_1, \theta_2, \theta_2, \theta_4\}$. The central plot shows an example of what can happen if the proposal distribution allows for large steps in the parameter space. In this case, the chain gets stuck since the proposals are unlikely. The right-hand side plot shows the opposite case of the central plot, where the proposal distribution allows for very narrow steps and the sampling of the parameter space is very slow.

Nested sampling

An alternative choice to sampling the posterior distribution with Metropolis-Hastings algorithms is to use the nested sampling algorithm, developed by Skilling [198]. This algorithm is based on the computation of the evidence Z, which was introduced in eq. 1.138, by expressing it as:

$$Z = \int L dX, \tag{1.142}$$

where L is the likelihood function and $dX = \Pi(\theta)d\theta$. In this equation, the prior and the posterior from eq. 1.138 are normalised to unity. Thus, the evidence is expressed as the integral of the likelihood function over the parameter space weighted by the prior distribution. Even though the nested sampling algorithm was developed to compute the evidence for model comparison studies, the sampling of the posterior distribution is a by-product of this computation.

This algorithm works in the following way:

1. Start with a set of samples uniformly distributed across the parameter space and allowed by the prior distribution. The set of samples in each step are called *live points*.

2. Compute the likelihood for each of the live points.

3. The live point with the lowest likelihood (L_0) is removed and replaced with a new set of live points, uniformly distributed in the parameter space and with higher likelihood than L_0. This reduces the region of the parameter space and explores points closer to the maximum of the likelihood.

4. These steps are repeated until convergence, which can be established when the *live evidence*, computed as:

$$Z_{\text{live}} = \sum L_{\text{live}} X_{\text{live}}, \tag{1.143}$$

   with Z_{live}, L_{live} and X_{live} the evidence, likelihood and number of live points, respectively, does not increase the cumulative evidence from all the previous iterations.

Finally, the posterior distribution is obtained from all the explored points, including the discarded ones, such that each point (i) is assigned a weight:

$$p(\theta) = \frac{w_i L(\theta)}{Z}, \tag{1.144}$$

where the weights are estimated as $w \sim \Delta X$, that is, the shift in the parameter space, weighted by the prior distribution, between two iterations.

Chapter 2

Imaging surveys

Over the last decades, wide-field galaxy surveys, such as the Sloan Digital Sky Survey (SDSS) [220], the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) [49], the Kilo-Degree Survey (KiDS) [58], the Dark Energy Survey (DES) [52] and the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP) [2], amongst others, have provided the community with numerous galaxy images, which allowed to perform a range of cosmological statistical analyses, such as estimating correlation functions for galaxy clustering or cosmic shear, which might help to infer the nature of DE and DM [217].

The latest generation of surveys, with instruments like Euclid [69], the Legacy Survey of Space and Time (LSST) [114] and the Dark Energy Spectroscopic Instrument (DESI) [59] aim at increasing the observed number of objects by an order of magnitude, reaching fainter magnitudes and deeper redshifts with the objective of improving the cosmological parameter constraints, allowing us to disentangle the origin, composition and evolution of the universe.

During this thesis, we have used data from two important imaging surveys, PAUS and Euclid. We describe them and outline their contributions to cosmology in Section 2.1 and Section 2.2, respectively.

2.1 PAUS

The Physics of the Accelerating Universe Survey (PAUS) is a photometric survey conducted at the William Herschel Telescope (WHT) at El Roque de Los Muchachos, in the Canary Islands. It observed the sky for 200 nights between 2015 and 2019. The PAUCamera [165] used a set of 40 narrow-band (NB) filters ranging from 4500Å to 8500Å in steps of 100Å with a Full Width at Half Maximum (FWHM) of 130Å. This exceptional filter arrangement is designed to provide high-precision photo-z, outperforming the precision achievable with

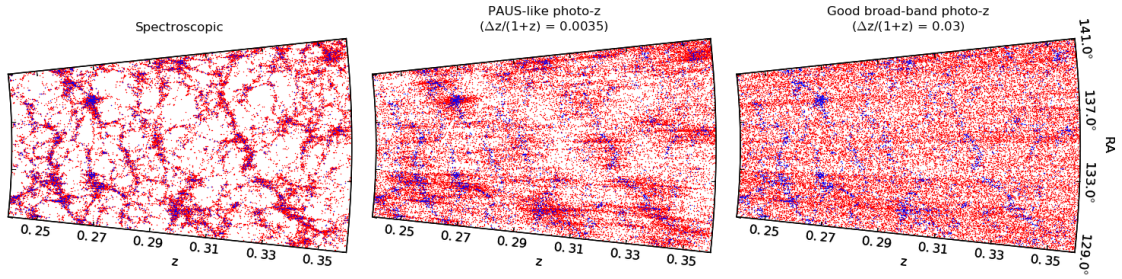

Fig. 2.1 Redshift distribution over R.A of different types of surveys, from left to right, spectroscopic, PAUS-like and BB survey. Note that the accuracy in the photometric redshifts of a PAUS-like survey lies between a spectroscopic and a good BB survey. Figure extracted from [205].

broad-band (BB) observations. In combination with the 40 NB, PAUS observations are complemented with BB photometry provided by CFHTLenS and KiDS, which are also used as reference samples. Fig. 2.1 shows the distribution of redshift as a function of the Right ascension (R.A) for spectroscopic-like (left), PAUS-like (centre) and BB-like (right) redshifts, where one can see that the accuracy of PAUS photometric redshifts lies between both the spectroscopic and the good BB cases.

PAUS targets are comprised of the COSMOS field [190], which is mainly used for calibration and validation processes, the W1, W3 and W4 Wide Fields from the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) [65, 98] and the Galaxy And Mass Assembly (GAMA) G09 field, which overlays with KiDS. The overlap between the W1, W3 and G09 fields with PAUS observations compose the PAUS deep wide fields, which we will refer to as PAUS wide fields for brevity. Fig. 2.2 shows the position on the sky of the PAUS targets. In this thesis, we exclude the W4 field, since the observations made by PAUS in it are rather scarce.

PAUS covers an area of $\sim$43 deg^2 in all 40 NB and $\sim$51 deg^2 with a coverage of at least 30 NB, up to $i_{\rm AB} = 23$ and with $\sim$1.8 million objects observed. The number density of PAUS objects with accurately measured photo-z is $\simeq 3 \times 10^4$ per square degree, making it ideal for some studies, such as the intrinsic alignments of galaxies. This number is large compared to other wide spectroscopic surveys, such as GAMA [62] or VIMOS Public Extragalactic Redshift Survey (VIPERS) [189], with number densities $\sim 10^3$, which are not as deep as PAUS and usually not complete in magnitude. Although spectroscopic surveys can have more precise redshifts and larger area coverage, PAUS regime is useful when we need larger

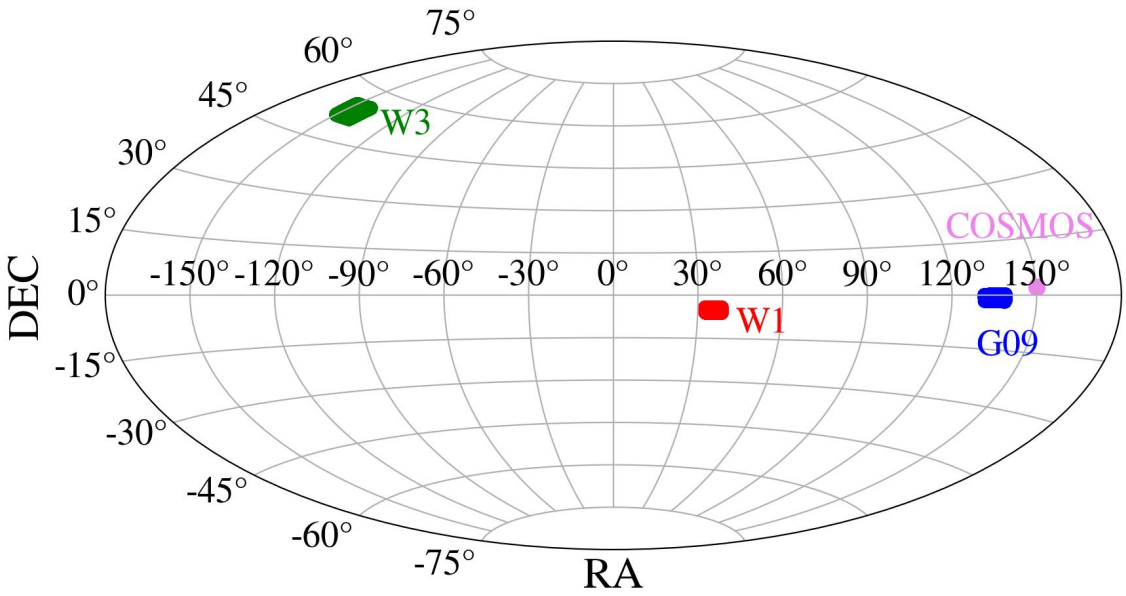

Fig. 2.2 Position on the sky of the W1, W3, G09 and COSMOS fields used in this study.

densities over wide fields, with photo-z that are more accurate than the ones obtained from BB.

Fig. 2.3 shows a comparison of PAUS with other spectroscopic redshift surveys, indicating the area (top) and the number of redshifts (bottom) as a function of the limiting magnitude. PAUS is deeper than GAMA and SDSS, and larger than VIPERS, VIMOS VLT Deep Survey (VVDS) and DEEP2, so that it yields the highest number of redshifts, almost one order of magnitude larger than any of these spectroscopic surveys. Note that this figure corresponds to initial estimations of the PAUS area, the final area being 50deg^2 instead of 100deg^2. Nevertheless, the conclusions from this plots are still valid. The latest generation of spectroscopic surveys, such as DESI, are not included in this plot and, although they observe more area and number of redshifts than PAUS, their number density is still smaller, allowing PAUS to still offer some insights, in particular in the range of non-linear scales.

PAUS fluxes are measured by doing forced photometry over the reference samples. This technique takes the positions and galaxy shapes in the reference samples and measures the NB fluxes at these fixed positions. The objects in the PAUS wide fields are observed an average of 3 times, with average exposure times of 2–3 minutes. These observations are later coadded at the catalogue level, obtaining the flux and its error per object, which will be used for science. For a detailed explanation of this procedure and the PAUS flux calibration, we refer the reader to [193] and [36].

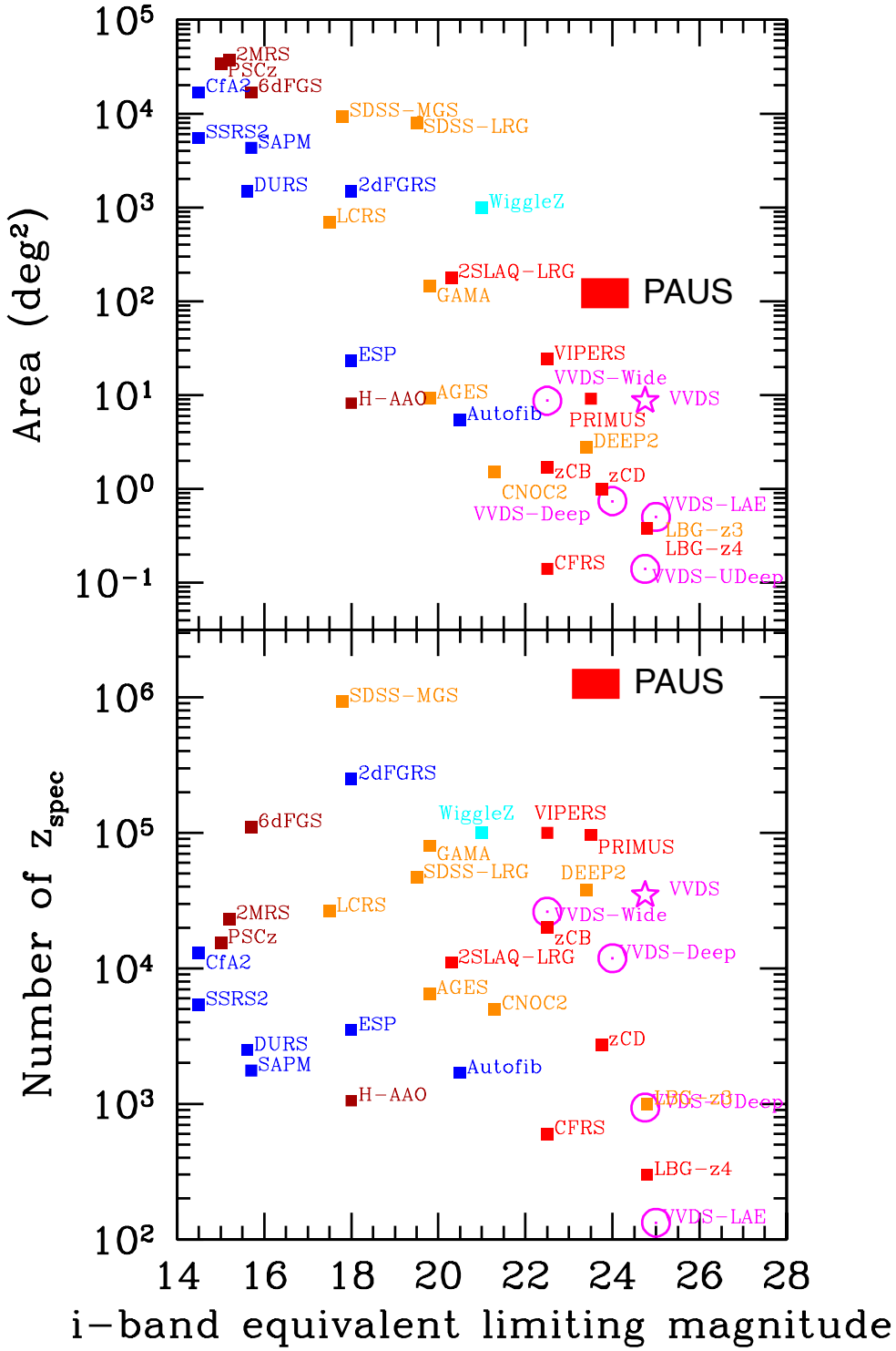

Fig. 2.3 Comparison of PAUS with respect to other surveys. The area (top plot) and the number of redshifts (bottom) versus the i-band limiting magnitude are shown. Fig. extracted from [165].

The treatment of the data is carried out by the PAU data management (PAUdm) team, located at Port d'Informació Científica (PIC). The responsibilities of PAUdm include the storage, data reduction and accessibility of PAUS measurements to its members. The PAU database, where the data and metadata of PAUS measurements and results are stored, is of utmost importance. For a detailed description of the design and responsibilities of PAUdm we refer the reader to [209].

2.2 Euclid

Euclid is a medium-class mission of the European Space Agency (ESA) that will cover around 14000deg^2 at optical and near-infrared wavelengths. Its main science cases are the study of galaxy clustering and weak gravitational lensing, which are primary cosmological probes that allow to study the evolution of the LSS of the universe and the nature of DM and DE.

The origin of Euclid comes from the combination of two different DE mission proposals to ESA in 2007, the Dark Universe Explorer (DUNE) [178] and the Spectroscopic All-Sky Cosmic Explorer (SPACE) [41], which were finally combined to conceive Euclid. In 2011, the main science case [140] was presented and Euclid was approved as an ESA mission in 2012. Since then, the design, assembly and verification of the software and hardware has been performed and Euclid was successfully launched from Cape Canaveral in a Falcon-9 rocket the 1st of July 2023. Its destination was the second Lagrange point of the Sun-Earth system (L2), where it shares location with other important telescopes, such as Gaia [82] and the James Webb Space Telescope (JWST) [175].

One of the main scientific goals of Euclid is to study the accelerated expansion of the universe. This expansion is usually explained by the presence of DE, whose nature is one of the main open questions in cosmology. DE is expected to have a negative equation of state parameter, although we do not know if it can be explained by a cosmological constant or by a dynamical field. Another option is to explain the acceleration of the universe not with DE, but with modified gravity theories. Euclid's aim is to obtain a Figure of Merit (FoM) for DE, defined as the inverse square root of the covariance matrix determinant corresponding to DE parameters, greater than 400, which roughly corresponds to the determination of the equation of state with a precision of 1%. Another key objective for Euclid is to study the composition of the universe. This can be done by analysing the growth of LSS, since this depends on the different components of the universe. In particular, by studying the dimensionless growth

rate, $f_g(z)$, which can be described as:

$$f_g(z) \sim [\Omega_m(z)]^{\gamma_g}, \tag{2.1}$$

where Ω_m is the fraction of matter with respect to the critical density. Euclid aims to determine γ_g with a precision of at least 0.02, which would allow distinguishing between General Relativity and modified gravity scenarios.

In order to achieve these two objectives, Euclid will use two powerful cosmological probes: weak gravitational lensing and galaxy clustering. Both of these probes and their cross-correlation allow studying the expansion of the universe, the growth of the LSS and the relation between baryonic matter and DM. On the one hand, for the case of weak gravitational lensing, Euclid will measure the shapes and the photometric redshifts of around 1.5 billion galaxies. On the other hand, for the galaxy clustering measurements, spectroscopic redshifts for more than 25 millions galaxies, in the range $0.9 < z < 1.8$, will be determined.

Euclid is equipped with two main scientific instruments: the Visible Imaging Instrument (VIS) and the Near Infrared Spectrometer and Photometer (NISP), as depicted in Fig. 2.4.

The VIS is an imager with a Field of View (FoV) of 0.54 deg^2, observing in the I_E passband and covering a wavelength of 530-920 nm, which is the one usually covered by the SED of the majority of galaxies. The limiting sensitivity of the I_E passband is $m_{\rm AB} = 26.7$, which corresponds to a $S/N = 10$ for objects with $m_{\rm AB} = 25.0$, in a 1".3 diameter aperture and a FWHM of 0".3. This instrument allows measuring the shapes required for the weak lensing analyses, besides providing information to estimate photometric redshifts. Considering the high precision at which Euclid operates, understanding the systematics in the measurements is a primordial task. In that sense, the VIS has been designed to provide optimal stability and support efficient calibration. Euclid's large FoV is covered by 36 Charge-Coupled Device (CCD), each of them with 4132x4096 pixels in four quadrants, accounting for 144 different quadrants. Once the pixels are read, all at the same time to reduce the noise, they are sent to the Control and Data Processing Unit (CDPU), where the images are produced, compressed and later sent to Earth.

As for the case of the NISP, it provides multiband photometry and spectroscopy, using a slitless grism, in the wavelength range of 920-2020nm. On the one hand, it observes in three photometric (NISP-P) passbands Y_E (949.9-1212.3 nm), J_E (1167.6-1567.0 nm) and H_E (1521.5-2021.4 nm), complementing the observations of the I_E passband from the VIS. These filters have a designed 5σ depth of $m_{\rm AB} = 24.0$. On the other hand, slitless spectra of objects are obtained with the NISP spectroscopic channel (NISP-S). This allows to measure

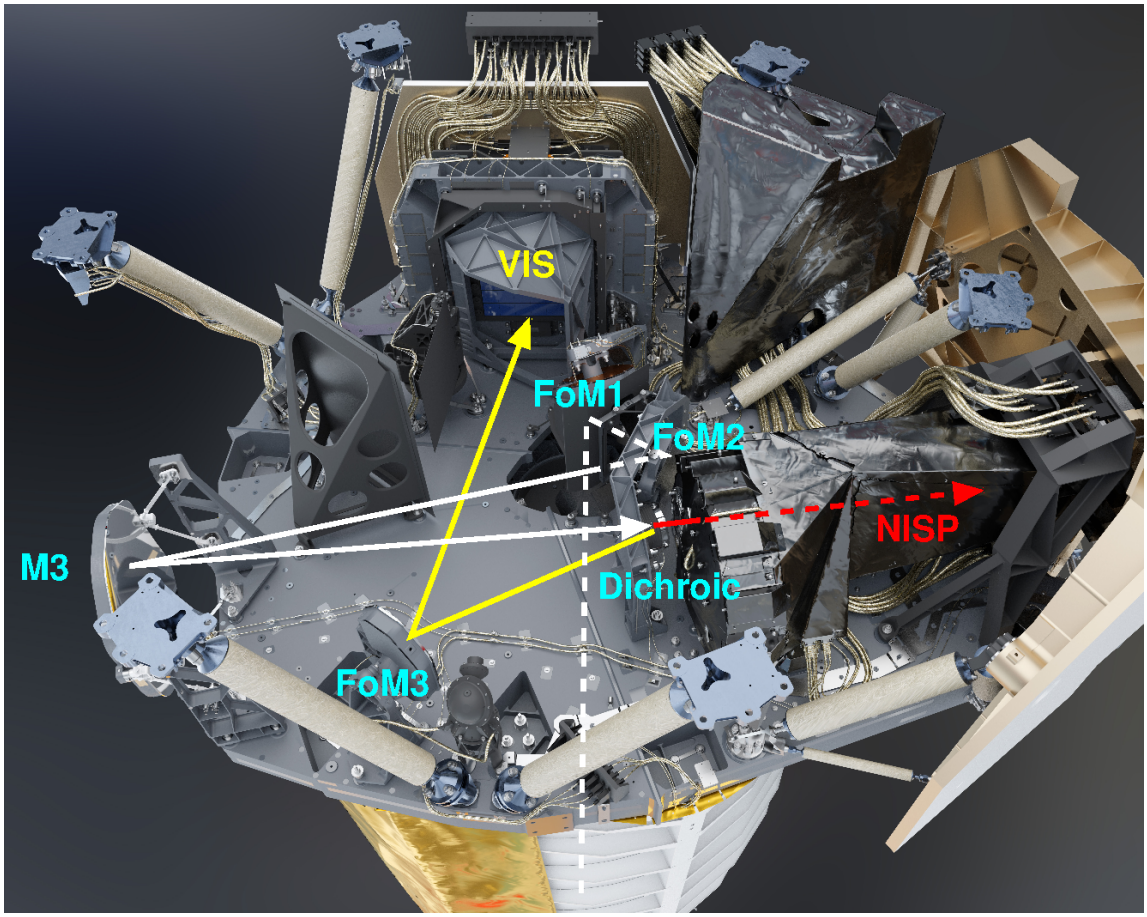

Fig. 2.4 Diagram of the instrument cavity showing the VIS and the NISP. The observation is towards the bottom of the rendering and lines indicate light paths. Fig. extracted from [69].

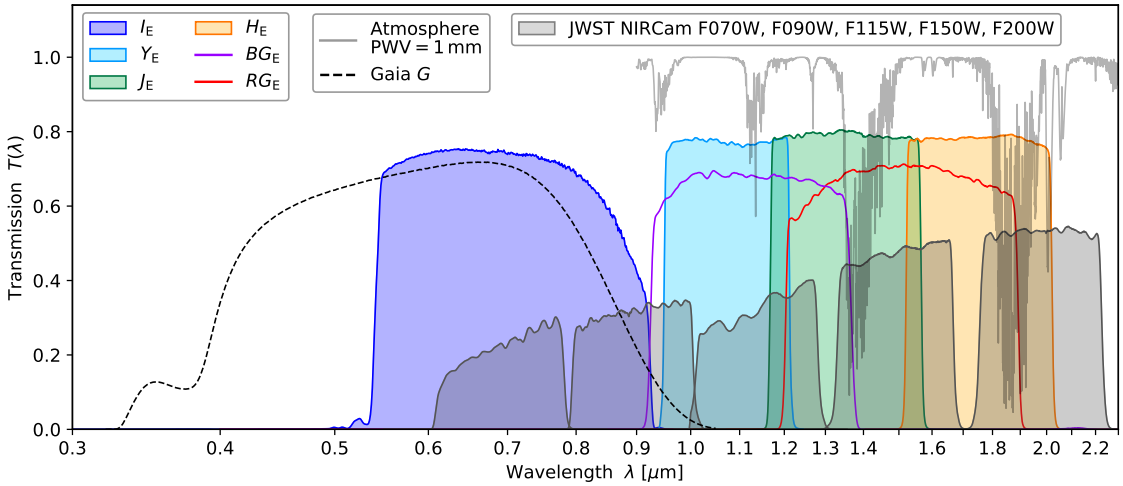

Fig. 2.5 Transmission as a function of wavelength for the Euclid photometric (VIS: I_E, NISP-P: Y_E, J_E, H_E) and spectroscopic (NISP-S: BG_E, RG_E) passbands. As a reference, the atmospheric transmission for a precipitable water vapour level of 1 mm, the Gaia G passband and some JWST NIRCam passbands are also shown. Fig. extracted from [69].

spectra for thousands of objects over the FoV with uniform quality. In the case of the Euclid Wide Survey (EWS), which will be explained later, the redder RG_E passband, covering a wavelength range of 1206-1892 nm, is used. This enables the detection of Hα emitter in the redshift range $z = 0.84 - 1.88$ with a redshift accuracy of $\sigma_z < 0.001(1+z)$. For the case of the Euclid Deep Fields (EDF) and Euclid Auxiliary Fields (EAF), a bluer passband, BG_E, covering a wavelength range of 926-1366 nm allows observing Hα emitters down to redshift $z = 0.41$. Fig. 2.5 shows the transmission as a function of the wavelength λ (in units of μm) of the VIS and the NISP passbands, together with the atmospheric transmission for a precipitable water vapour level of 1 mm. The Gaia G passband and some JWST Near Infrared Camera (NIRCam) passbands are also included for reference.

The survey planning of Euclid includes the EWS and the Euclid Deep Survey (EDS). In addition, some calibration observations are used to validate the performance of the instruments.

The area of the EWS was selected to have minimal contamination from our Galaxy, the objects from the Solar System and the zodiacal background. This leads to four separate regions in the sky, which can be seen in Fig. 2.6, delimited by blue lines. The sum of the area of the four regions accounts for 14679 deg^2. The EWS regions were designed to have an optimal overlap with the ground-based surveys, which are needed to model the Point Spread Function (PSF) and to assist in the photometric redshift estimation.

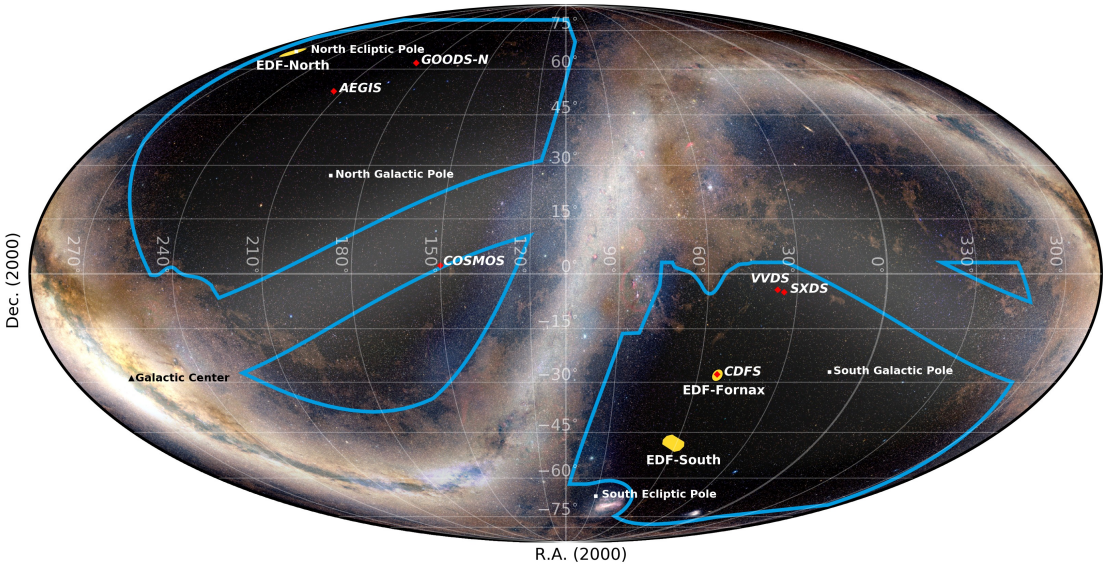

Fig. 2.6 Observed regions by Euclid. Blue delimited area correspond to the EWS, which account for 14679 deg^2. Yellow regions correspond to the EDF and red marks to the EAF. Fig. extracted from [69].

The EDS is composed of the EDF and the EAF. The goal of the EDS is to obtain deeper data than the EWS, by increasing the SNR by a factor of 6, which translates into reaching about 2 magnitudes deeper. This allows the characterisation of the typical EWS source population, which results in a spectroscopic sample with a 99% completeness and 99% purity, together with morphological measurements of galaxies, that enable the characterization of the systematic effects on the shape measurements. In addition, the EDS assists on the calibration of the telescope, the cameras and the instruments associated to them. Euclid has defined three EDF: the EDF North (EDF-N), the EDF South (EDF-S) and the EDF Fornax (EDF-F), which are depicted in yellow in Fig. 2.6. In contrast, the EAF reach a depth of 4–5 times that of the EWS and help with the calibration of photometric redshifts and the measurement of galaxy shapes in the presence of chromatic PSFs. The EAF are formed by the All-Wavelength Extended Groth Strip International Survey (AEGIS) [56], Chandra Deep Field South (CDFS) [88], COSMOS [191], Great Observatories Origins Deep Survey (GOODS)-North [89], Subaru/XMM-Newton Deep Survey (SXDS) [80] and VVDS [141] fields, which are marked in red in Fig. 2.6.

Chapter 3

Photometric redshift estimation in the PAUS wide fields

3.1 Motivation

In this Chapter 3, we compute photometric redshift estimates for $\sim$1.8 million objects in 51 deg^2 down to $i_{\rm AB} < 23$, with redshifts from 0 to 2. We take advantage of the photometry from the Physics of the Accelerating Universe Survey (PAUS) [165], presented in detail in Section 2.1, which is composed of 40 narrow-band (NB) with a width of $\sim$13nm. With bands one order of magnitude narrower than that of BB, the spectra recovered by PAUS have a resolution between spectroscopic and BB photometric surveys, which allows us to compute photometric redshifts (photo-z) with unprecedented precision. We use a SED template-fitting code called BCNz2 [67], which we will refer to as BCNZ to simplify the notation throughout the Chapter, specifically designed to deal with the 40 NB of PAUS.

Previous studies of photometric redshifts in PAUS have been performed in the COSMOS field with SED template-fitting codes [67, 3] and machine learning algorithms [66, 202, 30], covering an area of $\sim$1.5 deg^2, down to $i_{\rm AB} < 22.5$ in the case of Eriksen et al. [67], Soo et al. [202] and Eriksen et al. [66] and $i_{\rm AB} < 23$ in the case of Alarcon et al. [3] and Cabayol, L. et al. [30]. Here, we compute photometric redshifts for the wide fields from PAUS.

The photometric redshifts we present have already been used for scientific purposes, such as the study of the D4000 spectral break index in PAUS [179], the identification of close galaxy pairs and the determination of their mean mass [91], the effect in the accuracy of the photo-z for different calibration methods in the PAUS NB photometry [193] and the evaluation of the capability prediction of semi-analytical galaxy formation models using the spectroscopic samples presented in this Chapter 3 [150]. In Wittje et al. (in prep.) they

Table 3.1 Area with a minimum coverage of 30 NB, number of objects up to $i_{\rm AB} < 23$, R.A and DEC limits of the PAUS targets.

Field	Area 30 NB (deg^2)	# objects	RA limits (deg)	DEC limits (deg)
W1	12.04	401815	[32.3,38.5]	[-6.1, -4]
G09	15.7	663535	[131,139]	[-1.7, 0.5]
W3	22.64	792664	[209,219.5]	[51.4, 55.6]
KiDZ-COSMOS [1]	1	13380	[149.5,150.6]	[1.7, 2.7]
Total	51.38	1871394	-	-

simulate the PAUS fluxes using the Flagship simulation [176] and include a comparison between the performance of the photometric redshifts computed from the simulation and the estimates presented in this Chapter 3. Here, we aim to describe the process of obtaining the new photo-z for the deep wide fields of PAUS and study their performance as a function of magnitude, redshift, colour and photometric quality.

The structure of this Chapter 3 is as follows. In Section 3.2, we present the data used to obtain the fluxes of the objects and the spectroscopic information to validate the photometric redshifts. In Section 3.3, we explain the methodology employed to obtain the photo-z, describing the SED template-fitting code we have used and some improvements and adaptations that we implemented to it. In Section 3.4, we present and validate the photo-z, we study its performance as a function of the galaxy colours and we analyse the $p(z)$ distributions.

The content of this Chapter 3 has been mainly extracted from the published paper Navarro-Gironés et al. [162].

3.2 Data

3.2.1 PAUS

We use the PAUS data, described in Section 2.1, in order to extract the NB fluxes needed to compute the photometric redshifts. Table 3.1 shows the fields under study, including the W1, G09, W3 and KiDZ-COSMOS fields, together with the area covered by a minimum of 30NB, the number of objects and the angular limits, described in terms of R.A and DEC.

[1] KiDZ-COSMOS refer to objects from the COSMOS field with spectroscopic information, as will be explained in Section 3.2.3

3.2.2 Broad-band photometry

As explained in Section 2.1, BB photometry is needed to define the positions in PAUS and also to add extra information coming from the BB fluxes. These BB photometric catalogues are also called parent catalogues or reference catalogues, given that they define the samples to be studied by PAUS. The Wide Fields from CFHTLenS and the GAMA G09 field from KiDS constitute the parent catalogues used by PAUS.

The CFHTLS-Wide uses the wide-field imager, MEGACAM [22], installed at Mauna Kea. Its field of view covers 1 deg^2 in the *ugriz* (*y* after the *i* filter broke) BB filters up to a 5σ limiting magnitude of $i_{\rm AB} \approx 25.5$. The CFHTLS-Wide observes in the W1, W2, W3 and W4 fields, covering 157 deg^2. In particular, the CFHTLenS team was formed to conduct weak-lensing studies and the measuring of the galaxy shapes was implemented by the CFHTLenS shape measurement pipeline [159]. The best observing conditions were reserved for the *i*-band, making it the survey detection band. CFHTLenS multi-band photometry was extracted from PSF-homogenised [134] stacks with SExtractor [15] in dual-image mode, yielding high-quality colours for photo-z estimates [99].

KiDS is a wide-field imaging survey that uses OmegaCam [135], which is installed at the European Southern Observatory (ESO) Very Large Telescope (VLT) Survey Telescope [32] at the Paranal observatory. OmegaCam has a field of view of 1 deg^2 in the *ugri* bands, with the *r*-band being the one used in best conditions to enable the precise measurement of galaxy shapes. Matching objects between the fields observed by KiDS and those observed by the VISTA Kilo-degree INfrared Galaxy (VIKING) survey [64], allows the addition of VISTA's five near-infrared BB $ZYJHK_{\rm s}$. KiDS DR4 covers around 1000 deg^2 in the KiDS fields, KiDS-S and KiDS-N. In particular, KiDS-N overlaps with the G09 field of GAMA [62], which we will use in this study. The KiDS multi-band photometry is extracted with the GAaP method [134], which first convolves each image with a shapelet-based kernel to yield a Gaussian PSF and then measures fluxes in Gaussian-weighted elliptical apertures [133].

Fig. 3.1 shows the filter response as a function of the wavelength of the NB used by PAUS (lower panel) and the BB used by both CFHTLenS (top panel) and KiDS (middle panel). The wide coverage in wavelength of both broad and narrow bands used in this study is one key component that allows highly accurate determination of the photometric redshifts. The NB wavelength range ($450-850$nm) overlaps with the *g*, *r*, *i*, *y* and, partially, *z* CFHTLenS BB and by the *g*, *r*, *i* and *Z* KiDS BB, while the rest of the BB widen the wavelength coverage of PAUS NB.

The CFHTLenS *i*-band filter broke and was replaced by a similar filter that was labelled as the *y*-band. For objects that were not measured with the *i*-band (which amount to 6.3%

and 11.5% in the W1 and W3 fields, respectively), we will use y-band measurements and employ the y filter to compute the photometric redshifts. Nevertheless, when studying the performance of the photometric redshifts as a function of the magnitude, we will refer to both cases as $i_{\rm AB}$, that is, not differentiating between i and y magnitudes, which is a good approximation taking into account how similar the transmission curves are for the i and y bands in the top panel of Fig. 3.1. Another important aspect to take into account is that the filter response functions and the wavelength range covered by the i-bands used in CFHTLenS and KiDS are not the same. As a result, the $i_{\rm AB}$ of both systems is defined slightly differently. In order to select a similar population for all fields, we need to find a relation between $i_{\rm AB,\,KiDS}$ and $i_{\rm AB,\,CFHTLenS}$ (the $i_{\rm AB}$ magnitude in the KiDS and the CFHTLenS systems, respectively). This relation is established by studying the number counts (N) of the PAUS wide fields and determining what is the selection cut to be applied at $i_{\rm AB,\,KiDS}$ that corresponds to the same number count at $i_{\rm AB,\,CFHTLenS} = 23$. Fig. 3.2 shows the number counts for the three PAUS wide fields as a function of the magnitude $i_{\rm AB,\,CFHTLenS}$, which from now on we will refer to as $i_{\rm AB}$, if not specified otherwise. We find that applying a cut at $i_{\rm AB,\,KiDS} = 23.1$ and redefining $i_{\rm AB} \equiv i_{\rm AB,\,CFHTLenS} = i_{\rm AB,\,KiDS} - 0.1$ gives reasonably similar number count values between the G09 and the W1 and W3 fields over the whole magnitude range. However, there are still some differences between the G09 counts and those in the W1 and W3 fields that may be due to the different definitions used for the i-band (given that the area of the fields is relatively small, this could also be due to sample variance). This indicates that a comparison of the photo-z performance between fields as a function of $i_{\rm AB}$ is not straightforward, as we can see in Appendix A. In order to justify that the cut we applied at $i_{\rm AB,\,KiDS} = 23.1$ is the most appropriate, we show alternative cuts at $i_{\rm AB,\,KiDS} = 23$ and $i_{\rm AB,\,KiDS} = 23.2$ (redefining $i_{\rm AB} = i_{\rm AB,\,KiDS} - 0.2$ in this case), in dashed lines, where the former cut exhibits lower number counts than the W1 and W3 fields for all $i_{\rm AB}$ and the latter cut presents a higher number count than W1 and W3 at $i_{\rm AB} > 21.5$, which is the region with the largest number of objects.

To perform the analysis, we remove from the reference catalogues the objects with bad quality photometry or those classified as stars. In the case of CFHTLenS, we exclude stars by setting the star_flag $= 0$ and the mask from CFHTLenS less or equal than 1. In the case of KiDS, we set sg_flag $= 0$ and sg2dphot $= 1$ to remove stars. We also apply a series of masks related to the detection band r, which perform a stellar masking, mask due to saturation, trim and account for chip gaps, void mask and asteroids. Fig. 3.3 shows the distribution of galaxies as a function of R.A and DEC in the W1, W3 and G09 fields coloured by the fluctuations in the number density of objects (defined as the ratio between the number of

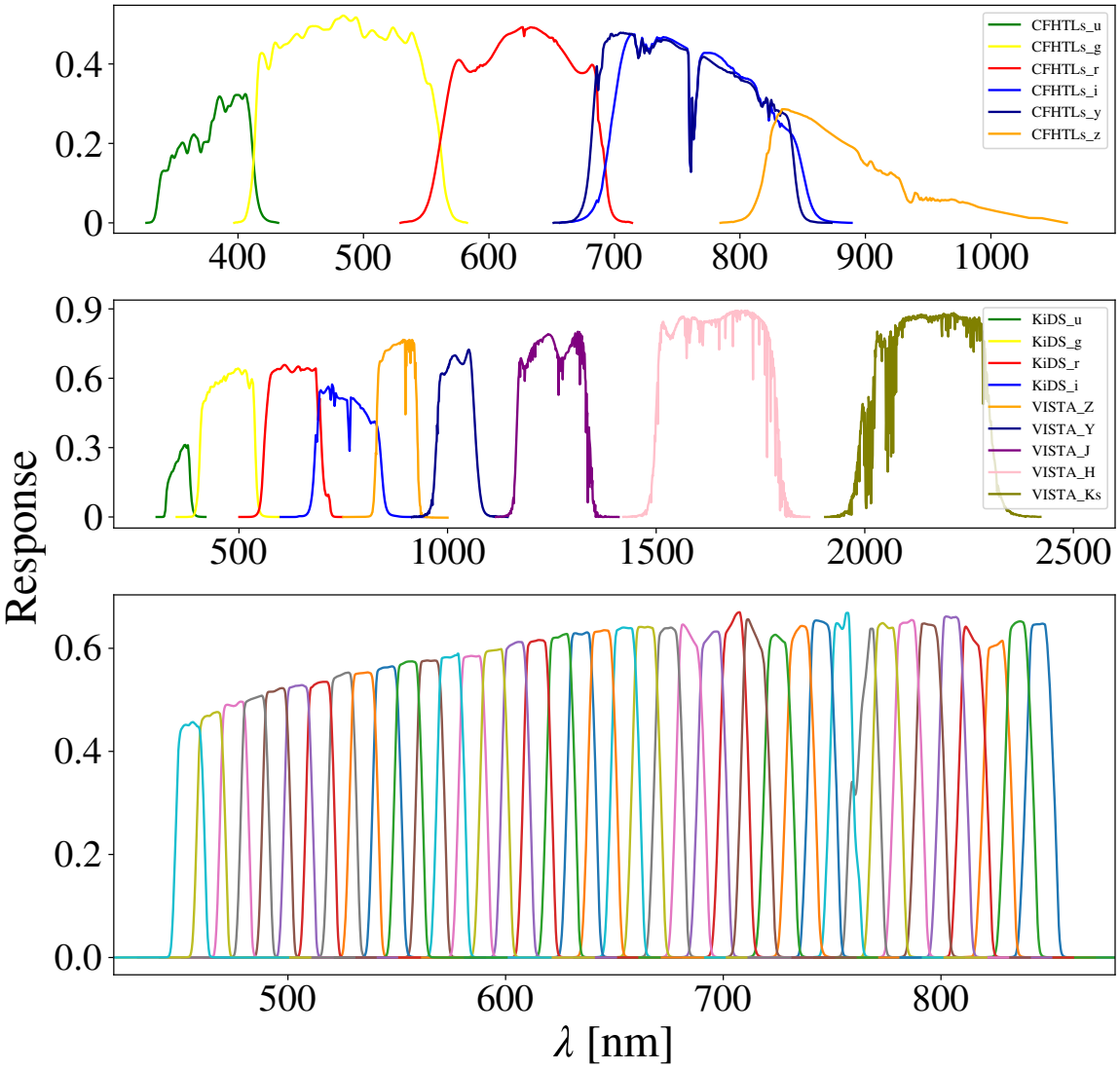

Fig. 3.1 Response of the filters used in this study as a function of the wavelength (nm) for the CFHTLenS (top) and the KiDS (middle) BB and for the PAUS NB (bottom).

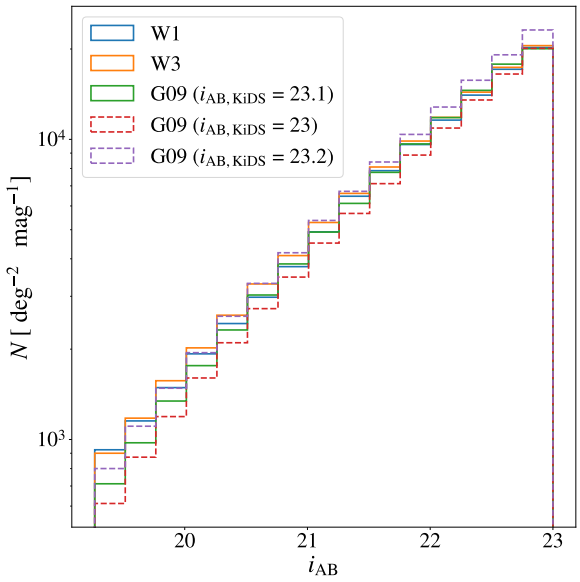

Fig. 3.2 Number counts in the three PAUS wide fields as a function of $i_{\rm AB}$, where $i_{\rm AB} \equiv i_{\rm AB,\,CFHTLenS} = i_{\rm AB,\,KiDS} - 0.1$ for the cut at $i_{\rm AB,\,KiDS} = 23.1$. The fact that the number counts for the G09 field are very similar to those for the W1 and W3 fields over the considered magnitude range, indicates that the limit imposed at $i_{\rm AB,\,KiDS} = 23.1$ in the selection is appropriate. Alternative cuts in the G09 field yielding different number counts than the W1 and W3 fields are shown in dashed lines.

objects and the area), after applying the reference catalogue masks and star flags. Here the fluctuations are computed as $\frac{1}{\sigma}\left(\frac{n}{\mu}-1\right)$, with σ and μ corresponding, respectively, to the σ_{68} and the median value of the number density, n. It is interesting to observe the angular clustering of galaxies through the overdensities (red regions) and voids (bluer regions).

We also apply an extinction correction to the fluxes (ϕ) that takes into account the Milky Way extinction:

$$\phi_{corr} = \phi_{uncorr} \cdot \frac{1}{C_0 \cdot E(B-V)^2 + C_1 \cdot E(B-V) + C_2}, \tag{3.1}$$

where the $E(B-V)$ values are extracted from the [170] dust map and C_0, C_1 and C_2 are band dependent extinction coefficients.

The fluxes from the reference catalogues were already extinction corrected by the CFHTLenS and KiDS teams, while the PAUS fluxes were not. To apply the same extinction method in all bands, we first added the extinction that was already corrected for in the BB and later corrected all NB and BB using the Milky Way extinction model defined in eq. 3.1.

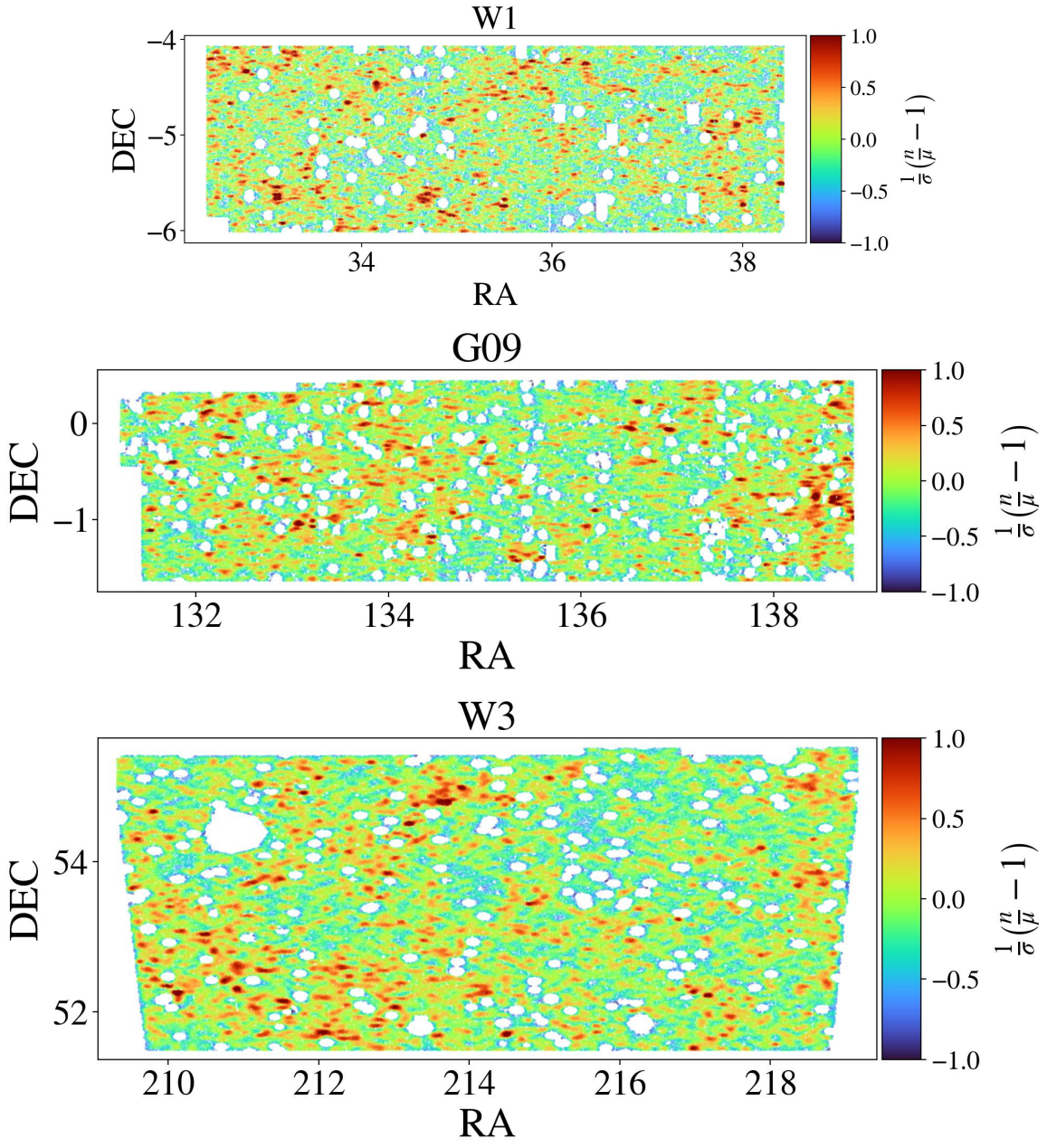

Fig. 3.3 Angular distribution of galaxies in the W1, G09 and W3 fields after applying the reference catalogue star flags and masks, which remove stars and objects not well identified or with bad photometry. The colour code indicates the fluctuations in the number density of objects (n), $\frac{1}{\sigma}\left(\frac{n}{\mu}-1\right)$, such that σ and μ correspond, respectively, to the σ_{68} and the median value of n.

3.2.3 Spectroscopic data

Spectroscopic redshifts (spec-z) are essential for validating the photo-z performance. This validation involves comparing the spectroscopic redshifts, which have greater accuracy, with their photometric counterparts. We refer to the subsamples with spec-z information as validation samples, which are defined after applying the survey mask and the stellar flags mentioned in Section 3.2.2. Due to the wide angular separation of the PAUS wide fields, we need spectroscopic data from different galaxy surveys. The properties of the main spectroscopic redshift surveys included in this analysis are described in the following lines:

1. The Sloan Digital Sky Survey (SDSS) is a wide-area spectroscopic survey conducted at the Apache Point Observatory (APO) and Las Campanas Observatory (LCO). We use data from the DR16 [1], with redshifts ranging from 0 to 1.1 and mag-i limit $\sim$22. We select galaxies based on the flag CLASS == 'GALAXY' and zWarning == 0, indicating that there are not unknown associated problems.

2. The Galaxy And Mass Assembly (GAMA) [107] spectroscopic survey observed galaxies over $\sim$286 deg^2 with a flux limit of mag-$r = 19.8$ and a redshift distribution that extends to $z = 0.5$. We use the DR3 [11], which mainly covers until redshift 0.5 and we select the best spectroscopic redshifts by setting the quality parameter $nQ >= 3$.

3. The VIPERS [189] was performed at ESO's VLT in Chile. The survey magnitude limit is $i_{\rm ab} = 22.5$ and covers a redshift range of $0.5 < z < 1.2$ (as targets are colour selected to lie in this range) over an area of $\sim$23.5 deg^2. The redshift quality flag $3 \leq$ zflg ≤ 4 is applied.

4. The DEEP2 redshift survey [55, 163] used the DEIMOS spectrograph at the Keck-II telescope. This survey covers $\sim$2.8 deg^2 in four fields and observes objects out to $z \sim 1$ up to a limiting magnitude $R_{\rm AB} = 24.1$. The quality flag $3 \leq$ zquality ≤ 4 is applied.

5. KiDZ-COSMOS objects are extracted from the KiDS DR5 (Wright et al., in press) spectroscopic sample in the COSMOS field. They are mainly provided from G10-COSMOS [54], a re-reduction of the zCOSMOS-bright sample [146, 145], which covers 1.7 deg^2 up to $i_{\rm AB} < 22.5$ and $0.1 < z < 1.2$. All objects have high quality spectroscopic redshifts with $nQ >= 3$.

6. The Two-degree Field Galaxy Redshift Survey (2dFGRS) [45] was observed from the Anglo-Australian Observatory and used the 2dF spectrograph [144], which covers a

Table 3.2 Main spectroscopic redshift surveys used in the W1, W3 and G09 fields. The first column gives the name of each spectroscopic survey. The second, third and fourth columns give the number of spectroscopic redshifts in each field.

Survey	W1	W3	G09
SDSS	5437	8018	1213
GAMA	8884	0	4704
VIPERS	21378	0	0
DEEP2	0	6969	0
KiDZ-COSMOS	0	0	11854
2dFGRS	2662	0	0
VVDS	2216	0	0
3DHST	933	707	0
Miscellaneous	1193	130	0
Total	42703	15824	17771

2 degree diameter field of view. It measured ∼250000 galaxies, covering an area of 2000 deg^2 with a limiting magnitude of $b_J \sim 19.45$ and a median redshift of $z = 0.11$.

7. The VVDS [81] is a magnitude limit spectroscopic redshift survey that has observed 34594 objects with spec-z from $0 \leq z \leq 6.7$ up to $i_{\rm AB} \sim 24.75$.

8. The 3D-Hubble Space Telescope (HST) [25] is a spectroscopic survey with the Hubble Space Telescope specially designed to study galaxy formation at $1 \leq z \leq 3.5$. It presents a 5σ SNR per resolution element up to $H_{140} \sim 23.1$, the F140W filter imaging.

Table 3.2 shows the number of objects of the main spectroscopic surveys from which the data is taken to validate the three PAUS wide fields. The W1 field from CFHTLenS is mainly covered by VIPERS, GAMA, SDSS, VVDS, 2dFGRS and 3D-HST. The G09 field is covered by GAMA and SDSS. However, as will be seen at the end of this section, some objects from the COSMOS field, which does not cover G09, are also used to validate that field. The reason for this is that both COSMOS and G09 photo-z are run within the same photometric system, as will be explained below. Finally, W3 overlaps with DEEP2, SDSS and 3D-HST. Another 1323 redshifts come from miscellaneous sources[2] not included in Table 3.2 for brevity.

Fig. 3.4 shows the distribution of $i_{\rm AB}$ and spec-z, in filled histograms, for the objects with spectroscopic redshifts for the three fields under study, compared with the $i_{\rm AB}$ and photo-z

[2]The miscellaneous sources are mainly composed by UDSz [23, 155], C3R2 [204], IMACS [60], VANDELS [83] and SAGA [85].

distributions, in empty histograms, for all the objects, that is, the photometric sample. The shape of the spectroscopic distributions is important when assessing the photometric redshift performance in each field as a function of $i_{\rm AB}$ or spec-z. Low counts in a given bin may lead to poor statistics in the determination of the performance. Ideally, the distribution of the validation samples should be very similar to the whole catalogue. However, this is not always possible due to the lack of spectroscopic redshifts available. In the case of the W1 and W3 fields, the spectroscopic distributions look quite similar in terms of $i_{\rm AB}$ and redshift, with the exception that there is a drop-off in the number of objects for the last magnitude bin ($i_{\rm AB} \sim 22.5$) in the W1 field that is not present in W3, where the number of objects above $i_{\rm AB} \sim 21$ remains fairly constant. In the case of the G09 validation sample, one can see that the coverage in both $i_{\rm AB}$ and redshift is quite poor.

The G09 validation sample barely goes beyond $i_{\rm AB} \sim 21$ and spectroscopic redshift $\sim$0.75. This poses a challenge when validating the performance of the photo-z in this field. To overcome this problem, we indirectly validate the G09 field by computing new photometric redshifts in the COSMOS field, using KiDS as a reference catalogue and using its BB photometry (we will refer to this catalogue as KiDZ-COSMOS). This validation is possible because both PAUS and KiDS observe in the COSMOS field. By doing that, we are using the same photometric system either in the G09 and KiDZ-COSMOS fields. We also compute these photo-z using the same magnitude depth in both cases. This validation process will be addressed in detail in Section 3.3.5. The number of KiDZ-COSMOS objects (listed in Table 3.2) nearly doubles the number of objects in the G09 validation sample.

Fig. 3.4 shows that KiDZ-COSMOS covers a range in $i_{\rm AB}$ and spec-z not covered before by the G09 field, allowing us to span the $i_{\rm AB}$ and the spectroscopic ranges of the W1 and W3 fields. However, the combined G09 and KiDZ-COSMOS validation samples, have a pronounced peak at $i_{\rm AB} \sim 19$ and the spec-z distribution is different from the W1 and W3 cases. This fact may result in a more intricate comparison of the G09 field's performance in relation to the W1 and W3 fields.

It is also important to note that the coverage in spec-z for all three fields decreases drastically beyond $z_{\rm s} = 1$. For this reason, we will restrict the validation to $z_{\rm s} < 1.5$ in Section 3.4. This is a challenge to validate our results, since the photo-z code we use allows us to compute redshifts until $z_{\rm b} = 2$. However, due to the magnitude limit at $i_{\rm AB} = 23$, there will not be that many objects at $z_{\rm b} > 1.5$.

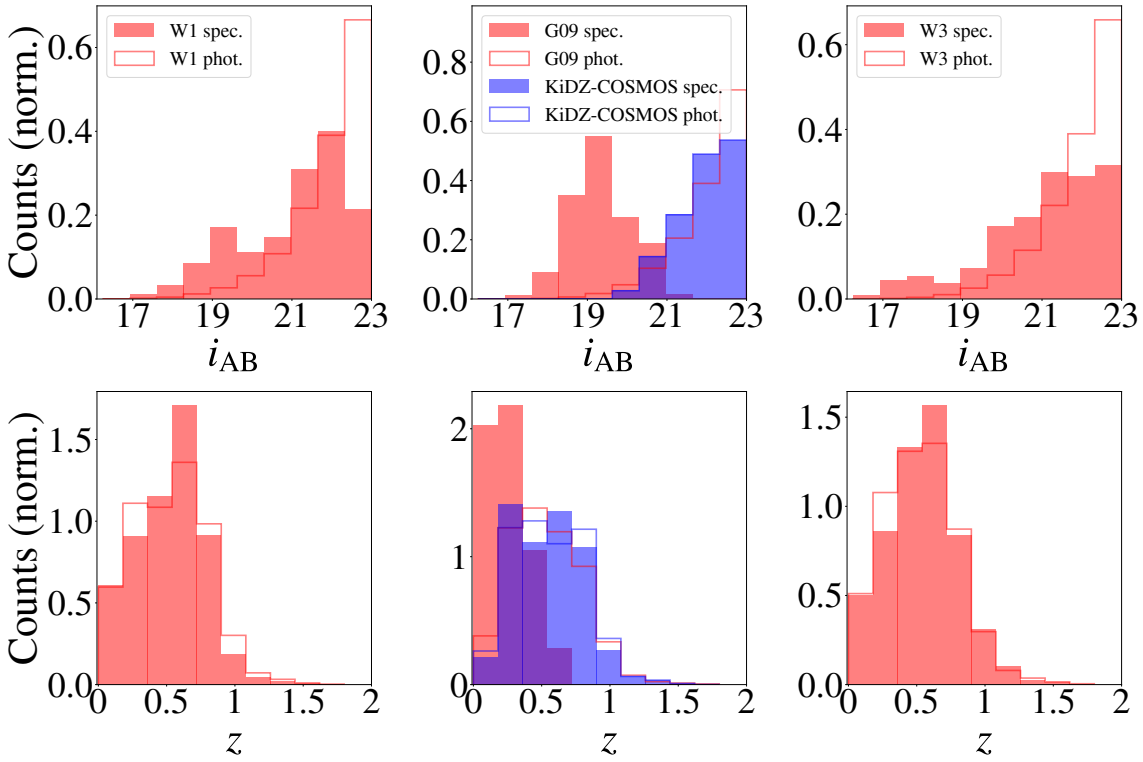

Fig. 3.4 $i_{\rm AB}$ magnitude (top) and redshift (bottom) distributions of the validation and the photometric samples for W1, G09, KiDZ-COSMOS and W3 (from left to right, as labelled in the top row panels). For the validation samples, the redshift shown in the lower panels corresponds to spec-z, while for the case of the photometric samples, it corresponds to photo-z. The validation sample of the G09 field does not cover the necessary range either in $i_{\rm AB}$ or spec-z, making it unsuitable for validating the G09 photometric redshifts by itself. The coverage in $i_{\rm AB}$ and spec-z for KiDZ-COSMOS complements that for the G09 field and allows it to cover deeper regions.

3.3 Methodology

3.3.1 BCNZ

The estimation of the photo-z is performed using a template based code called BCNZ [67], in which the observed flux of a galaxy is fitted against a linear combination of redshift-dependent templates. This code has been specifically designed to process the information coming from both the NB and BB data, also incorporating emission lines. In what follows, we will briefly explain how BCNZ computes the photometric redshifts [for a detailed explanation see 67].

For each galaxy, the probability redshift distribution is obtained via:

$$p(z) \propto \int_{\alpha_1 \geq 0} \mathrm{d}\alpha_1 \cdots \int_{\alpha_n \geq 0} \mathrm{d}\alpha_n e^{-0.5\chi^2(z,\alpha)} p_{\mathrm{prior}}(z,\alpha), \tag{3.2}$$

where α_i corresponds to the amplitude of template $i = 1, \ldots, n$, p_{prior} is the form of the priors and χ^2 is defined as:

$$\chi^2(z,\alpha) = \sum_{i,\mathrm{NB}} \left(\frac{\phi_i^{\mathrm{obs}} - l_i k \phi_i^{\mathrm{model}}}{\phi_{\mathrm{err},i}} \right)^2 + \sum_{i,\mathrm{BB}} \left(\frac{\phi_i^{\mathrm{obs}} - l_i \phi_i^{\mathrm{model}}}{\phi_{\mathrm{err},i}} \right)^2, \tag{3.3}$$

where ϕ_i^{obs} and ϕ_i^{model} are the observed and modelled fluxes in band i, respectively; $\phi_{\mathrm{err},i}$ is the error on ϕ_i^{obs} and k and l are calibration parameters.

On the one hand, k is intended to calibrate between the NB and BB fluxes for each galaxy and can be easily obtained by taking the derivative of χ^2 (eq. 3.3) with respect to k. On the other hand, the parameter l acts as a global calibration factor (zero-point) per band. This zero-point calibrates the offset between the observed fluxes and the best-fitting model and is defined as:

$$l_i = \mathrm{Median}[\phi_i^{\mathrm{model}} / \phi_i^{\mathrm{obs}}]. \tag{3.4}$$

In order to compute the photo-z, χ^2 is minimized as a function of a set of template amplitudes on a redshift grid covering $0.01 < z < 2$, with a resolution of 0.001, so each galaxy has a corresponding best fitting template from which the redshift probability density $p(z)$ is computed. The photometric redshift that is assigned to each object corresponds to the peak of the $p(z)$ distribution, which we label as z_{b}.

The templates we employ in the photo-z estimation are the same as in Eriksen et al. [67]. They include templates for elliptical and red spiral galaxies, star-bust galaxies are introduced following BC03 [28] models, with ages ranging from 0.03 Gyr to 3 Gyr. Additional BC03 templates with different ages and metallicities are also introduced. Emission lines are

modelled for star-bust galaxies and their galactic extinction is accounted for following the Calzetti law [31].

BCNZ provides some photo-z quality parameters that allow the galaxies with the best photo-z to be selected, which might be advantageous for some applications. For a full description of the different quality parameters, see [67]. Here, we have chosen to use the parameter Q_z, (eq. 3.5), since it is a combination of other quality parameters:

$$Q_z \equiv \frac{\chi^2}{N_f - 3}\left(\frac{z^{99}_{\rm quant} - z^{1}_{\rm quant}}{\rm ODDS}\right), \tag{3.5}$$

where χ^2 corresponds to eq. 3.3, N_f is the number of filters, $z^n_{\rm quant}$ are the nth percentile of the posterior distribution and ODDS is defined as:

$$\rm ODDS = \int_{z_b-\Delta z}^{z_b+\Delta z} dz\, p(z), \tag{3.6}$$

where $\Delta z = 0.035$. This last parameter quantifies the probability that is located around the peak of $p(z)$, that is, $z_{\rm b}$.

We present a study of the performance of the PAUS wide fields as a function of Q_z in Section 3.4.5, showing that the photo-z quality is correlated with Q_z.

3.3.2 New calibration of l

One improvement that has been made to the BCNZ code is related to the estimation of l, the zero-point per band in eq. 3.4. In order to compute l, a comparison between the observed flux and the best fitting model is needed, so one has to evaluate the best fitting model over the whole redshift range. Since this is computationally expensive, in Eriksen et al. [67] they defined a calibration sample for objects with spectroscopic redshifts and good photometry and only evaluated the best fitting model at the spectroscopic redshift of these objects, to later apply l to the whole sample. We will refer to this method as the spectroscopic calibration. This restricts the technique to be applied only when spectroscopic redshift information is available. Another disadvantage of this method arises from the potential selection bias between the spectroscopic sample and the full catalogue. This could lead to an overestimation or underestimation of the proportion of objects, either in terms of magnitude or redshift, which might result in the application of a non-representative zero-point to the whole sample. Finally, the spectroscopic sample is used both to calibrate l and to validate the photo-z, establishing a potential dependency in both steps. In order to alleviate these problems, we propose a new methodology to compute the zero-point.

Instead of computing the best fitting model at the spectroscopic redshift of each source in the calibration sample, the new technique uses an iterative approach. In the first iteration, l is set to one for all the bands and the photo-z are computed. In the next iteration, the best fitting model is evaluated at the photo-z computed in the previous iteration and BCNZ is run again, setting l according to eq. 3.4. This iteration is performed five times in total, and it has been shown to be sufficient to reproduce the photo-z accuracy obtained by the previous spectroscopic calibration, shown in Fig. 3.10 (which will be discussed in Section 3.4.1). In order to better determine the zero-points, we also split the catalogue into 6 $i_{\rm AB}$ bins containing equal numbers of objects, and we compute the zero-points for each of these subsamples. This way, we can better take into account the differences in the sample in terms of magnitude. We tried different approaches to divide the sample, such as using different numbers of magnitude and/or redshift bins, but found that the best solution is to divide it in 6 $i_{\rm AB}$ bins.

The calibration sample used in this new iterative approach is defined so that the objects with poorer photometry are discarded. For each field, the survey mask of the reference catalogue is applied, stars are rejected and only objects observed in the 40 NB are kept for the calibration step. To reduce the computational time, we also downsample the calibration sample to consider 10% of the objects in the whole catalogue. This is done except for the last iteration, where we apply l to the full sample and compute all the photometric redshifts, extending the sample by not accounting for the survey mask and including objects previously classified as stars and with a coverage of at least 30 NB.

Fig. 3.5 shows the zero-points for the iterative (solid lines) and the spectroscopic (dashed line) calibration as a function of the 40 NB and the BB (differentiated from the NB with a background grey colour) of the W1 field (similar trends are obtained for the other fields analysed). In the case of the iterative calibration, the distribution of the zero-points for the different magnitude bins follows a local trend per band, as expected, while preserving their differences. The peaks and off-peaks illustrate the corrections that need to be made in each band in order to fit the modelled flux. A similar local trend for the l factor can be seen in the spectroscopic calibration method, which validates the new method. Note that the spectroscopic case closely follows the line corresponding to the brightest magnitude bin (blue line), since that is the bin containing the majority of objects in the spectroscopic calibration sample.

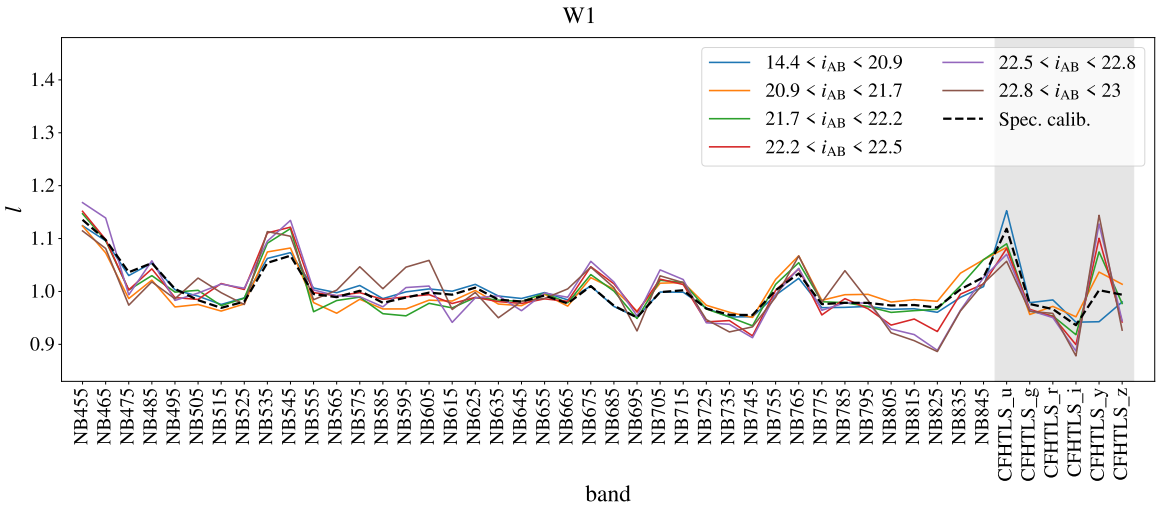

Fig. 3.5 NB and BB (marked with a background grey colour) l zero-points computed with the iterative (solid lines) and the spectroscopic (dashed line) calibration for the W1 field. The local trend per band for both cases is very similar. The spectroscopic calibration zero-point is much alike to the zero-point of the brightest bin in magnitude in the iterative calibration, since the majority of the objects in the spectroscopic calibration sample are in that range. In the iterative calibration method, variations in l for the different $i_{\rm AB}$ bins show that this method is able to capture the particular behaviour of each $i_{\rm AB}$ subsample.

3.3.3 Metrics

The metrics that quantify the performance of the photometric redshifts are presented in this section. These metrics are based on Δ_z, which is defined as:

$$\Delta_z = \frac{z_{\rm b} - z_{\rm s}}{1 + z_{\rm s}}, \tag{3.7}$$

where $z_{\rm b}$ and $z_{\rm s}$ correspond to the photometric and spectroscopic redshifts, respectively. This quantity quantifies how accurately the photo-z are estimated in comparison with the spec-z. Note that the metrics associated with Δ_z will only be available for objects with spec-z information.

We define the σ_{68} of the quantity Δ_z as:

$$\sigma_{68} = \frac{P[84] - P[16]}{2}, \tag{3.8}$$

where P[x] corresponds to the percentile x of the Δ_z distribution.

An object is considered to be an outlier if it satisfies the following condition:

$$|\Delta_z| > 0.1. \tag{3.9}$$

Finally, the bias shows the systematic difference between the spectroscopic and photometric redshifts and is defined as the median of the difference:

$$\mu = \mathrm{med}(z_\mathrm{b} - z_\mathrm{s}). \tag{3.10}$$

The centralised scatter, σ_{68}, and the outlier fraction of the quantity Δ_z, and the bias of $z_\mathrm{b} - z_\mathrm{s}$ will be quantified as a function of the i_AB, the spectroscopic and the photometric redshift.

3.3.4 Weighted photo-z

We obtain that, for very faint objects ($i_\mathrm{AB} > 22.5$), the performance of the photo-z computed using the PAUS NB+BB does not improve compared to the photo-z computed using BB photometry only. This is explained by the low SNR of the PAUS photometry for these faint objects. This low SNR is due to the fact that the noise in the NB is dominated by (Poisson) sky noise above $i_\mathrm{AB} > 22.5$, whereas the BB fluxes correspond to much deeper exposures. Additionally, the BB collect light over a wavelength range an order of magnitude higher, integrating more signal. All this is illustrated by Fig. 3.6, where the SNR, the flux (ϕ) and the flux error (ϕ_err) for the W3 field are shown (the case of W1 is similar, although not shown here for brevity). On the one hand, the errors in the NB flux stay almost constant for objects fainter than $i_\mathrm{AB} \sim 20$, while the errors in the BB decrease over the full i_AB range. On the other hand, the fluxes for both NB and BB continuously decrease as a function of i_AB. This combination makes the SNR of the NB to decrease as we reach fainter magnitudes, dropping to SNR values $\sim$1-3 beyond $i_\mathrm{AB} \sim 22.5$, while the SNR of the BB stays almost constant. For the brighter objects, the slope of the NB SNR is shallower because there are more objects measured as extended. The case for the G09 field is slightly different regarding the BB flux errors. We include a study on the different SNR and photo-z performance between the W1/W3 and the G09 fields in Appendix A.

In order to see how the differences in SNR between NB and BB affect the estimation of the photo-z, the top panels of Fig. 3.7 show the σ_{68} as a function of i_AB and z_b for the photo-z estimated from NB+BB (dashed black line) and BB-only (blue line) for the W3 field (similar trends are observed for the W1 and G09 fields). At high SNR (i.e. for bright galaxies) the

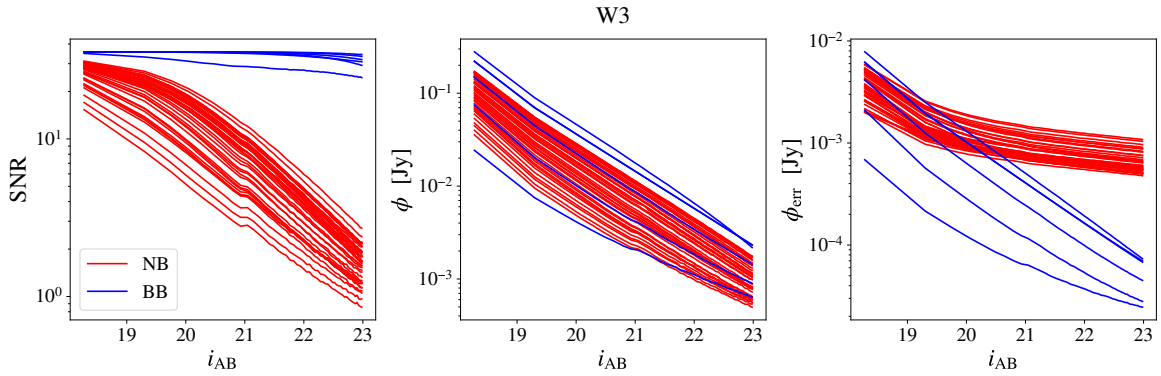

Fig. 3.6 SNR, flux (ϕ) and flux error ($\phi_{\rm err}$) (from left to right) of the W3 field for the BB (blue lines) and the NB (red lines). The SNR of the BB stays fairly constant for the entire $i_{\rm AB}$ range, since both the BB flux and flux errors continuously decrease. In the case of the NB, the SNR decreases as the objects become fainter, which is caused by the constant NB flux errors at faint magnitudes.

accuracy depends linearly on the width of the filters, reducing the σ_{68} by a factor of 10 (from few 10^{-2} to few 10^{-3}) when using NB, which are a factor of 10 narrower than the BB. Also, note how the left panel of Fig. 3.6 and the top left panel of Fig. 3.7, which respectively show the SNR in the different filters and the σ_{68} as a function of $i_{\rm AB}$, look like the reverse of each other. This indicates that the degradation of the photo-z performance with magnitude is mostly driven by the SNR decrease. We have verified that this is indeed the case by doing different data reductions with a different number of exposures (see Fig. 26 in Serrano et al. [193]).

It is important to highlight that the BB photo-z were estimated by the CFHTLenS and KiDS teams [99, 133] using a different photo-z code (BPZ, [13]) which was optimized for BB filters (for a detailed comparison with BCNZ see Eriksen et al. [67]). We will refer to the photo-z computed using only BB as $z_{\rm b,\,BPZ}$, and the photo-z computed with NB+BB as $z_{\rm b,\,BCNZ}$. Even when the BB and NB+BB estimates share some of the same data, the choice of templates, method and optimization makes the two estimates fairly independent. This is illustrated in Fig. 3.7, which also shows the inverse variance weighted photo-z (solid black line), combining the BB and NB+BB cases, defined as:

$$z_{\rm b,\,BCNZw} = \frac{z_{\rm b,\,BCNZ}\cdot w_{\rm BCNZ} + z_{\rm b,\,BPZ}\cdot w_{\rm BPZ}}{w_{\rm BCNZ} + w_{\rm BPZ}}, \tag{3.11}$$

where the weight $w = 1/\sigma^2_{68,i_{\rm AB}}$ is given by the corresponding σ_{68} values of the photo-z of BCNZ or BPZ as a function of $i_{\rm AB}$ in Fig. 3.7. The weighting of $z_{\rm b,\,BCNZ}$ and $z_{\rm b,\,BPZ}$ allows us to obtain a new $z_{\rm b,\,BCNZw}$, which closely follows the scatter of the photo-z from

$z_{\mathrm{b,\,BCNZ}}$ for objects brighter than $i_{\mathrm{AB}} = 22.5$. When the performance of $z_{\mathrm{b,\,BCNZ}}$ substantially decreases, $z_{\mathrm{b,\,BCNZw}}$ follows the performance of $z_{\mathrm{b,\,BPZ}}$. Also, at the intersection between the NB+BB and the BB cases, the new $z_{\mathrm{b,\,BCNZw}}$ shows slightly better σ_{68} than either of the other two cases. Note that the inverse variance weighting method is the optimal combination of two (or more) variables with independent errors, thus minimizing the variance of the weighted estimate. This assumption of independence is an approximation, since some correlation may exist between the BCNZ and BPZ estimates.

The bottom panel in Fig. 3.7 shows the bias as a function of i_{AB} and as a function of z_{b}. We note that $z_{\mathrm{b,\,BPZ}}$ presents a larger bias than $z_{\mathrm{b,\,BCNZ}}$. In order not to transfer that bias into the weighted estimate, we subtract the bias from $z_{\mathrm{b,\,BPZ}}$ before applying the inverse weighting technique. By doing that, we end up with a similar bias between $z_{\mathrm{b,\,BCNZ}}$ and $z_{\mathrm{b,\,BCNZw}}$.

Fig. 3.8 shows the scatter plot of photo-z vs spec-z of the three PAUS wide fields for the two photo-z versions, unweighted (BCNZ, left) and weighted (BCNZw, right), of the NB+BB estimates. An issue that is solved by the inverse weighted photo-z is the fact that, a small fraction ($\sim$2% in the W1 and W3 fields and $\sim$3% in the G09 field) of the photo-z computed with BCNZ having $z_{\mathrm{s}} > 0.75$ for W1 and W3 and $z_{\mathrm{s}} > 0.85$ for G09, have a wrongly assigned value close to $z_{\mathrm{b}} \simeq 0.72$ or $z_{\mathrm{b}} \simeq 0.89$, respectively. This creates a horizontal stripe (or focusing) in the z_{b} vs. z_{s} plots (left panel in Fig. 3.8 and upper panels in Fig. A.1 to see each field in detail). This redshift focusing effect occurs when the prior dominates the posterior probability distribution in the low SNR case. In this case, numerous objects are assigned the peak value of the prior, which leads to artificial peaks in the redshift histograms [99]. Even though some small focusing effect is expected when computing photo-z, the cases at $z_{\mathrm{b}} \simeq 0.72$ and $z_{\mathrm{b}} \simeq 0.89$ become an issue for us. Because of the different priors and libraries of SED and emission lines (see Eriksen et al. [67] for details), the focusing effect is different in the two photo-z codes. Therefore, the focusing issue is much dissipated in the newly defined $z_{\mathrm{b,\,BCNZw}}$ (compare left and right panels of Fig. 3.8).

However, for the G09 and the KiDZ-COSMOS fields, the peak of objects at $z_{\mathrm{b}} \simeq 0.89$ is not completely dissipated by $z_{\mathrm{b,\,BCNZw}}$, so that an artificial excess of objects at that redshift remains. In the case of the G09 field, we found that some of those outliers correspond to objects for which the BB photo-z are higher than 4, so we removed from the sample all objects with $z_{\mathrm{b,\,BPZ}} > 4$, which accounted for $\sim$1% of the G09 objects. The remaining outliers ($\sim$2% and $\sim$1% of the total sample in the G09 and KiDZ-COSMOS fields, respectively) were not removed, but instead their photometric redshift were substituted by the BB photo-z provided by KiDS. Since those redshifts have a resolution of 0.01 and BCNZ has a resolution of

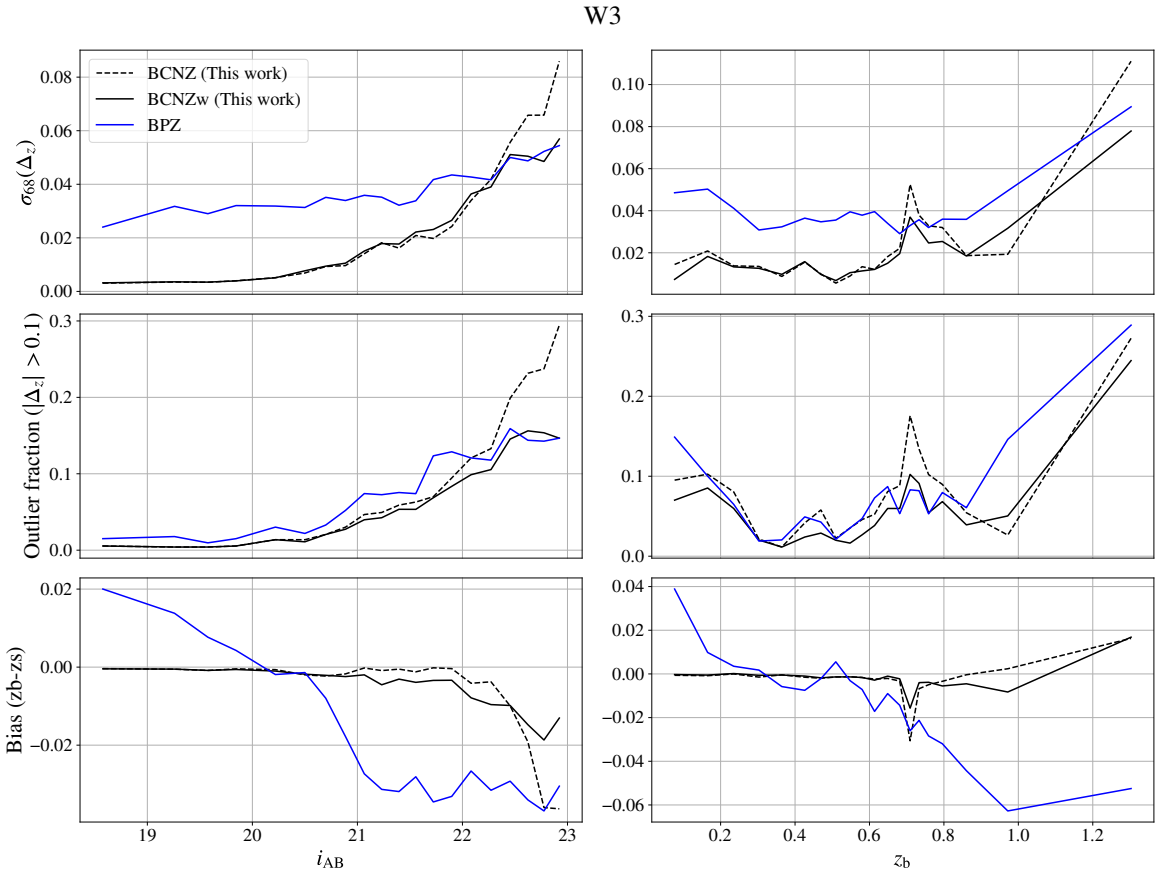

Fig. 3.7 σ_{68} (top), outlier fraction (middle) and bias (bottom) as a function of i_{AB} (left) and z_b (right) for the photometric redshifts in the W3 field. The z_b computed from BPZ (using only BB) are shown in blue, the ones computed by BCNZ (using NB+BB) in dashed black lines and the weighted combination of BCNZ and BPZ photo-z (BCNZw) in solid black lines. Each bin is defined to contain an equal number of objects. For bright objects, $z_{b,\,BCNZ}$ have better performance in terms of smaller values of σ_{68}, outlier fraction and bias. However, for fainter objects ($i_{AB} > 22.5$), $z_{b,\,BPZ}$ has better accuracy in terms of σ_{68} and outlier fraction. The weighted combination of BCNZ and BPZ yields the best overall photo-z performance in terms of σ_{68}, outlier fraction and bias.

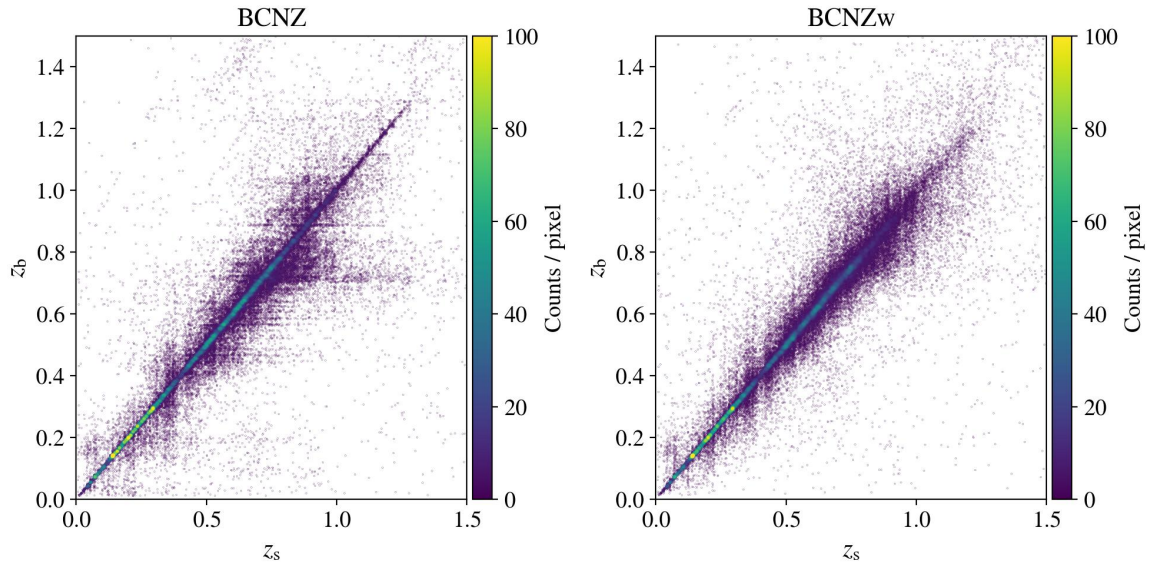

Fig. 3.8 Photometric redshift vs spectroscopic redshift for the BCNZ photo-z (left) and the BCNZw photo-z (right), that is, when the inverse variance weighting is applied, for the 3 PAUS wide fields combined. The colour bar indicates the density of objects. The horizontal stripes at $z_b \approx 0.72$ and $z_b \approx 0.89$ are dissipated when weighting with the BPZ photo-z, computed only with BB.

0.001, we applied a normal random distribution with a $\sigma = 0.01$ to disperse them around their initial value, matching the resolution of BCNZ-like photo-z.

Taking these two factors into account, that is, the low SNR in the NB for faint objects and the horizontal stripe around $z_{b,\,\mathrm{BCNZ}} \sim 0.72$ and $z_{b,\,\mathrm{BCNZ}} \sim 0.89$, we will study the performance of the BCNZ photo-z and the BCNZ weighted photo-z and compare them.

3.3.5 Validating the G09 field with KiDZ-COSMOS

As mentioned in Section 3.2.3, the validation sample in the G09 field is not adequate neither in terms of i_{AB} nor spec-z coverage, and is therefore very limited to assess the performance of the G09 photo-z. For that reason, a new validation sample is introduced named KiDZ-COSMOS, which covers a range of i_{AB} magnitude and spec-z not covered before. These objects lie in the COSMOS field and were observed by KiDS, so their photometry is equivalent to the objects observed in the G09 field.

For this test, the positions of the objects in COSMOS are defined by using forced photometry with the KiDZ-COSMOS catalogue, which allows us to measure the PAUS NB fluxes and their errors. Although in the COSMOS field PAUS usually does an average of 5 single exposures before coadding, for the purpose of this test we limit the average number of

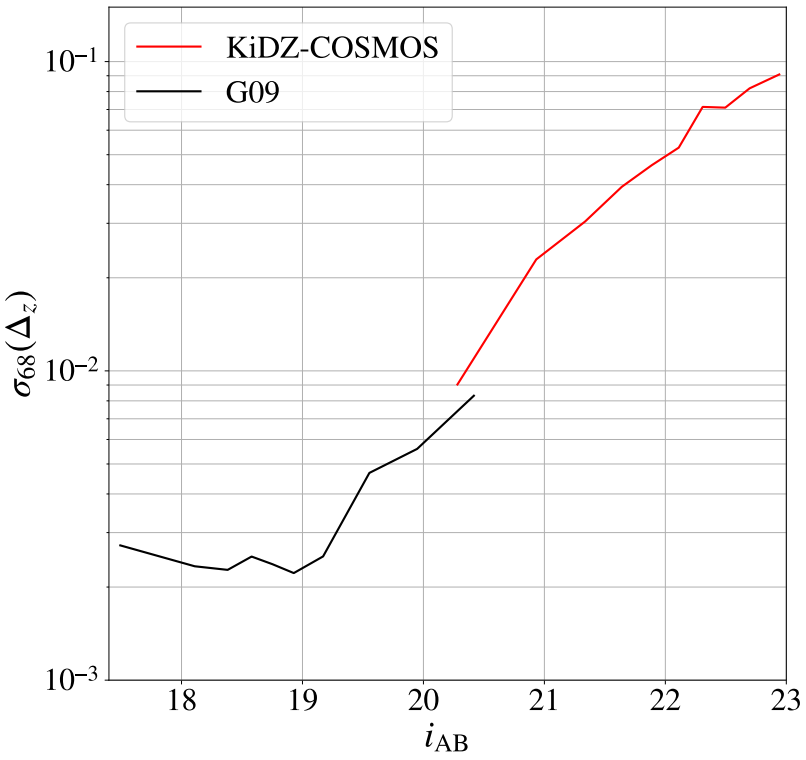

Fig. 3.9 $\sigma_{68}(\Delta_z)$ as a function of $i_{\rm AB}$ for the G09 and the KiDZ-COSMOS validation samples. While G09 can be used to validate the bright end of $i_{\rm AB}$, KiDZ-COSMOS validates the faintest objects. In the magnitude range where both validation samples meet ($i_{\rm AB} \sim 20.5$), the σ_{68} of both cases intersect, showing similar performances.

single exposures to 3, since this is the average number for the PAUS wide fields. Once the forced photometry is done, the photo-z are measured using the NB and BB ($ugrizYJHK_s$) from KiDZ-COSMOS.

Fig. 3.9 shows the performance of the photometric redshifts as a function of $i_{\rm AB}$ for the G09 and the KiDZ-COSMOS validation samples. Since the $i_{\rm AB}$ range of both samples is complementary (G09 covers most of the bright objects, while KiDZ-COSMOS covers the faintest), each one allows us to determine the performance for different magnitude ranges. At the magnitude where both validation samples meet, they present similar σ_{68} values, as expected for objects observed with the same photometric survey.

Finally, from now on we will use the G09 + KiDZ-COSMOS validation sample as the standard for G09 and will refer to it only as the G09 validation sample, unless indicated otherwise.

Table 3.3 Column name and its description for the published catalogue in CosmoHub.

Column name	Description
ref_id	PAUdm reference id (unique per PAUS wide field)
field	PAUS wide field
RA	right ascension (deg)
DEC	declination (deg)
zb_BCNZ	photometric redshift from BCNZ
zb_BCNZw	weighted photometric redshift from BCNZ
odds	BCNZ ODDS quality parameter
chi2	BCNZ minimum χ^2
nb_bands	number of narrow bands
qz	BCNZ Q_z quality parameter
mag_i	$i_{\rm AB}$ magnitude (not corrected for the difference between the CFHTLenS and KiDS magnitudes)
star_flag	Only for objects from CFHTLenS: Star-galaxy separator (0 =galaxy, 1 =star)
mask_cfhtlens	Only for objects from CFHTLenS: CFHTLenS mask value at the object's position
sg_flag	Only for objects from KiDS: Star/Gal Classifier (1 for galaxies)
sg2dphot	Only for objects from KiDS: 2DPhot Star/Galaxy classifier (1 for high confidence star)
mask_kids	Only for objects from KiDS: 9-band mask information
flux_x	Coadded flux of the x band
flux_error_x	Coadded flux error of the x band

3.4 Photo-z catalogues

In this section, the photometric catalogues are presented in detail. First, in Section 3.4.1, the new calibration technique is validated. Later, in Section 3.4.2, a comparison of the radial distribution of the PAUS wide fields and the performance of the photo-z are shown. Next, in Section 3.4.3 a detailed analysis is performed as a function of the galaxy colours. Finally, in Section 3.4.4 we study the $p(z)$ distribution obtained by BCNZ.

We publish the photo-z studied in this Chapter 3 in CosmoHub [207, 35], which can be accessed through this link: https://cosmohub.pic.es/catalogs/315. Table 3.3 shows the name and the description of the columns included in this catalogue.

3.4.1 Iterative vs. spectroscopic method

In Section 3.3.2, we introduced a new calibration technique of the zero-point l (see eq. 3.4) that consisted of an iterative approach, where l was set to 1 in the first iteration and updated by the photometric redshifts computed in the previous iteration. This method was in contrast with the earlier one used by BCNZ, where the calibration was performed via a subset with spectroscopic redshifts. Thus, this new technique allows to calibrate the photo-z estimation even when there are not spectroscopic redshifts available. It is also important to note that the

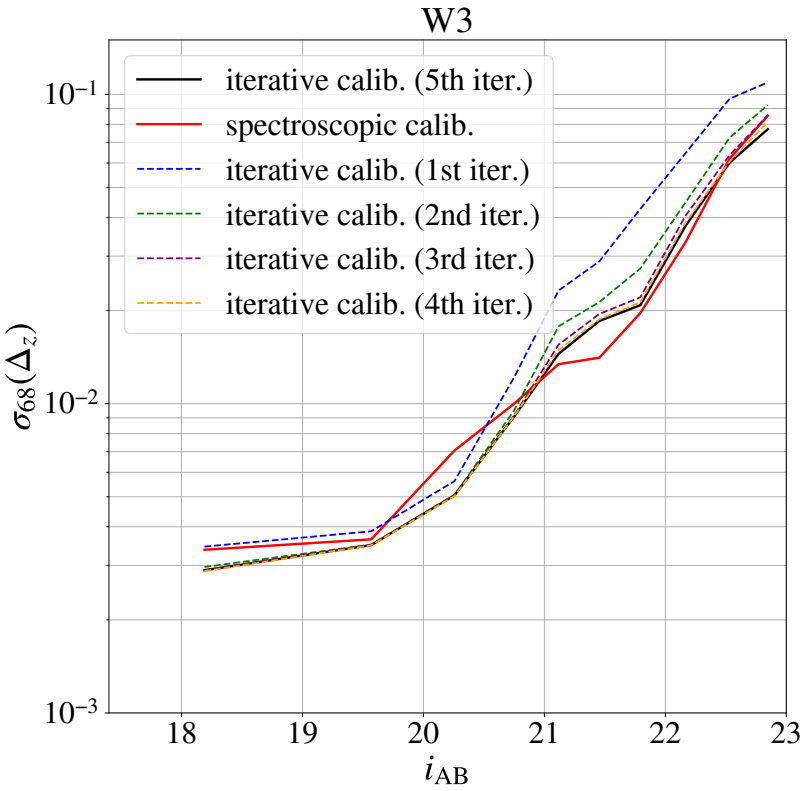

Fig. 3.10 Comparison of the σ_{68} as a function of $i_{\rm AB}$ for the iterative and the spectroscopic calibration. Solid lines show the final σ_{68} for the iterative and the spectroscopic calibration, while dashed lines show how the intermediate results for the iterative calibration converge to the final iteration. The performance achieved by both techniques is equivalent.

previous calibration method used the same objects with spectroscopic redshifts to calibrate and validate the photo-z, so that both steps of the photo-z estimation process were not independent.

To ensure that this iterative technique gives, at least, the same performance as the spectroscopic one, we show the $\sigma_{68}(\Delta_z)$ for both methods in the W3 field in Fig. 3.10, where the solid lines are the final results of both calibration techniques and the dashed lines are the intermediate results computed at each iteration in the iterative technique. From the solid lines, it is noticeable that both methods agree in performance, since the σ_{68} values are equivalent for the whole $i_{\rm AB}$ range. On the other hand, the dashed lines indicate that, iteration by iteration, the iterative calibration converges to the final result, while also indicating that 5 iterations are sufficient for convergence.

Table 3.4 Number of objects with photometric redshifts, area, number density and average σ_{68} of the z_b BCNZ and BCNZw for the PAUS wide fields observed by PAUS. These values correspond to objects observed with at least 30 NB, after applying the mask and rejecting stars.

# objects	1241047
Area [deg^2]	40.99
Number density [deg^{-2}]	30277
BCNZ σ_{68}	0.020
BCNZw σ_{68}	0.021

3.4.2 PAUS wide fields photometric redshifts comparison

The aim of this section is to present the radial distribution of the PAUS wide fields photo-z catalogues, study the photo-z performance and compare the different estimates of the photo-z, that is, $z_{b,\,BCNZ}$ and $z_{b,\,BCNZw}$.

The analysis is done after excluding the objects with bad photometry or those classified as stars, following the flags and masks defined on Section 3.2.2. Also, BCNZ allows to compute photometric redshifts for objects with different NB coverages and we opt to use objects with a coverage of 30 NB or more, as their performance is very similar to that obtained using objects with a coverage of only 40 NB (see Section 3.4.6 for more details). As a result, we gain almost 300 000 objects with a coverage below 40 NB. Finally, as detailed in Section 3.3.4, some objects of the G09 field with $z_{b,\,BPZ} > 4$ are considered outliers, so we remove them from the catalogue.

Table 3.4 shows the number of objects with photo-z information, the area, the number density and the average σ_{68}, after applying the mask, for both photo-z estimates, BCNZ and BCNZw. The number densities attained by PAUS are much higher than those reached with the spectroscopic surveys defined in Section 3.2.3, allowing us to have high quality redshift information between BB photometric and spectroscopic redshifts for high density regions on the sky. The mean σ_{68} of both estimates is essentially the same, although the differences appear when studying it as a function of i_{AB}, z_b and z_s, as it will be discussed in detail in Fig. 3.13.

The photometric redshift distributions divided by the area as a function of the photometric redshifts are presented in Fig. 3.11 for the combined PAUS wide fields, for the BCNZ original photo-z (dashed black line) and for the weighted cases (solid black line). It can be seen that there are some artificial peaks around $z_b \sim 0.7 - 0.8$ in the $z_{b,\,BCNZ}$, wrongly assigned by

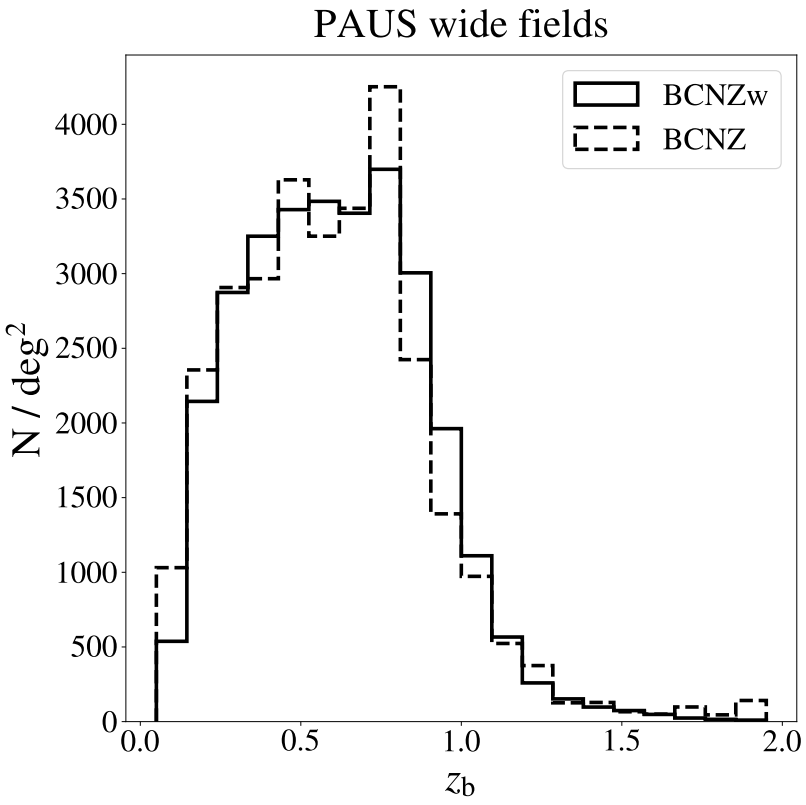

Fig. 3.11 Photometric redshift distributions divided by the area as a function of z_b for the PAUS wide fields. The output of BCNZ and the inverse variance weighting (BCNZw) cases are shown. The BCNZw photo-z do not show the artificial peaks seen in the redshifts from BCNZ.

the photo-z code, as discussed in Section 3.3.4. By weighting with the BB photo-z, these artificial peaks are mostly removed, leading to a smoother distribution.

Fig. 3.12 shows the distribution of the weighted photometric redshifts for a subsample of 2 deg^2 in DEC as a function of R.A, coloured by the fluctuation of the number density of objects for all three wide fields. This fluctuation in density is defined as $\frac{1}{\sigma_z}\left(\frac{n_z}{\mu_z}-1\right)$, where σ and μ correspond, respectively, to the σ_{68} and the median value of the number density, n, and the subscript z reflects that we slice in redshift bins. In this case, we compute the quantities in redshift bins, as opposed to the case in Fig. 3.3, since we want to focus in the fluctuations at each redshift. In order to select the best objects, we keep the 50% of objects with better photo-z quality based on the Q_z parameter. These figures illustrate the good determination of the photo-z in PAUS, highligting lower and higher overdensity regions. We can see some photo-z errors around some of the overdense regions, which can be identified at fixed R.A

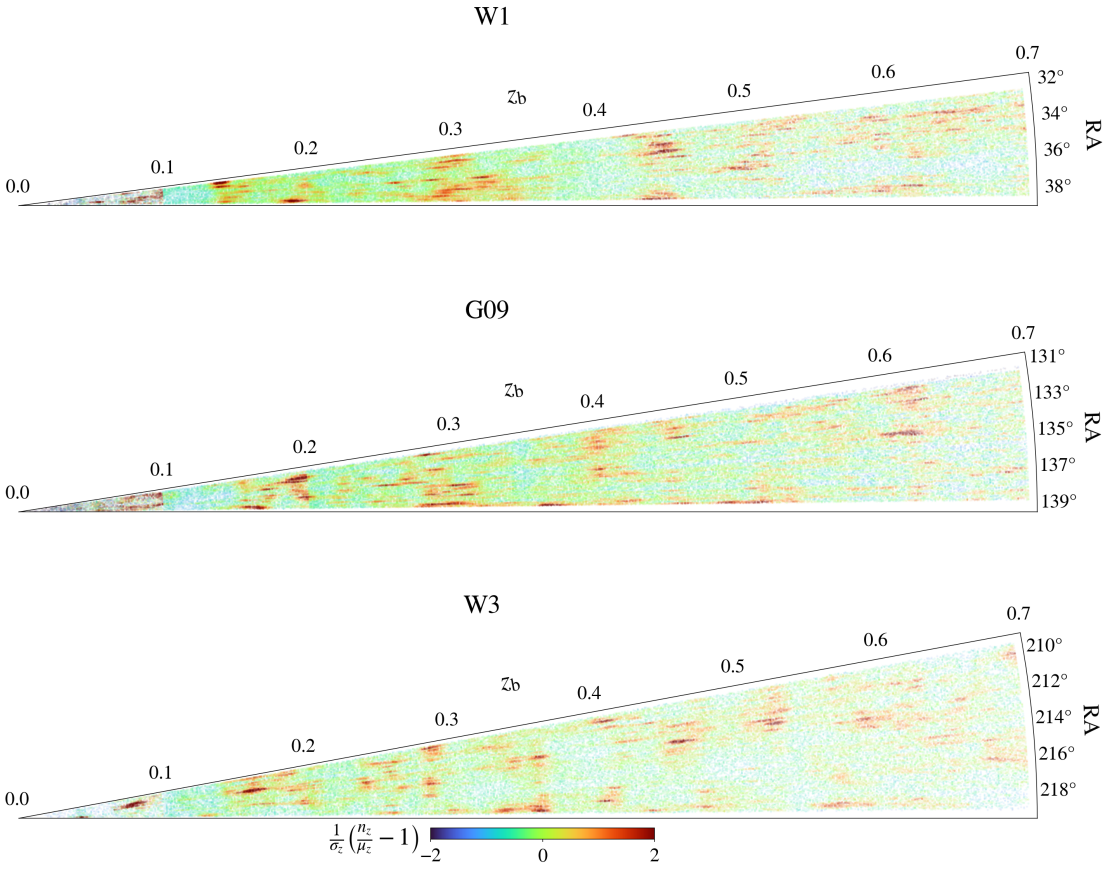

Fig. 3.12 Distribution of the photometric redshifts of the W1, G09 and W3 fields as a function of R.A for a cut in DEC, selecting the best 50% of objects based on Q_z. The colour bar indicates the fluctuation in the number density of objects per redshift bin (n_z), $\frac{1}{\sigma_z}\left(\frac{n_z}{\mu_z}-1\right)$, such that σ_z and μ_z correspond to the σ_{68} and the median of n_z.

as a function of redshift. This can be seen, for example, in the W1 field at R.A $\sim$38° and $z_b\sim$ 0.2 or in the G09 field at R.A $\sim$138° and $z_b\sim$ 0.3.

Fig. 3.13 shows the photo-z performance in the PAUS wide fields as a function of i_{AB}, z_b and z_s. From top to bottom, the σ_{68}, the outlier fraction and the bias are shown. Solid black lines show the BCNZ weighted photo-z (BCNZw), dashed black lines show the BCNZ photo-z and solid blue lines show the BPZ photo-z. As shown before in Fig. 3.7, the σ_{68} as a function of i_{AB} is lower for faint objects for the weighted photo-z case than for the BCNZ photo-z; the improvement in σ_{68} of the BCNZw photo-z starts to become apparent at $i_{AB} > 22.5$, near where the BCNZ and the BPZ lines cross. However, this is not necessarily the case as a function of z_b or z_s, where the σ_{68} does not show a clear preference for either of the BCNZ estimates. For the weighted photo-z, σ_{68} ranges between $\sim$0.003 and $\sim$0.06 for $i_{AB} \sim$ 19 and 23, respectively. Meanwhile, the BCNZ case increases up to $\sigma_{68} \sim 0.09$

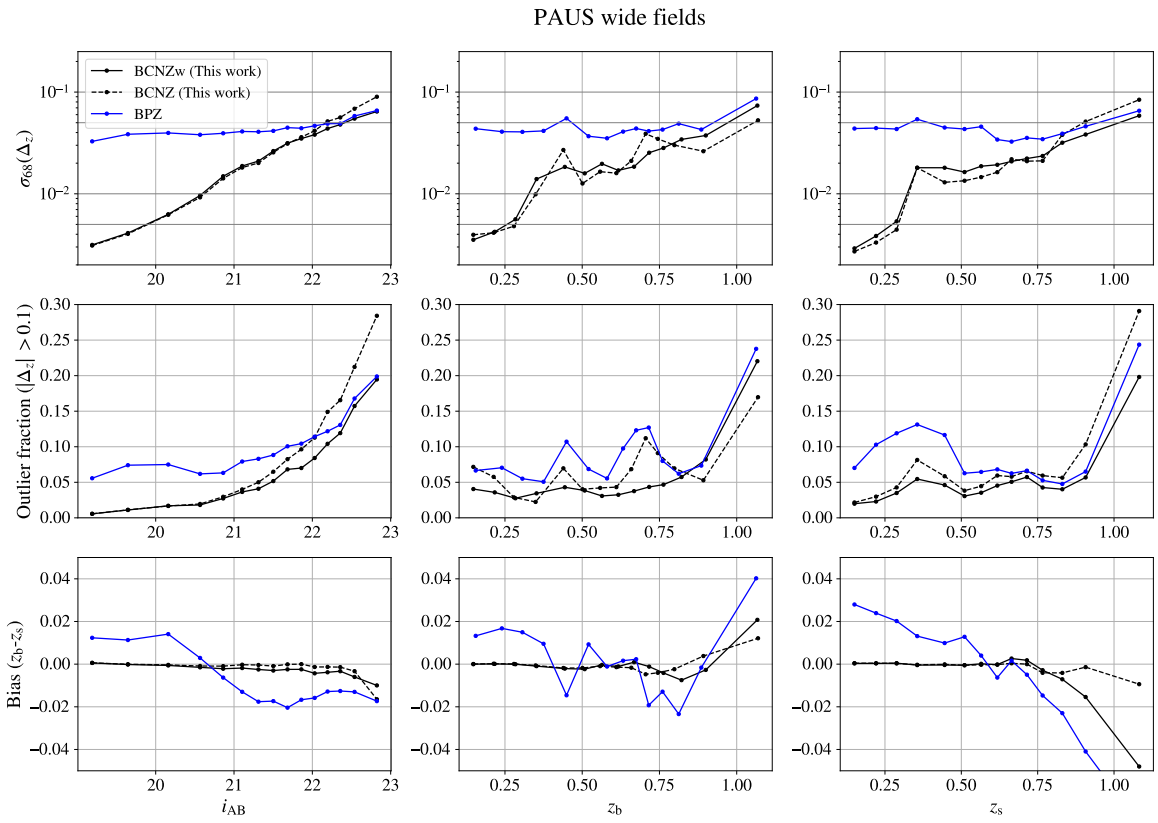

Fig. 3.13 Performance of the PAUS wide fields for the weighted BCNZw photo-z (solid black line), the BCNZ photo-z (dashed black line) and the BB photo-z computed from BPZ (blue line). From top to bottom, the σ_{68}, the outlier fraction and the bias are shown as a function of i_{AB} magnitude, photometric redshift and spectroscopic redshift. Each bin is defined to contain an equal number of objects. In general, BCNZ photo-z estimates give superior performance in all metrics considered, in comparison with BPZ photo-z.

at the faintest magnitude bin. As for the outlier fraction, we note that, for all the i_{AB} and most of the z_b and z_s ranges, the fraction of outliers is lower in the weighted case, reaching a maximum of $\sim$0.20. When studying the bias, it reaches $\sim$-0.02 at the faintest magnitudes in the BCNZ case, while it oscillates very close to 0 as a function of i_{AB} in the weighted case. The bias increases up to $\sim$0.02 and $\sim$-0.04 as a function of z_b and z_s, respectively, in the BCNZw photo-z for the last bin, although it stays very close to 0 in all the other redshift range. At the last redshift bin, the BCNZ photo-z shows a lower bias than BCNZw. Finally, the performance of the BPZ photo-z is worse than the BCNZ photo-z estimates in almost all the metrics analysed.

3.4.3 Colour separation

One of the most relevant properties of galaxies is its morphological type, which is broadly divided into elliptical and spiral galaxies. Elliptical galaxies tend to be massive and dominated by old stellar populations, while spiral galaxies are less massive and more gas rich, allowing star formation to happen [109, 61]. The morphology of galaxies is related to their colour, so that elliptical galaxies tend to be redder and spiral galaxies bluer. Colour can therefore be used as a proxy for morphology. Studying the dependence of the photo-z performance with respect to colour is useful since it is then possible to study science cases based on the morphology of the galaxies.

We derive rest-framed magnitudes from the SED fitting including BB and NB, as used for the photo-z estimations. The SED fitting is performed with the Code Investigating GALaxy Emission (CIGALE) [20], a state-of-the-art SED fitting tool based on the principles of the energetic balance between dust-absorbed stellar emission in the ultraviolet and optical bands and its re-emission in the infrared. We use a grid of stellar population models with delayed star formation history, a Chabrier [37] initial mass function (IMF) and assumed solar metallicity. We take into account the dust extinction using the reddening law of Charlot and Fall [38] and model the infrared (IR) emission from dust (reprocessed from the absorbed UV/optical stellar emission), using the dust templates provided by Dale et al. [50]. We also use the standard nebular emission model from Inoue [112]. A similar prescription was previously used in Tortorelli et al. [211], Johnston et al. [123] and Renard et al. [179], except now a correction factor is added. SEDs of these models are firstly fitted to the PAUS galaxy SEDs with spec-z. The correction factor for each photometric band, which is estimated based on the ratio of observed and modelled fluxes, is applied to the full dataset and CIGALE is run over the full dataset with photo-z. For galaxies with low SNR of NB, CIGALE fails to find proper models and the physical properties are derived only based on broad-band photometry. The quality of the fit is expressed by the reduced χ^2, which is used to select a sample of galaxies with good fits. We use rest-framed magnitudes for galaxies with reduced χ^2<5 to ensure reliable stellar properties (e.g. [152, 29]). A full detailed description of the SED fitting procedure is presented in Siudek et al, in prep.

Fig. 3.14 shows the U-V and V-J colours computed by CIGALE and illustrates the cuts we make in this colour-colour plane depending on the redshift of the galaxy (eq. 3.12), as was implemented in Williams et al. [218], where they studied star-forming galaxies up to $z \sim 2$:

$$U - V \geq 0.88 \cdot V - J + \alpha, \tag{3.12}$$

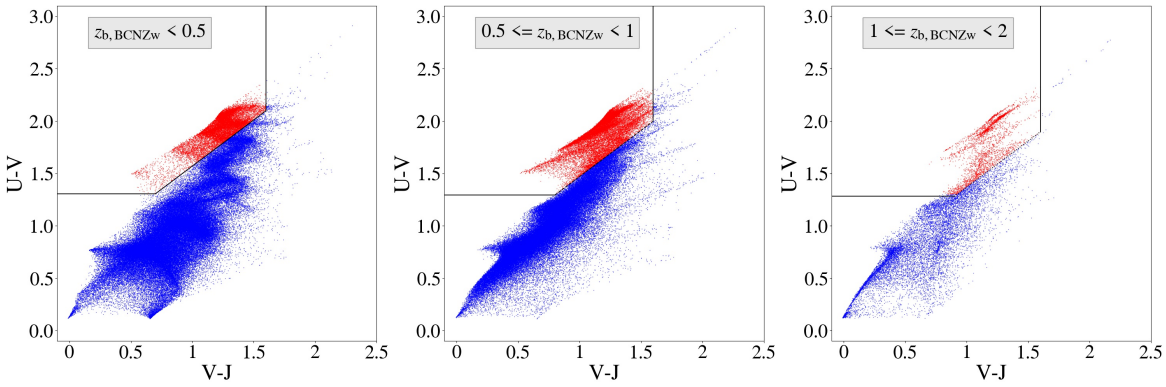

Fig. 3.14 U-V vs. V-J planes used to define blue and red galaxies. The diagonal cut varies across different photo-z ranges, shifting towards higher U-V values as the photo-z increase.

such that $\alpha = 0.69$ for galaxies with $z_{b,\,BCNZw} < 0.5$, $\alpha = 0.59$ for $0.5 \geq z_{b,\,BCNZw} < 1$ and $\alpha = 0.49$ for $z_{b,\,BCNZw} \geq 1$, where $z_{b,\,BCNZw}$ corresponds to the weighted photo-z of the galaxy, which is used to better divide the sample in colour. Additional conditions $U-V > 1.3$ and $V-J < 1.6$ were implemented in order to avoid contamination from unobscured and dusty star-forming galaxies [218]. All galaxies fulfilling these conditions are defined as red galaxies, while the rest of them are considered as blue. Using this definition and the χ^2<5 condition, we obtain 199749 red galaxies and 873047 blue ones.

Fig. 3.15 shows, from top to bottom, the σ_{68}, the outlier fraction, the bias and the SNR as a function of the i_{AB}, the photometric redshift and the spectroscopic redshift (from left to right) for red, blue and all galaxies for the PAUS wide fields. We include the SNR in this analysis to better understand the photo-z performance as a function of z_b and z_s. The percentage of red galaxies with associated spec-z oscillates between $\sim 15\% - 35\%$ in all the redshift bins and between $\sim 50\% - 10\%$ in the magnitude bins, being lower the percentage of red galaxies for faint and high redshift objects. The panel with σ_{68} as a function of i_{AB} shows that red galaxies have better photo-z statistics than blue galaxies for bright objects, while for faint objects, red galaxies perform worse, showing larger photo-z scatter. For low redshift galaxies, the performance is also better for red objects, while it is slightly worse for those at higher redshifts. The outlier fraction for red galaxies is lower for objects with $i_{AB} < 21$ while, as a function of the redshift, the outlier fraction for red galaxies is in general lower for the whole redshift range, specially at $z_b < 0.75$. Regarding the bias, red galaxies present larger values than blue galaxies for fainter galaxies ($i_{AB} > 21$), while for the rest of the i_{AB} and redshift range the bias is comparable between both types of galaxies. Finally, the SNR shows a clear difference between red and blue galaxies as a function of z_b and z_s, with blue galaxies having lower SNR at low redshift bins. This can be explained by the fact

that, for a given redshift bin, red galaxies are more massive and luminous than their blue counterparts, leading to higher SNR values and better photo-z accuracy. In order to test the dependence on redshift without the effect of the SNR, we divided the sample in 3 bins of SNR: SNR < 3, 3<SNR<6, 6<SNR<12. We only found better performance for red galaxies in the highest SNR bin for all z, while in the other bins the performance was very similar and independent of z.

3.4.4 Validation of the $p(z)$

BCNZ provides us with the redshift probability distribution, $p(z)$, for each galaxy, where the photo-z is determined by the maximum of that probability. However, a correct estimation of $p(z)$ allows us to use the photometric redshift not just as a point determined by the maximum, but as a probability density across the whole redshift space. This is useful for some science applications, such as [161] and [8], where they estimate galaxy clustering from the full $p(z)$ distribution. Nevertheless, the probability distribution obtained from SED template-fitting codes may be biased by degeneracies in the colour-redshift relation, non-representative SED templates, focusing effects and other inaccuracies [68]. Thus, in this section we try to correct the $p(z)$ that BCNZ outputs and we compare its performance with the photo-z given by the maximum of the $p(z)$.

The usual way of determining the validity of the $p(z)$ is by studying the probability integral transform (PIT, [57]) which, in the case of redshift calibration, we define as the cumulative probability distribution up to the spectroscopic redshift of a given galaxy:

$$\zeta(z_s) = \int_0^{z_s} p(z)\mathrm{d}z, \tag{3.13}$$

where the integral of the redshift probability distribution is carried out from 0 to the spectroscopic redshift (z_s).

For accurate $p(z)$, it is expected that the PIT values (N_{PIT}) will be uniformly distributed over the range from 0 to 1. However, for badly estimated $p(z)$, the N_{PIT} will be biased. If the $p(z)$ of the galaxies are too narrow in comparison to the distribution of the spectroscopic redshifts, N_{PIT} will be concave. In the opposite case, the $p(z)$ will be too broad and N_{PIT} will be convex [174]. It is possible to calibrate the $p(z)$ and obtain a uniform distribution of N_{PIT} by following the procedure in Bordoloi et al. [21], where they correct the biased $p(z)$ by multiplying it by the cumulative probability distribution at each z, obtaining a corrected $p(z)$ ($p_{\mathrm{corr}}(z)$):

$$p_{\mathrm{corr}}(z) = p(z) \cdot N_{PIT}(\zeta(z)). \tag{3.14}$$

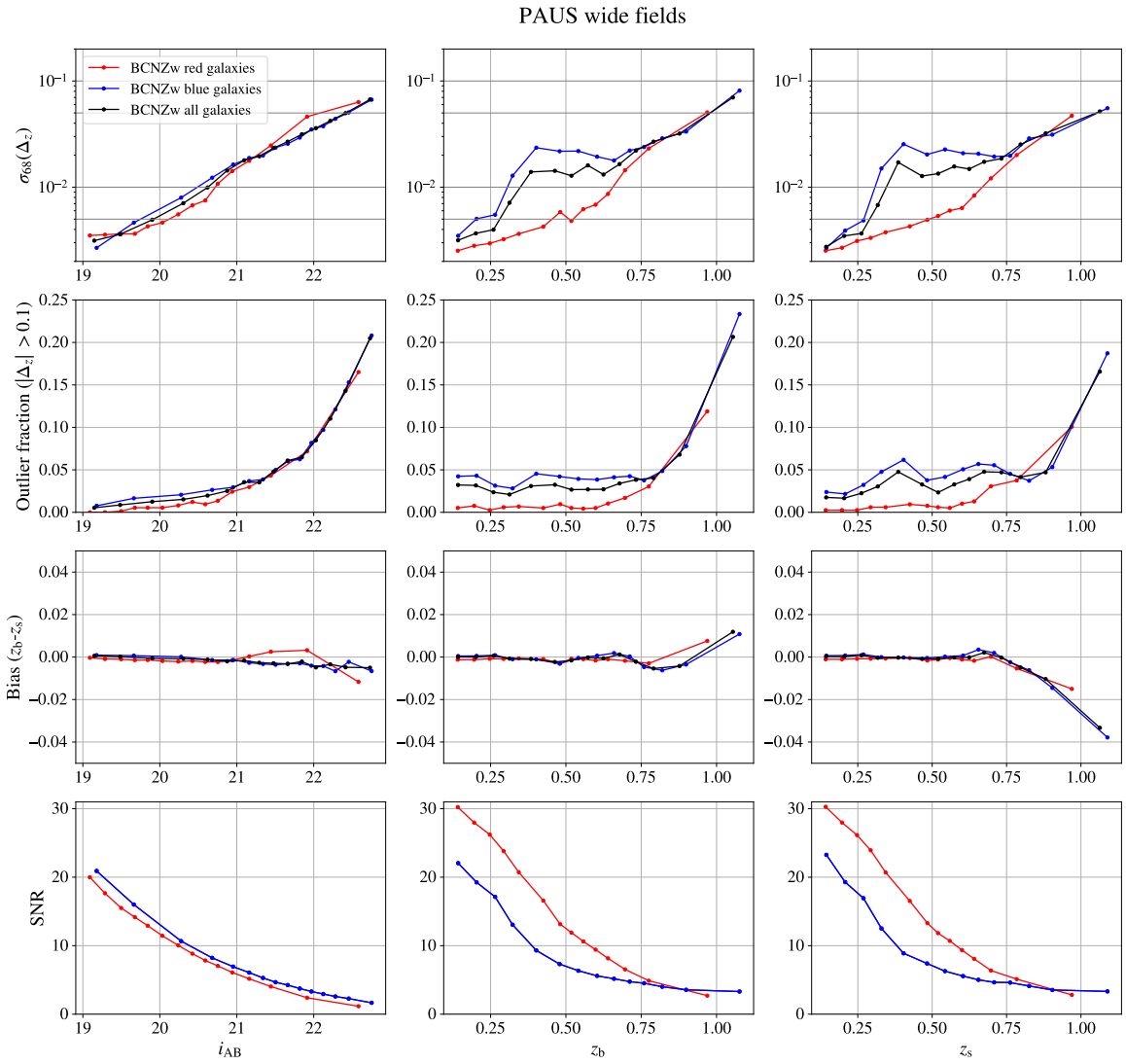

Fig. 3.15 Performance of the PAUS wide fields for red, blue and all galaxies. From top to bottom, the σ_{68}, the outlier fraction, the bias and the SNR are shown as a function of the i_{AB}, the photometric and the spectroscopic redshift. Each bin is defined to contain an equal number of objects. Bright and low redshift red galaxies exhibit superior performance compared to their blue counterparts. Conversely, faint and high redshift blue galaxies demonstrate better performance.

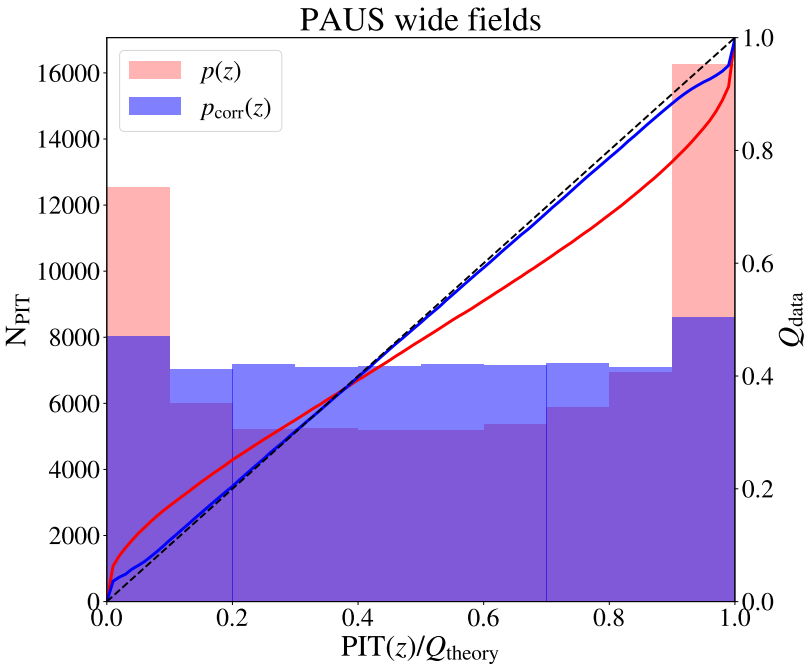

Fig. 3.16 Distribution of the PIT values computed for the $p(z)$ before (red) and after (blue) being corrected, following Bordoloi et al. [21], for the PAUS wide fields. The Quantile-Quantile plots, showing the fraction of spectroscopic redshift below a certain PIT value, are also shown. Before correcting the $p(z)$, the distribution of the PIT values show peaks at the edges, while the distribution becomes uniform after correcting it, as expected. The Quantile-Quantile plots also follow the diagonal line after the correction.

Even though this correction of $p(z)$ is only strictly valid for the subsample with spec-z information, it is reasonable to assume that applying it to the whole sample will still be valid if the spectroscopic sample is representative of the full catalogue.

Fig. 3.16 shows the $N_{\rm PIT}$ obtained from the $p(z)$ before and after being corrected for the PAUS wide fields. This figure also shows the Quantile-Quantile (QQ) plot, representing for each PIT value the fraction of spectroscopic redshifts found below it. The black dashed line shows the ideal case. We can see that the new $N_{\rm PIT}$, computed from the corrected $p(z)$, are now uniformly distributed and the QQ plots are much closer to the diagonal. It is also noticeable that, before being corrected, the $N_{\rm PIT}$ presents peaks at the edges of the PIT values, indicating that the uncorrected $p(z)$ are narrower than the distribution of the spectroscopic redshifts.

Fig. 3.17 shows the normalized n(z) distributions of the maximum of $p(z)$ (blue histogram), which corresponds to the $z_{\rm b,\,BCNZ}$ estimate, and the sum of all the individual $p_{\rm corr}(z)$ (black line). We note the change in the distribution when using a different estimate of the photo-z. In the case of comparing the performance of the photo-z from the maximum

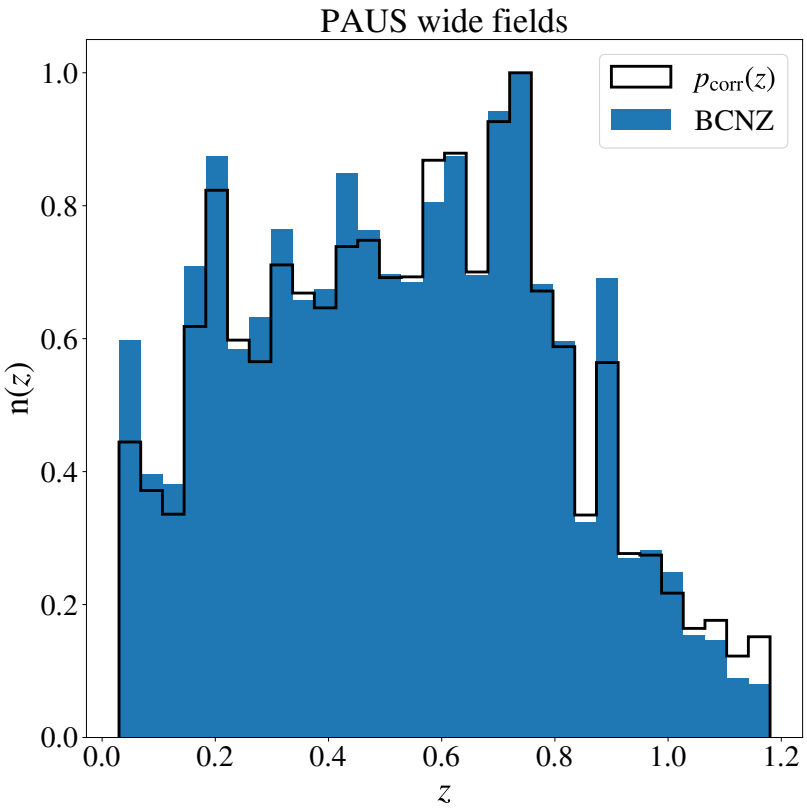

Fig. 3.17 Normalized n(z) for the PAUS wide fields as a function of the photometric redshift computed using the corrected $p(z)$ (black lines) and the estimation from the maximum of the distribution (BCNZ, blue histogram).

of the distribution or from the whole distribution, we only consider objects with photo-$z <$ 1.2, since some edge effects were being obtained if the analysis was extended until $z = 2$. We believe that the reason for this to happen is that the number of objects in the validation sample is so low in the range of $z = [1.2, 2]$ that the correction performed in eq. 3.14 fails.

Fig. 3.18 shows the performance of $p(z)$ and $p_{\rm corr}(z)$ (blue lines) in comparison with the one obtained from the maximum of $p(z)$ (names as BCNZ, black lines). Again, from top to bottom, we study the σ_{68}, the outlier fraction and the bias as a function of $i_{\rm AB}$, $z_{\rm b}$ and $z_{\rm s}$, from left to right, respectively. In this case, the metrics for $p(z)$ and $p_{\rm corr}(z)$ have been computed differently than before, since we do not define Δ_z (eq. 3.7) from the maximum of the probability distribution, but for each z in $p(z)$ and $p_{\rm corr}(z)$. As a consequence, we can transform $p_{\rm corr}(z)$ ($p(z)$) to $p_{\rm corr}(\Delta_z)$ ($p(\Delta_z)$), with $\Delta_z = 0$ corresponding to the $z_{\rm s}$ of each object. Then, we stack the $p_{\rm corr}(\Delta_z)$ ($p(\Delta_z)$) of all the objects, i, in each bin under study and obtain $\Sigma_i p_{i,{\rm corr}}(\Delta_z)$ ($\Sigma_i p_i(\Delta_z)$), from where we compute the desired metrics. From the σ_{68} plots, we see that $p_{\rm corr}(z)$ presents higher values than $p(z)$. The reason for this is that the

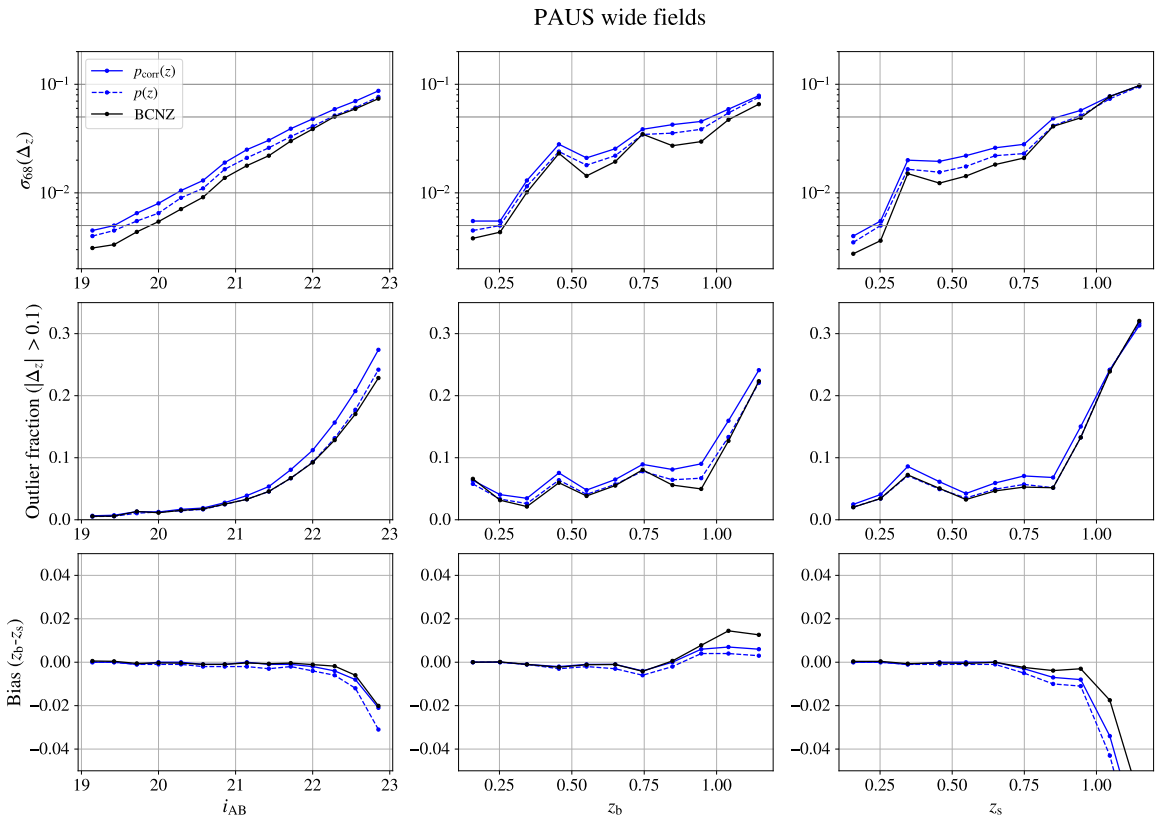

Fig. 3.18 Performance of the PAUS wide fields from the maximum of the $p(z)$ (BCNZ), from $p(z)$ and from the corrected $p(z)$ ($p_{corr}(z)$). From top to bottom, the σ_{68}, the outlier fraction and the bias are shown as a function of the i_{AB}, the photometric and the spectroscopic redshift. Each bin is defined to contain an equal number of objects. Estimating the z_b from the maximum of the $p(z)$ distribution demonstrates better performance than using the full $p_{corr}(z)$.

correction applied to $p(z)$ may distribute the probability away from its maximum, giving a more realistic $p_{corr}(z)$. Given that the σ_{68} measures the width around z_s, it is expected that this width will be larger when distributing the probability away. In the case of the maximum of $p(z)$, the σ_{68} presents lower values as a function of all the variables, with the exception of the faint and high redshift bins, where the σ_{68} of $p(z)$ and BCNZ are very similar. In the case of the outlier fraction, the maximum of $p(z)$ also presents lower values than $p_{corr}(z)$. As for the bias, it is more pronounced as a function of i_{AB} and z_s for $p(z)$, while the BCNZ and $p_{corr}(z)$ cases are more similar in all the variables, with the bias of $p_{corr}(z)$ being closer to 0 as a function of z_b. As a conclusion, this analysis shows that the photo-z accuracy reached by using the maximum of the $p(z)$ is higher than using the full $p_{corr}(z)$ distribution.

3.4.5 Q_z separation

BCNZ provides photometric redshift quality parameters that allow us to select subsamples of galaxies with the "best" photo-z. As stated in Section 3.3, we choose to use the parameter Q_z (eq. 3.5), since it is a combination of other quality parameters.

Fig. 3.19 shows the performance as a function of the quality parameter Q_z for the PAUS wide fields. Four different percentages of galaxies were selected based on Q_z, such that we retain 100%, 80%, 50% or 20% of objects in each of the bins under study with best photometric redshift estimates. Note that Q_z acts reliably for objects until $i_{\rm AB} < 21.5$ and $z < 1$, since a more restrictive Q_z cut in the catalogue yields better performance. However, for faint and high redshift objects, the performance does not improve much with the quality parameter cut. This can be seen in Fig. 3.20, where the dependence of Q_z as a function of $i_{\rm AB}$, $z_{\rm b}$ and $z_{\rm s}$ is shown. Fig. 3.19 can be useful in order to obtain a catalogue with a better performance, taking into account the loss in the number of objects.

3.4.6 Narrow-band coverage

Due to the observing strategy in PAUS, some objects may lack measurements in some bands. Thus, the recovered SED may be less precise, since some emission lines could be lost and the general shape of the SED could be less defined. However, this effect might not be relevant if the number of bands with no measurements is low. Since BCNZ does not require flux measurements in all the bands to compute the photo-z, the performance can be studied as a function of the NB coverage.

On the one hand, the left panel of Fig. 3.21 shows the number of objects as a function of the NB coverage. We decided to study the effect of a coverage of 30 NB in order not to lose much of the SED. Note that most of those objects have measurements in all 40 NB, while around $\sim$300000 objects lack some of the bands. On the other hand, the right panel of Fig. 3.21 shows the σ_{68} as a function of $i_{\rm AB}$ for two cases: objects with NB coverage greater or equal than 30 NB (red) and objects with coverage in all bands (blue). The degradation in the σ_{68} is almost negligible when reducing the coverage of NB.

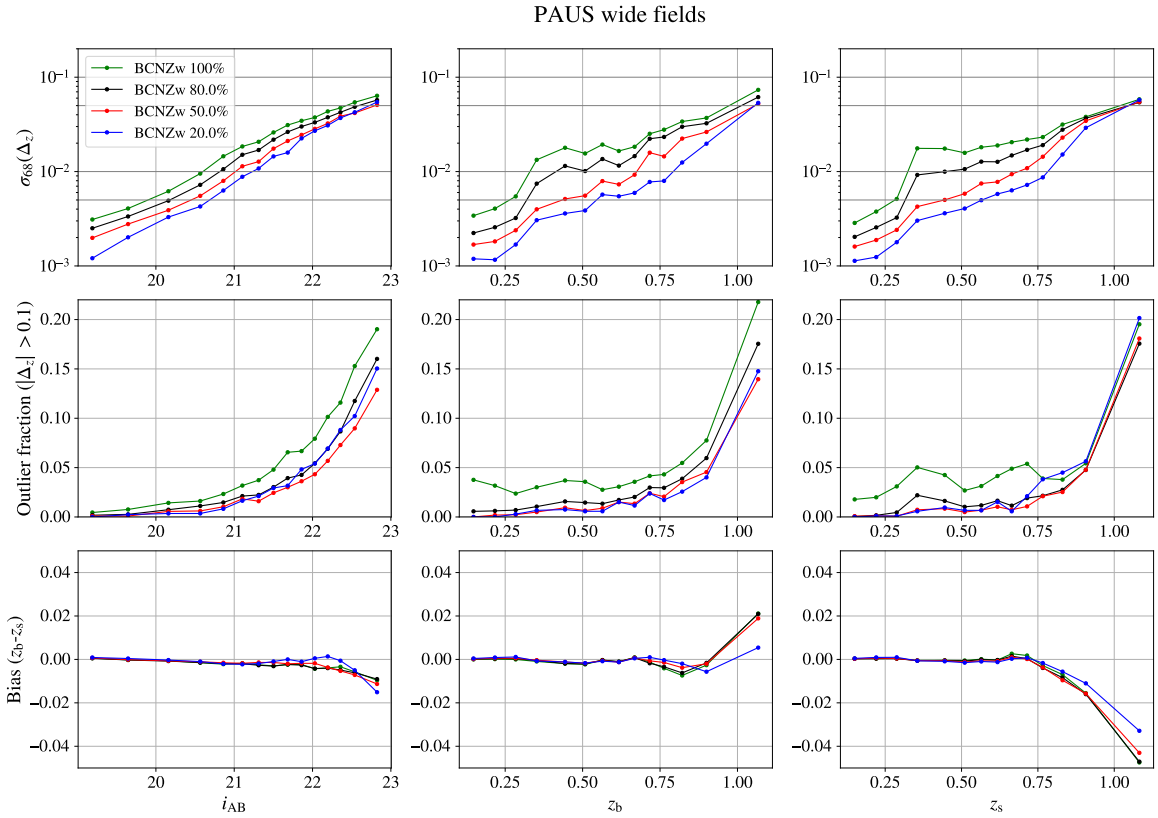

Fig. 3.19 Performance of the PAUS wide fields as a function of the quality parameter Q_z for the BCNZw photo-z. From top to bottom, the σ_{68}, the outlier fraction and the bias are shown as a function of i_{AB}, the photometric and the spectroscopic redshift. The performance is shown for the 20, 50, 80 and 100% best galaxies taking into account the quality parameter Q_z.

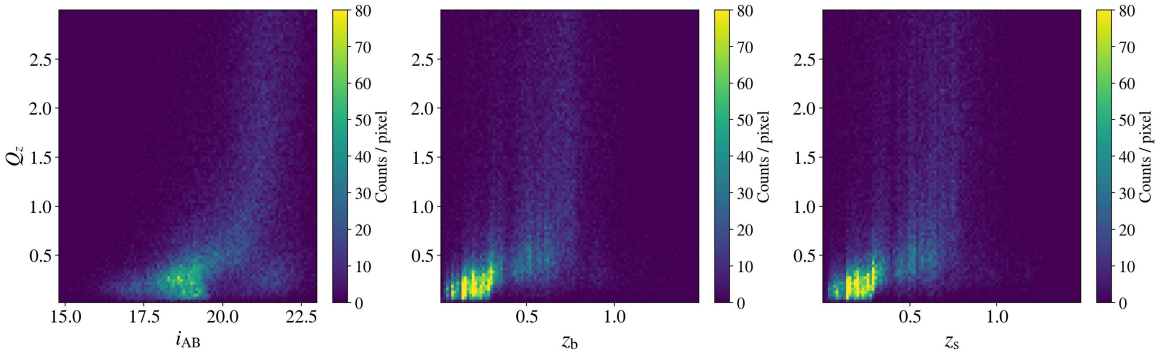

Fig. 3.20 Dependence of the quality factor Q_z on i_{AB} (left), z_b (middle) and z_s (right). At high magnitudes ($i_{AB} > 22$) and high redshifts ($z \sim 1$), the values of Q_z rise fast and are not that much correlated with the photo-z accuracy.

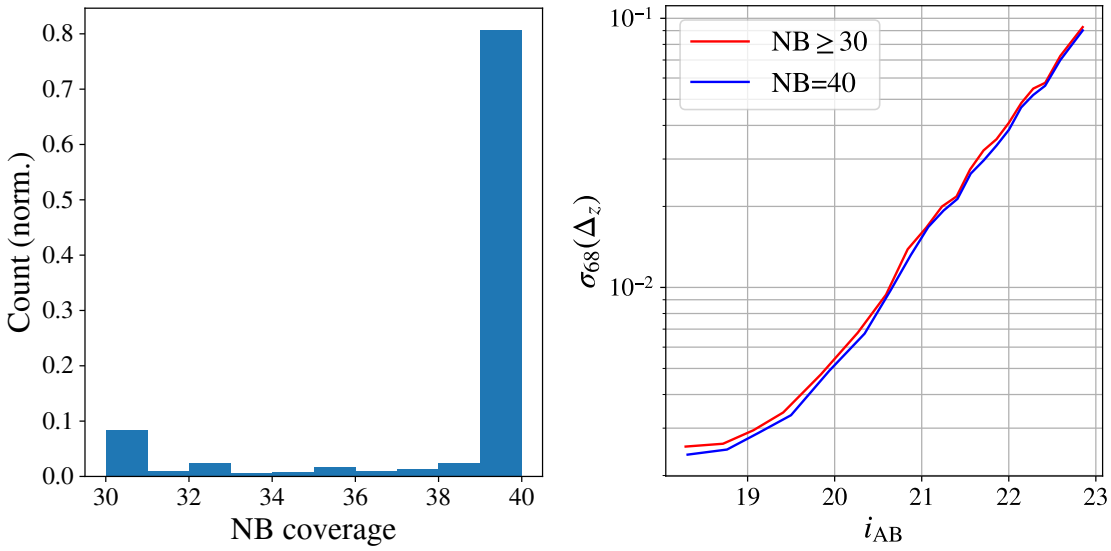

Fig. 3.21 Left Panel: Normalised number of objects as a function of the NB coverage. Around 80% of the objects with NB coverage equal or greater than 30 have measurements in all 40 NB. Right Panel: σ_{68} as a function of i_{AB} for objects in the PAUS wide fields with a NB coverage greater or equal than 30 NB (red) and for a NB coverage of 40 NB (blue). The degradation in the σ_{68} is negligible in comparison with the number of objects recovered by including a more relaxed NB coverage condition.

Chapter 4

Measuring galaxy clustering and intrinsic alignments in the PAUS wide fields

4.1 Motivation

In this Chapter 4, we study the Intrinsic Alignments (IA) and the GC in the PAUS fields. In order to do that, we use the photo-z computed in the previous Chapter 3 and the galaxy shapes from the BB surveys. We perform the measurements with a division in colour, redshift and luminosity and fit the GC and the IA parameters with the NLA model.

This is an important project because it allows to study the IA effect for intermediate and non-linear scales, with precise photo-z and at number densities that are difficult to reach with other surveys. The constraints on the IA parameters determined in this Chapter 4 may be useful to set informative priors in other last generations surveys, such as Euclid, LSST or DESI. The evolution of the IA amplitude with redshift and luminosity is also studied and compared with previous literature.

The Chapter is organized as follows. In Section 4.2, we describe the data used in the analysis, including the galaxy shape estimation, the colour selection and the MICE simulation. In Section 4.3, we define the estimators to measure the GC and the IA. Next, in Section 4.4, we describe the theoretical modelling employed in order to fit photometric correlation functions. Then, in Section 4.5 we explain the consistency tests performed during the analysis, related with the measurements and modelling of the observables. Finally, in Section 4.6 we show the results we obtain, as a function of colour, redshift and luminosity.

4.2 Data

4.2.1 Galaxy shapes

The galaxy shape estimation is a crucial point in any IA analysis. In this chapter, we follow the galaxy shear calibration technique described in Kaiser et al. [125], with modifications in the weight function by Hoekstra et al. [103] and the implementation of a multiplicative bias by Hoekstra et al. [104]. For a detailed description of this process, we refer the reader to those papers. As a summary, the following lines will describe the main aspects necessary to understand the shear estimation in our analysis.

The shapes of the objects are quantified by their polarization, which are defined from the central second moments of the galaxy images:

$$\varepsilon_1 = \frac{I_{11} - I_{22}}{I_{11} + I_{22}} \quad \text{and} \quad \varepsilon_2 = \frac{2I_{12}}{I_{11} + I_{22}}, \tag{4.1}$$

where the second moments are

$$I_{ij} = \frac{1}{I_0} \int \mathrm{d}^2\mathrm{x}\, x_i x_j W(\mathrm{x}) f(\mathrm{x}), \tag{4.2}$$

with I_0 the weighted monopole moment, x_i the i position of the galaxy image, $f(\mathrm{x})$ the observed galaxy image and $W(\mathrm{x})$ a weight function to reduce the sky noise. In order to choose the width of the weight function, we follow the same configuration used in Johnston et al. [123], that is, 1.75 times the observed half-light radius of the galaxy.

The shift in the polarization due to the shear can be expressed as:

$$\delta\varepsilon_\alpha = P^{\mathrm{sh}}_{\alpha\beta}\gamma_\beta, \tag{4.3}$$

where $P^{\mathrm{sh}}_{\alpha\beta}$ is the shear polarizability, which is a tensor that captures the response of an object to a shear γ.

However, the polarization might also be affected by an anisotropic PSF, distorting the galaxy images as a function of the position, and it needs to be corrected before applying the shear polarizability in eq. 4.3. The change in polarization due to an anisotropic PSF is shown in eq. 4.4:

$$\delta\varepsilon_\alpha = P^{\mathrm{sm}}_{\alpha\beta}p_\beta, \tag{4.4}$$

where $P^{\rm sm}_{\alpha\beta}$ corresponds to the smear polarizability, which captures the response of an object due to the convolution with an anisotropic PSF, and p_β is a measure of the PSF anisotropy. The latter can be measured by using the known polarization and smear polarizability of stars:

$$p_\alpha = \frac{\varepsilon^*_\alpha}{P^{{\rm sm}\,*}_{\alpha\alpha}}. \tag{4.5}$$

Finally, to correct for the circularization of the galaxy images due to their convolution with the PSF, the preseeing shear polarizability is computed following eq. 4.6:

$$P^\gamma = P^{\rm sh} - \frac{P^{\rm sh}_*}{P^{\rm sm}_*} P^{\rm sm}. \tag{4.6}$$

Even though P^γ is a 2x2 tensor, we assume it is a diagonal matrix with both elements having the same value, since this symmetrises the seeing correction. We also select objects with $P^\gamma > 0.1$ to remove noisy shape measurements.

Combining eq. 4.3, 4.4 and 4.6, we obtain the polarization after correcting for an anisotropic PSF:

$$\varepsilon^{\rm corr}_i = \frac{\varepsilon_i - P^{\rm sm}_{ii} p_i}{P^\gamma} \tag{4.7}$$

Although the polarizations obtained from eq. 4.7 yield good estimates, an incomplete correction for the PSF contamination might lead to multiplicative (μ_i) and/or additive bias (c_i) [97, 104] as expressed in eq. 4.8:

$$\varepsilon'_i = (1+\mu_i)\varepsilon^{\rm corr}_i + c_i \tag{4.8}$$

As shown in Johnston et al. [123], the additive bias is 0 in our case since the PSF is axisymmetric. As for the multiplicative bias, we follow the same procedure as in Hoekstra et al. [104] and Johnston et al. [123], where they determine this value from simulated images, mainly based on the size of the galaxy and its SNR. We refer the interested readers to these papers for further details. Since μ_1 and μ_2 are compatible with one another, we take the mean value for each galaxy. We compute the mean multiplicative bias for each field and sample selection we study in Section 4.6, which include splits in red and blue galaxies, in redshift and in luminosity bins.

An important remark to take into account is that we aim to perform IA measurements with a cut in $i_{\rm AB}$ magnitude. This poses the challenge that the CFHTLenS and the KiDS surveys use different bands to measure the galaxy shapes, being the $i_{\rm AB}$ band the one employed by CFHTLenS and the $r_{\rm AB}$ the one employed by KiDS. This leads to different completeness

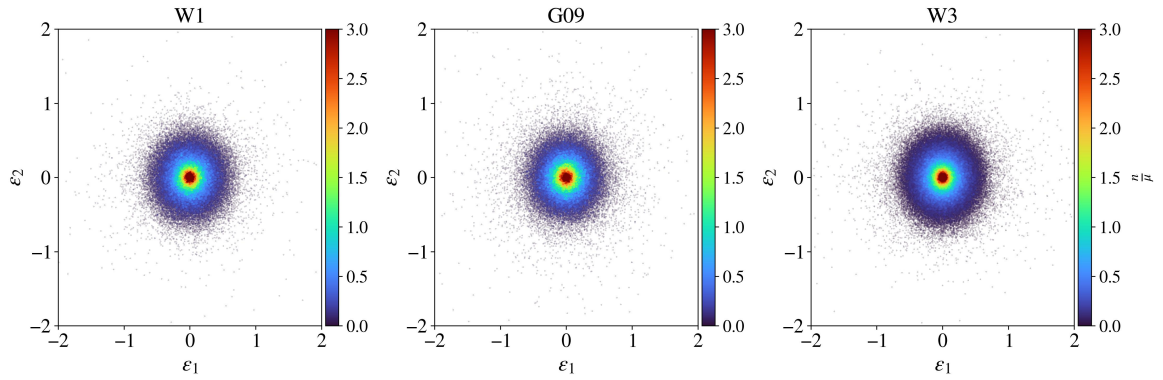

Fig. 4.1 Distribution of the polarizations, ε_1 and ε_2, for the W1, G09 and W3 fields after having corrected for the PSF and the multiplicative bias.

in the galaxy shape estimation as a function of the limiting magnitude $i_{\rm AB}$ for both surveys. In the case of CFHTLenS, galaxy shape measurement up to $i_{\rm AB} = 22.5$ are available and complete. Nevertheless, for the KiDS sample, galaxy shape measurements with good quality are only complete up to $i_{\rm AB} = 22$. For that reason, we will perform measurements of IA for two scenarios. The baseline scenario is the measurement of the combined CFHTLenS (W1 and W3) and KiDS (G09) fields up to $i_{\rm AB} = 22$. The second scenario extends the measurements to deeper magnitudes ($i_{\rm AB} = 22.5$) but only using the CFHTLenS fields.

Fig. 4.1 shows the distribution of ellipticities for the W1, G09 and W3 fields up to $i_{\rm AB} = 22$, after having corrected for the PSF and the multiplicative bias. The similarity in the distributions is an indication that the ellipticities of the three fields are consistent with one another. The comparison of the ellipticities for the case of the fainter CFHTLenS samples was also tested and it also yields similar distributions.

4.2.2 Restframe magnitudes and colours

A common approach in the measurements of IA is to characterize them in red and blue galaxies. Galaxy colours and morphology tend to be closely related, with red galaxies mainly corresponding to elliptical and passive types and blue galaxies to spiral and active.

Considering this, we use the PAUS physical properties derived in the previous Chapter 3 by performing SED fitting with CIGALE [20], a code that takes into account the dust absorption affecting stellar emission in the UV and optical bands and its re-emission in the IR. As stated in the previous Chapter 3, CIGALE does not perform well for PAUS objects with low SNR in the NB. This particularly impacts the estimation of the physical properties we will use to separate between active and passive galaxies in this Chapter 4. As a consequence,

we use the physical properties estimated with BB-only fits, as they perform better in the separation. We also perform a cut for objects with a reduced $\chi^2 > 5$, where χ^2 is related to the quality of the CIGALE fit.

We split our samples between active and passive populations using a (NUV-r) vs. (r-K) (NUVrK from now on) diagram, as it was done in Arnouts, S. et al. [7]. On the one hand, the NUV-r colour traces the specific star formation rate (sSFR), given the capability of the NUV and the r bands to track the star formation of young and old populations, respectively [181]. On the other hand, the $r-K$ colour accounts for dust attenuation in active galaxies and helps to break degeneracies related to this effect in the NUV-r cut [7]. The redshift range that was tested in Arnouts, S. et al. [7] was $0.2 \leq z \leq 1.3$, which mainly corresponds to the redshift range we consider in this analysis, which is $0.1 \leq z \leq 1$. The separation of active and passive galaxies following a NUVrK diagram is similar to the one performed in Williams et al. [219] following a $(U-V)$ vs. $(V-J)$ diagram, which was also used in the separation between active and passive galaxies in Chapter 3 in order to study the performance of photo-z as a function of colour. In this case, we opt for the NUVrK diagram, since it allows expanding the covered wavelength range analysed in the separation. We note that the terms "active/blue" and "passive/red" are used interchangeably throughout the rest of this chapter.

Following Davidzon, I. et al. [53], we define 2 cuts in the colour-colour space to separate between active and passive galaxies. In their case, they divide the NUVrK space in 3 regions: active (galaxies fulfilling eq. 4.9), green-valley (galaxies fulfilling eq. 4.10 but not eq. 4.9) and passive (the rest of the diagram):

$$\mathrm{NUV}r > 3.75 \ \ \& \ \ \mathrm{NUV}r > 1.37rK + 3.2 \ \ \& \ \ rK < 1.3 \tag{4.9}$$

$$\mathrm{NUV}r > 3.15 \ \ \& \ \ \mathrm{NUV}r > 1.37rK + 2.6 \ \ \& \ \ rK < 1.3 \tag{4.10}$$

However, in our case we can not aim to reach that level of precision when separating between active and passive galaxies, since we are using photo-z instead of spec-z. Nevertheless, given that the region located in the space defined by eq. 4.9 and 4.10 corresponds to galaxies which are reducing their sSFR, and may be in the transition between active and passive galaxies, we assign objects as active or passive so as to obtain a similar percentage of red and blue galaxies in all our PAUS fields. In particular, we assign the W1 and W3 objects that lie in this intermediate colour-colour space as passive and the ones from G09 as active. This way, we find that the percentage of passive galaxies is 18.1% (16.0%), 23.8% (20.5%) and 19.5% (17.2%) for the W1, G09 and W3 fields, respectively, at $i_{\mathrm{AB}} = 22.0$ ($i_{\mathrm{AB}} = 22.5$).

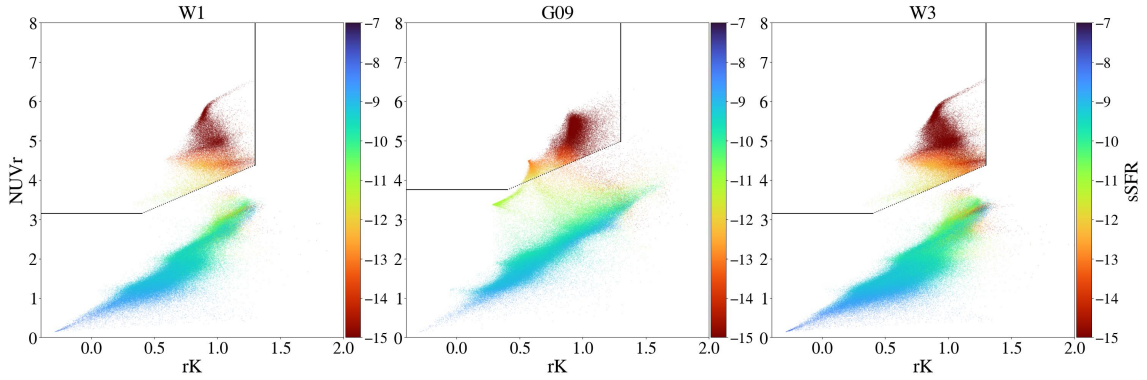

Fig. 4.2 Division in active and passive galaxies following a NUVrK and a $T_{\rm BPZ}$ selection, coloured by the sSFR obtained with CIGALE. The top regions delimited by the lines correspond to passive galaxies, while the complementary regions correspond to active galaxies. This separation is also confirmed by the values of the sSFR.

In order to secure our separation between active and passive galaxies, we also use the $T_{\rm BPZ}$ parameter employed in CFHTLenS and KiDS, allowing to separate objects by their spectral type, which is obtained by performing SED fitting using the BPZ photometric redshift estimate. This way, we are also conducting the separation based on morphological information. With this new parameter, we define red objects as those having $T_{\rm BPZ} \leq 1.5$ [200] and blue for the rest. For red galaxies, the difference in the number of objects after applying $T_{\rm BPZ}$ does not change much. The agreement in the objects classified as red with the NUVrK diagram and those classified as red following $T_{\rm BPZ}$ is 99.98%, 93.23% (94.47% at $i_{\rm AB} = 22.5$) and 99.98% of galaxies, for the W1, G09 and W3 fields, respectively. However, the case of blue galaxies is different, with $\sim 90\%$ ($\sim$87% at $i_{\rm AB} = 22.5$) of objects classified as blue by the NUVrK cut and still being considered blue by the $T_{\rm BPZ}$ criteria for all the fields.

Fig. 4.2 shows the NUVrK diagram used to separate between red and blue galaxies for the W1, G09 and W3 fields (from left to right). As stated before, the diagonal lines are different between fields, with eq. 4.9 and eq. 4.10 delimiting the region for active galaxies in the case of the W1/W3 and the G09 fields, respectively. It can be noted that there is a lack of objects just below the diagonal lines, which is produced by the additional cut performed in $T_{\rm BPZ}$. As a visual confirmation, it is also useful to note that active galaxies have a larger sSFR (computed by CIGALE) than passive galaxies, as indicated by the colour bar of these diagrams. The percentage of red objects after applying both the NUVrK and the $T_{\rm BPZ}$ filters is 20.2% (17.5%), 24.7% (21.4%) and 21.9% (19.0%) for the W1, G09 and W3 fields, respectively, at $i_{\rm AB} = 22.0$ ($i_{\rm AB} = 22.5$).

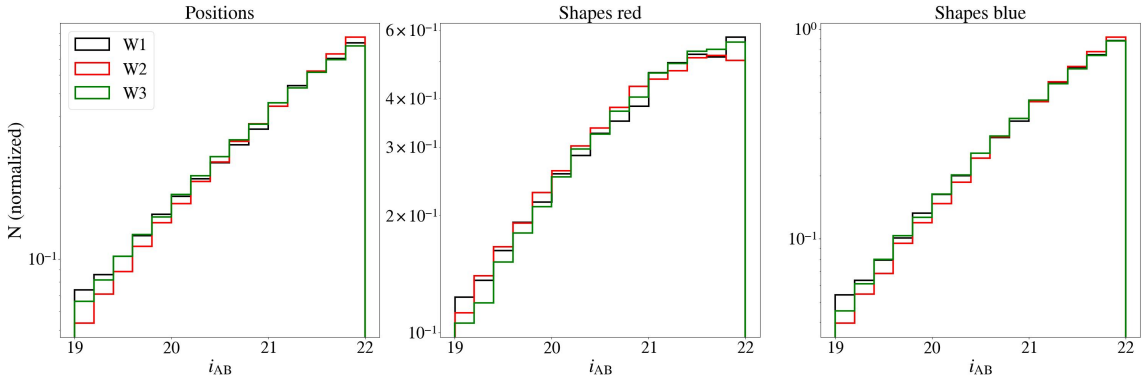

Fig. 4.3 i_{AB} distribution of the galaxy positions (left), the red galaxy shapes (middle) and the blue galaxy shapes (right) for each of the PAUS wide fields.

Fig. 4.3 shows the i_{AB} distribution for the galaxy positions (left), the red galaxy shapes (middle) and the blue galaxy shapes (right) for the three wide fields under study. The observed magnitude distribution in all fields is very similar for both the positions and the shapes separated by colour, which is an important aspect to take into account when combining the measurements of GC and IA, as we will do in our results in Section 4.6. In this case, we show the shape samples up to $i_{AB} = 22$ for conciseness. Nevertheless, the fainter shape samples of the CFHTLenS fields also have very similar i_{AB} distributions.

4.2.3 MICE

The Marenostrum Institut de Ciències de l'Espai Grand-Challenge (MICE-GC) [77, 48, 78, 34, 105] is an N-body simulation ran with the public code GADGET-2 [203]. It contains 4096^3 dark matter particles in a comoving volume of $(3\text{Gpc}h^{-1})^3$ and assumes a flat ΛCDM cosmology with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.044$, $n_s = 0.95$, $\sigma_8 = 0.8$ and $h = 0.7$.

Galaxies are introduced by combining the HOD [116, 192, 188] and the SHAM techniques [213, 46, 212]. The galaxy mock is built to follow the observed colour distributions and luminosity function, as well as the spatial GC, which is calibrated with data from SDSS [220].

Intrinsic alignments are introduced in MICE-GC by assigning intrinsic shapes and orientations to the MICE simulation up to redshift $z = 1.4$ [106]. This is done by following a semi-analytic IA model, where the intrinsic shapes and orientations are assigned based on the galaxy colour and the galaxy type (central or satellite). Red central galaxies have their 3D principal axes aligned with their host halo, while blue central galaxies have their minor axis aligned with the angular momentum of the host halo and their major axis randomly

oriented in the perpendicular plane of the minor axis. In the case of satellite galaxies, their major axes are oriented towards the host halo and their minor axes are randomly oriented in the perpendicular plane of the major axis. These alignment configurations are based on the following assumptions, that is, red central galaxies are pressure supported objects affected by the same tidal field of their host halo, blue central galaxies are rotationally supported and do not present IA and both red and blue satellite galaxies are preferentially oriented towards the centre of their host halo. The colour separation in MICE is defined with a $u-r = M_u - M_r > 0.94$ cut, where M_u and M_r correspond to the absolute rest-frame magnitudes in the CFHT-u and Subaru-r bands, respectively. Together with the semi-analytic IA model, the IA parameters are calibrated against the COSMOS [136, 93] and BOSS LOWZ [196] surveys.

We use this galaxy mock in order to validate and perform some consistency checks, which will be explained in detail in Section 4.5. Our first goal is to construct a subsample of the MICE-GC so that it resembles the PAUS wide fields. First, we cut the galaxy mock such that $19 < i_{\rm AB} < 22.5$ using the MICE synthetic observed magnitude in the COSMOS CFHT-i band, which is the one that most resembles the CFHTLenS and KiDS $i_{\rm AB}$ magnitudes. We chose to adopt the cut of the fainter samples ($i_{\rm AB} = 22.5$), instead of the brighter samples cut at $i_{\rm AB} = 22$ since the validations we want to perform are dependent on the photometric redshift quality, which worsens with fainter magnitudes. Thus, if we are able to validate our method for fainter magnitudes, the brighter case is also expected to work. Later, we introduce noise in the spectroscopic redshifts available in the MICE catalogue to generate PAUS-like photo-z in the simulation. For that, we use the galaxy mock designed by Anna et al. (in prep.), which is created using the Flagship simulation [44]. This galaxy mock generates objects with PAUS-like fluxes and computes the photo-z using BCNZ and BPZ. We combine both of these redshift estimates to obtain the weighted photo-z presented in the previous Chapter 3 and inject them in the MICE simulation. In order to do so, we bin the galaxy mock presented in Anna et al. in spectroscopic redshift bins of width $\Delta z_{\rm s} = 0.01$. Then, for each bin of spectroscopic redshifts, we have a distribution of photo-z, from which we sample over the same spectroscopic redshift bins in the MICE simulation and assign the corresponding photo-z.

Fig. 4.4 shows the distribution of the photo-z versus the spec-z for the MICE simulation (left) and for the PAUS objects that have spectroscopic redshifts in the W3 field. Note how the distribution of both plots agree well. In particular, the spread along the diagonal line at large redshifts.

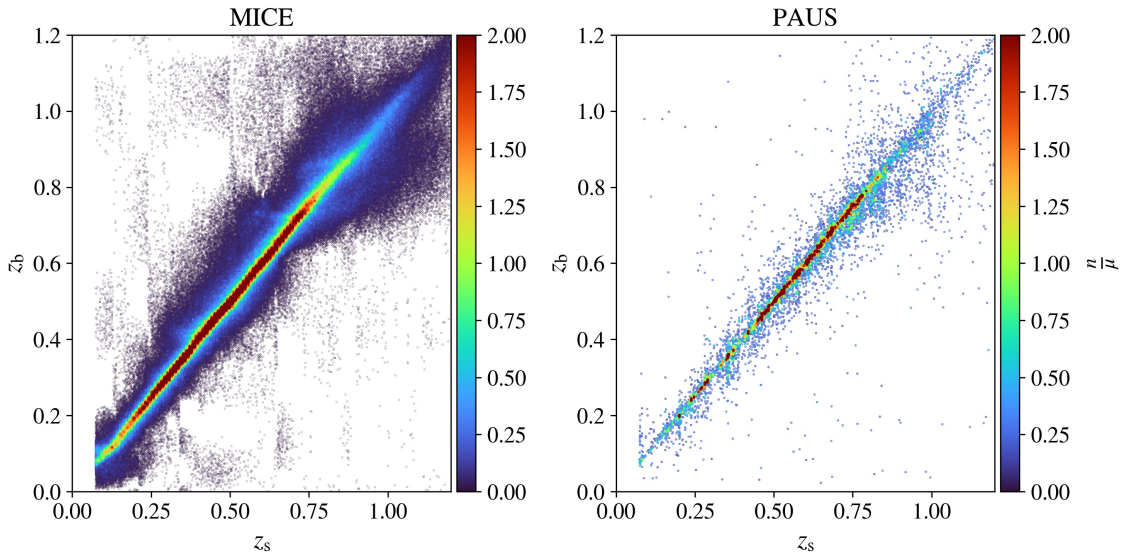

Fig. 4.4 Photometric redshift vs. spectroscopic redshift of the galaxy mock catalogue in the MICE simulation (left) and in the objects from the PAUS W3 field that have spectroscopic redshifts (right).

Fig. 4.5 shows the comparison of the $i_{\rm AB}$ and $z_{\rm b}$ distributions between the PAUS and the MICE galaxy mock, where it can be seen that the redshift and the magnitude distributions are very similar, indicating that the populations selected for both PAUS and MICE agree and that the galaxy mock is suitable to perform validation and consistency checks for our analysis.

In our analysis, we will perform measurements by combining the 3 PAUS wide fields. Thus, in order to have statistically significant results and be as realistic as possible, we divide the full octant of MICE into 180 patches, such that we create 60 combinations of 3 patches with the same area as W1, G09 and W3. This way, we will be able to obtain 60 realizations in MICE of the GC and IA measurements, which will be compared to the PAUS measurements. Fig. 4.6 shows the angular distribution of PAUS-like patches in the MICE full octant, coloured by the normalized number density of objects with respect to the median number density of the full octant.

4.3 Estimators

We compute the three-dimensional position-position correlation function using the Landy-Szalay estimator [138], binning in perpendicular and line-of-sight separations, r_p and Π, respectively:

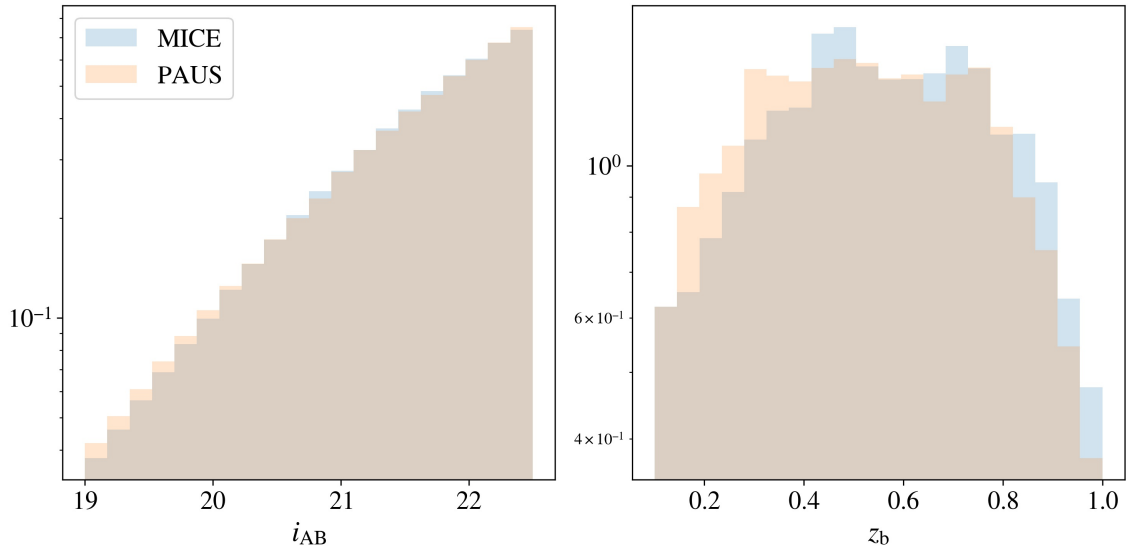

Fig. 4.5 Comparison of the distribution of i_{AB} and z_b for the PAUS wide fields and the MICE full octant.

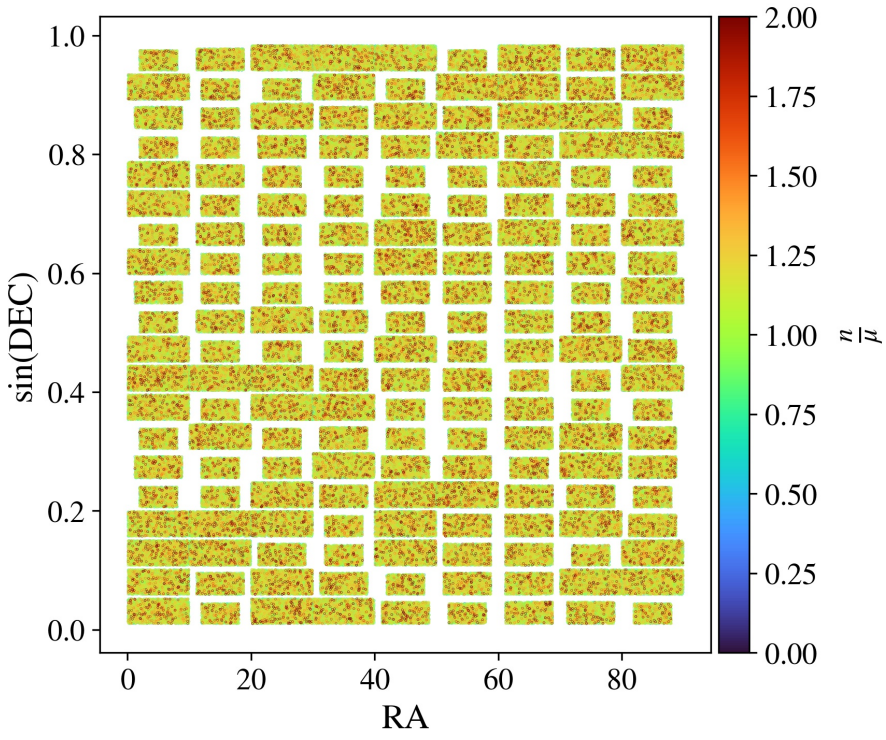

Fig. 4.6 R.A. vs DEC of the PAUS-like MICE patches. We created 60 combinations of PAUS-like patches, where each combination has one patch with a similar area as the W1, G09 and W3 PAUS wide fields. The colourbar shows the normalized number of objects with respect to the median of the full octant.

$$\xi_{gg}(r_p,\Pi) = \frac{D_iD_j(r_p) - D_iR_j(r_p) - D_jR_i(r_p) + R_iR_j(r_p)}{R_iR_j(r_p)}, \tag{4.11}$$

where D traces the positions of the density samples and R corresponds to random samples that follow the angular and radial distributions of the density samples. The length of the random catalogues is set to be 10 times that of the density sample, in order to reduce shot noise. Even though we will perform our IA analysis by separating the shape sample in red and blue galaxies, we measure the position-position correlation functions of the full density sample, since we have checked that this reduces the errors in the IA measurements, given that the number of objects being correlated is larger.

The three-dimensional galaxy-intrinsic shear correlation function is a generalization of the Landy-Szalay estimator for GC [149] and is defined as:

$$\xi_{gp}(r,\Pi) = \frac{S_+D - S_+R_D}{R_SR_D}, \tag{4.12}$$

with

$$S_+D = \sum_{i\neq j|r_p,\Pi} \frac{\varepsilon_+(j|i)}{\mathscr{R}}. \tag{4.13}$$

S_+ corresponds to the shape sample, D is the density sample and R_S and R_D are the random catalogues that trace the shape and density distributions, respectively. Here, $\varepsilon_+ = \mathrm{Re}(\varepsilon e^{2i\varphi})$, where $\varepsilon = \varepsilon_1 + i\varepsilon_2$ and φ is the polar angle connecting the correlated pair of galaxies. Again, only the shape sample is separated in galaxy colours. The expression in eq. 4.13 is the sum of the ellipticity components of galaxies i with respect to the galaxies in the density sample j, which is normalised by the shear responsivity $\mathscr{R} \sim 1 - \sigma_\varepsilon^2$, where σ_ε is the shapes' sample dispersion. As indicated in Hoekstra et al. [104], for each galaxy we weight the ellipticities in eq. 4.13 with w_i, defined as:

$$w_i = \frac{1}{\langle 0.25\rangle^2 + \left(\frac{\sigma_{\varepsilon,i}}{P_i^\gamma}\right)^2}, \tag{4.14}$$

where 0.25 is the value we adopt for the intrinsic variance of the galaxy ellipticity [102] and the subscript i refers to each galaxy in the shape sample.

An analogous estimator to eq. 4.12, ξ_{gx}, is also defined by rotating the polarizations by 45°, with $\varepsilon_x = \mathrm{Im}(\varepsilon e^{2i\varphi})$. This quantity is a measure of the direction of the curl in the shape sample and is expected to be 0, being a common way to check for systematics in the measurements.

The three-dimensional correlation functions defined in eq. 4.11 and 4.12 are projected along the line-of-sight, following:

$$w_{ab}(r_p) = \int_{-\Pi_{\max}}^{\Pi_{\max}} \xi_{ab}(r_p, \Pi) d\Pi, \tag{4.15}$$

where *ab* corresponds to the possible three-dimensional correlation functions, that is, gg, gp or gx.

The correlation functions are measured using TreeCorr [115], a package that allows for efficient computation of 2 and 3-point correlation functions. The bin slop parameter is set to 0 in our analysis. We define 12 angular bins in log space from 0.1-18 h^{-1}Mpc and define the radial binning following eq. 4.16, which is defined to have positive and negative values. This expression was also used in Johnston et al. [123], and allows accounting for the spread in the radial direction caused by photo-z, by enlarging the width of the radial separation as we move away from the object being correlated. This choice is justified to more extent in Section 4.5.2.

$$|\Pi| = 0, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233 h^{-1}\text{Mpc} \tag{4.16}$$

The errors associated to our measurements are computed using the jackknife (JK) method (eq. 4.17):

$$\text{Cov}_{\text{JK}} = \frac{N_{\text{JK}} - 1}{N_{\text{JK}}} \sum_{i=1}^{N_{\text{JK}}} (w_{ab,i} - \overline{w}_{ab})(w_{ab,i} - \overline{w}_{ab})^T, \tag{4.17}$$

where N_{JK} corresponds to the number of JK regions defined, $w_{i,ab}$ is the *ab* correlation function after removing the signal of the i-th JK region and $\overline{w}_{ab}$ is the mean of all the JK regions for the *ab* correlation function.

We define the JK regions so that we divide each of our fields in roughly equal area patches, corresponding to 4 regions in W1, 6 regions in G09 and 8 regions in W3. This number of angular JK regions was defined following Johnston et al. [123]. Additionally, we also define regions by dividing the redshift range. This division will depend on the case under study, since these have different redshift ranges. Table 4.2 includes a column indicating the number of JK patches defined in each of our studies.

4.4 Modelling

The goal of this section is to establish the theoretical expressions that allow to model the joint data vector w_{gg} + w_{gp}. Throughout this section, we will describe the modelling of both

of these expressions and the methodology we follow to constraint the galaxy bias and the IA parameters of the PAUS data. Sections 4.4.1 and 4.4.2 will describe the galaxy power spectrum and the IA power spectrum that will enter in our models, respectively. Next, we will describe how we compute the projected spectroscopic and photometric correlation functions in Sections 4.4.3 and 4.4.4, respectively. Section 4.4.5 will describe the contaminants introduced to our observables due to magnification and lensing. Finally, Section 4.4.6 will discuss the likelihood analysis performed to constrain the galaxy bias and the IA parameters, including a description of the scale cuts and the priors.

To compute the various power spectra, we use PyCCL [39] and FAST-PT [154, 72]. The linear matter power spectrum is computed using CAMB [143] and the non-linear matter power spectrum is computed with Halofit [206], both of them implemented in PyCCL.

4.4.1 Galaxy power spectrum

The relation between matter and galaxies was reviewed in Section 1.5, where we described the linear galaxy bias, eq. 1.70, and we showed an expression, eq. 1.71, that relates both fields at higher orders.

In this section, we express the galaxy overdensity field by expanding the density (δ_m) and the tidal field (s_{ij}) as [153, 10, 180]:

$$\delta_g = b_1\delta_m + \frac{1}{2}b_2(\delta_m^2 - \langle\delta_m^2\rangle) + \frac{1}{2}b_{s^2}(s^2 - \langle s^2\rangle) + b_{3\mathrm{nl}}\psi, \tag{4.18}$$

where $s^2 = s_{ij}s_{ji}$, ψ corresponds to the sum of third order non-local terms with the same scaling [180], b_1 is the linear galaxy bias, b_2 the local quadratic bias, b_{s^2} the tidal quadratic bias and b_{3nl} the third order non-local bias. One can express the galaxy-galaxy power spectrum from eq. 4.18 as [131]:

$$\begin{aligned} P_{gg}(k) = b_1^2 P_{\delta\delta}(k) &+ b_1 b_2 P_{b_1}P_{b_2}(k) \\ &+ b_1 b_{s^2} P_{b_1 s^2}(k) + b_1 b_{3nl} P_{b_1 b_{3nl}}(k) \\ &+ \frac{1}{4}b_2^2 P_{b^2b^2}(k) + \frac{1}{2}b_2 b_{s^2} P_{b_2 s^2}(k) + \frac{1}{4}b_s^2 P_{s^2s^2}(k), \end{aligned} \tag{4.19}$$

where $P_{\delta\delta}$ is the non-linear matter power spectrum, the power spectrum kernels $P_{b_1}P_{b_2}$, etc., are defined in Saito et al. [180] and we use the following co-evolution relations $b_{s^2} = -4/7(b_1 - 1)$ and $b_{3nl} = b_1 - 1$ [180, 166].

The galaxy power spectrum that enters the modelling of w_{gg} is the one expressed in eq. 4.19. However, for the case of w_{gp}, the IA model we use assumes linear galaxy bias. Thus, the galaxy-intrinsic power spectrum that enters in w_{gp} is given by:

$$P_{gI}(k,z) = b_1 P_{\delta I}(k,z), \tag{4.20}$$

where $P_{\delta I}$ is defined in the next Section 4.4.2. Nevertheless, we do not expect this to have much effect, since the galaxy bias parameters are mainly constrained by w_{gg}, due to its higher SNR.

4.4.2 IA power spectrum

We model the IA following the NLA model, which was introduced in Section 1.8.3. The TATT model (also introduced in Section 1.8.3) was also used to constraint the IA parameters, but there were high degeneracies present in our TATT constraints, probably due to the low SNR in most of the cases under study, that made us decide to focus on the NLA model.

From the different IA amplitudes for the NLA model (A_1) defined in Section 1.8.3, PyCCL implements the one from eq. 1.118, which we rewrite here:

$$A_1(z) = -\frac{\bar{C}_1 \rho_{m,0} (1+z)}{\bar{D}(z)}, \tag{4.21}$$

where $\rho_{m,0}$ is the matter density field today, $\bar{D}(z) = (1+z)D(z)$ and $D(z)$ is the growth factor.

The intrinsic-intrinsic and matter-intrinsic IA power spectra for the NLA model are defined, respectively, as

$$P_{II}(k,z) = A_1^2(z) P_{\delta\delta}(k,z) \tag{4.22}$$

and

$$P_{\delta I}(k,z) = A_1(z) P_{\delta\delta}(k,z), \tag{4.23}$$

with $P_{\delta\delta}$ the non-linear matter power spectrum.

4.4.3 Correlation functions with spectroscopic redshifts

From the power spectra computed in the previous sections, one can compute projected correlation functions along the line of sight, as the ones computed in Section 4.3. If we

assume the Limber approximation, we can define these projected correlation functions using Hankel transforms:

$$w_{gg}(r_p) = \int dz \mathscr{W}(z) \int \frac{dk_\perp dk_\perp}{2\pi} J_0(k_\perp r_p) P_{gg}(k_\perp, z), \tag{4.24}$$

$$w_{gp}(r_p) = -\int dz \mathscr{W}(z) \int \frac{dk_\perp dk_\perp}{2\pi} J_2(k_\perp r_p) P_{gI}(k_\perp, z), \tag{4.25}$$

where J_0 and J_2 are the 0th and 2nd order Bessel functions of the first kind, $k_\perp$ is the perpendicular wavelength and $\mathscr{W}$ is the projection kernel defined in Mandelbaum et al. [148] as:

$$\mathscr{W}(z) = \frac{n^i(z) n^j(z)}{\chi^2(z) d\chi/dz} \mathrm{x} \left[\int dz \frac{n^i(z) n^j(z)}{\chi^2(z) d\chi/dz} \right]^{-1}, \tag{4.26}$$

with n^i the redshift distribution of the i sample and $\chi^2(z)$ the comoving distance along the line of sight at redshift z. In the case of eq. 4.24, the redshift distributions used are the ones in the density sample, so n^i=n^j. Conversely, in the case of eq. 4.25, two samples are used, the density and the shape samples. Note that the expressions of eq. 4.24 and 4.25 are valid when we know the exact positions of the objects, that is, for the case of spectroscopic redshifts.

4.4.4 Correlation functions with photometric redshift

The modelling of projected correlation function with photo-z is more complex than when using spec-z. The effect photo-z have on the correlation functions is to scatter galaxies along the line of sight. The amplitude of this scatter will depend on how accurate are the photo-z. Even though PAUS photo-z have better accuracy than typical photo-z computed from other BB surveys, the scattering is still not negligible. One approach to account for this scatter is to increase the range of line of sight integration, as we did when defining eq. 4.16, so that the scattered objects are brought back into the correlation. Nevertheless, this is a subtle exercise, since one can not increase this range arbitrarily given that uncorrelated objects might enter the correlation function and reduce the SNR. During this Section 4.4.4 we follow the procedure in Joachimi, B. et al. [121] and Samuroff et al. [184].

For the case of the projected photometric GC correlation function, w_{gg}, the Limber integral of the galaxy power spectrum is defined as:

$$C_{gg}^{ij}(l \mid z_1, z_2) = \int_0^{\chi_{hor}} d\chi' \frac{p_g^i(\chi' \mid \chi(z_1)) \cdot p_g^j(\chi' \mid \chi(z_2))}{\chi'^2} \cdot P_{gg}(k = \frac{l+0.5}{\chi'}, z(\chi')), \tag{4.27}$$

where $p_g(\chi' \mid \chi(z_1))$ quantifies the error distribution, which corresponds to the spectroscopic redshift distribution at the comoving distance χ determined by the photometric redshift z_1. Here, the spectroscopic redshifts are the ones used in Chapter 3 to quantify the accuracy of the photo-z.

The expression in eq. 4.27 is transformed to angular space via:

$$\xi_{gg}^{ij}(\theta \mid z_1, z_2) = \frac{1}{2\pi} \int_0^{\infty} dl l J_0(l\theta)) C_{gg}^{ij}(l \mid z_1, z_2), \tag{4.28}$$

where again J_0 is the 0th order Bessel functions of the first kind. Finally, as indicated in Joachimi, B. et al. [121], a change of coordinates is performed, such that $z_m = (z_1 + z_2)/2$, $r_p \sim \theta\chi(z_m)$ and $\Pi \sim c(z_2 - z1)/H(z_m)$. With this, we can project the expression in eq. 4.28 along the line of sight as:

$$w_{gg}^{ij}(r_p) = \int_{-\Pi_{max}}^{\Pi_{max}} \int dz_m \mathscr{W}^{ij}(z_m) \xi_{gg}^{ij}(r_p, \Pi, z_m), \tag{4.29}$$

where now the redshift distributions entering in $\mathscr{W}(z)$ correspond to photo-z distributions.

In order to obtain the result of w_{gg} in eq. 4.29, we need a three-dimensional grid of $\xi_{gg}^{ij}(r_p, \Pi, z_m)$. This is done by selecting different Π and z_m values along our range of study, transforming them into z_1 and z_2 and performing the computations from eq. 4.27 and 4.28.

The procedure for the case of the w_{gp} computation is very similar and is expressed in eq. 4.30-4.32. In that case, the error distributions in eq. 4.30 correspond to the dense (p_g^i) and shape samples (p_p^j).

$$C_{gp}^{ij}(l \mid z_1, z_2) = \int_0^{\chi_{hor}} d\chi' \frac{p_g^i(\chi' \mid \chi(z_1)) \cdot p_p^j(\chi' \mid \chi(z_2))}{\chi'^2} \cdot P_{gI}(k = \frac{l+0.5}{\chi'}, z(\chi'))) \tag{4.30}$$

$$\xi_{gp}^{ij}(\theta \mid z_1, z_2) = \frac{1}{2\pi} \int_0^{\infty} dl l J_2(l\theta)) C_{gp}^{ij}(l \mid z_1, z_2) \tag{4.31}$$

$$w_{gp}^{ij}(r_p) = \int_{-\Pi_{max}}^{\Pi_{max}} \int dz_m \mathscr{W}^{ij}(z_m) \xi_{gp}^{ij}(r_p, \Pi, z_m) \tag{4.32}$$

4.4.5 Contaminants to the correlation functions via magnification and galaxy-galaxy lensing

The quantities defined in eq. 4.29 and eq. 4.32 do not take into account other possible two-point correlation functions that act as contaminants to our quantities of interest. In the case of position-position correlation functions, magnification acts as a contaminant by disturbing the galaxy number density in a region of the sky. As a consequence, besides the term described in eq. 4.27, the magnification-magnification term (mm) and magnification-galaxy position terms (gm and mg) need to be taken into account.

The terms contributing to the source-source correlation function are:

$$C_{nn}^{ij}(l) = C_{gg}^{ij}(l) + C_{gm}^{ij}(l) + C_{mg}^{ij}(l) + C_{mm}^{ij}(l), \tag{4.33}$$

where $C_{gm}^{ij} = C_{mg}^{ij}$.

On the one hand, the galaxy-magnification contribution is defined by:

$$\begin{aligned} C_{gm}^{ij}(l \mid z_1, z_2) = C_{mg}^{ij}(l \mid z_1, z_2) = 2(\alpha - 1) \int_0^{\chi_{hor}} d\chi' \\ \frac{p_n^i(\chi' \mid \chi(z_1)) \cdot q_e^j(\chi' \mid \chi(z_2))}{\chi'^2} \cdot P_{g\delta}(k = \frac{l+0.5}{\chi'}, z(\chi')), \end{aligned} \tag{4.34}$$

where q_x is the lensing weight function:

$$q_x(\chi) = \frac{3H_0^2\Omega_m}{2c^2} \frac{\chi}{a(\chi)} \int_0^{\chi_{hor}} d\chi' p_x(\chi') \frac{\chi' - \chi}{\chi'} \tag{4.35}$$

and we define α, following the theory of magnification in magnitude limited samples ([12, 117, 76]), to be proportional to the slope of the faint-end of the logarithmic galaxy count ($\log n$) over a given magnitude range, in our case:

$$\alpha(r_{\mathrm{AB}}) = 2.5 \frac{d \log[n(r_{\mathrm{AB}})]}{d r_{\mathrm{AB}}} \tag{4.36}$$

Table 4.2 shows the $\alpha(r_{\mathrm{AB}})$ values for the different configurations studied in this analysis.

On the other hand, the magnification-magnification contribution is defined by:

$$C^{ij}_{mm}(l \mid z_1, z_2) = 4(\alpha-1)(\alpha-1)\int_0^{\chi_{hor}} d\chi' \\ \frac{q^i_n(\chi' \mid \chi(z_1)) \cdot q^j_e(\chi' \mid \chi(z_2))}{\chi'^2} \cdot P_\delta(k = \frac{l+0.5}{\chi'}, z(\chi')), \tag{4.37}$$

In the case of position-shape correlation functions, the sources of contamination are the magnification and the galaxy-galaxy lensing, which introduces a signal produced by the lensing of a background galaxy by a foreground galaxy. We can decompose all the terms contributing to the position-shape correlations as:

$$C^{ij}_{ne}(l) = C^{ij}_{gI}(l) + C^{ij}_{gG}(l) + C^{ij}_{mI}(l) + C^{ij}_{mG}(l) \tag{4.38}$$

The contribution from the galaxy-shear is defined as:

$$C^{ij}_{gG}(l \mid z_1, z_2) = \int_0^{\chi_{hor}} d\chi' \\ \frac{p^i_n(\chi' \mid \chi(z_1)) \cdot q^j_e(\chi' \mid \chi(z_2))}{\chi'^2} \cdot P_{g\delta}(k = \frac{l+0.5}{\chi'}, z(\chi')), \tag{4.39}$$

The contribution from the magnification-shear is:

$$C^{ij}_{mG}(l \mid z_1, z_2) = 2(\alpha-1)\int_0^{\chi_{hor}} d\chi' \\ \frac{q^i_n(\chi' \mid \chi(z_1)) \cdot q^j_e(\chi' \mid \chi(z_2))}{\chi'^2} \cdot P_\delta(k = \frac{l+0.5}{\chi'}, z(\chi')), \tag{4.40}$$

Finally, the contribution from the magnification-intrinsic shear is:

$$C^{ij}_{mI}(l \mid z_1, z_2) = 2(\alpha-1)\int_0^{\chi_{hor}} d\chi' \\ \frac{p^i_n(\chi' \mid \chi(z_1)) \cdot q^j_e(\chi' \mid \chi(z_2))}{\chi'^2} \cdot P_{\delta I}(k = \frac{l+0.5}{\chi'}, z(\chi')), \tag{4.41}$$

4.4.6 Likelihood analysis

With the theoretical predictions (Section 4.4.4) and the measurements (Section 4.3) of our observables, we carry out a likelihood analysis to constraint the galaxy bias and the IA

parameters. As it was introduced in Section 1.9, if we assume that the parameters (θ) we want to constraint follow a Gaussian distribution, we can compute the likelihood of a data vector (D) with a certain covariance matrix (C) with respect to a given model (M) as:

$$-2\ln L(\theta) = \chi^2_{\rm fit} = \sum_{i=1}^{n}\sum_{j=1}^{n}\left[D_i - M_i(\theta)\right] C^{-1}_{ij} \left[D_j - M_j(\theta)\right]^T, \tag{4.42}$$

where n is the number of data points.

Nevertheless, given the limited area used for this analysis, we are restricted to few JK regions for each of the configurations we study. This leads to an also limited resolution in the covariance matrix, which is proportional to [84]:

$$\Delta\bar{C} \sim \sqrt{\frac{2}{N_{\rm JK}}}, \tag{4.43}$$

where $\bar{C}$ is the normalized to unit variance covariance matrix. As a consequence, the smaller the number of JK regions, the more degeneracies in $\bar{C}$ might appear. This can lead to numerical instabilities when computing the inverse of the covariance matrix in eq. 4.42. To solve this problem, we perform a singular value decomposition (SVD) of the normalized covariance matrix

$$\bar{C} = (U_{ik})^T D_{kl} V_{lj} \tag{4.44}$$

where $D_{ij} = \lambda_i^2 \delta_{ij}$ is a diagonal matrix with λ_i^2 values in the diagonal, which correspond to the singular values of the decomposition, and U_{ij} and V_{ij} are orthogonal matrices that decompose $\bar{C}$ into D_{ij}. The singular values represent the independent number of modes in the covariance matrix, and values too close to 0 may cause degeneracies in $\bar{C}$. When computing the inverse covariance in eq. 4.42, we keep the dominant singular values set by the condition:

$$\lambda_i^2 > \sqrt{\frac{2}{N_{\rm JK}}}, \tag{4.45}$$

which corresponds to the resolution limit specified in eq. 4.43.

The $\chi^2_{\rm fit}$ in eq. 4.42 in the case of applying a SVD is redefined as:

$$\chi^2_{\rm fit,SVD} = \sum_{i=1}^{n}\sum_{j=1}^{n} \bar{\Delta}_i \bar{C}^{-1}_{{\rm SVD},ij} \bar{\Delta}^T_j, \tag{4.46}$$

Table 4.1 Priors of the NLA model.

Parameter	Type	Prior
b_1	Flat	[0, 2]
b_2	Gaussian	Mean:0, std:0.5
A_1	Flat	[-12, 12]

where $\bar{C}^{-1}_{\rm SVD}$ is the inverse of the covariance matrix after performing SVD and we have defined the difference $\Delta_i \equiv D_i - M_i(\theta)$ and its normalization as $\bar{\Delta}_i = \Delta_i/\sigma_{C,i}$, with $\sigma_{C,i}$ corresponding to the square root of the diagonal of the covariance matrix evaluated at position i.

In order to sample the posterior distribution, we employ emcee [75], which is a python implementation of Goodman and Weare's Affine Invariant Markov chain Monte Carlo (MCMC) ensemble sampler [92]. We use the integrated autocorrelation time (τ_f) to quantify the Monte Carlo error. The idea behind the integrated autocorrelation time is that the samples of the chain are not independent, and one has to estimate the effective number of independent samples. This number can be quantified as N/τ_f, with N the total number of sampled points. From this, a convergence criterion can be established, where we impose that $N/\tau_f > 100$ and that τ_f changes by less than 1% every 100 iterations. Additionally, when plotting the results from the chain, we set the emcee *thin* parameter to $\tau_f/2$ to select independent points. We note that all these specifications may vary depending on the case under study, and that we follow the recommendations set by the available documentation from emcee. As a consistency test, we ran some of our chains using Nautilus [139], which is a nested sampler (Section 1.9.2) and found the same constraints.

Table 4.1 shows the priors used for b_1, b_2 and A_1 when fitting the joint $w_{gg} + w_{gp}$ data vector. In the case of b_1 and A_1, we set similar flat priors as those described in Samuroff et al. [184]. The differences are that the b_1 prior was reduced from [0, 3] to [0,2], since our samples usually present a galaxy bias closer to $b_1 \sim 1$, while the A_1 prior was increased from [-8, 8] to [-12, 12], because we were hitting the priors otherwise. For the 2nd order galaxy bias, b_2, we set a Gaussian prior with 0 mean and standard deviation of 0.5. A flat prior for this parameter was not enough, since there were some degeneracies centred in $b_2 = 0$ and $b_2 < -1$, with the second solution being less feasible.

The scale cuts in our combined data vectors are fixed to $r_{p,\rm min} = 2.0\,h^{-1}$Mpc for w_{gg}, which is typically the minimum separation we can reach with the non-linear galaxy bias model presented in Section 4.4.1. For the case of w_{gp}, we tried different options, which are reviewed in more detail in Appendix B, but decided to use the same $r_{p,\rm min} = 2.0\,h^{-1}$Mpc

as in w_{gg} for the principal constraints in this chapter. Note that this is a different approach from what is commonly used in previous literature for the NLA model, where they usually set $r_{p,\min} = 6.0\,h^{-1}$Mpc, such as in Samuroff et al. [184], Johnston et al. [122] and Fortuna, Maria Cristina et al. [76], amongst others. Nevertheless, this scale cut is more justified by the limitation in modelling the w_{gg} estimator, which is estimated from linear galaxy bias in these previous works, rather than by a limitation of the NLA model itself. This is proven in Romain et al. (in prep.), where they include an analysis on the dependence on the scale cuts when using linear and non-linear galaxy bias in the Flagship simulation. Additionally, given the reduced number of data points for the PAUS measurements, a scale cut of $r_{p,\min} = 2.0\,h^{-1}$Mpc is preferable over a scale cut of $r_{p,\min} = 6.0\,h^{-1}$Mpc, since it increases the available points to fit in our data vector, from 3 to 5, respectively.

4.5 Consistency tests

In this Section 4.5, we explain and justify the methods we develop in order to optimally measure and model the GC and the IA of galaxies in the PAUS wide fields. For this, we use the MICE simulation, described in Section 4.2.3. Given that in the final PAUS measurements we combine the three PAUS wide fields, in the following consistency tests we also combine triplets of MICE PAUS-like fields. First, we describe the generation of the random catalogues necessary to perform the measurements. Next, we show that we are able to obtain the same galaxy bias and IA parameters when using photometric or spectroscopic redshifts. Finally, we study the errors associated with the measurements using different estimators for the covariance matrix.

4.5.1 Randoms

When measuring the GC and the IA from eq. 4.11 and 4.12, we need a set of random catalogues that represent a universe with random positions. To obtain these random catalogues is not a trivial exercise, since we want to obtain catalogues where the GC is null but that, at the same time, follow the number count distribution as a function of redshift.

Since in this chapter we study projected correlation functions, we want random catalogues both in the angular and the radial directions. To obtain randoms in the angular direction is not a difficult task if we assume that our data has a homogeneous distribution in R.A. and DEC. This can be done in this analysis, since we are cutting at an $i_{\rm AB}$ magnitude much lower than the limiting magnitudes of the reference catalogue surveys. If that was not the case, we

should develop more advanced techniques, such as the ones described in Johnston, Harry et al. [124], where they develop randoms in the angular direction using SOMs. However, in this Chapter 4, the challenging task comes from developing a set of random catalogues in the radial direction. Previous papers, such as [42, 73, 123], have applied an algorithm that estimates the radial distribution of the randoms by computing a maximum volume for each galaxy, centred on its current position, around which it can be distributed. This method is designed to fit the luminosity and the overdensity functions as a function of the redshift for magnitude limited samples.

In this Chapter 4, we choose to develop a new technique that relies on the MICE simulation. It consists of comparing the GC and the IA signals for two different sets of random catalogues. In the first case, the radial distribution of the random catalogue is generated by sampling over the radial distribution of the full octant of MICE. This corresponds to the average distribution of matter without the effect of local clustering, as a result of the considerable size of the MICE area. This way, we obtain a representation of a random catalogue that does not carry galaxy clustering signal. In the second case, the radial distribution of the random catalogue is obtained by smoothing the radial distribution of the triplets of MICE PAUS-like patches with a tophat filter of a given number of Mpc. This case is the one we can reproduce when measuring the observables with the PAUS data, since the previous one is only accessible with the MICE simulation.

Once the two sets of random catalogues are defined, the GC and the IA signals are measured for both of them and the difference between the signals is analysed. An indication that both sets of random catalogues are similar is that the signals are consistent. We test different levels of radial smoothing for the second case, going from a smoothing of 20 Mpc until 420 Mpc (which corresponds to the radial distance at redshift 0.1, that is, the minimum redshift we consider to measure our correlation functions), finding that the best match corresponds to a smoothing of 420 Mpc.

Fig. 4.7 shows the comparison of the normalized radial distribution of three cases: the combination of three randomly selected PAUS-like patches, the randoms generated when smoothing these three PAUS-like patches with a tophat filter of 420Mpc and the randoms generated from sampling over the MICE full octant. Note how the radial distributions of the three cases are very similar, indicating that a smoothing of 420Mpc yields random catalogues analogous to those when sampling the MICE full octant radial distribution.

The comparison of the projected GC and IA signals from different versions of randoms is depicted in Fig. 4.8. On the one hand, the top plots of the figure show the comparison of the mean w_{gg} (left) and w_{gp} (right) signals, obtained from the combination of the 60 triplets

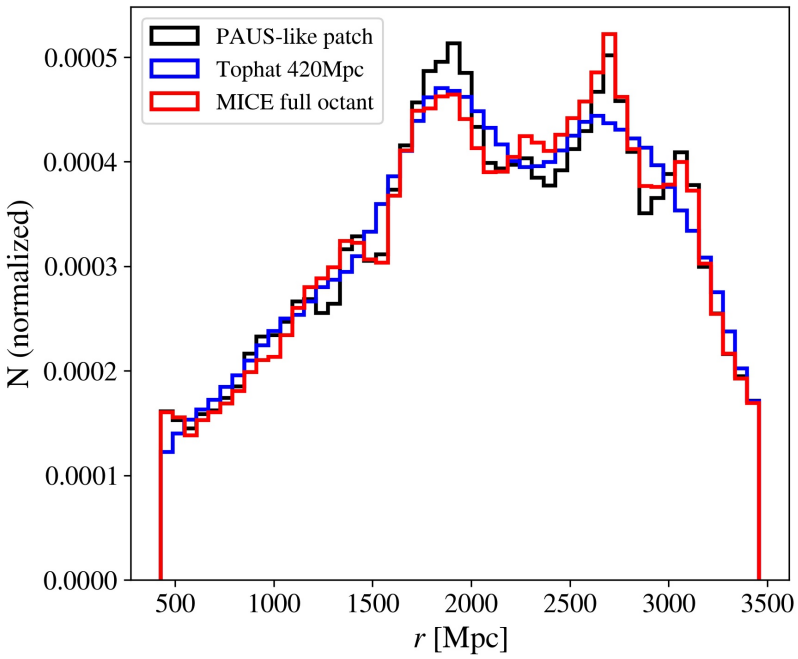

Fig. 4.7 Normalized radial distribution of the MICE data, the random catalogues following the radial distribution of 3 patches with a smoothing of 420 Mpc and the radial distribution of the full octant of MICE.

of MICE PAUS-like patches. On the other hand, the bottom plots show the difference of the signal divided by the error, which has been computed as the square root of the diagonal of the ensemble covariance of the 60 triplets (see Section 4.5.3 for more details). It can be seen that the signal obtained with both random catalogues is consistent with a difference below 1σ. Two important remarks can be extracted from this figure. First, the GC signal is more affected than the IA signal when using different random catalogues. This is expected, since the estimator of ξ_{gp} (eq. 4.12) correlates shapes with positions, while the estimator of ξ_{gg} (eq. 4.11) correlates positions with positions, which is more sensible to the random catalogue used. Second, although the signals are consistent, the case of the smoothed randoms with a tophat filter yield slightly less signal than the randoms from the MICE full octant. The reason for this is that the smoothed randoms still carry some clustering, which reduces the total signal from GC and IA. However, we do not expect this to affect our constraints, as shown in the next Fig. 4.9.

The importance of constructing suitable random catalogues comes from the fact that we want to recover the true galaxy bias parameters, since they are fundamental in order to obtain the true IA parameters (see eq. 4.20 and 4.30). Following with the discussion on the difference between the GC and IA signals, in Fig. 4.9 we show the constraints on the galaxy

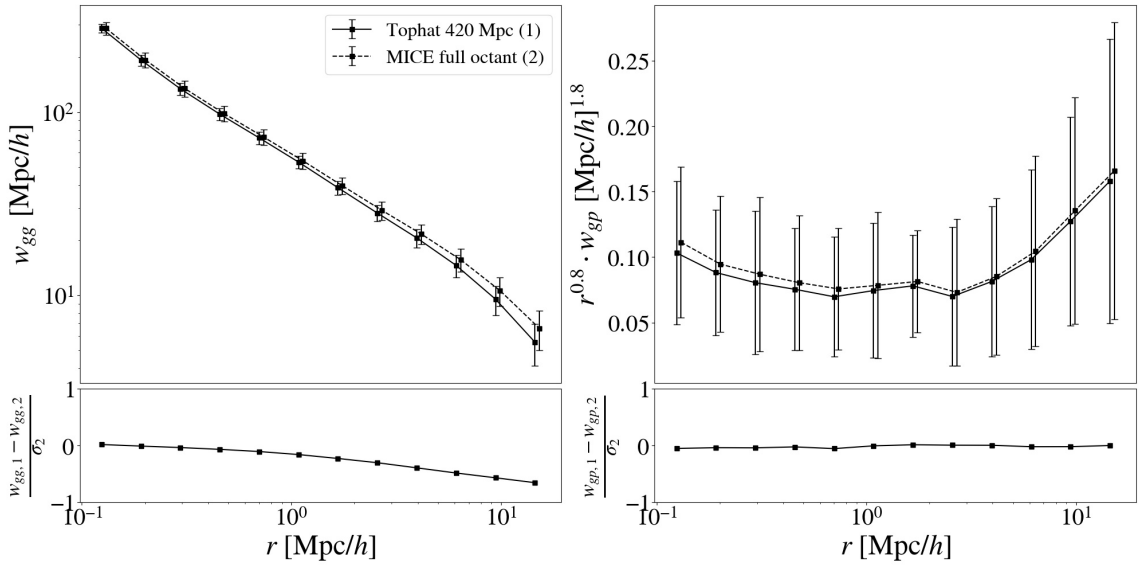

Fig. 4.8 (Left): Galaxy clustering measurements in MICE with different random catalogues. Both types of randoms have consistent measurements, with a difference of less than 1 σ as seen in the bottom panel. (Right): Same as left plots, but showing the IA measurements. In this case, the difference between both types of randoms in terms of σ is negligible.

bias and the IA parameters for different sets of random catalogues. Blue contours show the case when using the randoms sampled from the MICE full octant distribution, while red contours show the randoms obtained when smoothing the radial distribution with a tophat filter. The recovered parameters are consistent between one another, specially for the IA amplitude, A_1.

4.5.2 Comparison of the photometric and spectroscopic signals

In this section, we compare the correlation functions in MICE when using photo-z or spec-z and test if the constrained GC and IA parameters are consistent for both cases.

It is expected that both the GC and the IA signals are smeared out along the radial direction when using photometric instead of spectroscopic redshifts. This is explained by the fact that the positions of the objects are quantified with a lower precision when using photo-z, and so the correlation between objects decrease. One way to correct for that is to increase the line of sight integration in eq. 4.15, in order to recover the pairs of correlated objects that have been spread over this radial direction. However, it is important not to define an excessively wide radial range, since this might turn into an increase of the noise. We define two radial binnings, one for the spectroscopic and one for the photometric cases. On the one hand, we take the spectroscopic radial binning from previous works, such as Georgiou et al.

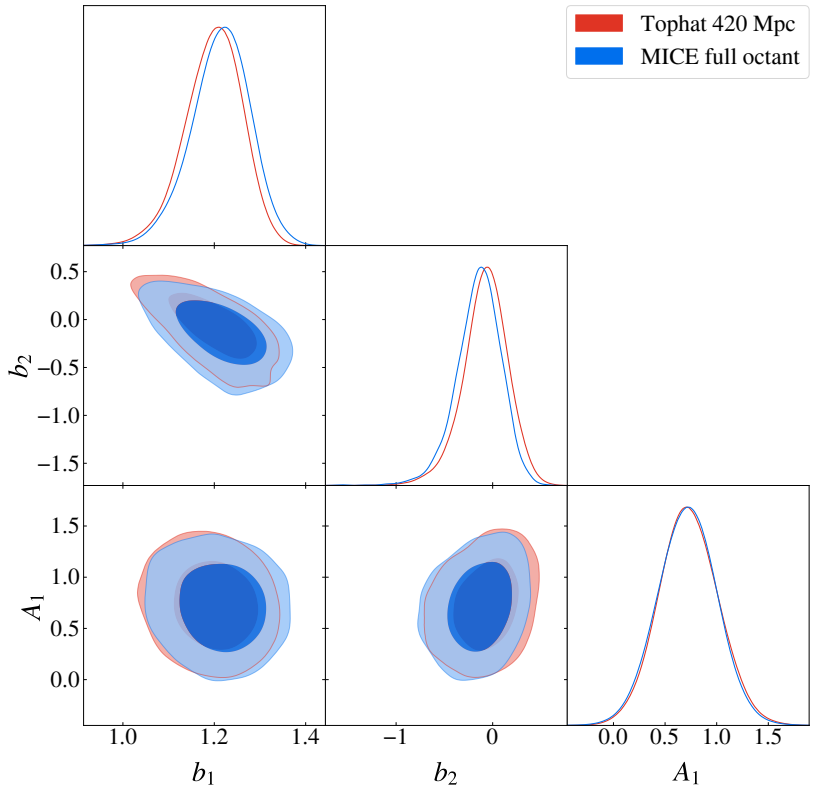

Fig. 4.9 Galaxy bias and IA parameters contour plots obtained from fitting the GC and the IA signals in MICE, using the randoms from the MICE full octant and from the smoothed version of the randoms.

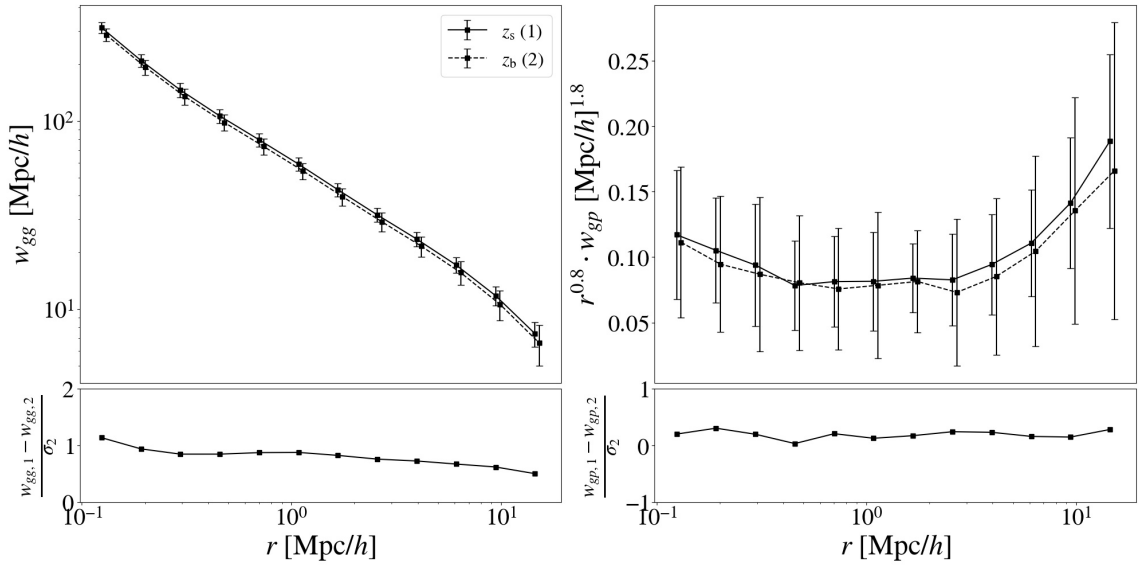

Fig. 4.10 (Left): Galaxy clustering measurements in MICE with different estimates of the radial position, spectroscopic and photometric redshifts. Both estimates have a consistent measurement. (Right): Same as left plots, but showing the IA measurements. In that case, the difference between both types of randoms in terms of σ is negligible.

[87] and Johnston et al. [122], where they define it in the range -60h^{-1}Mpc to 60h^{-1}Mpc in steps of 4h^{-1}Mpc. On the other hand, we define the photometric radial binning as in eq. 4.16, following Johnston et al. [123].

In order to test if the spectroscopic and photometric radial binnings recover the same signal, we measure and compare the GC and the IA signals in the MICE simulation for both options, following a similar procedure as in previous Section 4.5.1. The top plots of Fig. 4.10 show the comparison of the GC (left) and IA (right) signals obtained when using spectroscopic and photometric redshifts in the MICE simulation. In turn, the bottom plots of the figure show the difference between the signals divided by the error from the ensemble covariance (bottom). On the one hand, for the GC signal, the difference that results between both redshift estimates is of $\sim 1\sigma$ at small scales, while it approaches $\sim 0.4\sigma$ at larger scales. Note how the signal measured using spectroscopic redshifts is slightly larger, as expected because of the smearing out caused by photo-z. However, both results can be considered consistent. On the other hand, for the case of IA, the difference in terms of σ is almost negligible. This seems to indicate, at least in the MICE simulation, that IA are less affected than GC when using less precise redshift estimates.

Fig. 4.11 shows the GC and IA bias parameters obtained when fitting both the spectroscopic and the photometric cases. Note that the constraining power in the case of employing

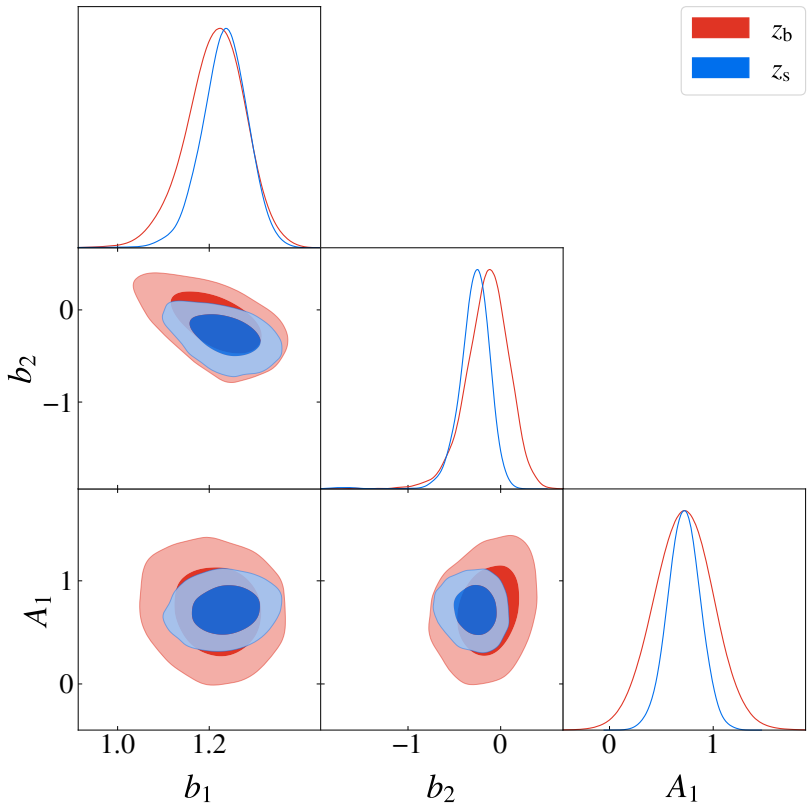

Fig. 4.11 Galaxy bias and IA parameters contour plots obtained from fitting the GC and the IA signals in MICE, using spectroscopic redshifts and photometric redshifts.

spectroscopic redshifts is higher, given that the errors associated to spectroscopic correlation functions are lower. Still, the recovered values are consistent, specially for the A_1 parameter. This consistency test also serves as an indication that the pipeline we designed to model photometric correlations is correct.

Since the measurements in the data will be also split in redshift bins, we repeat the tests in Fig. 4.10 and Fig. 4.11 but dividing the redshift space in three bins, as it will be the case for the PAUS measurements. The objective of this exercise is to check that the results from the previous figures can be extrapolated for redshift bins with different photo-z accuracies. The three redshift bins in MICE are defined to cover the following ranges: z=[0.1-0.4], z=[0.4-0.7] and z=[0.7-1]. Fig. 4.12 shows the GC (top) and the IA (bottom) measurements in MICE for the three redshift bins. Given that the photometric redshift error increases for higher redshifts, we are not able to recover neither the GC nor the IA measurements for the highest redshift bin. Again, the difference is more notable in the case of GC than in the case of IA, as seen in the previous Fig. 4.10. Nevertheless, as seen in Fig. 4.13, where the GC and the IA constraints are shown for the three redshift bins, this does not translate into excessively

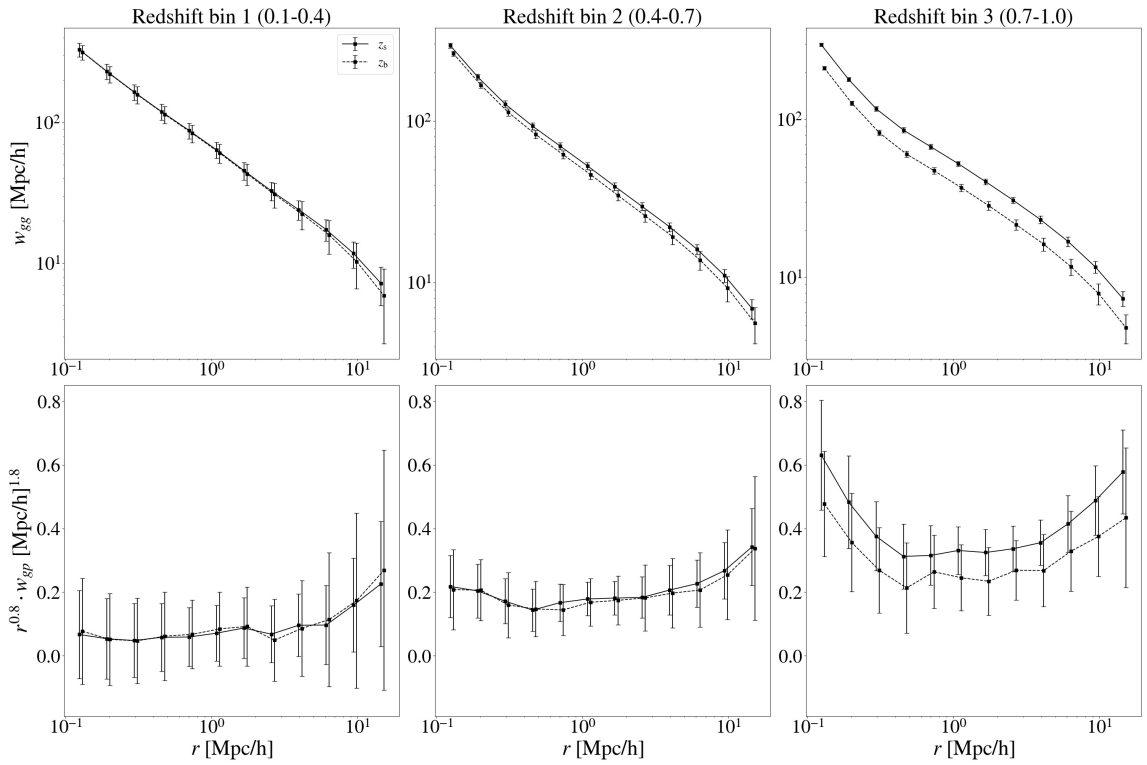

Fig. 4.12 (Top): Galaxy clustering measurements in MICE for different redshift bins with different estimates of the radial position, spectroscopic and photometric redshifts. (Bottom): Same as top plots, but showing the IA measurements.

shifted constraints. This is because the photometric redshift error is also accounted for in the modelling. However, we note again that the constraints between the spectroscopic and the photometric scenarios are different in terms of area, with higher redshift bins leading to broader constraints in the case of photometric redshifts. This is caused by larger errors in the signal due to the decrease in the photo-z accuracy.

4.5.3 Error estimation

As discussed in Section 4.3, we estimate the errors for our correlation functions using the JK method (eq. 4.17). However, considering that the amount of JK regions is not very large, due to the limited area in PAUS, we also need to validate our errors with the MICE simulation. In order to do that, we compare the errors derived from the JK method with the errors from the ensemble covariance of the MICE simulation (eq. 4.47), which are a representation of the true errors of the full octant:

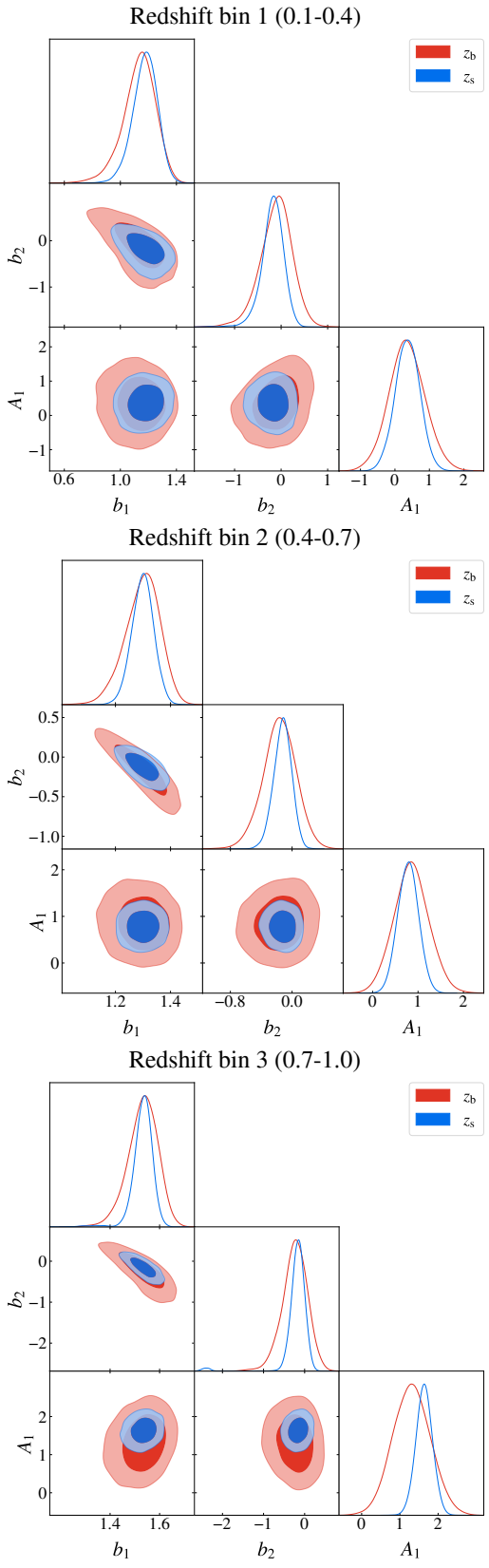

Fig. 4.13 Same as Fig. 4.11 but splitting the samples in redshift bins, to check that our method recovers the bias and IA parameters for all redshifts. From top to bottom: the redshift increases.

$$Cov^{ens}_{ab,i,j} = \sum_{k=1}^{N} = \frac{(w_{ab,k,i} - \bar{w}_{ab,i}) * (w_{ab,k,j} - \bar{w}_{ab,j})}{N}, \tag{4.47}$$

where k corresponds to each combination of 3 PAUS-like patches, i and j denote the r_p angular position of the data vector, N is the number of combinations and $\bar{w}_{ab,i}$ is the mean of the correlation function (w_{gg} or w_{gp}) at the position i.

Fig. 4.14 shows a comparison between the errors obtained from the ensemble and the JK covariances, for the w_{gg} (left) and the w_{gp} (right) estimators. The histograms in the plots show the distribution of the JK covariances (in terms of the square root of the diagonals) for the 60 triplets of the MICE PAUS-like patches. The vertical red and blue lines show the ensemble covariance and the mean JK error, respectively. This is shown at 4 different separations in terms of r_p, so as to capture the evolution as a function of the separation between galaxies. For the case of w_{gg}, it can be seen that the errors are underestimated when comparing the mean of the JK errors with the ensemble covariance. This can also be noted in the distribution of the individual JK errors, where in most cases they are located at values lower than the ensemble covariance, especially at low separations. In contrast, for the case of w_{gp}, the opposite trend is seen, where the mean JK errors are larger than the ensemble covariance for all separations, although with different variations. However, in any of the two cases, both error estimates agree quite well, since the ensemble covariance is well inside the range of the individual JK errors.

The fact that both error estimates agree well can also be seen in Fig. 4.15. On the one hand, the top panel shows the ensemble covariance and the mean JK error (also in terms of the square root of the diagonal) as a function of the r_p separation, for w_{gg} (left) and w_{gp} (right). The error bars depicted for the mean JK error correspond to its standard deviation. On the other hand, the bottom panel shows the difference between the ensemble covariance and the mean JK error, normalized by the standard deviation of the mean JK error. From this figure, the agreement between both error estimates is verified for all r_p separations, with a maximum of a $\sim 1\sigma$ difference for the case of the w_{gp} estimator and below $\sim 0.3\sigma$ for w_{gg}.

So far, we have only shown how the diagonal terms of the mean JK and the ensemble covariances relate with each other. Nevertheless, since the modelling of the PAUS data will use the whole covariance matrix, we are also interested on how the off-diagonal terms affect the modelling of our observables. Besides, even though there is a maximum of $\sim 1\sigma$ differences between both error estimates, we are interested on how these may affect our constraints. In that sense, Fig. 4.16 shows the galaxy bias and the IA bias constraints obtained when fitting the mean $w_{gg} + w_{gp}$ data vector using the mean JK error and the ensemble

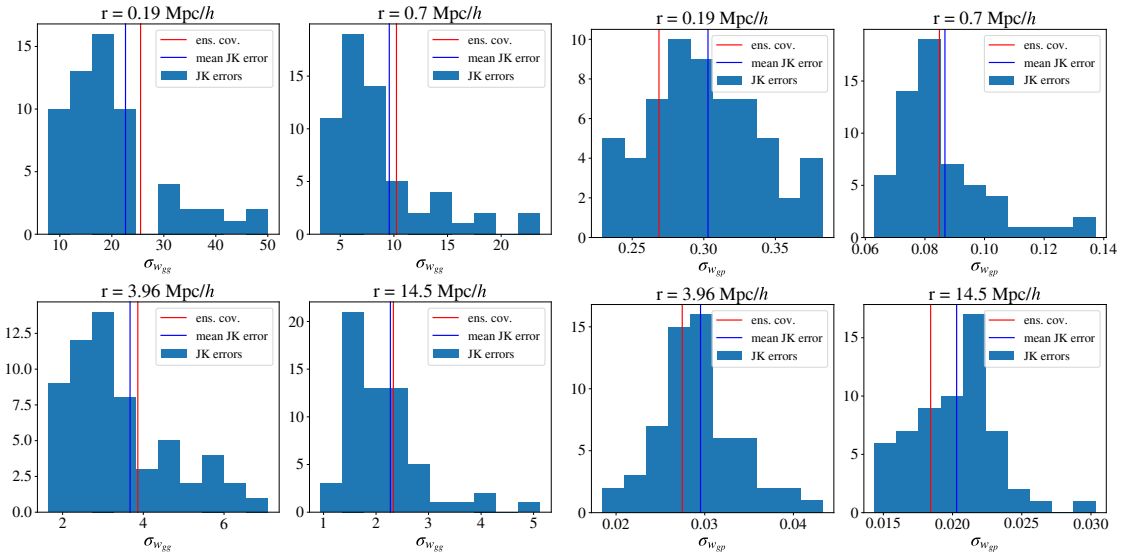

Fig. 4.14 Distribution of the individual JK errors of the 60 combinations of MICE regions at different separations. A red and blue vertical lines show the ensemble covariance and the mean JK covariance at that separation, respectively.

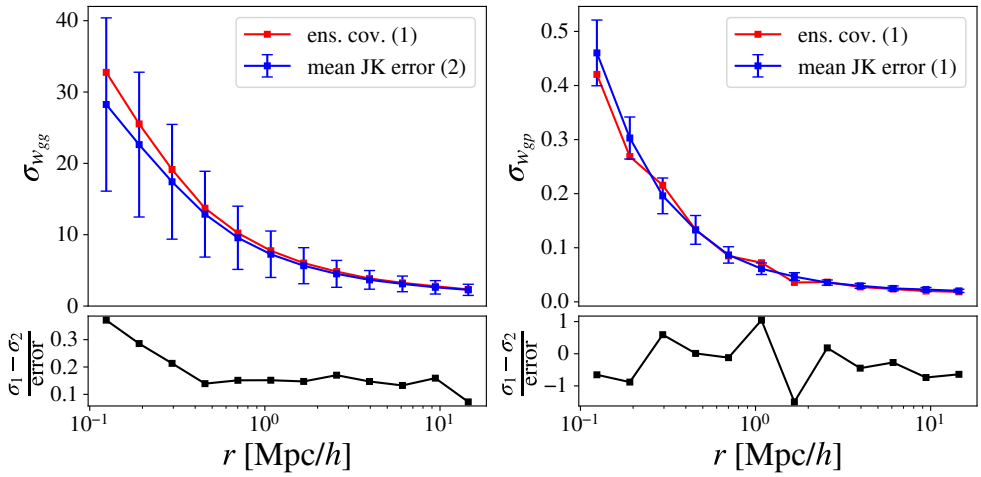

Fig. 4.15 (Top): Diagonal terms of the ensemble covariance matrix (red) and the mean JK error (blue) as a function of the r_p separation. The error bars of the mean JK error correspond to the standard deviation of the individual JK errors with respect to the mean. (Bottom): Difference between the ensemble covariance and the mean JK error, normalized by the standard deviation of the mean JK error.

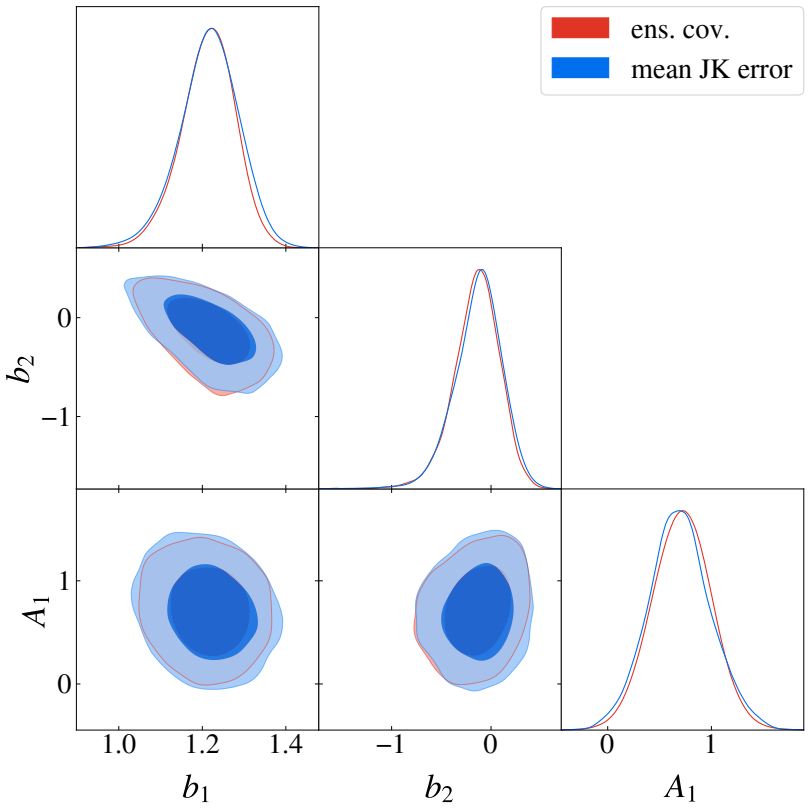

Fig. 4.16 Galaxy bias and IA parameters contours obtained when fitting the mean $w_{gg} + w_{gp}$ data vector with the ensemble covariance (red) and the mean JK error (blue). The galaxy bias and IA parameters obtained from both methods are very similar.

covariances, where it can be seen that there is a near-total agreement between the contour plots of both covariance matrices. An important note to take into account is that, in the PAUS data, we only have one realization of the results, while in the MICE simulation we have 60 triplets of PAUS-like patches, from which we compute the mean JK covariance. Thus, even though the mean JK and the ensemble covariances yield very similar constraints, it might happen that the PAUS data covariance differs more from the ensemble covariance, as it is the case for some of the individual combinations seen in the histograms of Fig. 4.14. Nevertheless, the majority of the individual JK errors agree quite well with the mean of the covariances, so the probability of significantly different errors in the data is expected to be low.

As a final validity test related with the error estimation, we show an analogous of Fig. 4.15, but this time as a function of the redshift bins defined in the previous Section 4.5.2. Again, as in the cases of Fig. 4.12 and 4.13, this is done to check the effect on the photo-z accuracy at different redshifts. In all the cases, the covariance estimates from the mean JK and the

ensemble agree quite well, with differences of $\sim 1\sigma$. We note that the error differences increase for higher redshift bins, which is an expected result given the lower accuracy of the photo-z.

4.6 Results

The results we obtain from the GC and IA measurements are shown in this section. Each of the measurements will be complemented by the modelled signal fitted to it, following the methodology in Section 4.4.4 and Section 4.4.6. In general, the results we show correspond to the bright sample (W1+G09+W3 below $i_{\rm AB} < 22$), although we show the evolution of A_1 as a function of the colour, the redshift and the luminosity for both the bright and faint samples (W1+W3 below $i_{\rm AB} < 22.5$).

In order to study the dependence of IA with different physical properties, we decided to split our measurements by colour, defining a red and a blue shape sample. Furthermore, together with the split in colour, we performed a separation in 3 equipopulated redshift bins and in 3 equipopulated luminosity bins. Note that we split in equipopulated bins before separating the objects by colour, which may result in red and blue galaxies having slightly different binning in redshift and luminosity. This leads to a total of 14 different scenarios, for which we have measured and modelled the photometric GC and the IA. As for the density samples, that is, those containing the positions, we also split them in redshift and luminosity bins. However, we choose not to separate them by colour, as doing so results in higher SNR. This does not represent a problem, since the purpose of measuring the GC in this Chapter 4 is to obtain the galaxy bias, which is required in the w_{gp} estimator. Even though the galaxy bias is dependent of the colour of the galaxy sample, when correlating the positions (without separating by colour) with the shape samples (separated by colour), the galaxy bias becomes independent of the colour of the shapes, albeit some small effects which will be discussed throughout this section.

Table 4.2 shows the 14 different configurations (7 cases x 2 colours) studied in this Chapter 4. For each case, we present the number of objects of the density sample ($\rm N_D$) and of the shape sample ($\rm N_S$) for red and blue galaxies. We show the number of JK regions defined in each of the cases. We also present the mean redshift, luminosity and stellar mass of the red and blue shape samples, together with the accuracy of the photo-z as a function of the $\sigma_{68}(\Delta z)$ (eq. 3.8).

The number of red objects for the shape samples is lower with respect to the blue case. This can also be seen in Fig. 4.2. As a consequence, we expect larger error bars in the

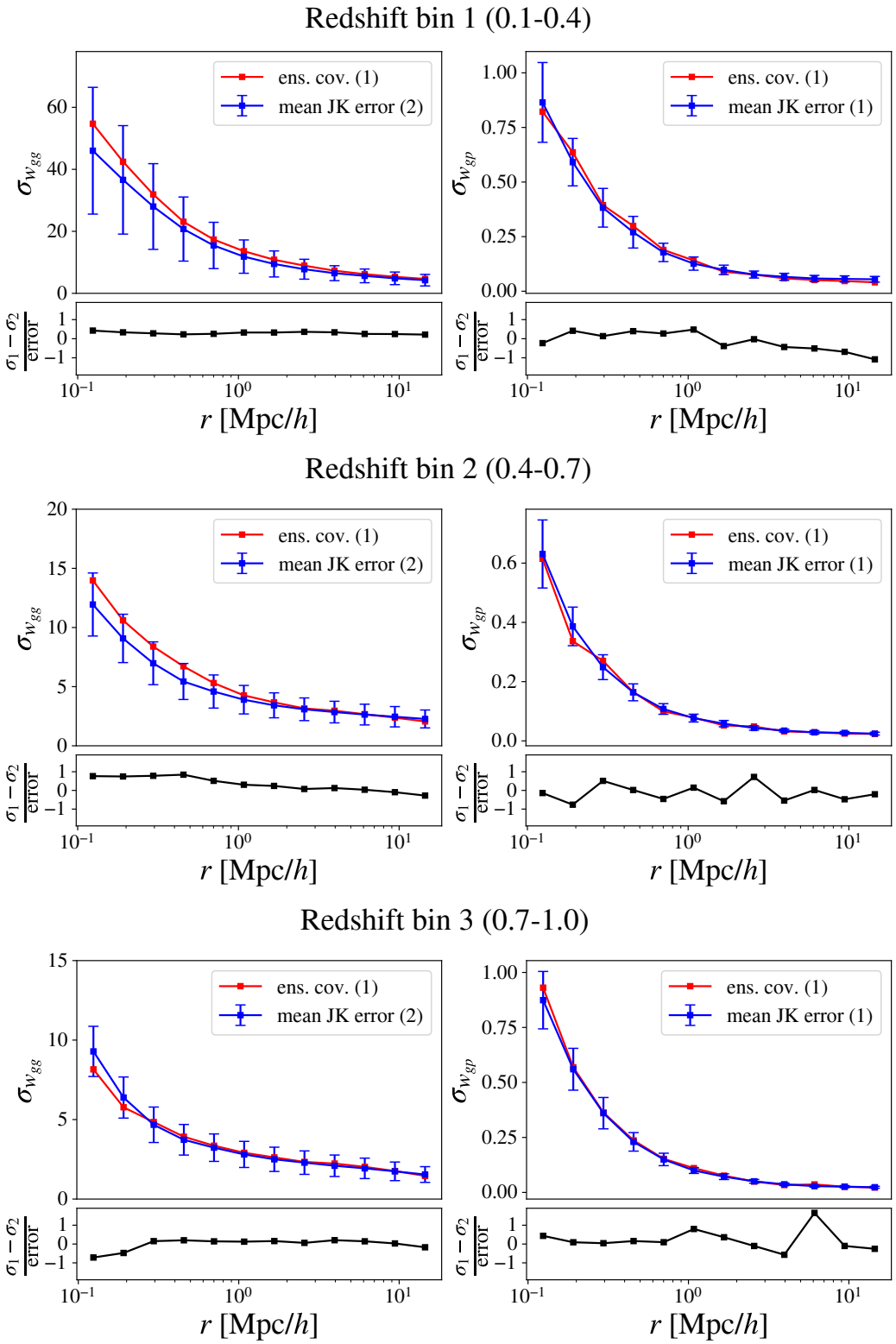

Fig. 4.17 Same configuration as Fig. 4.15, but analysing the effect for different redshift bins.

measurements of red objects. This low number is specially important in the lowest redshift and in the lowest luminosity bins, with ~ 15000 objects.

The number of angular JK regions is the sum of the angular regions of the three fields, which corresponds to 18 angular JK regions, as seen in Section 4.3. In the case where we only perform the analysis as a function of the colour, this is complemented by 3 cuts in redshift bins, defining JK regions in the radial direction as well, which account for a total of 54 JK regions. Given that the IA signal does depend on the redshift, the definition of JK regions in the radial direction might affect the covariance matrix, since the different regions might not be independent but conserve some evolution with redshift. This was reviewed in Section 4.5.3, where we found that the JK errors are consistent with the ensemble covariance matrix. The same division in redshift bins used to define JK regions is also used to define our samples divided by redshift. Thus, each of these cases have 18 JK regions. Finally, in the case where we split our samples by luminosity bins, we can not use the same redshift bins to define the JK regions. This is because each luminosity bin exhibits a different redshift distribution, making it impossible to define the same equipopulated redshift bins. In this case, we decided to divide the redshift space for each scenario in 2 equipopulated redshift bins, accounting for 36 JK regions in each case.

The $\sigma_{68}(\Delta z)$ is also shown for each configuration. As expected, in the case of the samples divided by redshift bins, the sample with the higher redshift cut presents a lower photo-z accuracy, since the photo-z accuracy decreases with redshift. In the case of the luminosity bins, we also note that the luminosity bins are closely related with redshift, with lower luminosity bins presenting lower redshift means. Nevertheless, this does not necessarily translate into worse $\sigma_{68}(\Delta z)$ values, as the luminosity bins are also linked to the apparent magnitude, so that fainter objects correspond to lower luminosity bins. Given that photo-z accuracy also depends on the apparent magnitude, with fainter objects having poorer photo-z accuracy, this counteracts the relation between luminosity and mean redshift. In that sense, for the case of red objects, the photo-z accuracy increases as we move towards higher luminosity bins, while in the case of blue objects, the photo-z accuracy is quite stable, with a minor worsening for the central luminosity bin.

The stellar mass is included in the table since previous studies, such as Joachimi et al. [120], van Uitert and Joachimi [214] and Piras et al. [169], have shown that the IA amplitude is expected to increase with DM halo mass, given that more massive DM halos will exert a higher gravitational pull on the LSS surrounding them, increasing its alignment. In our case, instead of the DM halo mass, we provide values for the stellar mass as a CIGALE output, which is proportional to that of DM. Note that, as expected, the stellar mass of red galaxies is

Table 4.2 Number of objects of the density and shapes samples, the latter divided by red and blue objects, number of JK regions, magnification, mean redshift, σ_{68}, mean luminosity and stellar mass for red and blue galaxies for the IA cases studied in this Chapter 4.

Sample	All	z bin 1	z bin 2	z bin 3	Lum. bin 1	Lum. bin 2	Lum. bin 3
N_D	482400	159903	161118	161379	130457	162390	181715
N_S red	83469	15084	34582	33803	15721	42191	22465
N_S blue	293988	109303	93354	91331	81899	82863	120645
N_{JK}	54	18	18	18	36	36	36
$\alpha(r_{AB})$	0.99	0.68	0.79	1.61	1.43	0.98	0.73
<z>red	0.55	0.3	0.49	0.73	0.43	0.58	0.64
<z>blue	0.48	0.26	0.49	0.74	0.26	0.47	0.66
σ_{68} red	0.008	0.004	0.006	0.013	0.028	0.016	0.005
σ_{68} blue	0.015	0.005	0.015	0.018	0.016	0.021	0.015
<L/L_0>red	-0.59	-0.78	-0.61	-0.49	-1.03	-0.62	-0.24
<L/L_0>blue	-0.6	-1.05	-0.47	-0.19	-1.25	-0.6	-0.15
$\log_{10}(M_\odot)$ red	10.62	10.24	10.56	10.86	10.07	10.63	11.00
$\log_{10}(M_\odot)$ blue	9.87	9.33	10.01	10.37	9.10	9.85	10.39

larger than that of blue galaxies for all 14 scenarios. Additionally, an increase of stellar mass with redshift and luminosity is also observed.

For each configuration presented in Table 4.2, we present in Fig. 4.18 what we refer to as the reduced χ^2 of SNR ($\chi^2_{\nu,\mathrm{SNR}}$) for the w_{gp} and w_{gx} estimates, defined as:

$$\chi^2_{\nu,\mathrm{SNR}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{D}_i\,\bar{C}^{-1}_{\mathrm{SVD},ij}\,\bar{D}^T_j}{N_{\mathrm{d.o.f.}}-1}, \tag{4.48}$$

where $\bar{D}_i = D_i/\sigma_{C,i}$ is the normalized data vector and $N_{\mathrm{d.o.f.}}$ is the number of degrees of freedom after performing the SVD (see Section 4.4.6). The χ^2 employed in this equation is similar to the one presented in eq. 4.46 but this time the model is replaced by the null value, in order to check for detection of IA signal or systematics. The separations by colour, redshift and luminosity are differentiated in Fig. 4.18 by contour gray areas to facilitate the reading. Horizontal lines representing a $\chi^2_{\nu,\mathrm{SNR}} = 2$ and $\chi^2_{\nu,\mathrm{SNR}} = 3$ are indicated in dashed and dotted lines, respectively. It is important to note that this test helps to identify detection of signal (w_{gp}) or systematics (w_{gx}), but it is not a conclusive test. The reason for this is that, in some cases, specially for the systematics, the value of $\chi^2_{\nu,\mathrm{SNR}}$ increases because of uncorrelated points of w_{gx}. Thus, in order to confirm a detection of IA or of systematics, the signal has to be correlated for different angular separations. Otherwise, we risk misinterpreting noise in the measurements as signal or systematics. However, this test is useful for identifying

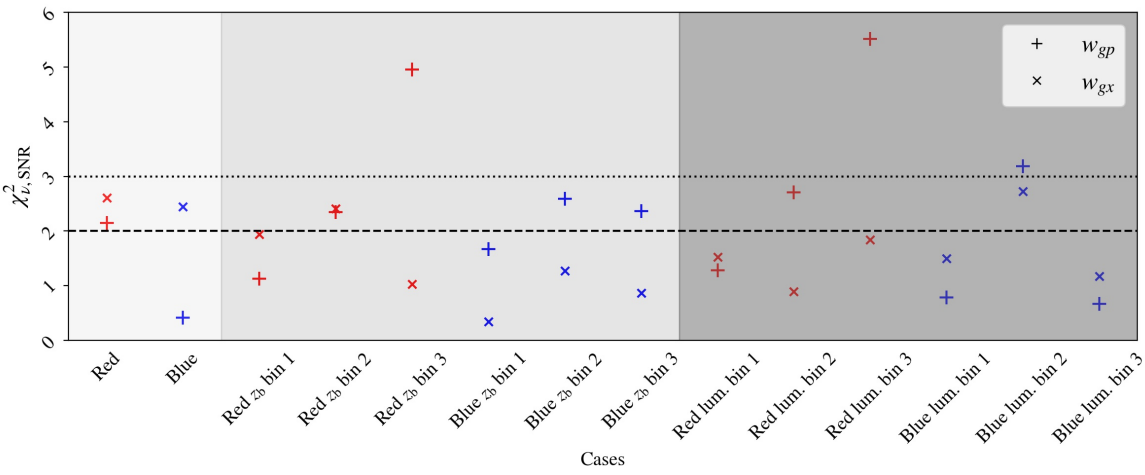

Fig. 4.18 $\chi^2_{\nu,\mathrm{SNR}}$ for the w_{gp} and the w_{gx} correlation functions for the different configurations analysed in this chapter. High values of w_{gp} may indicate detection of IA, while high values of w_{gx} may indicate possible systematics present for that configuration.

non-detection of signal or systematics, in the sense that a low value of $\chi^2_{\nu,\mathrm{SNR}}$ indicates correlations consistent with 0.

In the case of the separation by colour, both w_{gp} and w_{gx} have a $\chi^2_{\nu,\mathrm{SNR}} > 2$ for the red objects, so the detection is not clear from this plot and we will need more information from other analyses in order to confirm it. For the case of blue objects, the $\chi^2_{\nu,\mathrm{SNR}}$ for w_{gp} is clearly below 1, while the systematics are $\chi^2_{\nu,\mathrm{SNR}} > 2$, so it seems clear that there is no detection of IA in blue galaxies when we do not separate in redshift or luminosity bins.

In the case of the separation by redshift bins, for red objects the $\chi^2_{\nu,\mathrm{SNR}}$ of w_{gp} increases with redshift, arriving to $\chi^2_{\nu,\mathrm{SNR}} > 5$ for the highest redshift bin. The $\chi^2_{\nu,\mathrm{SNR}}$ of w_{gx} is ~ 2 for the lowest and intermediate redshift bins and ~ 1 for the highest redshift bin. From this, the data seems to indicate that there is a clear detection for the highest redshift bin, a detection consistent with systematics for the intermediate redshift bin and a non-detection for the lowest redshift bin, with the $\chi^2_{\nu,\mathrm{SNR}}$ of the systematics being larger than the signal. For blue objects, the systematics in all redshift bins have $\chi^2_{\nu,\mathrm{SNR}} \leqslant 1$, while $\chi^2_{\nu,\mathrm{SNR}} \sim 2$ for w_{gp}. This may indicate a mild detection, although further analyses need to be done to confirm this.

Finally, in the case of the separation by luminosity bins, for red objects the situation is similar to the case of the separation in redshift bins, with the $\chi^2_{\nu,\mathrm{SNR}}$ of w_{gp} increasing with luminosity and reaching $\chi^2_{\nu,\mathrm{SNR}} > 5$ for the highest luminosity bin. In this case, the systematics are even lower than in the redshift separation scenario, being $1 < \chi^2_{\nu,\mathrm{SNR}} < 2$ for w_{gx}. The figure seems to indicate that there is no detection for the lowest luminosity bins, mild detection for the intermediate one and a clear detection for the most luminous bin. For

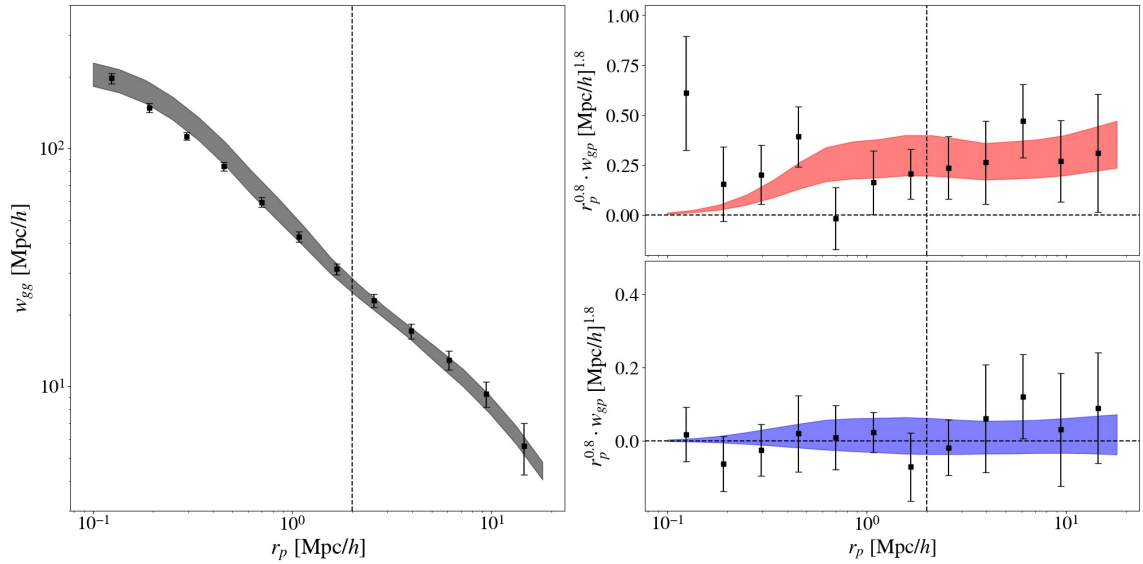

Fig. 4.19 Galaxy clustering (left, w_{gg}) and IA (right, w_{gp}) measurements. The IA are divided in red and blue galaxies. The dashed contours show the best fit NLA.

blue objects, the $\chi^2_{\nu,\mathrm{SNR}}$ for both w_{gp} and w_{gx} is ~ 1 for the lowest and highest luminosity bin, indicating a null detection, while for the intermediate luminosity bin the $\chi^2_{\nu,\mathrm{SNR}} \sim 3$ for the signal and the systematic, so that further analyses need to be done to disentangle if there is any correlation between the data points below or above 0 and be able to assess a detection of IA.

In Appendix C, a comparison of the w_{gp} and w_{gx} measurements can be found, for the cases where it is unclear the presence of w_{gp} signal.

4.6.1 Division by colour

Fig. 4.19 shows the GC and the IA measurements, in terms of w_{gg} and w_{gp}, respectively, for red and blue galaxies as black dot points, with the error bars corresponding to the square root of the diagonal of the JK covariance. These measurements are accompanied by dashed contour lines, representing the 68% best fit model to the measurements. The dashed vertical lines show the $r_p = 2\,h^{-1}$Mpc scale below which we do not fit the model. Despite the $\chi^2_{\nu,\mathrm{SNR}}$ value for red galaxies in Fig. 4.18, which was similar to that of the systematics, we find a clear signal of IA for red galaxies, given that the w_{gp} measurements are correlated for different angular separations, while there is null alignments for blue galaxies. These findings are in line with those found in the previous analyses by Johnston et al. [123].

Fig. 4.20 shows the contour plots for the galaxy bias parameters b_1 and b_2 and for the IA amplitude A_1 for red and blue galaxies. In terms of the A_1 parameter, its amplitude peaks above 2 for the red sample, while for the blue sample it is consistent with 0, as was also seen in Fig. 4.19. In the case of the b_1 and b_2 terms, since we are not splitting the dense sample by colour, we obtain very similar values for both red and blue galaxies. The differences we observe might come from the fact that the SNR of the w_{gp} estimator is different for red and blue galaxies, given that there are 3–4 times more blue than red objects. So, even though the constraining power of the galaxy bias parameters mostly comes from w_{gg} (given its higher SNR with respect to w_{gp}), the w_{gp} estimator also affects the constraints through eq. 4.32. This effect will be slightly different depending on the number of objects in the shape sample. Another consideration that might affect the difference between the galaxy bias parameters for red and blue galaxies is related with the term $p_p(\chi' \mid \chi(z_i))$ in eq. 4.30, which accounts for the error distribution associated to modelling correlation functions with photo-z. Ideally, this term should not depend on colour, redshift or luminosity. However, since it involves knowledge of spectroscopic redshifts and the availability of those does depend on these properties, the constraining power on b_1 and b_2 may also be affected by this factor. As a final remark on this issue, the galaxy bias parameter that eq. 4.32 constrains is b_1, since $P_{gI}(k,z) = b_1 P_{\delta I}(k,z)$ (eq. 4.20). Nevertheless, as an indirect consequence, the b_2 term might be affected through w_{gg} by the change in b_1.

Fig. 4.21 shows the A_1 amplitudes for red and blue galaxies, in this case including both the brighter and the fainter samples analysed in this Chapter 4. The IA amplitudes for both bright and faint samples are consistent with one another. This is an indication that the IA is quite stable when going from $i_{\rm AB} < 22$ to $i_{\rm AB} < 22.5$ when only splitting by colour. It is also an indication that the combination of fields is done in a robust way since, if we exclude the G09 field from the analysis, the results are stable. This means that the alignment in all the three fields is equivalent, as expected for similar galaxy populations.

Fig. 4.22 shows how the contaminants affect the w_{gp} measurements, indicated as the ratio of each contaminant over the total w_{gp} signal. We consider the different terms that affect the galaxy-shape correlation which are, as indicated in eq. 4.38, gI, gG, mI and mG. Note that these terms do not need to have the same sign, which makes it possible for the ratio of one of the terms to be negative or positive and to be greater or less than 1. In fact, in general the galaxy-shear term has opposite sign from the galaxy-intrinsic term, which leads to a decrease in the total value of w_{gp}. This is a strong reason on why it is important to include contaminants in our modelling. The contaminants on w_{gg} are also included in the fitting of the data but are not shown here, since they contribute to the galaxy-galaxy correlation at a percent

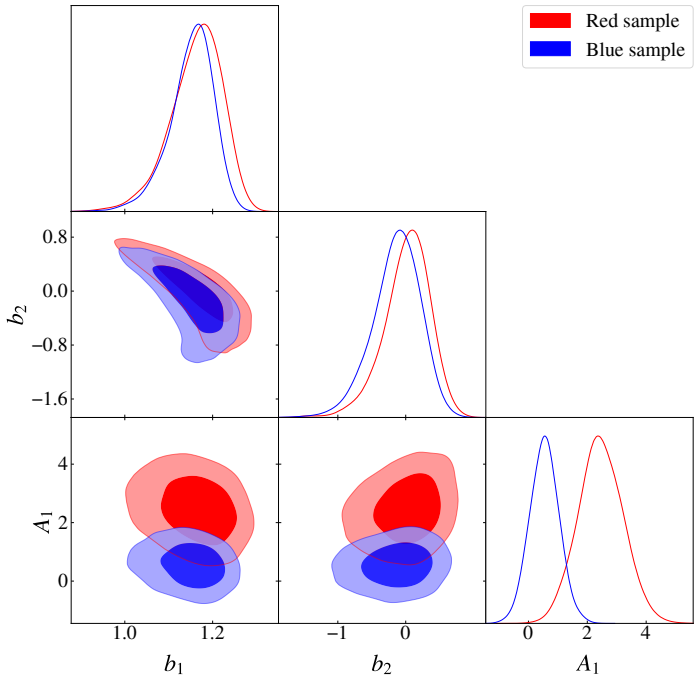

Fig. 4.20 Contour plots for the galaxy bias, b_1 and b_2, and the IA bias, a_1, parameters for red and blue galaxies with the NLA model.

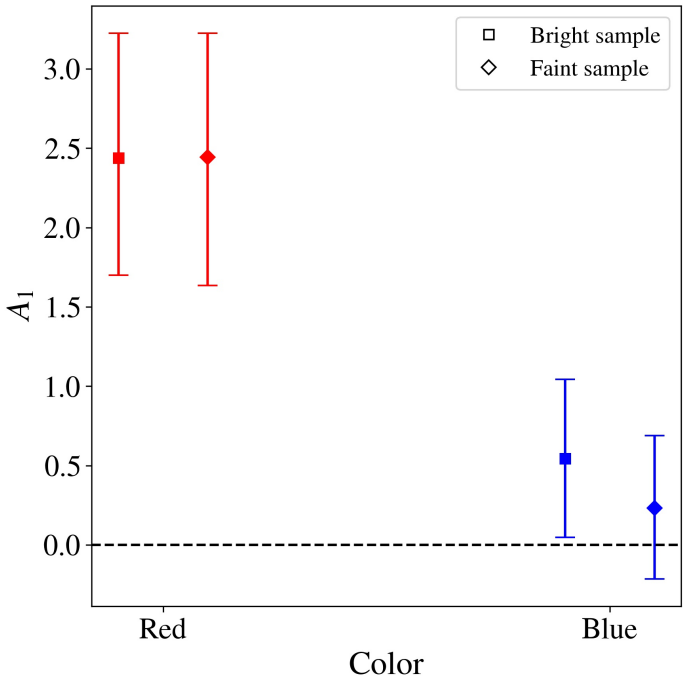

Fig. 4.21 Evolution of A_1 for red and blue galaxies in terms of the bright (W1+G09+W3 at $i_{\rm AB} < 22$) and faint (W1+W3 at $i_{\rm AB} < 22.5$) samples.

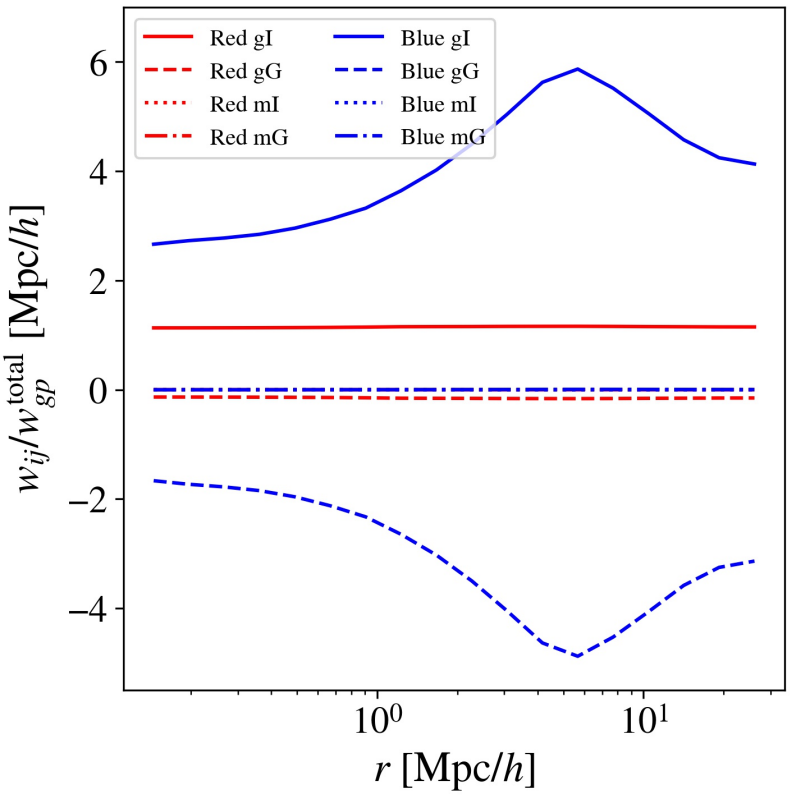

Fig. 4.22 Fraction of contamination to w_{gp} as a function of the colour for the different terms that contribute to eq. 4.38: gI (solid line), gG (dashed line), mI (dotted line) and mG (dash-dotted line).

level. For both red and blue galaxies, the contaminant terms depending on magnification, mI (dotted line) and mG (dash-dotted line), are negligible, given that $\alpha(r_{\rm AB})$ for this case is very close to 1, as seen in Table 4.2. The galaxy-intrinsic term, gI (solid line), is the most important one for both galaxy colours, as it is expected for measurements that are designed to focus on the IA effect, since these consider objects close in redshift. The galaxy-shear term, gG (dashed line), is the main contaminant to the IA measurements. In the case of red objects, the ratio of gG over w_{gp} is -0.2 (while the ratio of gI is 1.2), which accounts for a $\sim$ 10% of the signal. In the case of blue galaxies, the contribution from the galaxy-shear term is even higher, arriving to a $\sim$ 40%, since the IA signal is very low. Note that the percentages we described correspond to the ratio taken over all the r_p range. Nevertheless, the contribution from the gI and gG terms vary over that range, as seen in Fig. 4.22.

4.6.2 Division by colour and redshift

We analyse now the results obtained when splitting the samples by colour and redshift bins. As stated before, we generate three redshift bins, approximately covering the redshift range $z_b \sim [0.1, 0.4]$, $z_b \sim [0.4, 0.7]$ and $z_b \sim [0.7, 1.0]$, with means of $z_b \sim 0.3(0.26)$, $z_b \sim 0.49$ and $z_b \sim 0.73(0.74)$ for red (blue) galaxies. We refer to these redshift bins as z_b bin 1, z_b bin 2 and z_b bin 3, respectively.

Fig. 4.23 shows the w_{gg} and w_{gp} measurements and modelling, as in the case of Fig. 4.19, but this time including the division in three redshift bins. On the left-hand side plot, the evolution of the GC measurements as a function of the redshift is indicated by upside down triangles, squares and face up triangles for the z_b bin 1, z_b bin 2 and z_b bin 3, respectively. A general trend indicating an increase of GC with redshift is observed. This behaviour is expected, since high redshift samples tend to have higher GC than lower redshift samples belonging to similar populations. On the right-hand side plots, the IA measurements are depicted in terms of w_{gp}. The three upper plots correspond to measurements on the red samples, with increasing redshift bins as we move downward. The three bottom plots correspond to the blue samples, with the same order in the redshift bins as the red samples. As in the case of Fig. 4.19, the dashed vertical lines at $r_p = 2h^{-1}$Mpc indicate the minimum separation we include in the modelling of our measurements. For red galaxies, the w_{gp} measurements for the lowest redshift bin are consistent with 0 in most of the r_p bins, in accordance to Fig. 4.18. For the z_b bin 2 and z_b bin 3, there is a stronger positive alignment, which is very similar between both redshift bins 2 and 3 for $r_p > 2h^{-1}$Mpc. Nevertheless, for z_b bin 3, the values at scales lower than $r_p = 2h^{-1}$Mpc are higher than for the z_b bin 2, although it is not straightforward to disentangle this with noise. In fact, it is reasonable to think that the higher $\chi^2_{v,\mathrm{SNR}}$ seen in Fig. 4.18 for the z_b bin 3 comes from these points at small scales, which do not seem to be correlated. In the case of the blue samples, the lowest redshift bin w_{gp} measurement is consistent with 0 at all r_p separations. For the z_b bin 2, the alignment is also consistent with 0 at $r_p > 2h^{-1}$Mpc, although with higher variance than in the lower redshift bin case. Nevertheless, for $r_p < 2h^{-1}$Mpc, there seems to be a consistent preference towards negative IA, which is an interesting result. Finally, for z_b bin 3, at $r_p > 2h^{-1}$Mpc the signal slightly prefers negative IA, although the top error bars in most cases are consistent with 0. The modelling of the measurements at $r_p > 2h^{-1}$Mpc also indicates a negative alignment. In the case of the signal at $r_p < 2h^{-1}$Mpc, the hint of negative alignment seen in the z_b bin 2 case is almost diluted.

Fig. 4.24 shows the contour plots of b_1, b_2 and A_1 for red and blue galaxies and for the z_b bin 1, z_b bin 2 and z_b bin 3, from top to bottom, respectively. In the case of the z_b bin 1, the

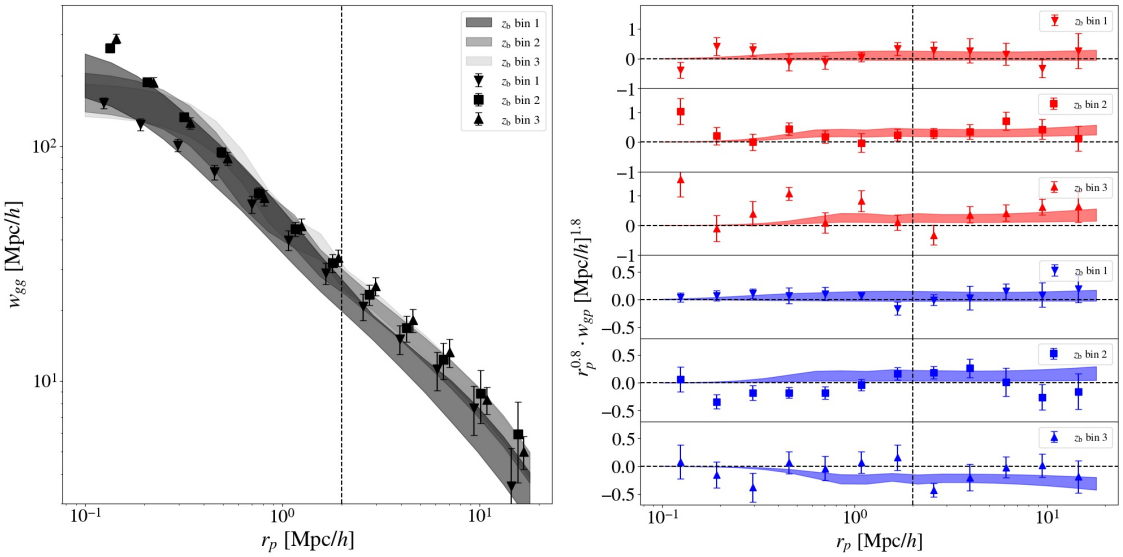

Fig. 4.23 Galaxy clustering (left, w_{gg}) and IA (right, w_{gp}) measurements for different redshift bins. The IA are also divided in red and blue galaxies. The dashed contours show the best fit NLA.

A_1 amplitude is consistent with 0 for both galaxy colours, with a tighter contour around 0 for blue galaxies. For the z_b bin 2, there is a clear A_1 amplitude for red galaxies centred at $A_1 \sim 3$, while there is also a hint of positive A_1 amplitude centred at ~ 1 for blue galaxies, although this is consistent with 0. Finally, for the z_b bin 3, the amplitude for red galaxies is also centred at $A_1 \sim 3$, while there is a trend of negative amplitude for blue galaxies. In terms of the b_1 and b_2 contours, they are in good agreement, specially for the z_b bin 1 and z_b bin 2. Nevertheless, they diverge more for the case of the z_b bin 3. One reason for that might be related with the lower number of spectroscopic objects for these high redshift bins.

The evolution of the A_1 amplitudes as a function of the redshift bins for the brighter (square markers) and fainter (diamond markers) samples is depicted in Fig. 4.25. For the case of red galaxies, the A_1 amplitudes for the brighter and fainter samples are consistent between them, where we observe that, although the mean of the fit is different in some cases, the constraints are between errors. For the case of blue galaxies, this is also the case, except for the z_b bin 3, where the A_1 amplitude is consistent with 0 for the fainter sample, but negative for the brighter sample, with the error bars not overlapping. These differences might come from different reasons: 1) there are indeed differences between the A_1 amplitudes of the bright and faint samples due to the different magnitude limits, which lead to different populations at higher redshift bins; 2) the fields used in both samples are not the same, since the faint sample does not include the G09 field; 3) the point closer to $r_p > 2\,h^{-1}$Mpc in the

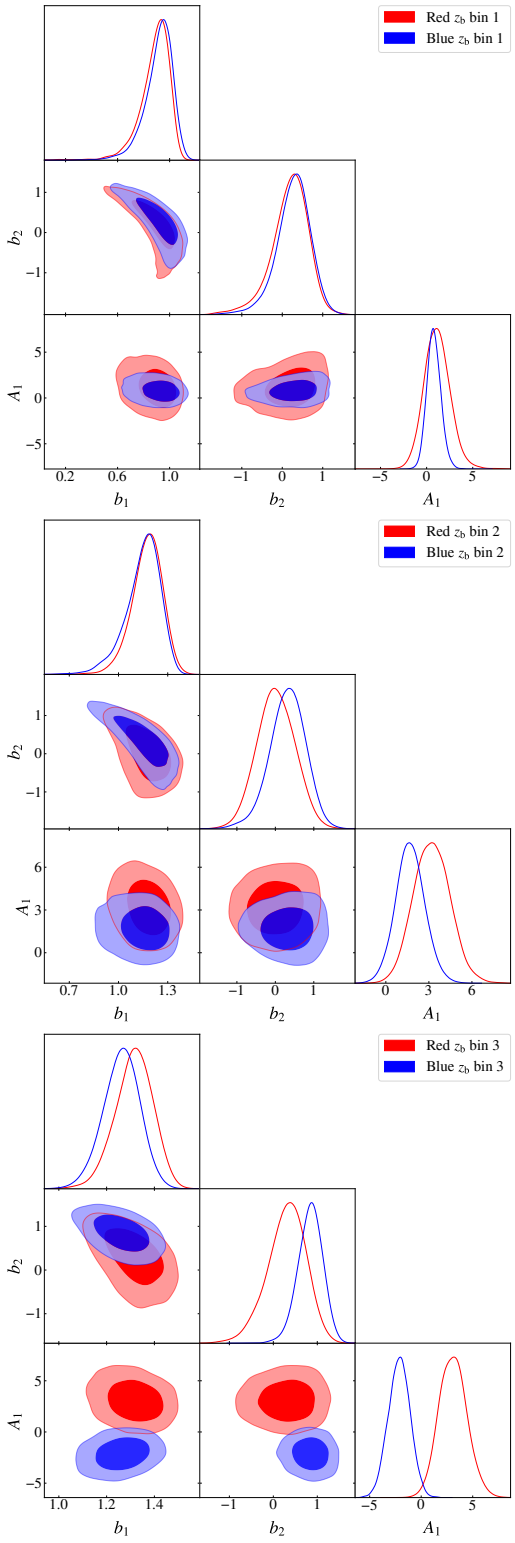

Fig. 4.24 Contour plots for the galaxy bias, b_1 and b_2, and the IA bias, a_1, parameters for red and blue galaxies and for different redshift bins with the NLA model, the redshift range increases from top to bottom.

bright sample (seen in Fig. 4.23) is an outlier that is biasing the constraint on A_1, although the other points at larger r_p also show a preference towards negative alignment; 4) a worse accuracy of the photo-z for the faint sample might lead to a suppression of the IA signal, which might explain the null amplitude of alignment for blue galaxies. However, we observe this last cause is not affecting the red population, even though the σ_{68} for the z_b bin 3 for red and blue galaxies (0.013 and 0.018, respectively) are similar, as shown in Table 4.2. Even though it is not straightforward to disentangle the reason for this difference, we find a negative alignment for blue galaxies at high redshift for the bright sample, which is a new result away from the typical null alignment usually obtained for blue galaxies, so it may be worth it to investigate this further.

Together with the PAUS results, we also include constraints from previous analyses. These constraints correspond to studies analysing the IA redshift evolution and come from the LOWZ ([197]), MegaZ LRG+SDSS ([119]), red and blue GAMA+SDSS ([122]), KiDS-1000 ([76]), DESY3 RML (redMaGiC low-z), DESY3 RMH (redMaGiC high-z), eBOSS LRGs and CMASS (the last four coming from [184]). As a remark, the fits from these previous analyses take as minimum r_p scales larger than $6h^{-1}$Mpc, in contrast to our choice of $2h^{-1}$Mpc. Note that all the points included from previous literature correspond to red samples, except for the blue GAMA+SDSS case. From these studies, none of the previous literature has found any redshift dependence. In particular, a common approach is to fit a power law to:

$$A_1(z) = A_z \left(\frac{1+z}{1+z_0} \right)^{\eta_1}, \tag{4.49}$$

where z_0 is the pivot redshift and previous constraints indicate a η_1 consistent with 0.

In our study, we do not try to fit a power law to our data for 2 reasons. First, the number of data points is reduced, with one point per redshift bin, so that the power law one can obtain would not be much constrained. Second, for some points, the error bars are quite large, difficulting also a power law fit. In any case, one can see that the PAUS results do not follow a clear redshift evolution, although it is true that we find more alignment for the two higher redshift bins in red galaxies. However, these two higher redshift bins have very similar values, indicating that a power law is not being followed. Instead of fitting a power law, our aim is to compare our constraints with those from previous literature. In the case of the lowest redshift bin, red galaxies present lower alignments than those from the red sample of GAMA+SDSS and LOWZ, while for blue galaxies the results are in agreement with those from the GAMA+SDSS blue sample. In the case of the intermediate and highest redshift bins, there are no data points from blue samples in the previous literature we included, so

the comparison can only be done with red samples. In that case, we see that the PAUS constraints are in agreement, at different levels, with KiDS-1000, DESY3 RML, DESY3 RMH, MegaZ LRG+SDSS, CMASS and eBOSS LRGs constraints. One important note is that the samples shown here from previous literature correspond to different populations, so the comparison between constraints is not straightforward. Instead, they can be used to provide an approximate comparison to our data.

Fig. 4.26 shows the contaminants affecting the three redshift bins under study. As in the case of Fig. 4.22, we show the gI, gG, mI and gG terms affecting the total source-shape correlation. Again, for all three redshift bins, we find that the terms accounting for magnification (mI and mG) are negligible, although in this case the values for $\alpha(r_{\rm AB})$ deviate further from 1 than in the colour-only separation scenario. The gG terms for the red (blue) objects account for a $\sim 18\%$ ($\sim 28\%$), $\sim 8\%$ ($\sim 14\%$), $\sim 12\%$ ($\sim 17\%$) for the $z_{\rm b}$ bin 1, $z_{\rm b}$ bin 2 and $z_{\rm b}$ bin 3, respectively. Again, it is important to remember that the contribution of the gG term for blue galaxies is more pronounced, since the gI term is lower for these objects. As a final note, the gG and gI have opposite signs in all cases except for the blue galaxies in the $z_{\rm b}$ bin 3, given that the gI contribution is negative for that configuration.

4.6.3 Division by colour and luminosity

In this section, we analyse the GC and IA signals, splitting the samples by colour and by luminosity bins. The luminosity is computed from the $r_{\rm AB}$-band apparent magnitude from CFHTLenS, in the case of the W1 and W3 fields, and from KiDS, in the case of the G09 field. First, we compute the absolute magnitude from:

$$M_r = m_r - 5\log_{10}(D_{\rm L}/10{\rm pc}) + 2.5\log_{10}\frac{\phi_z}{\phi_{z=0}}, \tag{4.50}$$

where M_r and m_r correspond to the absolute and apparent magnitudes, respectively. $D_{\rm L}$ is the luminosity distance (see Section 1.2.3) and the last term corresponds to the k-correction, with ϕ_z and $\phi_{z=0}$ being the flux of the galaxy at redshift z (the redshift at which the object is located) and $z=0$, respectively. The k-correction is computed with the help of the BCNZ code (Section 3.3.1), which outputs the flux of each galaxy for all the redshift range between $z=[0,2]$.

A caveat in the computation of the absolute magnitude is the fact that we are using the $r_{\rm AB}$-band magnitude from different surveys, which may lead to differences in the number counts for the fields under study. This was the case for Fig. 3.2, where there was the need of adding an offset value to the $i_{\rm AB}$ magnitude from KiDS in order to match that of CFHTLenS.

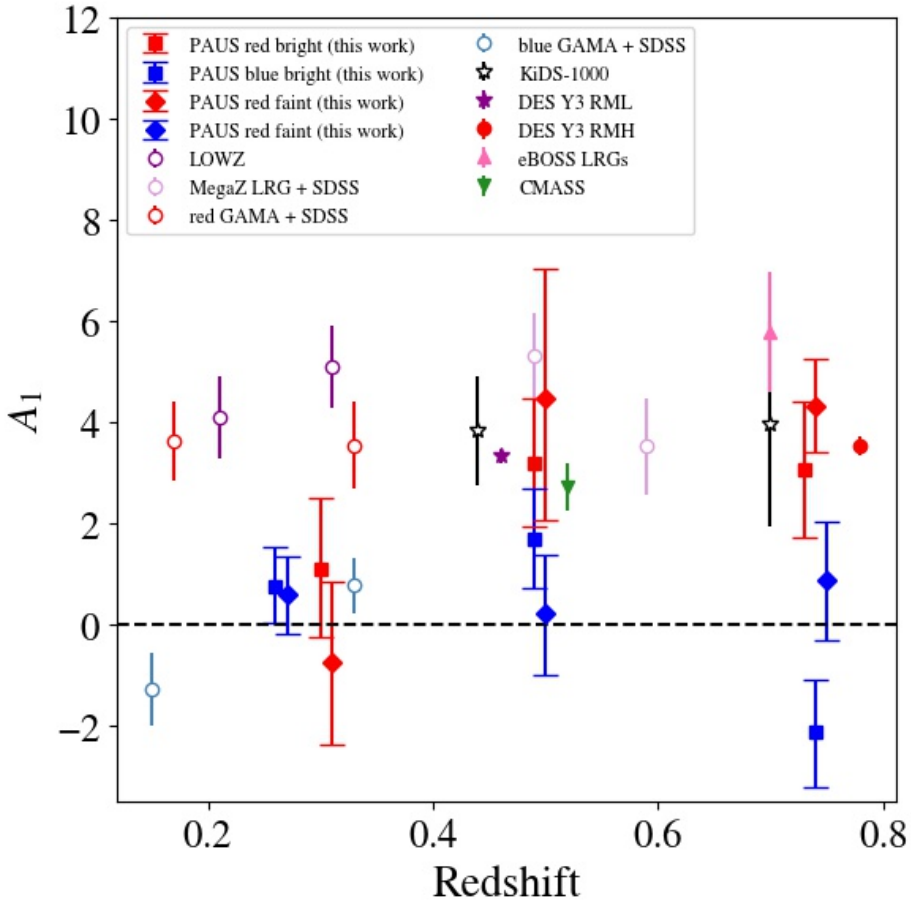

Fig. 4.25 Evolution of A_1 as a function of redshift for red and blue galaxies for the bright (W1+G09+W3 at $i_{\rm AB} < 22$) and faint (W1+W3 at $i_{\rm AB} < 22.5$) samples. Previous results from literature are also shown: LOWZ (dark purple circle, [197]), MegaZ LRG+SDSS (light purple circle, [119]), red and blue GAMA+SDSS (red and blue empty circles, respectively, [122], KiDS-1000 (black star, [76]) and DESY3 RML (which stands for redMaGiC low-z, purple star), DESY3 RMH (redMaGic high-z, filled red circle), eBOSS LRGs (pink triangle up) and CMASS (green triangle down) from [184].

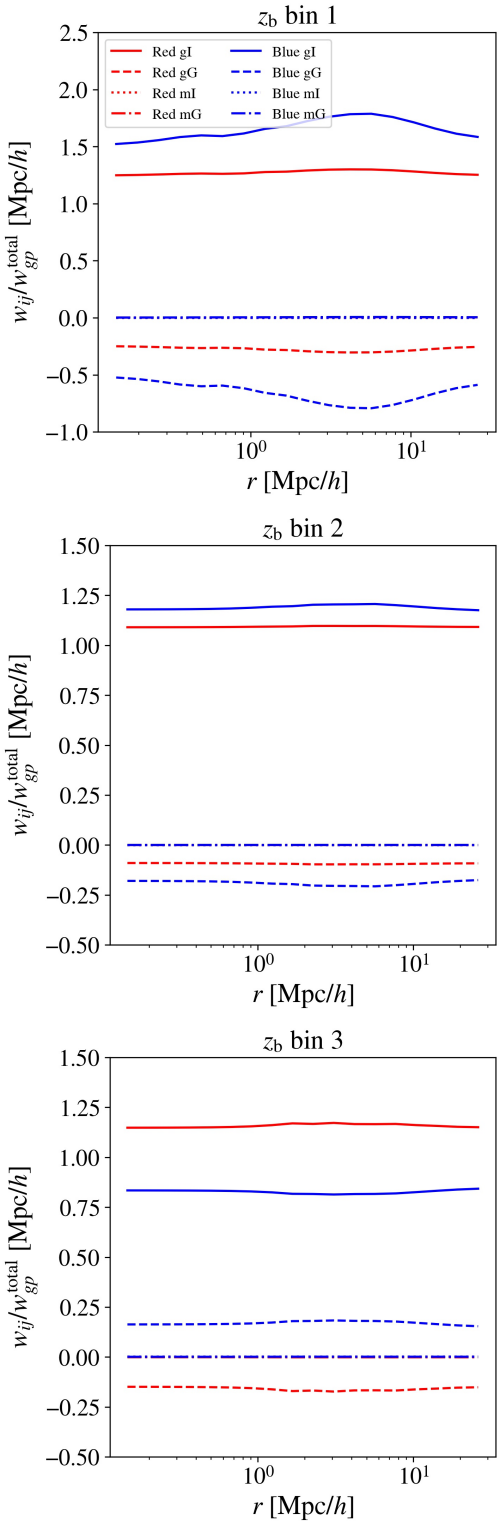

Fig. 4.26 Same as Fig. 4.22 but including a separation in three redshift bins.

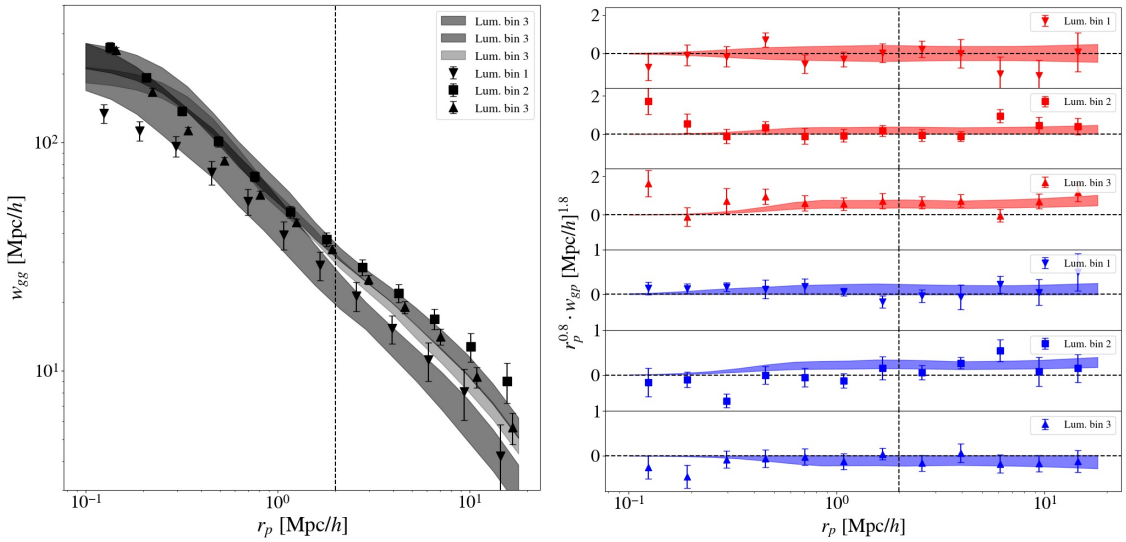

Fig. 4.27 Galaxy clustering (left, w_{gg}) and IA (right, w_{gp}) measurements for different luminosity bins. The IA are also divided in red and blue galaxies. The dashed contours show the best fit NLA.

A similar scenario happens in this situation, where we need to add an offset value of +0.4 in the $r_{\rm AB}$ absolute magnitude of the KiDS field to match the distribution of the CFHTLenS fields.

Once the absolute magnitude is computed through eq. 4.50, we compute the normalized luminosity with respect to the absolute magnitude $M_{r,0} = -22$, in order to compare with previous literature, so that:

$$M_r - M_{r,0} = -2.5 \log_{10} \frac{L_r}{L_0} \tag{4.51}$$

We generate three equipopulated luminosity bins, before splitting by colour, which generate samples with means of $\log_{10}(L_r/L_0) \sim -1.03(-1.25)$, $\log_{10}(L_r/L_0) \sim -0.62(-0.60)$ and $\log_{10}(L_r/L_0) \sim -0.24(-0.15)$ for the red (blue) samples. We refer to these luminosity bins as luminosity bin 1, luminosity bin 2 and luminosity bin 3, respectively.

Fig. 4.27 shows the w_{gg} and w_{gp} for red and blue objects as a function of luminosity. The distribution is the same as in Fig. 4.23. The GC shows lower values for the luminosity bin 1, corresponding to the faintest objects, than for the luminosity bins 2 and 3, which present similar GC values, with slightly lower values in the case of the luminosity bin 3. In the case of the IA measurements for red objects, the signal increases as the luminosity does, from the case of the luminosity bin 1 where the signal is consistent with null alignment to a progressive increase in the alignment in luminosity bins 2 and 3. This is in agreement with

the $\chi^2_{\nu,\mathrm{SNR}}$ in Fig. 4.18. In the case of blue galaxies, the alignment is consistent with 0 for the luminosity bin 1, although for the majority of the r_p separations there is a small trend of positive alignment. The luminosity bin 2 is also consistent with 0 for most of the scales, although there is a hint of positive alignment at $r_p > 2h^{-1}$Mpc, as indicated by the blue line that corresponds to the modelling. This was also indicated by the higher $\chi^2_{\nu,\mathrm{SNR}}$, in both w_{gp} and w_{gx}, with respect to luminosity bins 1 and 3, in Fig. 4.18. A further analysis of this particular configuration is performed in Appendix C. Finally, for the brightest luminosity bin, the luminosity bin 3, there is a trend of negative alignment for the majority of scales, although in most of the cases this alignment is also consistent with 0.

Fig. 4.28 shows the contour plots of the b_1, b_2 and A_1 for red and blue galaxies and for the luminosity bins 1, 2 and 3, from top to bottom, respectively. The IA amplitude of the luminosity bin 1 for red objects is centred at $A_1 \sim 0$, while at $A_1 \sim 1$ for blue galaxies, although still consistent with 0. For the case of the intermediate luminosity bin, the A_1 amplitude is centred at $A_1 \sim 1-2$ for the red and blue samples, although it is also consistent with 0, specially for the red population. Finally, for the brightest luminosity bin, there is a clear IA amplitude for red objects, centred at $A_1 \sim 5$, with the alignment being in agreement with 0 for blue galaxies.

The evolution of A_1 as a function of the luminosity is shown in Fig. 4.29 for the PAUS bright and faint samples. In all cases, the amplitudes from both samples are consistent with one another, although their central values do not agree. We also show constraints from previous analyses for comparison, as in the previous Section 4.6.2, but now using samples that were defined to study the luminosity dependence. Again, the fits to these data take as minimum r_p scales larger than $6h^{-1}$Mpc, in contrast to our choice of $2h^{-1}$Mpc. These previous fits indicate a $A_1 - L$ relation based on a broken power-law, where the slope for $\log(L/L_0) < -0.2$ is shallower than above it. This is indicated by the dotted blue line for low luminosities (which is fitted with the GAMA+SDSS results) and the solid purple line for high luminosities (fitted with the MegaZ LRG+SDSS measurements). Although this double power-law is not justified from first principles, so far it has been able to describe previous IA measurements. Instead of fitting a power law to the PAUS data, we opt again for comparing our measurements with previous literature, given the low number of data points and the fact that the errors in the measurements are large in some cases, as explained in Section 4.6.2. However, we note that the samples used in the previous literature correspond to red galaxies and vary in some physical properties with respect to the PAUS samples, so that it is expected that there are differences in the $A_1 - L$ relation they follow.

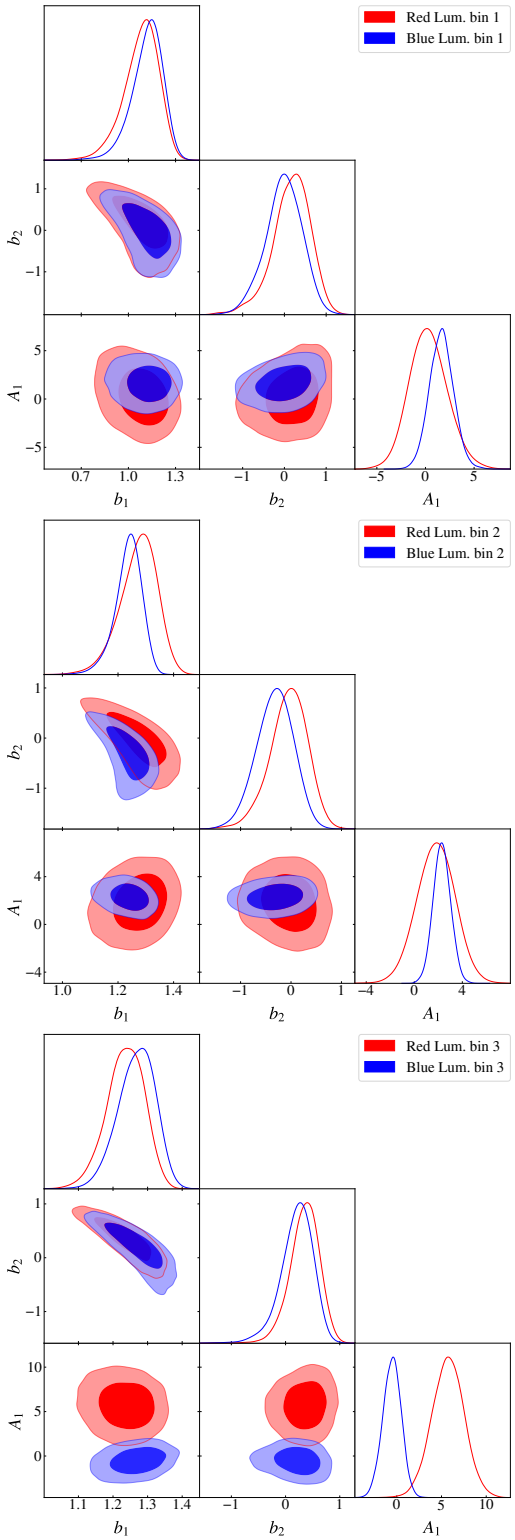

Fig. 4.28 Contour plots for the galaxy bias, b_1 and b_2, and the IA bias, a_1, parameters for red and blue galaxies and for different luminosity bins with the NLA model, from top to bottom the luminosity increases.

One of the most important goals from the study of IA within PAUS is to extend the $A_1 - L$ relation towards fainter objects. In that regard, the three luminosity bins we defined lie in the luminosity range dominated by the solid purple line, even extending the luminosity range with respect to previous measurements down to $\log(L/L_0) \sim -1$ for red objects and down to $\log(L/L_0) \sim -1.3$ for blue objects. For the case of the red bright samples, the two lowest luminosity bins agree with the power law from the GAMA+SDSS measurements, while the highest luminosity bin is above that power law. Nevertheless, this is not alarming since the luminosity bin in that case lies in the middle of both power-laws, with other measurements near in luminosity being compatible with it. In the case of the red faint samples, the intermediate luminosity bin does not follow the power law suggested by GAMA+SDSS measurements. However, this measurement is less than 2σ away from it and is also consistent with other nearby measurements. We also include the measurements of blue galaxies in this plot, although we do not necessarily expect them to follow the power laws from previous literature, since those were computed for red populations. As indicated in the previous Fig. 4.28, the alignment for blue galaxies is not centred at $A_1 \sim 0$ for the two lowest luminosity bins, while it is for the highest luminosity bin, for both the bright and the faint samples.

As in Fig. 4.26, we show the contaminants of the w_{gp} estimator in Fig. 4.30, but this time for the three luminosity bins. The magnification terms (mI and mG) are almost negligible, although more important than in the previous analyses, specially for the highest luminosity bin. In terms of the gG contribution, these are $\sim 38\%$ ($\sim 20\%$), $\sim 16\%$ ($\sim 13\%$), $\sim 7\%$ ($\sim 43\%$) for red (blue) objects in the luminosity bin 1, luminosity bin 2 and luminosity bin 3, respectively. The reason for the higher gG contribution in the luminosity bin 1 for red objects is most likely due to the lower IA signal.

Table 4.3 summarises the A_1 and the $\chi^2_{\nu,\mathrm{fit,SVD}}$ (eq. 4.46) for the different scenarios studied. Note that the $\chi^2_{\nu,\mathrm{fit,SVD}}$ varies from values below and above 1, with 1 being the expected value for a χ^2_ν if the fit of a data vector to a model is correctly performed. Nevertheless, checking the χ^2_ν values from previous literature analyses, such as the ones described in Fig. 4.25 and 4.29, we find $\chi^2_{\nu,\mathrm{fit,SVD}}$ values within the range of these previous studies. We also find similar values of $\chi^2_{\nu,\mathrm{fit,SVD}}$ for the MICE fits, indicating that it is unlikely that there is a problem with the PAUS data when performing the fits. An explanation on why χ^2_ν values below and above 1 may be acceptable can be found in Andrae et al. [4], where they analyse the effect of priors and non-linear models in the computation of χ^2_ν.

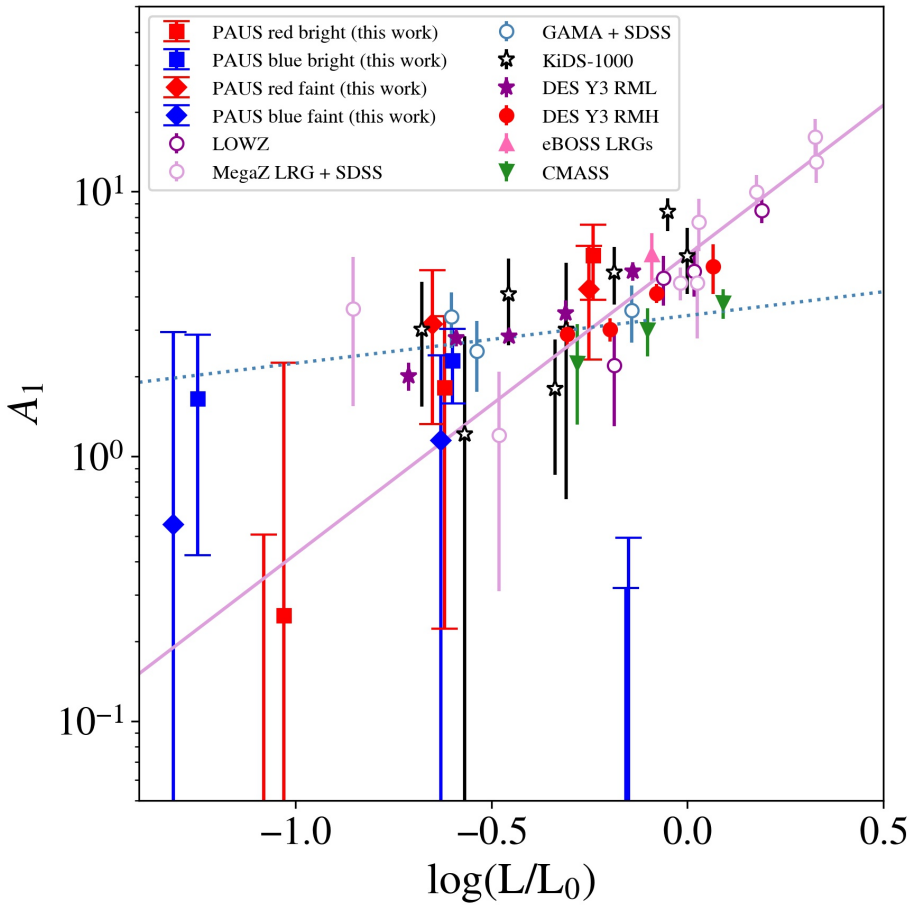

Fig. 4.29 Evolution of A_1 as a function of luminosity for PAUS red and blue galaxies, where the square and the diamond marks correspond to the bright and faint samples, respectively. Previous results from literature are also shown: LOWZ (dark purple circle, [197]), MegaZ LRG+SDSS (light purple circle, [119]), GAMA+SDSS (blue circle, [122], KiDS-1000 (black star, [76]) and DESY3 RML (which stands for redMaGiC low-z, purple star), DESY3 RMH (redMaGic high-z, red circle), eBOSS LRGs (pink triangle up) and CMASS (green triangle down) from [184]. The dotted blue line shows a power law from fitting the GAMA+SDSS results (low luminosity), while the solid purple line indicates a fit to the MegaZ LRG+SDSS results (high luminosity). Figure adapted and extended to include the PAUS measurements from [184].

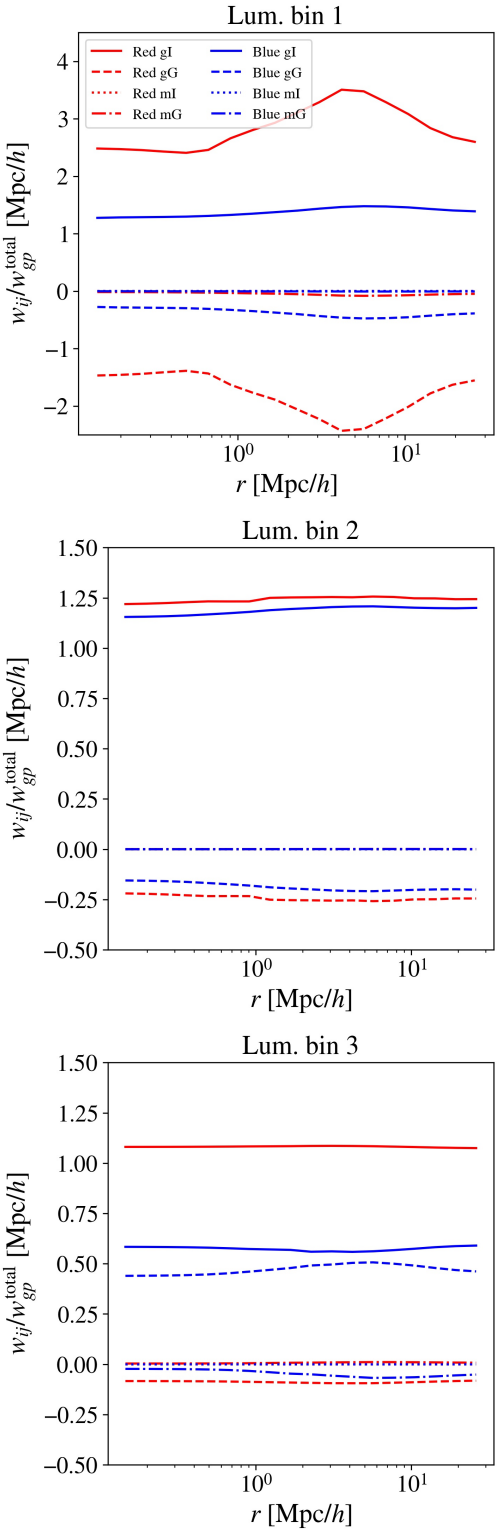

Fig. 4.30 Same as Fig. 4.22 but including a separation in three luminosity bins.

Table 4.3 A_1 and $\chi^2_{\nu,\mathrm{fit,SVD}}$ for red and blue galaxies for the IA cases studied in this Chapter 4.

Sample	A_1 red	A_1 blue	$\chi^2_{\nu,\mathrm{fit,SVD}}$ red	$\chi^2_{\nu,\mathrm{fit,SVD}}$ blue
All	$2.44^{+0.79}_{-0.74}$	$0.55^{+0.50}_{-0.50}$	0.33	0.38
z bin 1	$1.10^{+1.41}_{-1.35}$	$0.76^{+0.77}_{-0.72}$	0.99	0.47
z bin 2	$3.20^{+1.27}_{-1.27}$	$1.69^{+1.01}_{-0.97}$	0.77	1.93
z bin 3	$3.07^{+1.35}_{-1.34}$	$-2.12^{+1.04}_{-1.09}$	1.40	2.19
Lum. bin 1	$0.25^{+2.01}_{-1.88}$	$1.65^{+1.23}_{-1.23}$	1.18	1.02
Lum. bin 2	$1.82^{+1.55}_{-1.60}$	$2.29^{+0.73}_{-0.71}$	2.22	1.55
Lum. bin 3	$5.73^{+1.76}_{-1.82}$	$-0.51^{+1.00}_{-1.04}$	1.83	0.31

Chapter 5

Intrinsic alignment forecasts with Euclid

5.1 Motivation

In this Chapter 5, we describe the implementation of the TATT IA model in the theory and likelihood pipeline of Euclid, called the Cosmology Likelihood for Observables in Euclid (CLOE)[1]. This implementation is important since it allows the Euclid Collaboration to describe the IA effect with more than one model. In particular, it will allow understanding the IA model that best fits the Euclid data. Additionally, the TATT implementation is expected to allow for the inclusion of smaller scales in parameter inference analyses. This is explained by the fact that this model is expected to describe the IA effect better than the NLA model, which is the other model implemented in CLOE, specially at intermediate scales. Nevertheless, this comes with the price of adding, at least, 2 extra parameters.

The structure of this chapter is the following. In Section 5.2, we introduce Euclid's theory and likelihood pipeline, CLOE, enumerating its capabilities. Next, in Section 5.3 we describe the TATT implementation, while also reviewing the previous IA model implemented. Then, in Section 5.4, we validate the TATT implementation with the external code COSMOlogical Survey Inference System (CosmoSIS) to assess that both pipelines are consistent. In Section 5.5, we analyse the differences in terms of C_l observables of the NLA and the TATT models, while in Section 5.6 we study the contribution of IA to the total signal of C_l observables. Following that, we perform a case study where we analyse the impact in cosmological parameter inference of mismodelling the IA effect with a wrong model. Finally, we outline some ideas for future work.

[1]The results and materials of this thesis, related with the Euclid Consortium, are preliminary and have not yet been endorsed by the Euclid Consortium.

5.2 CLOE

In order to constrain cosmological parameters, Euclid needs to face key challenges so as so to compare its measurements with theoretical predictions. These challenges include accounting for small statistical uncertainties, large number of parameters, the difficulty of modelling the non-linear structure formation in the universe, the role that astrophysical processes play on the distribution of matter and the precise covariance matrices that account for correlations between the different cosmological probes.

CLOE is the theory and likelihood pipeline of Euclid that takes care of this task of constraining cosmological parameters. This pipeline is written in python3 and is a flexible and modular software, validated against other similar software packages, allowing to model for many observables and their systematics.

Parameter inference with CLOE takes place in the following way:

1. Computation of the theoretical models for the desired observables, given a set of cosmological and nuisance parameters.

2. Comparison of the theoretical predictions and the measurements, leading to a likelihood probability distribution.

3. When the convergence is reached, the pipeline outputs the posterior probability distribution for the set of cosmological and nuisance parameters. CLOE uses external codes such as Cobaya [210] and CosmoSIS [222] to sample the posterior probability distribution of the desired parameters, enabling the use of different MCMC sampling algorithms, such as Metropolis-Hastings (Section 1.9.2) and polychord [95].

CLOE depends on the Boltzmann solver codes CAMB [143] and/or CLASS [17] to predict the theoretical ingredients needed to compute the modelled observables. One key aspect provided by CLOE is the inclusion of modifications to the modelled observables resulting from the non-linear structure formation. This is essential since it allows including smaller scales in the constraining of the cosmological parameters, which would not be possible without the addition of these non-linear terms.

In the following, we summarise the main non-linear effects implemented in CLOE. First, the modelling of the non-linear galaxy power spectrum in redshift space, which enables to model the spectroscopic galaxy clustering at mildly non-linear scales. For this implementation, a new code has been developed [160] that accounts for 1-loop perturbation theory with counterterms from effective field theory (EFT) [51, 113]. Second, the Euclid Emulator 2 [70] and the bacco emulator [5], which allow computing the non-linear matter

power spectrum for non-linear scales, supplemented by BCEmu [90] and bacco [6] for the modelling of baryonic feedback effects. Besides these emulators, other common tools such as halofit [201, 206] and HMCode [157, 156] also allow accounting for non-linear effects. Third, the intrinsic alignment of galaxies can be predicted using the NLA and the TATT models, the last one being the focus of this chapter. Fourth, an analytical covariance has been implemented following [215], which includes terms from non-linear effects, a window function, the super-survey modes and the integral constraints. The inclusion of correlations between scales, tomographic bins and cosmological probes for a 3x2pt covariance matrix is an expected feature that is still in progress. Finally, in order to allow for the exploration beyond the ΛCDM model, recipes for the *w*CDM model and other massive neutrinos cosmologies have been included, while models for more exotic dark energy scenarios and modified gravity are work in progress.

5.3 TATT implementation in CLOE

With the goal of contributing to the inclusion of models that help on the mission of understanding the non-linear scales, we implemented the TATT model (see Section 1.8.3) in CLOE. Prior to the implementation, only the NLA model was available in CLOE to model IA, which is implemented as follows:

- Intrinsic-intrinsic power spectrum

$$P_{\rm II}(z,k) = [f_{\rm IA}(z)]^2 P_{\delta\delta}(z,k) \tag{5.1}$$

- Matter-intrinsic power spectrum:

$$P_{\delta{\rm I}}(z,k) = [f_{\rm IA}(z)] P_{\delta\delta}(z,k) \tag{5.2}$$

- Galaxy-intrinsic power spectrum:

$$P_{\rm gI}^{\rm photo}(z,k) = [f_{\rm IA}(z)] b_g^{\rm photo}(z) P_{\delta\delta}(z,k) \tag{5.3}$$

with

$$f_{\rm IA}(z) = -A_1 \bar{C}1 \, \frac{\Omega_m \cdot \rho_{crit}}{D(z)} \left(\frac{1+z}{1+z_0} \right)^{\eta_1} \, [\langle L \rangle(z)/L_\star(z)]^{\beta_1}, \tag{5.4}$$

where $\bar{C1} \cdot \rho_{crit} = 0.0134$, $D(z)$ is the growth factor, z_0 is the pivoting redshift, $b_g^{\rm photo}$ is the photometric galaxy bias and A_1, η_1, β_1 are free NLA parameters, the two latter ones indicating the redshift and luminosity dependence, respectively.

We can consider three alternatives in the NLA model. First, for $\eta_1 = 0$ and $\beta_1 = 0$, we recover the *minimal* NLA model. Second, for $\eta_1 \neq 0$ and $\beta_1 \neq 0$, we have the *extended* NLA model (eNLA), having a dependence in redshift and luminosity. Finally, for $\eta_1 \neq 0$ and $\beta_1 = 0$, we have a model with redshift dependence, which we refer to as zNLA.

5.3.1 Details of the implementation

In the case of the TATT implementation, we follow the expressions in Blazek et al. [18], which give the following power spectra:

- Intrinsic-intrinsic power spectrum:

$$\begin{aligned} P_{\rm II}^{EE}(z,k) = C_1^2 P_{\delta\delta}(z,k) + 2C_1 C_{1\delta} D(z)^4 [\mathrm{A_{0|0E}(k) + C_{0|0E}(k)}] \\ + C_{1\delta}^2 D(z)^4 \mathrm{A_{0E|0E}(k) + C_2^2 D(z)^4 A_{E2|E2}(k) + 2C_1 C_2 D(z)^4} \\ \mathrm{[A_{0|E2}(k) + B_{0|E2}(k)] + 2C_{1\delta} C_2 D(z)^4 D_{0E|E2}(k)} \end{aligned} \quad (5.5)$$

- Matter-intrinsic power spectrum:

$$\begin{aligned} P_{\delta \rm I}(z,k) = C_1 P_{\delta\delta}(z,k) + C_{1\delta} D(z)^4 [\mathrm{A_{0|0E}(k) + C_{0|0E}(k)}] \\ + C_2 D(z)^4 [\mathrm{A_{0|E2}(k) + B_{0|E2}(k)}] \end{aligned} \quad (5.6)$$

- Galaxy-intrinsic power spectrum:

$$\begin{aligned} P_{\rm GI}^{\rm photo}(z,k) = b_g^{\rm photo}(z) * [C_1 P_{\delta\delta}(z,k) + C_{1\delta} D(z)^4 [\mathrm{A_{0|0E}(k) + C_{0|0E}(k)}] \\ + C_2 D(z)^4 [\mathrm{A_{0|E2}(k) + B_{0|E2}(k)}]] \end{aligned} \quad (5.7)$$

with

$$C_1(z) = -A_1 \cdot \bar{C1} \cdot \frac{\Omega_m \cdot \rho_{crit}}{D(z)} \cdot \left(\frac{1+z}{1+z_0} \right)^{\eta_1}, \quad (5.8)$$

$$C_2(z) = 5 \cdot A_2 \cdot \bar{C1} \cdot \frac{\Omega_m \cdot \rho_{crit}}{D(z)^2} \cdot \left(\frac{1+z}{1+z_0} \right)^{\eta_2}, \quad (5.9)$$

Table 5.1 Fiducial values and priors for the CLOE chains.

Parameter	$\Omega_m h^2$	$\Omega_b h^2$	H_0	$\log A_s$	n_s	w	w_a	Σm_ν	$N_{\rm eff}$	τ	A_1	A_2	η_1	η_2	$A_{1\delta}$
Value	0.12	0.023	67.37	3.04	0.966	-1	0	0.06	3.046	0.05	0.16	0.4	1.66	1.5	2
Prior	[0.01,0.37]	-	[55,91]	[1.61,3.91]	-	-	-	-	-	-	[-5,5]	[-5,5]	[-5,5]	[-5,5]	[-5,5]

$$C_{1\delta(z)} = A_{1\delta} \cdot C_1(z) \tag{5.10}$$

where A_1, A_2, $A_{1\delta}$, η_1 and η_2 are the TATT parameters. Note that the parameters A_1 and η_1 are shared with the NLA model, which is recovered in the case where A_2=η_2=$A_{1\delta}$=0. The one loop terms in eq. 5.5, 5.6, 5.7 are defined in Blazek et al. [18]. As a remark, for the moment, no luminosity dependence has been included in the TATT implementation, as opposed to the case of NLA, where it is desired to have that flexibility given the luminosity dependence of the NLA model, as seen in Fig. 4.29.

The implementation of TATT was validated using PyCCL and CosmoSIS, by comparing the power spectra and the C_l observables of each code. Section 5.4 shows the details of the validation with CosmoSIS. Additionally to that validation, we also introduced new unit tests in CLOE to assess that the consistency of the code was guaranteed.

5.3.2 Chains with CLOE

In this section, we analyse the contour plots obtained when performing parameter inference using synthetic data vectors (DVs) in CLOE. This is an additional validation test, which complements the other ones we have already referenced before, since it revalidates the consistency of the code.

For this test, we consider a 3x2pt scenario with 13 redshift bins, which is the expected specification for the Euclid final data release analyses. Fig. 5.1 shows the $n(z)$ distribution for these 13 tomographic bins, which range from $z = [0-3]$, with the majority of the bins concentrated at $z < 1.5$. Note that these are not necessarily the tomographic bins that will be used in the final analyses and are subject to change. When performing the parameter inference, we do not vary neither the magnification bias nor the photo-z error, so as to simplify the exercise. We consider a l range of [10-3000] in the modelling of the three observables: GCphot, WL and WL-GCphot. We model the non-linear matter power spectrum with HMCode-2020 [156] and we use the *mcmc* sampler integrated in Cobaya. Table 5.1 shows the fiducial values employed to generate the synthetic DVs, together with the priors used.

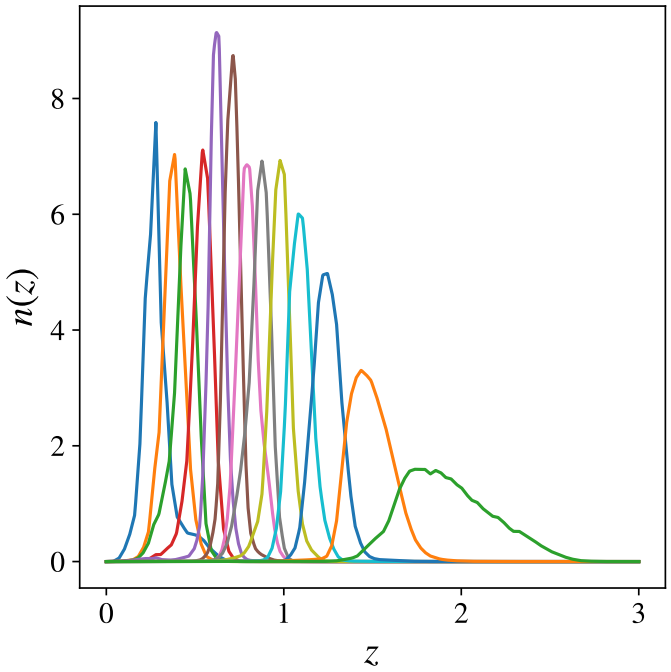

Fig. 5.1 $n(z)$ distribution for the Euclid configuration with 13 tomographic bins.

Fig. 5.2 shows the results, in empty blue contours, of the chain obtained when only varying the TATT parameters, while leaving fixed all the cosmological parameters. We are able to recover the fiducial parameters (indicated by black star markers), which demonstrates that the test was successful. In order to check for convergence, the *mcmc* sampler in Cobaya outputs the Gelman-Rubin statistic [86] R. This convergence criterion compares the variance within individual chains and the variance between chains to combine this information and compute the statistic R, which must be close to 1 for converged chains. In the case of this chain, a R-1 = 0.0115 is reached after 52183 steps. A most complex chain, which helps to further validate the TATT implementation, is also shown in the figure in red contours. In this case, besides from varying the TATT parameters, three cosmological parameters, A_s, Ω_m and H_0, are varied following the priors in Table 5.1. In order to speed up the convergence of this chain, the covariance matrix resultant from the same scenario with fixed TATT parameters is used. A R-1 = 0.00696 is reached after 91118 steps. Again, the fiducial values are recovered.

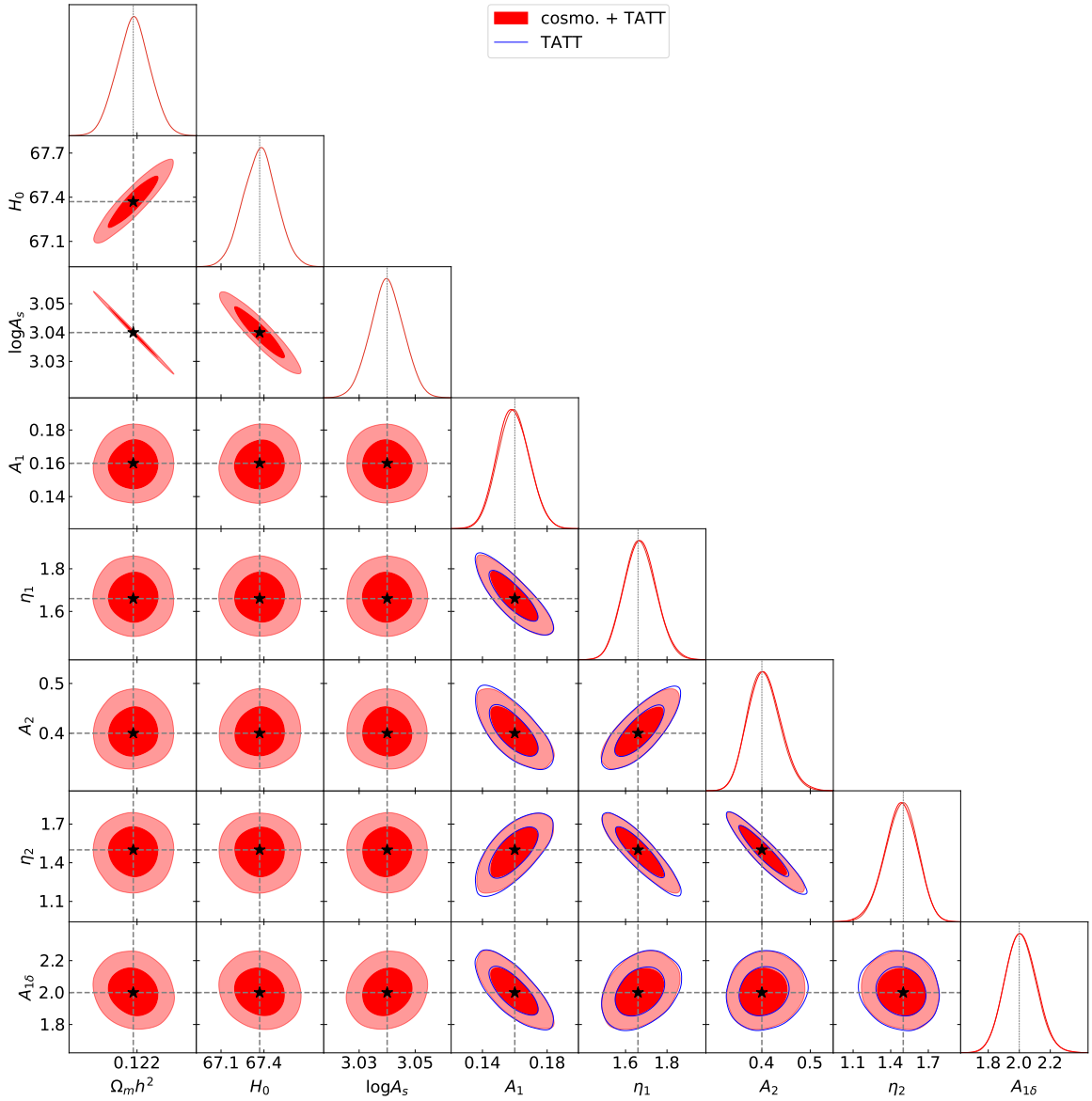

Fig. 5.2 Chains with CLOE. Blue: Contour plots varying the TATT parameters A_1, η_1, A_2, η_2 and $A_{1\delta}$. Red: Contour plots varying the cosmological parameters A_s, Ω_m and H_0 and the TATT parameters.

Table 5.2 Cosmological parameter values used in the validation with CosmoSIS.

Ω_m	Ω_b	Ω_k	H_0	n_s	w	w_a	A_s	Σm_ν	$N_{\rm eff}$	τ
0.32	0.05	0	67	0.96	-1	0	2.13e-9	0.06	3.046	0.0925

Table 5.3 Galaxy bias per tomographic bin used in the validation with CosmoSIS.

b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
1.10	1.22	1.27	1.32	1.36	1.40	1.44	1.50	1.56	1.74	1.74	1.74	1.74

5.4 Validation with CosmoSIS

The TATT implementation in CLOE is validated with the external code CosmoSIS. This is a cosmological parameter estimation code widely used in the scientific community. It is based on calculation modules, which enhance flexibility, re-usability, debugging, verifiability, and code sharing. Each module allows and/or contributes to add specific computations to the observables, such as the power spectra, the correlation functions or the systematics. Besides the computation of observables, CosmoSIS is able to perform parameter inference through different sampling methods, including MCMC samplers and nested samplers (see Section 1.9.2), amongst others. The validation with CosmoSIS of the TATT implementation in CLOE is beneficial in two ways. On the one hand, it allows checking that the implementation is robust, validating that both codes give the same results. On the other hand, some groups in the Euclid Consortium are performing parameter inference with CosmoSIS, so that validating CLOE with CosmoSIS is an essential process.

The cosmological parameters used for this validation are defined in Table 5.2, where the number of massive neutrinos is set to 1. For this validation, we assess that both the NLA and the TATT values in CLOE agree with the ones in CosmoSIS. For NLA, we use the parameters A_1: 0.16, η_1: 1.66 and β_1: 0.0, while for TATT, we use A_1: 0.16, A_2: 0.4, $A_{1\delta}$: 2, η_1: 1.66, η_2: 1.5. The pivoting redshift z_0 is set to 0. As in the previous Section 5.3.2, 13 tomographic bins are used, with the galaxy bias defined per bin in Table 5.3 (constant galaxy bias per bin are assumed). In order to simplify the validation, the magnification biases are set to 0 and no RSD is considered. In this case, the Takahashi non-linear model is used.

The implementation in CLOE and CosmoSIS differ in some aspects, detailed in the following items:

- The CosmoSIS implementation do not have the prefactors defined in the C_l of eq. 5.12-5.20.

- The k_0 used to normalize the growth factor is different, being $k_0 \sim 0.02$ for CosmoSIS and $k_0 = 0.05$ for CLOE.
- Some precision parameters related with the definition of redshift arrays vary between both codes.

As a consequence of these differences, in order to perform the validation, we add the prefactors to the C_l in the CosmoSIS output and we adjust the CosmoSIS k_0 and the precision parameters to match those in CLOE. The validation is done in two levels. First, we check that the power spectrum given by both codes is the same within errors and, later, we check the equivalence in the C_l.

5.4.1 Comparison of the power spectra

In this Section 5.4.1, we compare the power spectra from CLOE and CosmoSIS. First, we focus on the non-linear matter power spectrum, $P_{\delta\delta}$, since the consistency of this case is the basis for the consistency of all the others. Next, we focus on the power spectra that include IA terms, which are $P_{\delta I}$, P_{II} and P_{gI}. We also assessed that the P_{gg} given by both codes are consistent, although we do not show it here since it does not depend on the IA terms.

Fig. 5.3 shows the comparison of $P_{\delta\delta}$ for CLOE (solid line) and CosmoSIS (star markers) as a function of k at different four redshift bins: $z = 0$, $z = 0.4$, $z = 1.49$, $z = 2.02$. The bottom panels in each plot show the difference in % between both $P_{\delta\delta}$, normalized by the $P_{\delta\delta}$ of CosmoSIS. Contour areas in the bottom panels indicate a normalized difference of 0.1% (dark gray) and 0.2% (light gray), respectively. From this figure, we assess that the non-linear matter spectrum from both codes agree well with each other for all redshift bins and all the k range, with maximum differences of $\sim 0.2\%$.

Fig. 5.4, 5.5 and 5.6 show a similar structure as Fig. 5.3, but displaying the $P_{\delta I}$, P_{II} and P_{gI}, respectively, for the NLA (black) and TATT (blue) IA models. Again, the CLOE and CosmoSIS implementations yield the same results with maximum differences of $\sim 0.2\%$. Also, note that the NLA and the TATT models in CLOE show very similar differences with CosmoSIS, indicating that both models are correctly implemented.

5.4.2 Comparison of the C_l

In this Section 5.4.2, we study the consistency between the C_l affected by IA terms obtained by CLOE and CosmoSIS, which are the WL C_l and the WLxGCphot C_l. For completeness, we also checked that the GCphot C_l is consistent between codes, since this observable also

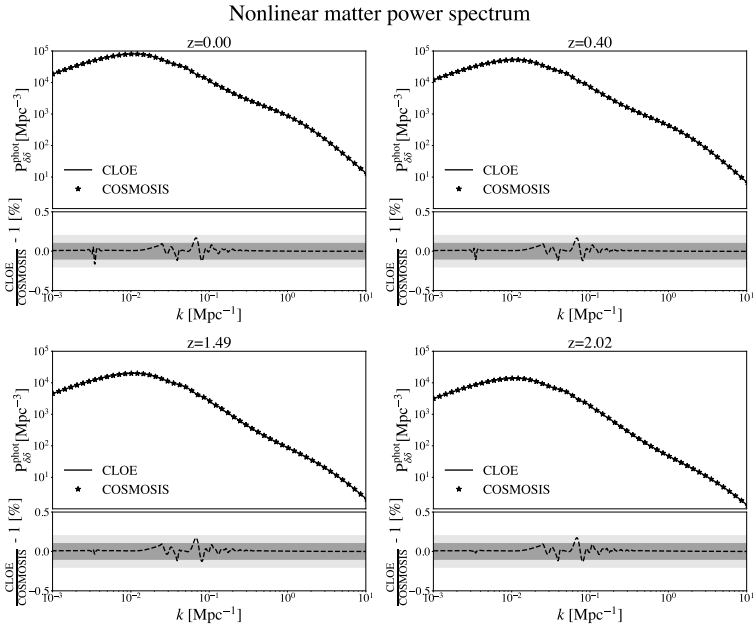

Fig. 5.3 Comparison of the CLOE and CosmoSIS $P_{\delta\delta}$ as a function of the wavelength (k) at different redshift bins. The bottom plots show the normalized percentage difference between both cases.

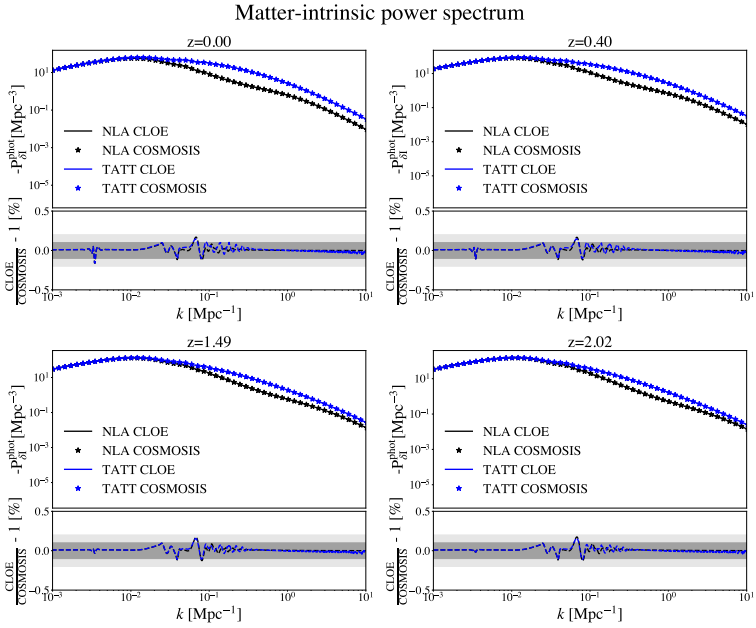

Fig. 5.4 Comparison of the CLOE and CosmoSIS $P_{\delta I}$ as a function of the wavelength (k) at different redshift bins. The bottom plots show the normalized percentage difference between both cases. The NLA and the TATT IA models are compared with different colours and different line styles.

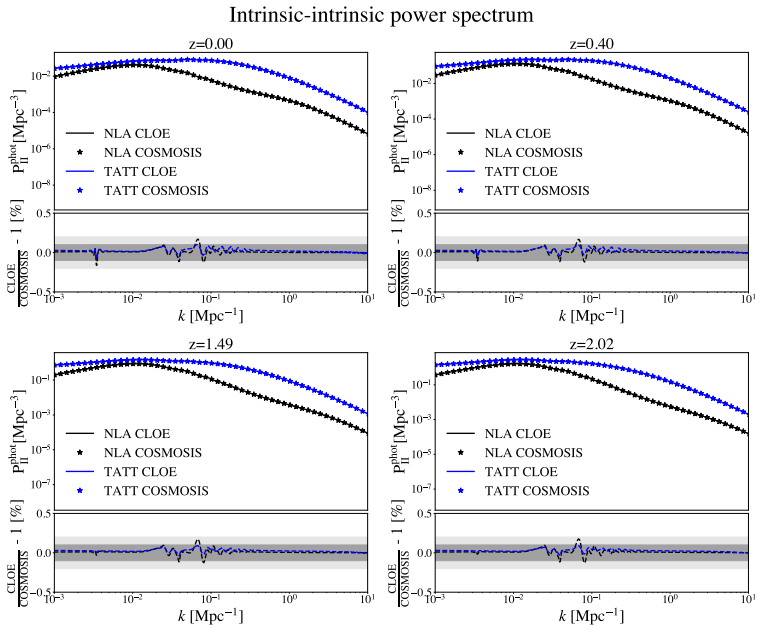

Fig. 5.5 Comparison of the CLOE and CosmoSIS P_{II} as a function of the wavelength (k) at different redshift bins. The bottom plots show the normalized percentage difference between both cases. The NLA and the TATT IA models are compared with different colours and different line styles.

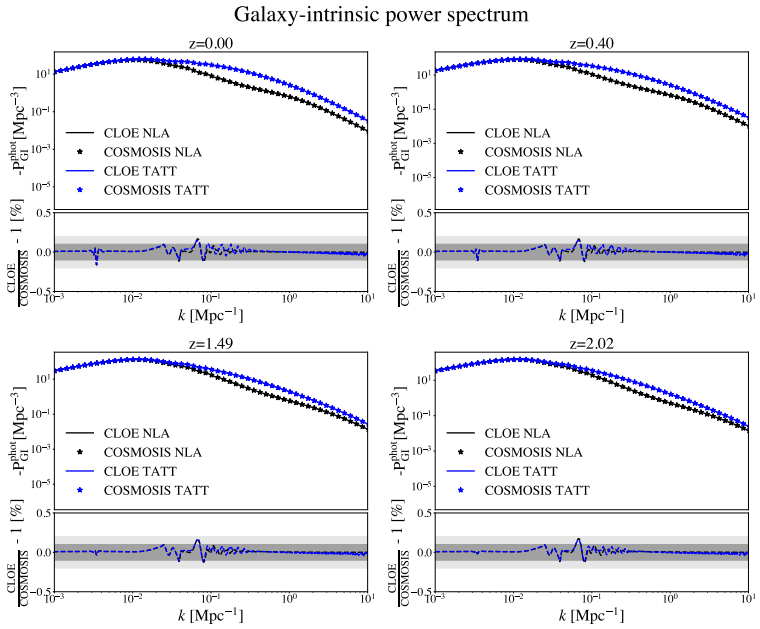

Fig. 5.6 Comparison of the CLOE and CosmoSIS P_{gI} as a function of the wavelength (k) at different redshift bins. The bottom plots show the normalized percentage difference between both cases. The NLA and the TATT IA models are compared with different colours and different line styles.

enters the 3x2pt analysis, but we do not show it here for conciseness, given that it does not depend on IA.

WL C_l

The WL C_l in CLOE for the correlation of tomographic bins i and j ($C_{ij}^{\rm LL}$) is composed by the contribution from the shear-shear ($C_{ij}^{\gamma\gamma}(\ell)$), shear-intrinsic ($C_{ij}^{\rm I\gamma}(\ell)$) and intrinsic-intrinsic ($C_{ij}^{\rm II}(\ell)$) terms:

$$C_{ij}^{\rm LL}(\ell) = C_{ij}^{\gamma\gamma}(\ell) + C_{ij}^{\rm I\gamma}(\ell) + C_{ij}^{\rm II}(\ell), \tag{5.11}$$

with

$$C_{ij}^{\gamma\gamma}(\ell) = \frac{(\ell+2)!}{(\ell-2)!}\left(\frac{2}{2\ell+1}\right)^4 \int {\rm d}z \frac{W_i^{\gamma}(z) W_j^{\gamma}(z)}{H(z) S_k^2\left[\chi(z)\right]} P_{\delta\delta}^{\rm photo}\left[k(z), z\right], \tag{5.12}$$

$$C_{ij}^{\rm I\gamma}(\ell) = \frac{(\ell+2)!}{(\ell-2)!}\left(\frac{2}{2\ell+1}\right)^4 \int {\rm d}z \frac{W_i^{\gamma}(z) W_j^{\rm I}(z) + W_i^{\rm I}(z) W_j^{\gamma}(z)}{H(z) S_k^2\left[\chi(z)\right]} P_{\delta{\rm I}}^{\rm photo}\left[k(z), z\right], \tag{5.13}$$

$$C_{ij}^{\rm II}(\ell) = \frac{(\ell+2)!}{(\ell-2)!}\left(\frac{2}{2\ell+1}\right)^4 \int {\rm d}z \frac{W_i^{\rm I}(z) W_j^{\rm I}(z)}{H(z) S_k^2\left[\chi(z)\right]} P_{\rm II}^{\rm photo}\left[k(z), z\right], \tag{5.14}$$

where the prefactors placed before the integrals are the ones not included in the CosmoSIS implementation, S_k is defined in eq. 1.6 and $W^{\gamma}(z)$ and $W^{\rm I}(z)$ are the weighting functions for the shear (eq. 5.15) and for the IA (eq. 5.16), respectively:

$$W_i^{\gamma}(z) = \frac{3}{2}\left(\frac{H_0}{c}\right)^2 \Omega_{\rm m,0}(1+z)\Sigma(z,k) S_k\left[\chi(z)\right] \int_z^{z_{\rm max}} {\rm d}z' n_i^{\rm L}(z') \frac{S_k\left[\chi(z') - \chi(z)\right]}{S_k\left[\chi(z')\right]} \tag{5.15}$$

$$W_i^{\rm I}(z) = n_i^{\rm L}(z)\,\frac{H(z)}{c} \tag{5.16}$$

where $\Sigma(z,k)$ accounts for Modified Gravity models (which we do not take into account in this thesis) and $n_i^{\rm L}$ is the redshift distribution of tomographic bin i for the lenses. Note that eq. 5.15 is the same as eq. 1.109, but expressed in terms of z and adding the Modified Gravity model option.

Given the fact that $C_{ij}^{\rm LL}(\ell)$ crosses the 0 value for some l, it is hard to validate the similarity of the results with CLOE and CosmoSIS using this observable, due to numerical instabilities. In order to avoid this problem, the comparison is done via the individual contributions in

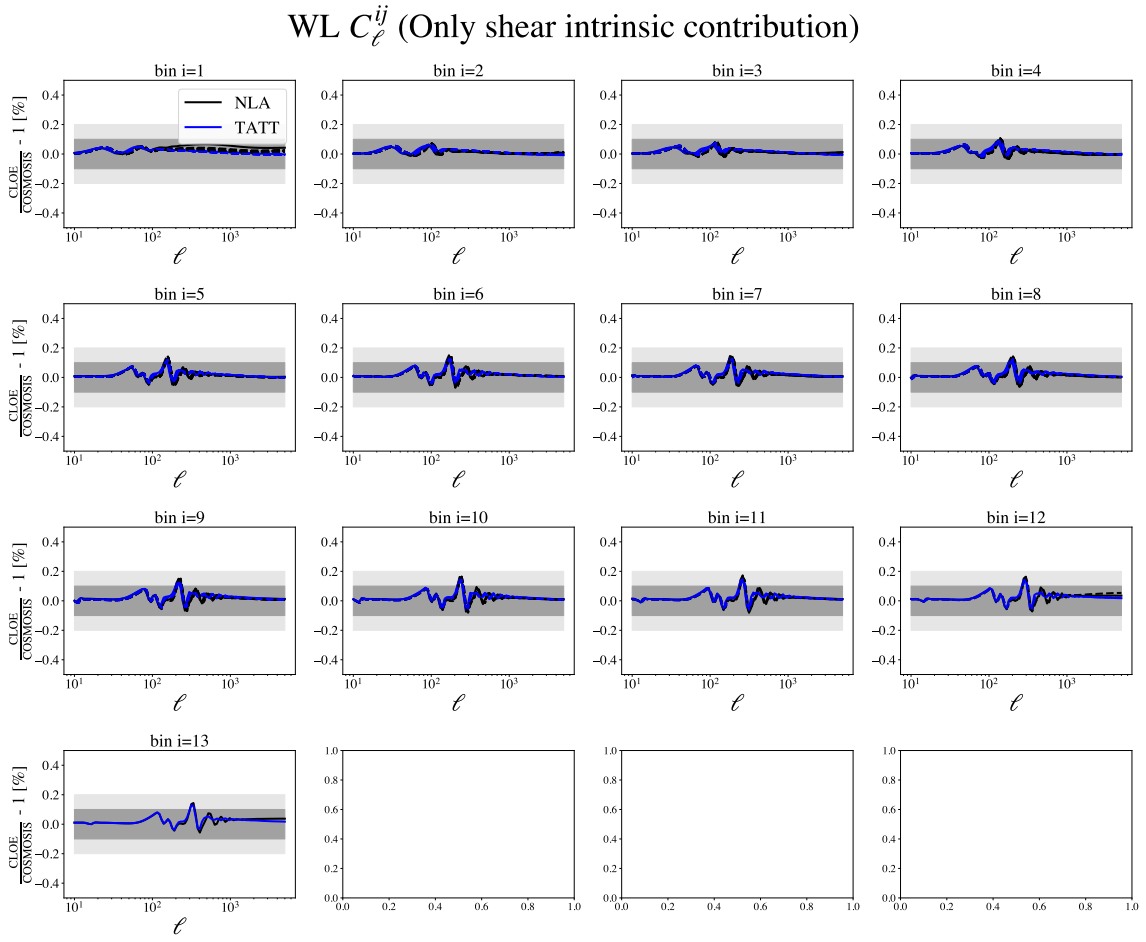

Fig. 5.7 Percentage difference for the NLA and TATT cases between the CLOE and the CosmoSIS shear-intrinsic contributions to the WL C_l, as a function of l for every combination of tomographic redshift bins.

eq. 5.11. In particular, we focus on the terms affected by the IA effect, which are $C_{ij}^{\mathrm{I}\gamma}(\ell)$ and $C_{ij}^{\mathrm{II}}(\ell)$. Nevertheless, the term $C_{ij}^{\gamma\gamma}(\ell)$ was also validated, but not shown here. Fig. 5.7 and 5.8 show the % differences of the shear-intrinsic and intrinsic-intrinsic terms, respectively, between CLOE and CosmoSIS, normalized by the CosmoSIS C_l. This is done for the 13 tomographic bins, where solid lines indicate auto-correlation of bins and dashed lines show cross-correlation of bins. The NLA (black) and the TATT (blue) models are shown. We note that the maximum differences are of $\sim 0.2\%$ for all cases. Also, the NLA and the TATT models implemented in CLOE agree with very similar precision to those in CosmoSIS.

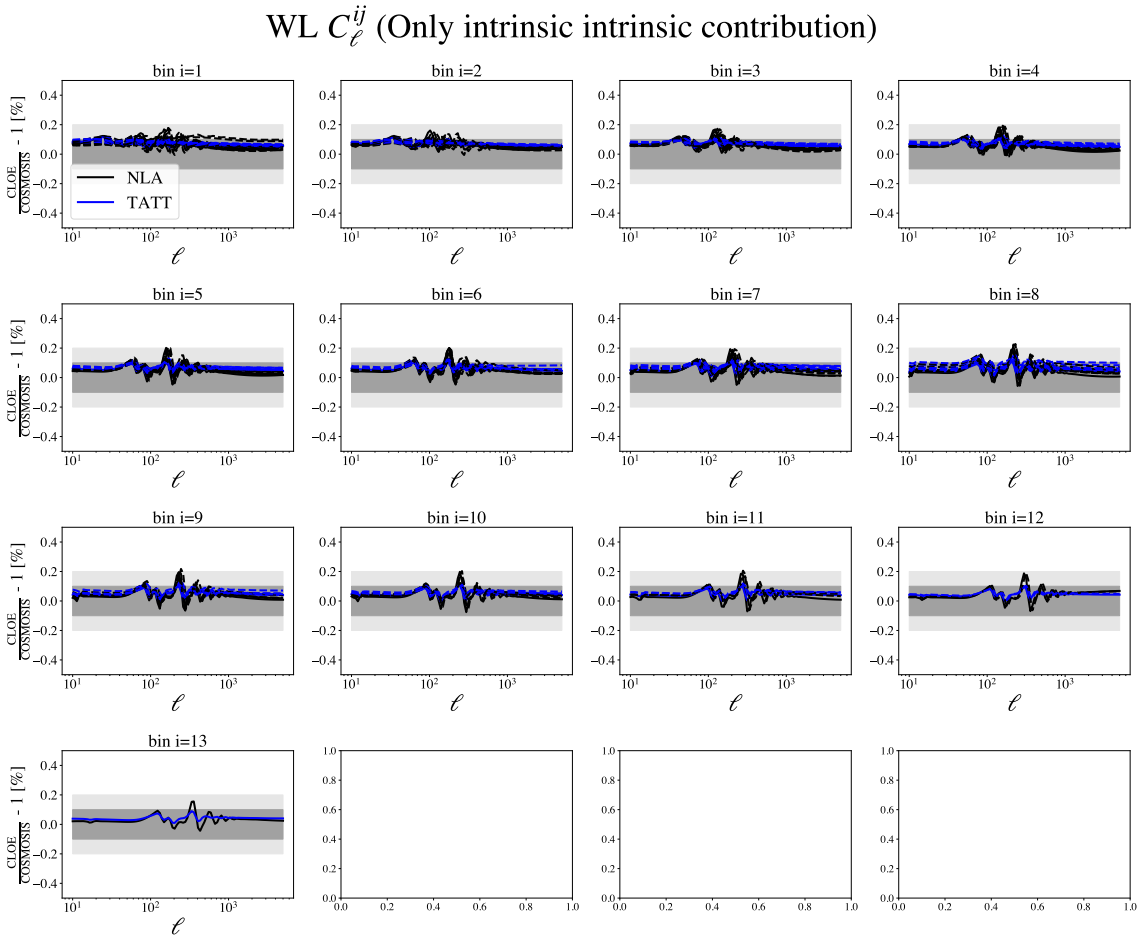

Fig. 5.8 Percentage difference for the NLA and TATT cases between the CLOE and the CosmoSIS intrinsic-intrinsic contributions to the WL C_l, as a function of l for every combination of tomographic redshift bins.

WLxGCphot C_l

The WLxGCphot C_l in CLOE for the correlation of tomographic bins i and j ($C_{ij}^{\rm LG}$) is composed by the contribution from the galaxy-shear ($C_{ij}^{\gamma{\rm g}}(\ell)$), galaxy-intrinsic ($C_{ij}^{\rm Ig}(\ell)$), magnification-shear ($C_{ij}^{\gamma\mu}(\ell)$) and magnification-intrinsic ($C_{ij}^{\rm I\mu}(\ell)$) terms:

$$C_{ij}^{\rm LG} = C_{ij}^{\gamma{\rm g}}(\ell) + C_{ij}^{\rm Ig}(\ell) + C_{ij}^{\gamma\mu}(\ell) + C_{ij}^{\rm I\mu}(\ell). \tag{5.17}$$

However, with the goal of simplifying the comparison, and since the magnification does not interfere with the IA implementation, we set the magnification terms to 0, so that eq. 5.17 can be redefined as:

$$C_{ij}^{\rm LG} = C_{ij}^{\gamma{\rm g}}(\ell) + C_{ij}^{\rm Ig}(\ell), \tag{5.18}$$

with

$$C_{ij}^{\gamma{\rm g}}(\ell) = \left(\frac{(\ell+2)!}{(\ell-2)!}\right)^{1/2}\left(\frac{2}{2\ell+1}\right)^{2}\int {\rm d}z \frac{W_i^{\gamma}(z)W_j^{\rm g}(z)}{H(z)S_k^2\left[\chi(z)\right]}P_{g\delta}^{\rm photo}\left[k(z),z\right] \tag{5.19}$$

$$C_{ij}^{\rm Ig}(\ell) = \left(\frac{(\ell+2)!}{(\ell-2)!}\right)^{1/2}\left(\frac{2}{2\ell+1}\right)^{2}\int {\rm d}z \frac{W_i^{\rm I}(z)W_j^{\rm g}(z)}{H(z)S_k^2\left[\chi(z)\right]}P_{\rm gI}^{\rm photo}\left[k(z),z\right] \tag{5.20}$$

Again, the prefactors that are in CLOE are not introduced in CosmoSIS. The galaxy weight function is defined as:

$$W_i^{G}(z) = n_i(z)\frac{H(z)}{c} \tag{5.21}$$

As in the validation of the WL C_l, we focus on comparing the terms that depend on IA in eq. 5.18, which is just $C_{ij}^{\rm Ig}(\ell)$. Fig. 5.9 shows the % differences of the galaxy-intrinsic contributions to the WLxGCphot C_l, with the same structure shown before in Fig. 5.7 and 5.8. Again, we are able to validate the agreement between the CLOE and the CosmoSIS implementations, with very similar levels of accuracy in both IA models.

5.5 Differences between NLA and TATT

As indicated in Section 5.3, the TATT model has been implemented in CLOE in order to be able to better understand the non-linear physics of IA, in the case they were better constrained

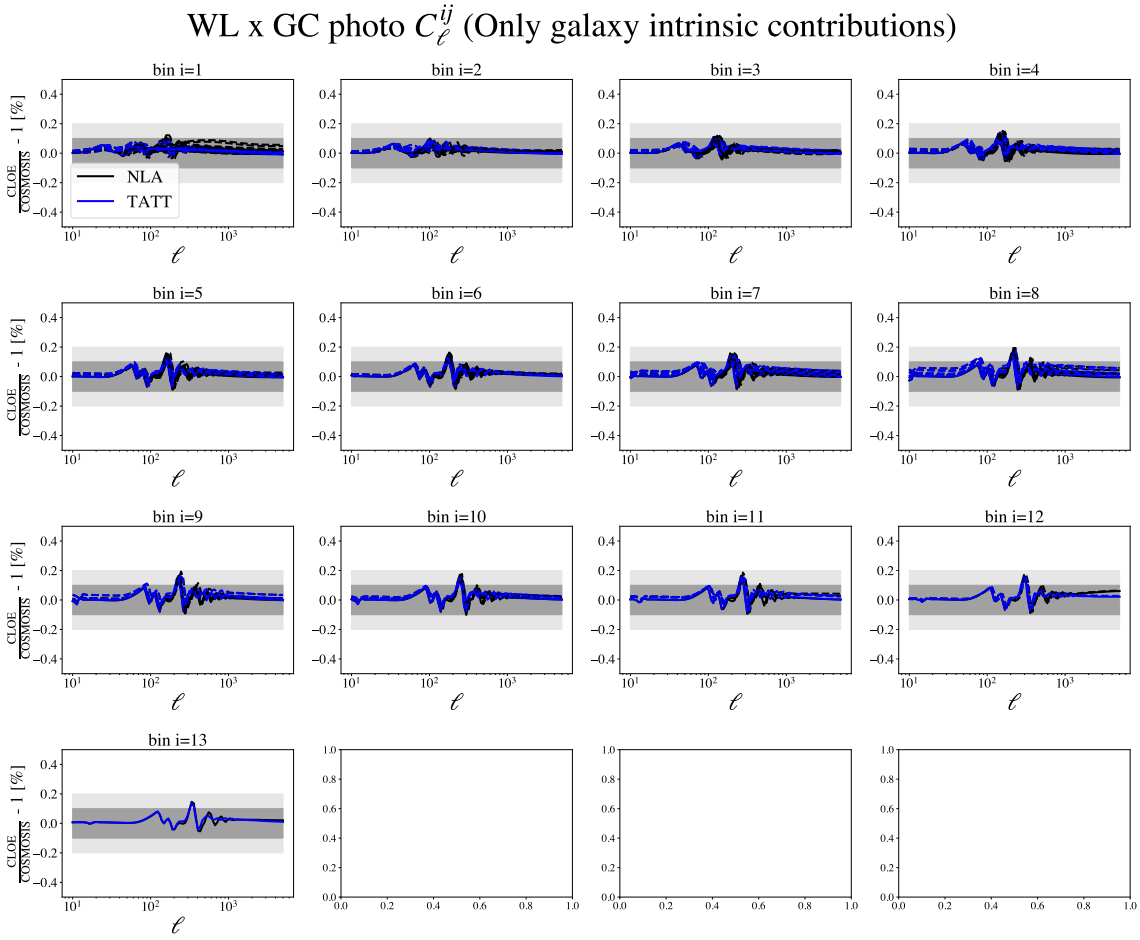

Fig. 5.9 Percentage difference for the NLA and TATT cases between the CLOE and the CosmoSIS galaxy-intrinsic WLxGCphot contributions C_l, as a function of l for every combination of tomographic redshift bins.

by TATT. In this Section 5.5, we study the differences in the WL C_l and the WLxGCphot C_l signals when using the NLA and the TATT models, with the objective of understanding how much of a difference in the observables these two models can carry. In order to do that, we use the parameters specified in Table 5.1, but modifying the IA parameters to $A_1 = 1.72$, $A_2 = 2$, $\eta_1 = -0.41$, $\eta_2 = 1$ and $A_{1\delta} = 1$. We perform this change so as to select a set of IA parameters that reflects well the differences between NLA and TATT. Note that different sets of parameters will lead to different scenarios and we do not intend to show here an exhaustive comparison in all the IA parameter space, but to show an illustrative example that allows to understand why the TATT implementation performed in CLOE is important.

Fig. 5.10 shows the differences in the WL C_l for the NLA (red) and the TATT (black) models at the auto-correlation of 4 different tomographic bins, from lower (bin 1) to higher redshifts (bin 10). The bottom panels show the ratio of the NLA over the TATT case. One can differentiate between two scenarios: large scales (low l) and small scales (large l). At large scales, the results for both NLA and TATT are consistent, which is expected since the TATT model is an expansion of NLA, accounting for second order terms in the expansion of the tidal field (eq. 1.123). However, at small scales, NLA and TATT show differences, given that TATT is able to model IA at smaller scales than NLA. This leads to a lack of signal in NLA at those scales, as shown in the figure. These differences are stronger for lower redshift bins where, in the case of bin = 1, the ratio of NLA over TATT drops close to 0. The reason for this is that the IA effect is larger at low redshift bins, as we will see in the next Section 5.6.

Fig. 5.11 shows the same configuration as Fig. 5.10, but showing the WLxGCphot C_l. As in the previous case, the IA models are consistent at large scales, having ratios closer to 1. However, as one moves to smaller scales, the ratio decreases until dropping to values close to 0 for all redshift bins. Thus, the difference between NLA and TATT for the WLxGCphot observable seems to be more pronounced at small scales, affecting all redshift bins similarly.

5.6 Contribution of IA to the total cosmic shear and galaxy-galaxy lensing signal.

Having analysed the differences in the WL and WLxGCphot observables when using NLA and TATT, we focus on the contribution of IA to the total C_l signal when using the TATT model. We follow eq. 5.11 and eq. 5.18 to describe the terms for both WL C_l and WLxGCphot C_l, respectively. Again, we set the magnification terms to 0 to simplify the analysis. For each

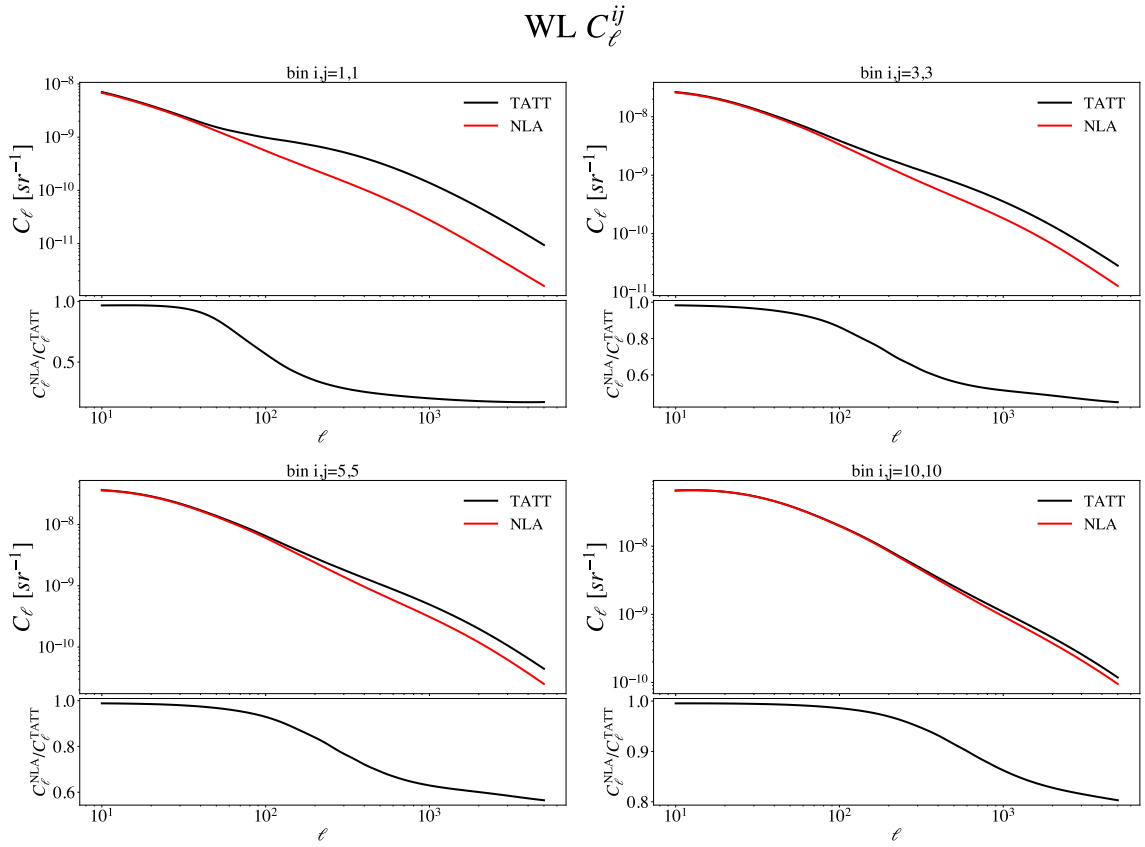

Fig. 5.10 Differences in the WL C_l signal for the NLA and the TATT models for different auto-correlations of tomographic bins.

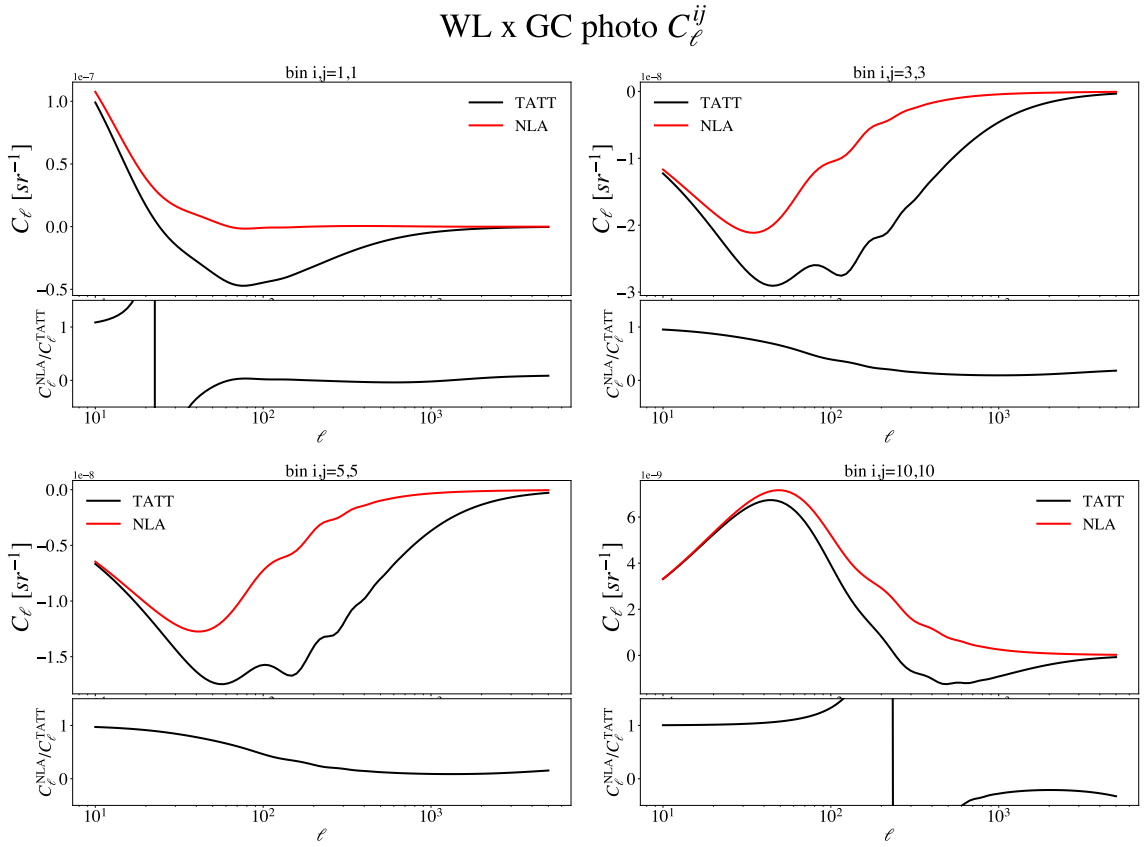

Fig. 5.11 Differences in the GCxWL C_l signal for the NLA and the TATT models for different auto-correlations of tomographic bins.

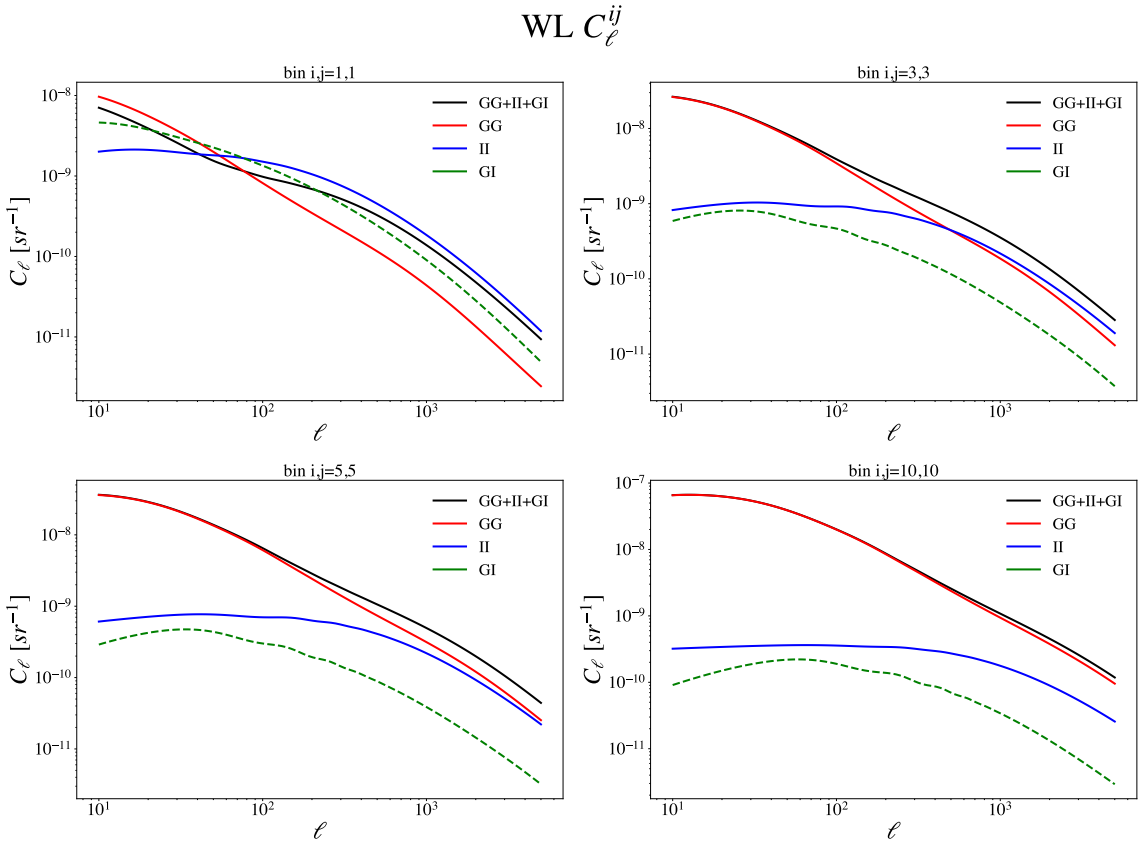

Fig. 5.12 WL contributions from the GG, II, GI to the total signal for auto-correlated bins.

of the observables, we show the auto and the cross correlations between tomographic bins. The IA parameters are the same as in Section 5.5.

Fig. 5.12 shows the GG (red), II (blue), GI (green) and the total (black) contributions to the WL C_l when autocorrelating 4 tomographic bins. The dashed lines indicate negative values. On the one hand, the II term decreases as we move towards higher redshift bins, with the smallest redshift bin constituting the main contribution to the total signal at scales smaller than $l = 50$. This decline with redshift of the II signal, in comparison to the GG term, affects progressively smaller scales, with the highest redshift bin being mainly dominated by the GG term. On the other hand, the GI term is only comparable to the GG and II terms at the lowest redshift bin, which may partially counteract the effect of the II term, given that they have different signals. However, this GI term is almost residual, in comparison to the GG term, for the other higher redshift bins.

Fig. 5.13 shows the same configuration of Fig. 5.12 but presenting the results when cross-correlating tomographic bins. The first two cases (top right and top left) keep one of the bins fixed to $j = 1$ while increasing by two the other bin, from $i = 3$ to $i = 5$. From

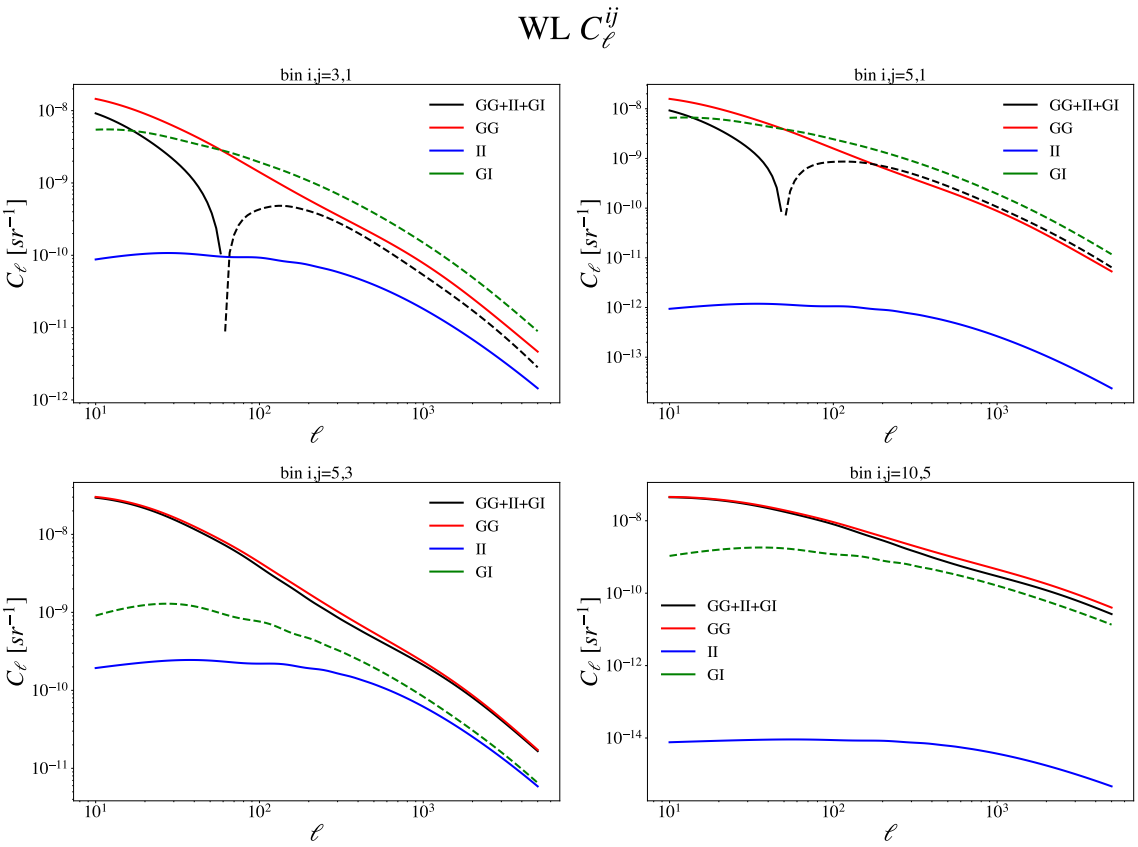

Fig. 5.13 WL contributions from the GG, II, GI to the total signal for cross-correlated bins.

the second to the third case (bottom left), the first bin is fixed, $i = 5$, and the second one is increased by two, from $j = 1$ to $j = 3$. Finally, for the fourth case (bottom right), the distance between both bins is the same as in the second case but at higher redshifts, $i = 10$ and $j = 5$. Since IA is a local effect, the II term when cross-correlating bins is almost negligible for all the configurations. This is not the case for the GI term, which is the main driver of the total signal for the two first cases. Note that from the first to the second scenarios, the tomographic bin accounting for shear in the GI term is increased, from $i = 3$ to $i = 5$, which leads to an increase in both GG and GI terms, as the lensing kernel (eq. 5.15) integrates a higher redshift space. In the third scenario, where the redshift bin accounting for IA in the GI term is increased from $j = 1$ to $j = 3$, this GI term decreases since the IA effect is less important, as seen in the previous Fig. 5.12 when going from the 1-1 to the 3-3 auto-correlation. Finally, the last case has much less impact from IA, with respect to the second case (which has the same separation between tomographic bins), since the IA effect is reduced when going from $j = 1$ to $j = 5$, as it was also seen in Fig. 5.12.

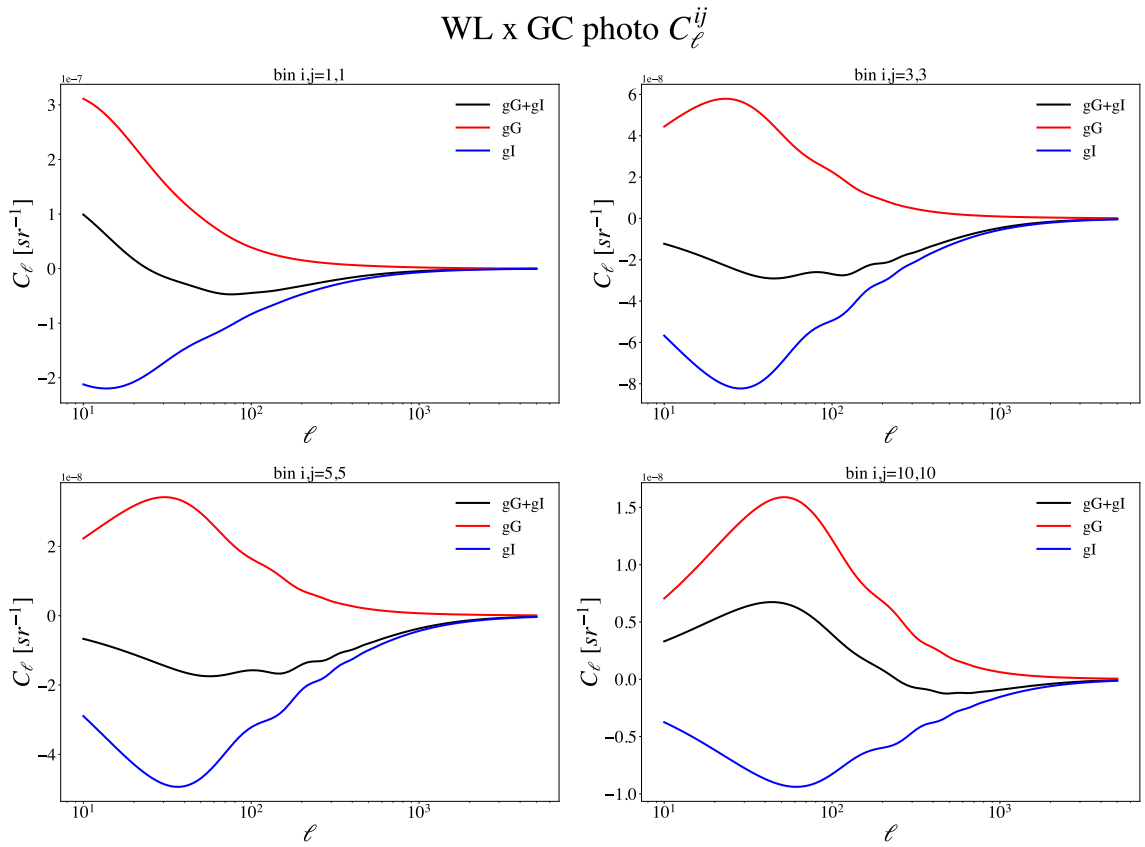

Fig. 5.14 WLxGC contributions from the gG, gI terms to the total signal for auto-correlated bins.

Now, we turn our attention to the contributions to the WLxGC C_l, which are the gG and gI terms.

Fig. 5.14 shows the results when auto-correlating the same tomographic bins as in Fig. 5.12. A few interesting remarks can be done about this case. First, the gG and the gI contributions have different signs, with the gG term having positive correlations and the gI term negative ones, although this depends on the shape convention one defines. Second, as opposed to the shape-shape auto-correlation in Fig. 5.12, the galaxy-shape auto-correlation yields similar values for both the gG and gI terms for all tomographic bins included here. Note how, even though there are some differences, the total signal is much reduced, in comparison to the gG term, in all cases. In other words, the effect of IA in the WLxGC C_l observable is highly significant. Third, related to the previous remark, with the auto-correlation of WLxGC maximising the IA effect, this kind of auto-correlations are the ones analysed in the measurements of IA. As an example of this, the measurements performed in the previous Chapter 4. Nevertheless, in that case the condition of proximity of galaxy pairs was more

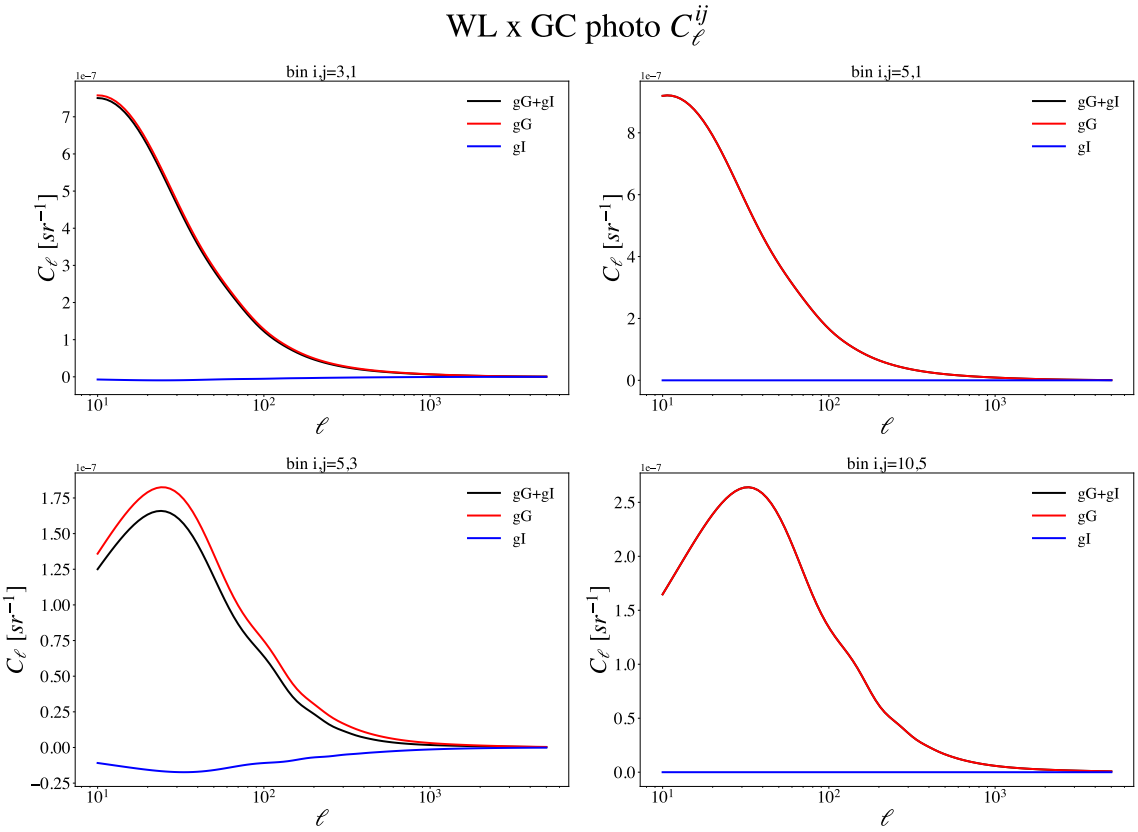

Fig. 5.15 WLxGC contributions from the gG, gI terms to the total signal for cross-correlated bins.

restrictive, since we did not use tomographic bins but estimates of photometric redshifts for each individual galaxy, which allowed to define a narrower redshift separation between pairs, maximising even more the effect of IA.

Fig. 5.15 shows the cross-correlation, with the same tomographic bins as in Fig. 5.13, of the WLxGC C_l. In all cases depicted here, the gI effect is almost negligible in all cases, indicating that the gI term is a local effect.

5.7 A case study: mismodelling IA

In this Section 5.7, we apply the lessons learned in the previous Sections 5.5 and 5.6. In particular, we aim at validating the effect it can have on the constraining power of cosmological and IA parameters to use the NLA or the TATT models, considering the level of contamination to the observables on the total signal.

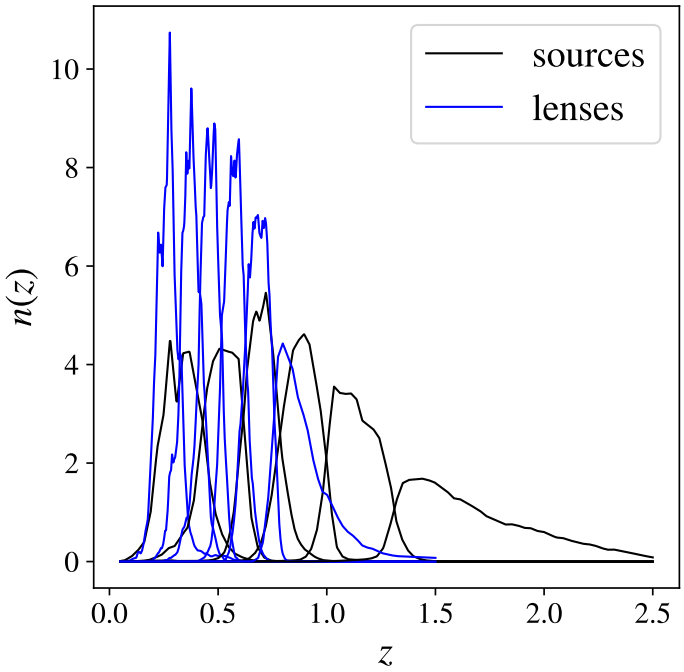

Fig. 5.16 $n(z)$ distribution for the Euclid configuration with 6 tomographic bins.

To that end, we analyse two different scenarios. On the first one, we assume that IA are described by the NLA model in our universe, so that we generate a DV with that model, and we test the biases in the recovered parameters when modelling it with either NLA or TATT. On the second scenario, we analyse the opposite case, where we generate a DV with TATT and we model it with both NLA and TATT. This exercise can be considered as a mismodelling of the IA effect. Investigating this is essential because the IA model governing this systematic is not well-known, and it is crucial to understand the potential biases that may arise from such mismodelling.

This study is performed for a 3x2pt analysis with 6 tomographic bins. We use 6 bins, instead of the 13 bins expected for the last data release, in order to simplify the analysis and to approach to studies similar to those expected in the first data release in Euclid (DR1). Fig. 5.16 shows the $n(z)$ for the 6 tomographic bins employed, where the distribution of the lenses and the sources is different, with the lenses being at closer redshifts, as it is usually done for these 3x2pt analyses. The limiting magnitude for the lenses is 22.25 in the Euclid VIS I_E band, while the sources go deeper, until I_E magnitude 24.

For this analysis, CosmoSIS is used for both generating the fiducial DVs and performing the parameter inference, the latter one using the Nautilus sampler [139], where we consider

Table 5.4 Fiducial values and prior for the IA mismodelling chains.

Parameter	Ω_m	Ω_b	H_0	$A_s(10^{-9})$	n_s	w	w_a	Σm_ν	$N_{\rm eff}$	τ	A_1	A_2	η_1	η_2	$A_{1\delta}$
Value	0.319	0.049	67	2.1	0.96	-1	0	0.0773	3.046	0.0697	0.16	0.4	1.66	1.5	2
Prior	[0.1,0.9]	[0.03,0.07]	[55,91]	[0.5,5.0]	[0.87,1.07]	-	-	-	-	-	[-5,5]	[-5,5]	[-5,5]	[-5,5]	[0,4]

Table 5.5 Fiducial values for galaxy bias and magnification parameters for the IA mismodelling chains.

Parameter	b_1	b_2	b_3	b_4	b_5	b_6	α_1	α_2	α_3	α_4	α_5	α_6
Value	1.09	1.28	1.28	1.40	1.49	1.79	0.27	0.27	0.31	0.36	0.48	0.87
Prior	[0.8,3.5]	[0.8,3.5]	[0.8,3.5]	[0.8,3.5]	[0.8,3.5]	[0.8,3.5]	-	-	-	-	-	-

systematic effects added by both IA and magnification. On the one hand, Table 5.4 shows the cosmological and IA parameters used for the generation of the DVs, which are also used as the fiducial values of the chain, together with the priors employed in the parameter inference. On the other hand, Table 5.5 shows the galaxy bias and the magnification parameters (expressed as α_i) for the 6 tomographic bins, where the galaxy bias parameters are also varied when constraining the parameters, as indicated by the specified priors.

Fig. 5.17 shows the constraints on Ω_m, H_0, Ω_b, n_s, A_s, A_1, A_2, η_1, η_2 and $A_{1\delta}$ when generating the DV with the NLA model and performing the constraints with TATT (red) and NLA (blue). Focusing on the cosmological parameters, the biases found when mismodelling a NLA DV with the TATT model are almost negligible, since the contour plots almost overlap. The reason for this is that TATT is an extension of NLA. Thus, we are able to recover similar, although less constrained, A_1 and η_1 parameters when using TATT to model the NLA DV. This inferior constraining power also arises from the fact that TATT has 3 more parameters than NLA, increasing the area of the contours when fitting similar DV. In contrast, the A_2 posterior from the TATT constraints is located around 0, as expected. The η_2 is unconstrained, since it corresponds to an exponent of A_2, which is centred around 0. Finally, the $A_{1\delta}$ also peaks around 0 since the $C_{1\delta}$ term (eq. 5.10) is null.

Fig. 5.18 shows the scenario when generating the DV with TATT and modelling with both TATT (red) and NLA (blue). This case is more interesting to study, given that we obtain biased cosmological parameters when modelling with NLA. Nevertheless, note that the contour plots in the cosmological parameters still show a large fraction of overlap, given that IA is, in general, a subdominant effect in WL C_l and WLxGCphot C_l, as seen in Section 5.6. An interesting remark about this Fig. 5.18 is the fact that, when modelling with NLA, the A_1 and η_1 parameters are also constrained, although at biased values. This is important because one might believe having constrained his measurements with the NLA model when, in fact, the true IA parameters that describe the DV do not correspond to the constrained parameters.

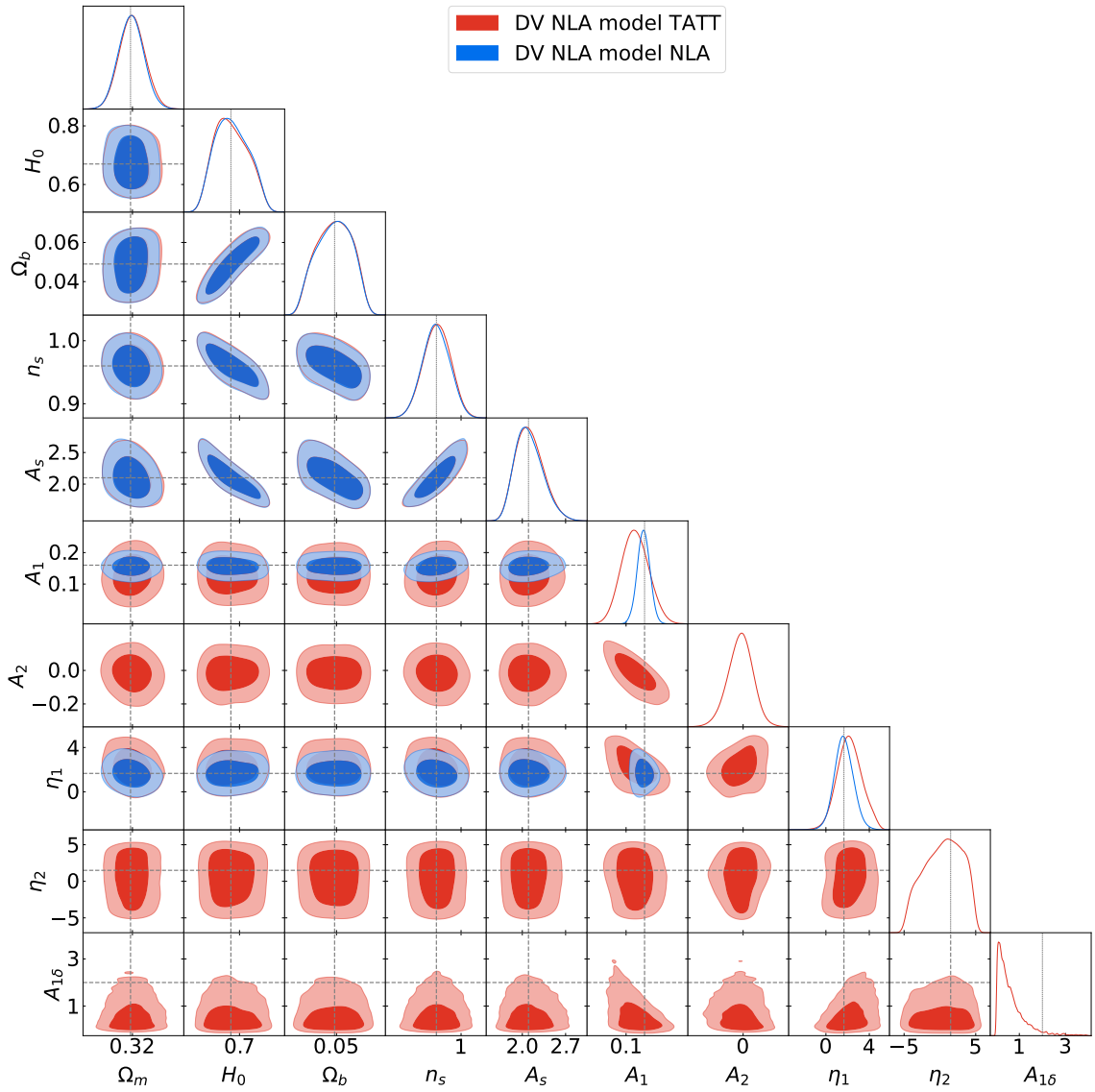

Fig. 5.17 Constraints on the cosmological and IA parameters when generating a DV with NLA and modelling it with TATT (red) and NLA (blue).

In this exercise, we do not intend to quantitatively study the biases in the parameters, but to qualitatively check if these biases exist. Besides, in both Fig. Fig. 5.17 and 5.18, the results may strongly depend on the chosen IA parameters, which may lead to different scenarios of bias.

5.8 Future work

In this Section 5.8, we briefly comment on future work that will make use of the TATT implementation described in this chapter. This future work is related to a key-project paper we recently started leading in the Euclid Consortium. The objective of the paper is to study how IA may affect the cosmological parameter inference in the Euclid analyses. For this, we will analyse numerous scenarios, including studies with different combinations of observables: 3x2p, 2x2pt (GCphot and WLxGCphot) and WL. This work will also use Flagship DVs, so that, instead of only fitting synthetic DVs generated with CosmoSIS, more realistic Flagship DVs will be employed. Different systematic effects will also be used in these analyses, besides IA and magnification, such as shear multiplicative bias and photo-z uncertainties. Additionally, we also expect to consider both cross-correlation between tomographic bins and go beyond the Limber approximation in the GC C_l (not included in this chapter).

All these analyses will be performed using both the NLA and the TATT models. In particular, an interesting test will be to check if we can move to smaller scale cuts when using TATT, and what are the benefits of doing this. Another study will focus on analysing how the cosmological constraints differ when employing NLA and TATT for the different combination of observables. Also, we aim at extending the mismodelling forecast we showed in Section 5.7 to different IA fiducial parameters, to understand the cosmological biases in a wider IA parameter space.

The results from this paper will focus on the Euclid DR1 configuration. With this, we aim to contribute to Euclid's first cosmological constraints and, in particular, to clarify the effect that IA has on them.

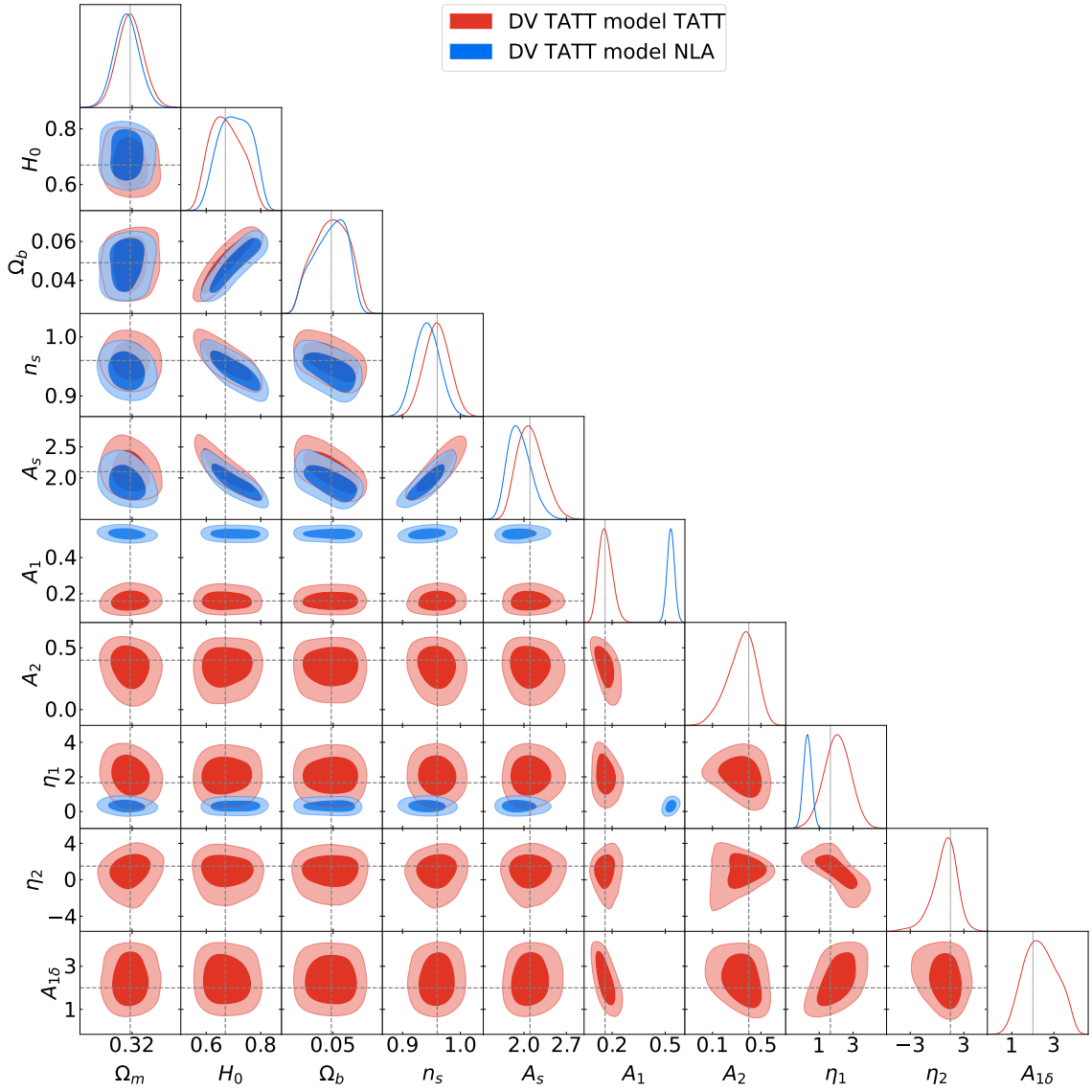

Fig. 5.18 Constraints on the cosmological and IA parameters when generating a DV with TATT and modelling it with TATT (red) and NLA (blue).

Chapter 6

Conclusions

This thesis focuses on the study of the large-scale structure of the universe using the Physics of the Accelerating Universe Survey (PAUS) and Euclid imaging surveys. In particular, we focus on the photometric redshift estimation and the Galaxy Clustering (GC) and Intrinsic Alignments (IA) measurements of the PAUS wide fields, two projects that are intrinsically related. In addition, we describe and validate the implementation of the Tidal Alignment Tidal Torquing (TATT) model in the likelihood pipeline of Euclid, together with an analysis on the effects that IA have on cosmological parameter estimation.

In Chapter 1, we review the theoretical background on which the content of this thesis relies. Then, in Chapter 2, we describe the PAUS and Euclid imaging surveys. Next, in Chapter 3, we estimate the photometric redshifts with PAUS fluxes, for around 1.8 million objects, together with the determination of the photometric redshift quality as a function of different physical properties. Subsequently, in Chapter 4, we measure the GC and IA observables, using the photo-z from Chapter 3 and the shapes from external broad-band surveys, and we analyse the dependence of these observables as a function of colour, redshift and luminosity. Finally, in Chapter 5, we describe the implementation of the TATT model in the likelihood pipeline of Euclid, including a study of the difference with another IA model and the impact of IA to the total cosmological observables. Additionally, we perform a forecast of mismodelling the IA systematic to understand the possible biases in Euclid's cosmological constraints.

Photometric redshift estimation in the PAUS wide fields

We present photometric redshifts from the PAUS wide fields, which cover a total area of $\sim$51 deg^2, for $\sim$1.8 million objects, down to $i_{\rm AB} = 23$ in the photo-z range of $0 < z_{\rm b} < 2$, with the vast majority of objects having $0 < z_{\rm b} < 1.2$.

We compute photometric redshifts with a SED template-based algorithm called BCNZ, introducing a modification in the zero-point calibration technique between observed and modelled fluxes, that allows us to be independent of spectroscopic samples, without degrading the performance of the photo-z.

Since the direct output from BCNZ degrades in performance for faint objects, where the NB photometry has low SNR, we introduce an innovative weighting scheme that combines the BCNZ estimates with the BB photo-z from BPZ. This weighted photo-z estimate (BCNZw) not only yields better accuracy for faint objects ($i_{\rm AB} > 22.5$), but also shows a reduced percentage of outliers.

The abundance of galaxies as a function of magnitude and the photo-z performance in each individual PAUS wide field are comparable. This enables us to study the performance of the combined fields in the main text, leaving for Appendix A the analysis of the individual fields. We obtain a σ_{68} of 0.003 for the brightest objects ($i_{\rm AB} \sim 19$) and a σ_{68} of $\sim$0.09 and $\sim$0.06 for the faintest ($i_{\rm AB} \sim 23$), in the BCNZ and BCNZw photo-z estimates, respectively. Using the weighted photo-z estimate introduced in this thesis, we manage to reduce the outlier fraction when studied as a function of $i_{\rm AB}$, $z_{\rm b}$ and $z_{\rm s}$.

We also study the dependence of the photo-z performance on galaxy colours, with colours determined from rest-frame absolute magnitudes. In general, we find that for bright ($i_{\rm AB} < 21$), low redshift objects ($z_{\rm b}$ and $z_{\rm s} < 0.75$), red galaxies have more accurate photo-z. In contrast, for faint objects, blue galaxies have slightly better photo-z estimates. However, the photo-z performance as a function of colour is relatively similar for both populations.

Finally, we validate and calibrate the probability density distribution, $p(z)$, given by BCNZ. We show how to properly correct $p(z)$ to follow the theoretical PIT distributions, indicating that they are well calibrated. This allows us to estimate the photo-z not only as a single-point measurement, but to assign a robust photo-z probability to each object. Nevertheless, we find better photo-z accuracy by taking the maximum of $p(z)$ as the photo-z estimate, rather than using the full $p(z)$ distribution.

Measuring galaxy clustering and intrinsic alignments in the PAUS wide fields

We measure the photometric GC and IA signals from the PAUS wide fields in the redshift range of $0.1 < z < 1.0$, down to $i_{\rm AB} < 22$, which accounts for approximately ~ 400000 objects. A deeper sample going down to $i_{\rm AB} < 22.5$ is also defined, but only for the CFHTLenS objects due to limitations in the shape estimation with KiDS. We analyse the GC and IA results by performing divisions in colour, redshift, and luminosity.

The GC and IA signals are measured by projecting three-dimensional correlation functions for both GC and IA, where these correlation functions employ the PAUS photo-z estimated in Chapter 3. As for the galaxy shapes, they are extracted from the KiDS and CFHTLenS external catalogues, and calibrated to account for the PSF and the multiplicative bias. The colour separation between red (passive) and blue (active) galaxies is defined by combining a NUVrK diagram, which traces the sSFR and dust attenuation, and a morphological parameter from the KiDS and CFHTLenS catalogues.

Together with the measurements, the modelling of the GC and IA signals is performed using photometric correlation functions, which allows obtaining constraints on the galaxy bias and IA parameters. This model accounts for the possible loss of signal due to the lower accuracy of photo-z, while also including contaminant terms coming from magnification and shear.

We include consistency tests to ensure that both our measurements and constraints are robust. For this purpose, we employ the MICE simulation, generating a galaxy mock that both resembles the galaxy populations obtained in PAUS and reproduces the PAUS-like photo-z. The consistency tests involve validating the random catalogues used in the measurements, checking that the use of photo-z does not bias our constraints compared to spec-z, and assessing that the error estimation in our PAUS measurements are suitably defined.

The measurements and the modelling are performed for three scenarios: 1) A division by colour 2) A division by colour and redshift 3) A division by colour and luminosity.

In the first case (colour-based division), we detect IA for red galaxies, with IA amplitudes consistent to $A_1 \sim 2.5$ while, for blue galaxies, we observe a null detection.

In the case of the analysis by redshift bins, the redshift space is split into three equipopulated regions. No detection of IA is seen for red objects in the lowest redshift bin while, for the intermediate and highest redshift bins, a similar IA amplitude of $A_1 \sim 3$ is observed, although with standard deviations close to $A_1 \sim 1$. For blue galaxies, measurements generally show null detection, but in the highest redshift bin, there is a tendency towards negative

alignments. A comparison of the A_1 amplitudes as a function of redshift with previous literature shows that our results deviate from a power law, but align with general trends.

Finally, for the analysis by luminosity, three equipopulated regions in luminosity are defined. For red galaxies, we observe an increase in IA with luminosity. The lowest luminosity bin shows no alignment, while the intermediate and highest luminosity bins exhibit $A_1 \sim 2$ and $A_1 \sim 5$, respectively. However, the constraints have large standard deviations. For blue galaxies, no clear alignment is observed in the lowest and highest luminosity bins, but there is a hint of positive IA in the intermediate bin. Our results extend the A_1-luminosity relation to lower luminosities and are generally consistent with previous studies.

Intrinsic alignment forecasts with Euclid

We present the implementation of the TATT model in the likelihood pipeline of Euclid, CLOE. This implementation is useful because it is expected to enable us to extend the modelling of IA towards smaller scales. The implementation is validated with CosmoSIS, obtaining excellent agreement between both codes.

A study of the differences in the WL and WLxGCphot C_l between the NLA and TATT models is presented for a set of IA parameters. This analysis indicates that, at small scales, TATT allows for the inclusion of some IA signal not present in the NLA model. Besides, in the case of WL, the differences in the IA signals between both models are more pronounced at lower redshifts.

We also study the contribution of IA to the total WL and WLxGCphot C_l, for both auto and cross-correlations of tomographic bins. In the case of WL, since the intrinsic-intrinsic term is a local effect, it is only non-negligible for the auto-correlation of bins, specially for low redshift bins, where the IA signal is larger. As for the shear-intrinsic term, it has opposite sign with respect to the other terms and its signal is also specially important at lower redshift bins, for both auto and cross-correlated bins. In the case of WLxGCphot, only the auto-correlation of bins yields significant IA signal, which counteracts, at some level, the galaxy-shear term.

We include an analysis on the biases that the mismodelling of IA can cause on the cosmological parameter inference. For this, we generate two synthetic DVs, one with NLA and the other one with TATT, and fit them with both models. In the case where the NLA DV is fitted with TATT, there are no observed biases in the cosmological parameters, given that TATT is an extension of NLA and can be reduced to it. However, when fitting a TATT DV with NLA, some biases in the cosmological parameters are introduced, since NLA can not reproduce the signal generated by TATT.

Finally, in future work, we will extend these analyses by performing forecasts on how the different IA models may affect the cosmological parameter inference for a broader range of scenarios. This future project will focus on Euclid's first data release, where we aim to understand the role of IA.

References

[1] Ahumada, R., Prieto, C. A., Almeida, A., Anders, F., Anderson, S. F., et al. (2020). The 16th data release of the sloan digital sky surveys: First release from the apogee-2 southern survey and full release of eboss spectra. *The Astrophysical Journal Supplement Series*, 249:3.

[2] Aihara, H., AlSayyad, Y., Ando, M., Armstrong, R., Bosch, J., et al. (2019). Second data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publications of the Astronomical Society of Japan*, 71(6):114.

[3] Alarcon, A., Gaztanaga, E., Eriksen, M., Baugh, C. M., Cabayol, L., et al. (2021). The PAU Survey: an improved photo-z sample in the COSMOS field. *Monthly Notices of the Royal Astronomical Society*, 501(4):6103–6122.

[4] Andrae, R., Schulze-Hartung, T., and Melchior, P. (2010). Dos and don'ts of reduced chi-squared.

[5] Angulo, R. E., Zennaro, M., Contreras, S., Aricò, G., Pellejero-Ibañez, M., and Stücker, J. (2021). The BACCO simulation project: exploiting the full power of large-scale structure for cosmology. *Monthly Notices of the Royal Astronomical Society*, 507(4):5869–5881.

[6] Aricò, G., Angulo, R. E., Contreras, S., Ondaro-Mallea, L., Pellejero-Ibañez, M., and Zennaro, M. (2021). The BACCO simulation project: a baryonification emulator with neural networks. *Monthly Notices of the Royal Astronomical Society*, 506(3):4070–4082.

[7] Arnouts, S., Le Floc'h, E., Chevallard, J., Johnson, B. D., Ilbert, O., et al. (2013). Encoding of the infrared excess in the nuvrk color diagram for star-forming galaxies. *Astronomy and Astrophysics*, 558:A67.

[8] Asorey, J., Carrasco Kind, M., Sevilla-Noarbe, I., Brunner, R. J., and Thaler, J. (2016). Galaxy clustering with photometric surveys using PDF redshift information. *Monthly Notices of the Royal Astronomical Society*, 459(2):1293–1309.

[9] Astier, P. and Pain, R. (2012). Observational evidence of the accelerated expansion of the universe. *Comptes Rendus. Physique*, 13(6-7):521–538.

[10] Baldauf, T., Smith, R. E., Seljak, U. c. v., and Mandelbaum, R. (2010). Algorithm for the direct reconstruction of the dark matter correlation function from weak lensing and galaxy clustering. *Phys. Rev. D*, 81:063531.

[11] Baldry, I. K., Liske, J., Brown, M. J. I., Robotham, A. S. G., Driver, S. P., et al. (2017). Galaxy And Mass Assembly: the G02 field, Herschel–ATLAS target selection and data release 3. *Monthly Notices of the Royal Astronomical Society*, 474(3):3875–3888.

[12] Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. *Physics Reports*, 340(4):291–472.

[13] Benítez, N. (2000). Bayesian Photometric Redshift Estimation. *The Astrophysical Journal*, 536(2):571–583.

[14] Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., et al. (2013). Nine-year wilkinson microwave anisotropy probe (wmap) observations: Final maps and results. *The Astrophysical Journal Supplement Series*, 208(2):20.

[15] Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement Series*, 117:393–404.

[16] Binggeli, B. (1982). The shape and orientation of clusters of galaxies. *Astronomy and Astrophysics*, 107:338–349.

[17] Blas, D., Lesgourgues, J., and Tram, T. (2011). The cosmic linear anisotropy solving system (class). part ii: Approximation schemes. *Journal of Cosmology and Astroparticle Physics*, 2011(07):034.

[18] Blazek, J. A., MacCrann, N., Troxel, M. A., and Fang, X. (2019). Beyond linear galaxy alignments. *Phys. Rev. D*, 100:103506.

[19] Boggess, N. W., Mather, J. C., Weiss, R., Bennett, C. L., Cheng, E. S., et al. (1992). The COBE Mission: Its Design and Performance Two Years after Launch. *The Astrophysical Journal*, 397:420.

[20] Boquien, M., Burgarella, D., Roehlly, Y., Buat, V., Ciesla, L., Corre, D., Inoue, A. K., and Salas, H. (2019). Cigale: a python code investigating galaxy emission. *Astronomy and Astrophysics*, 622:A103.

[21] Bordoloi, R., Lilly, S. J., and Amara, A. (2010). Photo-z performance for precision cosmology. *Monthly Notices of the Royal Astronomical Society*, 406(2):881–895.

[22] Boulade, O., Charlot, X., Abbon, P., Aune, S., Borgeaud, P., et al. (2003). Megacam: the new Canada-France-Hawaii Telescope wide-field imaging camera. In Iye, M. and Moorwood, A. F. M., editors, *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*, volume 4841, pages 72 – 81. International Society for Optics and Photonics, SPIE.

[23] Bradshaw, E. J., Almaini, O., Hartley, W. G., Smith, K. T., Conselice, C. J., et al. (2013). High-velocity outflows from young star-forming galaxies in the UKIDSS Ultra-Deep Survey. *Monthly Notices of the Royal Astronomical Society*, 433(1):194–208.

[24] Brainerd, T. G. (2005). Anisotropic Distribution of SDSS Satellite Galaxies: Planar (Not Polar) Alignment. *The Astrophysical Journal Letters*, 628(2):L101–L104.

[25] Brammer, G. B., van Dokkum, P. G., Franx, M., Fumagalli, M., Patel, S., et al. (2012). 3d-hst: A wide-field grism spectroscopic survey with the hubble space telescope*. *The Astrophysical Journal Supplement Series*, 200(2):13.

[26] Brown, M. L., Taylor, A. N., Hambly, N. C., and Dye, S. (2002). Measurement of intrinsic alignments in galaxy ellipticities. *Monthly Notices of the Royal Astronomical Society*, 333(3):501–509.

[27] Bruzual, G. and Charlot, S. (2003a). Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344(4):1000–1028.

[28] Bruzual, G. and Charlot, S. (2003b). Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344(4):1000–1028.

[29] Buat, V., Mountrichas, G., Yang, G., Boquien, M., Roehlly, Y., et al. (2021). Polar dust obscuration in broad-line active galaxies from the XMM-XXL field. *Astronomy and Astrophysics*, 654:A93.

[30] Cabayol, L., Eriksen, M., Carretero, J., Casas, R., Castander, F. J., et al. (2023). The pau survey and euclid: Improving broadband photometric redshifts with multi-task learning. *Astronomy and Astrophysics*, 671:A153.

[31] Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., and Storchi-Bergmann, T. (2000). The Dust Content and Opacity of Actively Star-forming Galaxies. *The Astrophysical Journal*, 533(2):682–695.

[32] Capaccioli, M. and Schipani, P. (2011). The VLT Survey Telescope Opens to the Sky: History of a Commissioning. *The Messenger*, 146:2–7.

[33] Carrasco Kind, M. and Brunner, R. J. (2014). SOMz: photometric redshift PDFs with self-organizing maps and random atlas. *Monthly Notices of the Royal Astronomical Society*, 438(4):3409–3421.

[34] Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., and Fosalba, P. (2014). An algorithm to build mock galaxy catalogues using MICE simulations. *Monthly Notices of the Royal Astronomical Society*, 447(1):646–670.

[35] Carretero, J., Tallada, P., Casals, J., Caubet, M., Castander, F., et al. (2017). CosmoHub and SciPIC: Massive cosmological data analysis, distribution and generation using a Big Data platform. In *Proceedings of the European Physical Society Conference on High Energy Physics. 5-12 July*, page 488.

[36] Castander, F. J., Serrano, S., Eriksen, M., Gaztañaga, E., Casas, R., et al. (2024). The PAU survey: photometric calibration of narrow band images. *Monthly Notices of the Royal Astronomical Society*, 531(4):5067–5083.

[37] Chabrier, G. (2003). Galactic Stellar and Substellar Initial Mass Function. *Publications of the Astronomical Society of the Pacific*, 115(809):763–795.

[38] Charlot, S. and Fall, S. M. (2000). A Simple Model for the Absorption of Starlight by Dust in Galaxies. *The Astrophysical Journal*, 539(2):718–731.

[39] Chisari, N. E., Alonso, D., Krause, E., Leonard, C. D., Bull, P., et al. (2019). Core Cosmology Library: Precision Cosmological Predictions for LSST. *The Astrophysical Journal Supplement Series*, 242(1):2.

[40] Chisari, N. E., Mandelbaum, R., Strauss, M. A., Huff, E. M., and Bahcall, N. A. (2014). Intrinsic alignments of group and cluster galaxies in photometric surveys. *Monthly Notices of the Royal Astronomical Society*, 445(1):726–748.

[41] Cimatti, A., Robberto, M., Baugh, C., Beckwith, S. V. W., Content, R., et al. (2009). SPACE: the spectroscopic all-sky cosmic explorer. *Experimental Astronomy*, 23(1):39–66.

[42] Cole, S. (2011). Maximum likelihood random galaxy catalogues and luminosity function estimation. *Monthly Notices of the Royal Astronomical Society*, 416(1):739–746.

[43] Coleman, G. D., Wu, C. C., and Weedman, D. W. (1980). Colors and magnitudes predicted for high redshift galaxies. *The Astrophysical Journal Supplement Series*, 43:393–416.

[44] Collaboration, E., Castander, F. J., Fosalba, P., Stadel, J., Potter, D., et al. (2024). Euclid. v. the flagship galaxy mock catalogue: a comprehensive simulation for the euclid mission.

[45] Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., et al. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts. *Monthly Notices of the Royal Astronomical Society*, 328(4):1039–1063.

[46] Conroy, C., Wechsler, R. H., and Kravtsov, A. V. (2006). Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time. *The Astrophysical Journal*, 647(1):201–214.

[47] Cooray, A. and Sheth, R. (2002). Halo models of large scale structure. *Physics Reports*, 372(1):1–129.

[48] Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., and Carretero, J. (2015). The MICE Grand Challenge lightcone simulation – II. Halo and galaxy catalogues. *Monthly Notices of the Royal Astronomical Society*, 453(2):1513–1530.

[49] Cuillandre, J.-C. J., Withington, K., Hudelot, P., Goranova, Y., McCracken, H., et al. (2012). Introduction to the CFHT Legacy Survey final release (CFHTLS T0007). In Peck, A. B., Seaman, R. L., and Comeron, F., editors, *Observatory Operations: Strategies, Processes, and Systems IV*, volume 8448, page 84480M. International Society for Optics and Photonics, SPIE.

[50] Dale, D. A., Helou, G., Magdis, G. E., Armus, L., Díaz-Santos, T., and Shi, Y. (2014). A Two-parameter Model for the Infrared/Submillimeter/Radio Spectral Energy Distributions of Galaxies and Active Galactic Nuclei. *The Astrophysical Journal*, 784(1):83.

[51] d'Amico, G., Gleyzes, J., Kokron, N., Markovic, K., Senatore, L., et al. (2020). The cosmological analysis of the sdss/boss data from the effective field theory of large-scale structure. *Journal of Cosmology and Astroparticle Physics*, 2020(05):005.

[52] Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., Aleksić, J., Allam, S., et al. (2016). The Dark Energy Survey: more than dark energy - an overview. *Monthly Notices of the Royal Astronomical Society*, 460(2):1270–1299.

[53] Davidzon, I., Cucciati, O., Bolzonella, M., De Lucia, G., Zamorani, G., et al. (2016). The vimos public extragalactic redshift survey (vipers) - environmental effects shaping the galaxy stellar mass function. *Astronomy and Astrophysics*, 586:A23.

[54] Davies, L. J. M., Driver, S. P., Robotham, A. S. G., Baldry, I. K., Lange, R., et al. (2014). Galaxy And Mass Assembly (GAMA): curation and reanalysis of 16.6k redshifts in the G10/COSMOS region. *Monthly Notices of the Royal Astronomical Society*, 447(1):1014–1027.

[55] Davis, M., Faber, S. M., Newman, J., Phillips, A. C., Ellis, R. S., et al. (2003). Science objectives and early results of the DEEP2 Redshift Survey. In Guhathakurta, P., editor, *Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*, volume 4834, pages 161 – 172. International Society for Optics and Photonics, SPIE.

[56] Davis, M., Guhathakurta, P., Konidaris, N. P., Newman, J. A., Ashby, M. L. N., et al. (2007). The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets. *The Astrophysical Journal Letters*, 660(1):L1–L6.

[57] Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, 147(2):278–292.

[58] de Jong, J. T., Kleijn, G. A. V., Kuijken, K. H., and Valentijn, E. A. (2013). The kilo-degree survey. *Experimental Astronomy*, 35:25–44.

[59] DESI Collaboration, Adame, A. G., Aguilar, J., Ahlen, S., Alam, S., et al. (2024). The Early Data Release of the Dark Energy Spectroscopic Instrument. *The Astronomical Journal*, 168(2):58.

[60] Dressler, A., Bigelow, B., Hare, T., Sutin, B., Thompson, I., et al. (2011). IMACS: The Inamori-Magellan Areal Camera and Spectrograph on Magellan-Baade. *Publications of the Astronomical Society of the Pacific*, 123(901):288.

[61] Driver, S. P., Allen, P. D., Graham, A. W., Cameron, E., Liske, J., et al. (2006). The Millennium Galaxy Catalogue: morphological classification and bimodality in the colour–concentration plane. *Monthly Notices of the Royal Astronomical Society*, 368(1):414–434.

[62] Driver, S. P., Hill, D. T., Kelvin, L. S., Robotham, A. S. G., Liske, J., et al. (2011). Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release. *Monthly Notices of the Royal Astronomical Society*, 413(2):971–995.

[63] Dyson, F. W., Eddington, A. S., and Davidson, C. (1920). A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London Series A*, 220:291–333.

[64] Edge, A., Sutherland, W., Kuijken, K., Driver, S., McMahon, R., et al. (2013). The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift. *The Messenger*, 154:32–34.

[65] Erben, T., Hildebrandt, H., Miller, L., van Waerbeke, L., Heymans, C., et al. (2013). CFHTLenS: the Canada–France–Hawaii Telescope Lensing Survey – imaging data and catalogue products. *Monthly Notices of the Royal Astronomical Society*, 433(3):2545–2563.

[66] Eriksen, M., Alarcon, A., Cabayol, L., Carretero, J., Casas, R., et al. (2020). The PAU Survey: Photometric redshifts using transfer learning from simulations. *Monthly Notices of the Royal Astronomical Society*, 497(4):4565–4579.

[67] Eriksen, M., Alarcon, A., Gaztanaga, E., Amara, A., Cabayol, L., et al. (2019). The PAU Survey: early demonstration of photometric redshift performance in the COSMOS field. *Monthly Notices of the Royal Astronomical Society*, 484(3):4200–4215.

[68] Euclid Collaboration, Ilbert, O., de la Torre, S., Martinet, N., Wright, A. H., et al. (2021). Euclid preparation. XI. Mean redshift determination from galaxy redshift probabilities for cosmic shear tomography. *Astronomy and Astrophysics*, 647:A117.

[69] Euclid Collaboration, Mellier, Y., Abdurro'uf, Acevedo Barroso, J. A., Achúcarro, A., et al. (2024). Euclid. I. Overview of the Euclid mission. *arXiv e-prints*, page arXiv:2405.13491.

[70] Euclid Collaboration, Knabenhans, M., Stadel, J., Potter, D., Dakin, J., et al. (2021). Euclid preparation: IX. EuclidEmulator2 – power spectrum emulation with massive neutrinos and self-consistent dark energy perturbations. *Monthly Notices of the Royal Astronomical Society*, 505(2):2840–2869.

[71] Faltenbacher, A., Jing, Y. P., Li, C., Mao, S., Mo, H. J., Pasquali, A., and van den Bosch, F. C. (2008). Spatial and Kinematic Alignments between Central and Satellite Halos. *The Astrophysical Journal*, 675(1):146–155.

[72] Fang, X., Blazek, J. A., McEwen, J. E., and Hirata, C. M. (2017). Fast-pt ii: an algorithm to calculate convolution integrals of general tensor quantities in cosmological perturbation theory. *Journal of Cosmology and Astroparticle Physics*, 2017(02):030.

[73] Farrow, D. J., Cole, S., Norberg, P., Metcalfe, N., Baldry, I., et al. (2015). Galaxy and mass assembly (GAMA): projected galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 454(2):2120–2145.

[74] Fioc, M. and Rocca-Volmerange, B. (1996). Spectrophotometric Evolution from the UV to the NIR. In Leitherer, C., Fritze-von-Alvensleben, U., and Huchra, J., editors, *From Stars to Galaxies: the Impact of Stellar Physics on Galaxy Evolution*, volume 98 of *Astronomical Society of the Pacific Conference Series*, page 67.

[75] Foreman-Mackey, D., Conley, A., Meierjurgen Farr, W., Hogg, D. W., Lang, D., et al. (2013). emcee: The MCMC Hammer. Astrophysics Source Code Library, record ascl:1303.002.

[76] Fortuna, Maria Cristina, Hoekstra, Henk, Johnston, Harry, Vakili, Mohammadjavad, Kannawadi, Arun, et al. (2021). Kids-1000: Constraints on the intrinsic alignment of luminous red galaxies. *Astronomy and Astrophysics*, 654:A76.

[77] Fosalba, P., Crocce, M., Gaztañaga, E., and Castander, F. J. (2015). The MICE grand challenge lightcone simulation – I. Dark matter clustering. *Monthly Notices of the Royal Astronomical Society*, 448(4):2987–3000.

[78] Fosalba, P., Gaztañaga, E., Castander, F. J., and Crocce, M. (2014). The MICE Grand Challenge light-cone simulation – III. Galaxy lensing mocks from all-sky lensing maps. *Monthly Notices of the Royal Astronomical Society*, 447(2):1319–1332.

[79] Frieman, J. A., Turner, M. S., and Huterer, D. (2008). Dark energy and the accelerating universe. *Annual Review of Astronomy and Astrophysics*, 46(Volume 46, 2008):385–432.

[80] Furusawa, H., Kosugi, G., Akiyama, M., Takata, T., Sekiguchi, K., et al. (2008). The subaru/xmm-newton deep survey (sxds). ii. optical imaging and photometric catalogs*. *The Astrophysical Journal Supplement Series*, 176(1):1.

[81] Fèvre, O. L., Cassata, P., Cucciati, O., Garilli, B., Ilbert, O., et al. (2013). The vimos vlt deep survey final data release: a spectroscopic sample of 35 016 galaxies and agn out to z ∼6.7 selected with $17.5 \leq i_{AB} \leq 24.75$. *Astronomy and Astrophysics*, 559.

[82] Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., Vallenari, A., et al. (2016). The gaia mission. *Astronomy and Astrophysics*, 595:A1.

[83] Garilli, B., McLure, R., Pentericci, L., Franzetti, P., Gargiulo, A., et al. (2021). The vandels eso public spectroscopic survey - final data release of 2087 spectra and spectroscopic measurements. *Astronomy and Astrophysics*, 647:A150.

[84] Gaztañaga, E. and Scoccimarro, R. (2005). The three-point function in large-scale structure: redshift distortions and galaxy bias. *Monthly Notices of the Royal Astronomical Society*, 361(3):824–836.

[85] Geha, M., Wechsler, R. H., Mao, Y.-Y., Tollerud, E. J., Weiner, B., et al. (2017). The SAGA Survey. I. Satellite Galaxy Populations around Eight Milky Way Analogs. *The Astrophysical Journal*, 847(1):4.

[86] Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457 – 472.

[87] Georgiou, C., Johnston, H., Hoekstra, H., Viola, M., Kuijken, K., et al. (2019). The dependence of intrinsic alignment of galaxies on wavelength using KiDS and GAMA. *Astronomy and Astrophysics*, 622:A90.

[88] Giacconi, R., Rosati, P., Tozzi, P., Nonino, M., Hasinger, G., Norman, C., et al. (2001). First results from the x-ray and optical survey of the chandra deep field south*. *The Astrophysical Journal*, 551(2):624.

[89] Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., Dickinson, M., Alexander, D. M., et al. (2004). The Great Observatories Origins Deep Survey: Initial Results from Optical and Near-Infrared Imaging. *The Astrophysical Journal Letters*, 600(2):L93–L98.

[90] Giri, S. K. and Schneider, A. (2021). Emulation of baryonic effects on the matter power spectrum and constraints from galaxy cluster data. *Journal of Cosmology and Astroparticle Physics*, 2021(12):046.

[91] Gonzalez, E. J., Rodriguez, F., Navarro-Gironés, D., Gaztañaga, E., Siudek, M., et al. (2023). The PAU survey: close galaxy pairs identification and analysis. *Monthly Notices of the Royal Astronomical Society*, 522(4):5655–5668.

[92] Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80.

[93] Griffith, R. L., Cooper, M. C., Newman, J. A., Moustakas, L. A., Stern, D., et al. (2012). The Advanced Camera for Surveys General Catalog: Structural Parameters for Approximately Half a Million Galaxies. *The Astrophysical Journal Supplement Series*, 200(1):9.

[94] Hambly, N., Irwin, M., and MacGillivray, H. (2001). The SuperCOSMOS Sky Survey — II. Image detection, parametrization, classification and photometry. *Monthly Notices of the Royal Astronomical Society*, 326(4):1295–1314.

[95] Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015). polychord: nested sampling for cosmology. *Monthly Notices of the Royal Astronomical Society*, 450:L61–L65.

[96] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

[97] Heymans, C., Van Waerbeke, L., Bacon, D., Berge, J., Bernstein, G., et al. (2006). The Shear Testing Programme – I. Weak lensing analysis of simulated ground-based observations. *Monthly Notices of the Royal Astronomical Society*, 368(3):1323–1339.

[98] Heymans, C., van Waerbeke, L., Miller, L., Erben, T., Hildebrandt, H., et al. (2012). CFHTLenS: the Canada–France–Hawaii Telescope Lensing Survey. *Monthly Notices of the Royal Astronomical Society*, 427(1):146–166.

[99] Hildebrandt, H., Erben, T., Kuijken, K., van Waerbeke, L., Heymans, C., et al. (2012). CFHTLenS: improving the quality of photometric redshifts with precision photometry. *Monthly Notices of the Royal Astronomical Society*, 421(3):2355–2367.

[100] Hirata, C. M., Mandelbaum, R., Ishak, M., Seljak, U., Nichol, R., et al. (2007). Intrinsic galaxy alignments from the 2SLAQ and SDSS surveys: luminosity and redshift scalings and implications for weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 381(3):1197–1218.

[101] Hirata, C. M. and Seljak, U. c. v. (2004). Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *Phys. Rev. D*, 70:063526.

[102] Hoekstra, H., Franx, M., and Kuijken, K. (2000). Hubble Space Telescope Weak-Lensing Study of the z=0.83 Cluster MS 1054-03. *The Astrophysical Journal*, 532(1):88–108.

[103] Hoekstra, H., Franx, M., Kuijken, K., and Squires, G. (1998). Weak Lensing Analysis of CL 1358+62 Using Hubble Space Telescope Observations. *The Astrophysical Journal*, 504(2):636–660.

[104] Hoekstra, H., Herbonnet, R., Muzzin, A., Babul, A., Mahdavi, A., et al. (2015). The Canadian Cluster Comparison Project: detailed study of systematics and updated weak lensing masses. *Monthly Notices of the Royal Astronomical Society*, 449(1):685–714.

[105] Hoffmann, K., Bel, J., Gaztañaga, E., Crocce, M., Fosalba, P., and Castander, F. J. (2014). Measuring the growth of matter fluctuations with third-order galaxy correlations. *Monthly Notices of the Royal Astronomical Society*, 447(2):1724–1745.

[106] Hoffmann, K., Secco, L. F., Blazek, J., Crocce, M., Tallada-Crespí, P., et al. (2022). Modeling intrinsic galaxy alignment in the mice simulation. *Phys. Rev. D*, 106:123510.

[107] Hopkins, A. M., Driver, S. P., Brough, S., Owers, M. S., Bauer, A. E., et al. (2013). Galaxy and mass assembly (gama): Spectroscopic analysis. *Monthly Notices of the Royal Astronomical Society*, 430:2047–2066.

[108] Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173.

[109] Hubble, E. P. (1926). Extragalactic nebulae. *The Astrophysical Journal*, 64:321–369.

[110] Hung, C.-L. and Ebeling, H. (2012). Galaxy alignments in very X-ray luminous clusters at $z > 0.5$. *Monthly Notices of the Royal Astronomical Society*, 421(4):3229–3237.

[111] Huterer, D. and Shafer, D. L. (2018). Dark energy two decades after: observables, probes, consistency tests. *Reports on Progress in Physics*, 81(1):016901.

[112] Inoue, A. K. (2011). Rest-frame ultraviolet-to-optical spectral characteristics of extremely metal-poor and metal-free galaxies. *Monthly Notices of the Royal Astronomical Society*, 415(3):2920–2931.

[113] Ivanov, M. M., Simonović, M., and Zaldarriaga, M. (2020). Cosmological parameters from the boss galaxy power spectrum. *Journal of Cosmology and Astroparticle Physics*, 2020(05):042.

[114] Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., et al. (2019). LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873(2):111.

[115] Jarvis, M., Bernstein, G., and Jain, B. (2004). The skewness of the aperture mass statistic. *Monthly Notices of the Royal Astronomical Society*, 352(1):338–352.

[116] Jing, Y. P., Mo, H. J., and Börner, G. (1998). Spatial Correlation Function and Pairwise Velocity Dispersion of Galaxies: Cold Dark Matter Models versus the Las Campanas Survey. *The Astrophysical Journal*, 494(1):1–12.

[117] Joachimi, B. and Bridle, S. L. (2010). Simultaneous measurement of cosmology and intrinsic alignments using joint cosmic shear and galaxy number density correlations. *Astronomy and Astrophysics*, 523:A1.

[118] Joachimi, B., Cacciato, M., Kitching, T. D., Leonard, A., Mandelbaum, R., et al. (2015). Galaxy Alignments: An Overview. *Space Science Reviews*, 193(1-4):1–65.

[119] Joachimi, B., Mandelbaum, R., Abdalla, F. B., and Bridle, S. L. (2011). Constraints on intrinsic alignment contamination of weak lensing surveys using the MegaZ-LRG sample. *Astronomy and Astrophysics*, 527:A26.

[120] Joachimi, B., Semboloni, E., Hilbert, S., Bett, P. E., Hartlap, J., et al. (2013). Intrinsic galaxy shapes and alignments – II. Modelling the intrinsic alignment contamination of weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 436(1):819–838.

[121] Joachimi, B., Mandelbaum, R., Abdalla, F. B., and Bridle, S. L. (2011). Constraints on intrinsic alignment contamination of weak lensing surveys using the megaz-lrg sample. *Astronomy and Astrophysics*, 527:A26.

[122] Johnston, H., Georgiou, C., Joachimi, B., Hoekstra, H., Chisari, N. E., et al. (2019). KiDS+GAMA: Intrinsic alignment model constraints for current and future weak lensing cosmology. *Astronomy and Astrophysics*, 624:A30.

[123] Johnston, H., Joachimi, B., Norberg, P., Hoekstra, H., Eriksen, M., et al. (2021). The PAU Survey: Intrinsic alignments and clustering of narrow-band photometric galaxies. *Astronomy and Astrophysics*, 646:A147.

[124] Johnston, Harry, Wright, Angus H., Joachimi, Benjamin, Bilicki, Maciej, Elisa Chisari, Nora, et al. (2021). Organised randoms: Learning and correcting for systematic galaxy clustering patterns in kids using self-organising maps. *Astronomy and Astrophysics*, 648:A98.

[125] Kaiser, N., Squires, G., and Broadhurst, T. (1995). A Method for Weak Lensing Observations. *The Astrophysical Journal*, 449:460.

[126] Kinney, A. L., Calzetti, D., Bohlin, R. C., McQuade, K., Storchi-Bergmann, T., and Schmitt, H. R. (1996). Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and Their Application to K-Corrections. *The Astrophysical Journal*, 467:38.

[127] Kitching, T. D. and Taylor, A. N. (2011). Path integral marginalization for cosmology: scale-dependent galaxy bias and intrinsic alignments. *Monthly Notices of the Royal Astronomical Society*, 410(3):1677–1686.

[128] Knebe, A., Draganova, N., Power, C., Yepes, G., Hoffman, Y., et al. (2008a). On the relation between the radial alignment of dark matter subhaloes and host mass in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 386(1):L52–L56.

[129] Knebe, A., Libeskind, N. I., Knollmann, S. R., Yepes, G., Gottlöber, S., and Hoffman, Y. (2010). The impact of baryonic physics on the shape and radial alignment of substructures in cosmological dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 405(2):1119–1128.

[130] Knebe, A., Yahagi, H., Kase, H., Lewis, G., and Gibson, B. K. (2008b). The radial alignment of dark matter subhaloes: from simulations to observations. *Monthly Notices of the Royal Astronomical Society*, 388(1):L34–L38.

[131] Krause, E., Fang, X., Pandey, S., Secco, L. F., Alves, O., et al. (2021). Dark Energy Survey Year 3 Results: Multi-Probe Modeling Strategy and Validation. *arXiv e-prints*, page arXiv:2105.13548.

[132] Kuhlen, M., Diemand, J., and Madau, P. (2007). The Shapes, Orientation, and Alignment of Galactic Dark Matter Subhalos. *The Astrophysical Journal*, 671(2):1135–1146.

[133] Kuijken, Heymans, C., Dvornik, A., Hildebrandt, H., de Jong, J. T. A., et al. (2019). The fourth data release of the kilo-degree survey: ugri imaging and nine-band optical-ir photometry over 1000 square degrees. *Astronomy and Astrophysics*, 625:A2.

[134] Kuijken, K. (2008). GaaP: PSF- and aperture-matched photometry using shapelets. *Astronomy and Astrophysics*, 482(3):1053–1067.

[135] Kuijken, K. (2011). OmegaCAM: ESO's Newest Imager. *The Messenger*, 146:8–11.

[136] Laigle, C., McCracken, H. J., Ilbert, O., Hsieh, B. C., Davidzon, I., et al. (2016). The cosmos2015 catalog: Exploring the 1<z<6 universe with half a million galaxies. *The Astrophysical Journal Supplement Series*, 224(2):24.

[137] Lamman, C., Tsaprazi, E., Shi, J., Šarčević, N. N., Pyne, S., Legnani, E., and Ferreira, T. (2024). The IA Guide: A Breakdown of Intrinsic Alignment Formalisms. *The Open Journal of Astrophysics*, 7:14.

[138] Landy, S. D. and Szalay, A. S. (1993). Bias and Variance of Angular Correlation Functions. *The Astrophysical Journal*, 412:64.

[139] Lange, J. U. (2023). nautilus: boosting Bayesian importance nested sampling with deep learning. *Monthly Notices of the Royal Astronomical Society*, 525(2):3181–3194.

[140] Laureijs, R., Amiaux, J., Arduini, S., Augueres, J. L., Brinchmann, J., et al. (2011). Euclid definition study report.

[141] Le Fèvre, O., Vettolani, G., Garilli, B., Tresse, L., Bottini, D., et al. (2005). The vimos vlt deep survey. first epoch vvds-deep survey: 11 564 spectra with $17.5 \leq$ iab ≤ 24, and the redshift distribution over $0 \leq z \leq 5$. *Astronomy and Astrophysics*, 439(3):845–862.

[142] Leclercq, F., Pisani, A., and Wandelt, B. D. (2014). Cosmology: from theory to data, from data to theory. *arXiv e-prints*, page arXiv:1403.1260.

[143] Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *The Astrophysical Journal*, 538(2):473–476.

[144] Lewis, I. J., Cannon, R. D., Taylor, K., Glazebrook, K., Bailey, J. A., et al. (2002). The Anglo-Australian Observatory 2dF facility. *Monthly Notices of the Royal Astronomical Society*, 333(2):279–298.

[145] Lilly, S. J., Le Brun, V., Maier, C., Mainieri, V., Mignoli, M., et al. (2009). The zCOSMOS 10k-Bright Spectroscopic Sample. *The Astrophysical Journal Supplement Series*, 184(2):218–229.

[146] Lilly, S. J., Le Fèvre, O., Renzini, A., Zamorani, G., Scodeggio, M., et al. (2007). zCOSMOS: A Large VLT/VIMOS Redshift Survey Covering $0 < z < 3$ in the COSMOS Field. *The Astrophysical Journal Supplement Series*, 172(1):70–85.

[147] Limber, D. N. (1953). The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *The Astrophysical Journal*, 117:134.

[148] Mandelbaum, R., Blake, C., Bridle, S., Abdalla, F. B., Brough, S., et al. (2010). The WiggleZ Dark Energy Survey: direct constraints on blue galaxy intrinsic alignments at intermediate redshifts. *Monthly Notices of the Royal Astronomical Society*, 410(2):844–859.

[149] Mandelbaum, R., Hirata, C. M., Ishak, M., Seljak, U., and Brinkmann, J. (2006). Detection of large-scale intrinsic ellipticity-density correlation from the Sloan Digital Sky Survey and implications for weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 367(2):611–626.

[150] Manzoni, G., Baugh, C. M., Norberg, P., Cabayol, L., van den Busch, J. L., et al. (2024). The PAU Survey: a new constraint on galaxy formation models using the observed colour redshift relation. *Monthly Notices of the Royal Astronomical Society*, 530(2):1394–1413.

[151] Maraston, C. (2005). Evolutionary population synthesis: models, analysis of the ingredients and application to high-z galaxies. *Monthly Notices of the Royal Astronomical Society*, 362(3):799–825.

[152] Masoura, V. A., Mountrichas, G., Georgantopoulos, I., Ruiz, A., Magdis, G., and Plionis, M. (2018). Disentangling the AGN and star formation connection using XMM-Newton. *Astronomy and Astrophysics*, 618:A31.

[153] McDonald, P. (2006). Clustering of dark matter tracers: Renormalizing the bias parameters. *Phys. Rev. D*, 74:103512.

[154] McEwen, J. E., Fang, X., Hirata, C. M., and Blazek, J. A. (2016). Fast-pt: a novel algorithm to calculate convolution integrals in cosmological perturbation theory. *Journal of Cosmology and Astroparticle Physics*, 2016(09):015.

[155] McLure, R. J., Pearce, H. J., Dunlop, J. S., Cirasuolo, M., Curtis-Lake, E., et al. (2013). The sizes, masses and specific star formation rates of massive galaxies at $1.3 < z < 1.5$: strong evidence in favour of evolution via minor mergers. *Monthly Notices of the Royal Astronomical Society*, 428(2):1088–1106.

[156] Mead, A. J., Brieden, S., Tröster, T., and Heymans, C. (2021). hmcode-2020: improved modelling of non-linear cosmological power spectra with baryonic feedback. *Monthly Notices of the Royal Astronomical Society*, 502(1):1401–1422.

[157] Mead, A. J., Heymans, C., Lombriser, L., Peacock, J. A., Steele, O. I., and Winther, H. A. (2016). Accurate halo-model matter power spectra with dark energy, massive neutrinos and modified gravitational forces. *Monthly Notices of the Royal Astronomical Society*, 459(2):1468–1488.

[158] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

[159] Miller, L., Heymans, C., Kitching, T. D., van Waerbeke, L., Erben, T., et al. (2013). Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey. *Monthly Notices of the Royal Astronomical Society*, 429(4):2858–2880.

[160] Moretti, C., Tsedrik, M., Carrilho, P., and Pourtsidou, A. (2023). Modified gravity and massive neutrinos: constraints from the full shape analysis of boss galaxies and forecasts for stage iv surveys. *Journal of Cosmology and Astroparticle Physics*, 2023(12):025.

[161] Myers, A. D., White, M., and Ball, N. M. (2009). Incorporating photometric redshift probability density information into real-space clustering measurements. *Monthly Notices of the Royal Astronomical Society*, 399(4):2279–2287.

[162] Navarro-Gironés, D., Gaztañaga, E., Crocce, M., Wittje, A., Hildebrandt, H., et al. (2023). The pau survey: Photometric redshift estimation in deep wide fields.

[163] Newman, J. A., Cooper, M. C., Davis, M., Faber, S. M., Coil, A. L., et al. (2013). The deep2 galaxy redshift survey: Design, observations, data reduction, and redshifts. *The Astrophysical Journal Supplement Series*, 208.

[164] Okumura, T., Jing, Y. P., and Li, C. (2009). Intrinsic Ellipticity Correlation of SDSS Luminous Red Galaxies and Misalignment with Their Host Dark Matter Halos. *The Astrophysical Journal*, 694(1):214–221.

[165] Padilla, C., Castander, F. J., Alarcón, A., Aleksic, J., Ballester, O., et al. (2019). The Physics of the Accelerating Universe Camera. *The Astronomical Journal*, 157(6):246.

[166] Pandey, S., Krause, E., Jain, B., MacCrann, N., Blazek, J., et al. (2020). Perturbation theory for modeling galaxy bias: Validation with simulations of the dark energy survey. *Phys. Rev. D*, 102:123522.

[167] Peacock, J. A. and Dodds, S. J. (1994). Reconstructing the linear power spectrum of cosmological mass fluctuations. *Monthly Notices of the Royal Astronomical Society*, 267(4):1020–1034.

[168] Pereira, M. J., Bryan, G. L., and Gill, S. P. D. (2008). Radial Alignment in Simulated Clusters. *The Astrophysical Journal*, 672(2):825–833.

[169] Piras, D., Joachimi, B., Schäfer, B. M., Bonamigo, M., Hilbert, S., and van Uitert, E. (2017). The mass dependence of dark matter halo alignments with large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 474(1):1165–1175.

[170] Planck Collaboration, Abergel, A., Ade, P. A. R., Aghanim, N., Alves, M. I. R., et al. (2014). Planck 2013 results. xi. all-sky model of thermal dust emission. *Astronomy and Astrophysics*, 571:A11.

[171] Planck Collaboration, Adam, R., Ade, P. A. R., Aghanim, N., Akrami, Y., et al. (2016). Planck 2015 results. I. Overview of products and scientific results. *Astronomy and Astrophysics*, 594:A1.

[172] Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., et al. (2020). Planck 2018 results. VI. Cosmological parameters. *Astronomy and Astrophysics*, 641:A6.

[173] Polletta, M., Tajer, M., Maraschi, L., Trinchieri, G., Lonsdale, C. J., et al. (2007). Spectral Energy Distributions of Hard X-Ray Selected Active Galactic Nuclei in the XMM-Newton Medium Deep Survey. *The Astrophysical Journal*, 663(1):81–102.

[174] Polsterer, K. L., D'Isanto, A., and Gieseke, F. (2016). Uncertain photometric redshifts.

[175] Pontoppidan, K. M., Barrientes, J., Blome, C., Braun, H., Brown, M., et al. (2022). The JWST Early Release Observations. *The Astrophysical Journal Letters*, 936(1):L14.

[176] Potter, D., Stadel, J., and Teyssier, R. (2017). PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Comput. Astrophys.*, 4(2).

[177] Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *The Astrophysical Journal*, 187:425–438.

[178] Réfrégier, A., Boulade, O., Mellier, Y., Milliard, B., Pain, R., et al. (2006). DUNE: the Dark Universe Explorer. In Mather, J. C., MacEwen, H. A., and de Graauw, M. W. M., editors, *Space Telescopes and Instrumentation I: Optical, Infrared, and Millimeter*, volume 6265, page 62651Y. International Society for Optics and Photonics, SPIE.

[179] Renard, P., Siudek, M., Eriksen, M. B., Cabayol, L., Cai, Z., et al. (2022). The PAU survey: measurements of the 4000 Å spectral break with narrow-band photometry. *Monthly Notices of the Royal Astronomical Society*, 515(1):146–166.

[180] Saito, S., Baldauf, T., Vlah, Z., Seljak, U. c. v., Okumura, T., and McDonald, P. (2014). Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum. *Phys. Rev. D*, 90:123522.

[181] Salim, S., Charlot, S., Rich, R. M., Kauffmann, G., Heckman, T. M., et al. (2005). New constraints on the star formation histories and dust attenuation of galaxies in the local universe from galex. *The Astrophysical Journal*, 619(1):L39.

[182] Salvato, M., Ilbert, O., and Hoyle, B. (2019). The many flavours of photometric redshifts. *Nature Astronomy*, 3:212–222.

[183] Samuroff, S., Campos, A., Porredon, A., and Blazek, J. (2024). Joint constraints from cosmic shear, galaxy-galaxy lensing and galaxy clustering: internal tension as an indicator of intrinsic alignment modelling error. *The Open Journal of Astrophysics*, 7:40.

[184] Samuroff, S., Mandelbaum, R., Blazek, J., Campos, A., MacCrann, N., et al. (2023). The Dark Energy Survey Year 3 and eBOSS: constraining galaxy intrinsic alignments across luminosity and colour space. *Monthly Notices of the Royal Astronomical Society*, 524(2):2195–2223.

[185] Sastry, G. N. (1968). Clusters Associated with Supergiant Galaxies. *Publications of the Astronomical Society of the Pacific*, 80(474):252.

[186] Schneider, M., Bridle, S., and Kirk, D. (2010). A Halo Model for Intrinsic Alignments of Galaxy Ellipticities. In *American Astronomical Society Meeting Abstracts #215*, volume 215 of *American Astronomical Society Meeting Abstracts*, page 376.01.

[187] Schneider, M. D., Cole, S., Frenk, C. S., Kelvin, L., Mandelbaum, R., et al. (2013). Galaxy And Mass Assembly (GAMA): galaxy radial alignments in GAMA groups. *Monthly Notices of the Royal Astronomical Society*, 433(4):2727–2738.

[188] Scoccimarro, R., Sheth, R. K., Hui, L., and Jain, B. (2001). How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *The Astrophysical Journal*, 546(1):20–34.

[189] Scodeggio, M., Guzzo, L., Garilli, B., Granett, B. R., Bolzonella, M., et al. (2018). The vimos public extragalactic redshift survey (vipers): Full spectroscopic data and auxiliary information release (pdr-2). *Astronomy and Astrophysics*, 609.

[190] Scoville, N., Abraham, R. G., Aussel, H., Barnes, J. E., Benson, A., et al. (2007a). COSMOS: Hubble Space Telescope Observations. *The Astrophysical Journal Supplement Series*, 172(1):38–45.

[191] Scoville, N., Aussel, H., Brusa, M., Capak, P., Carollo, C. M., et al. (2007b). The Cosmic Evolution Survey (COSMOS): Overview. *The Astrophysical Journal Supplement Series*, 172(1):1–8.

[192] Seljak, U. (2000). Analytic model for galaxy and dark matter clustering. *Monthly Notices of the Royal Astronomical Society*, 318(1):203–213.

[193] Serrano, S., Gaztañaga, E., Castander, F. J., Eriksen, M., Casas, R., et al. (2023). The Physics of the Accelerating Universe Survey: narrow-band image photometry. *Monthly Notices of the Royal Astronomical Society*, 523(3):3287–3317.

[194] Sheth, R. K. and Tormen, G. (1999). Large-scale bias and the peak background split. *Monthly Notices of the Royal Astronomical Society*, 308(1):119–126.

[195] Sifón, C., Hoekstra, H., Cacciato, M., Viola, M., Köhlinger, F., et al. (2015). Constraints on the alignment of galaxies in galaxy clusters from ∼14 000 spectroscopic members. *Astronomy and Astrophysics*, 575:A48.

[196] Singh, S. and Mandelbaum, R. (2016). Intrinsic alignments of BOSS LOWZ galaxies - II. Impact of shape measurement methods. *Monthly Notices of the Royal Astronomical Society*, 457(3):2301–2317.

[197] Singh, S., Mandelbaum, R., and More, S. (2015). Intrinsic alignments of SDSS-III BOSS LOWZ sample galaxies. *Monthly Notices of the Royal Astronomical Society*, 450(2):2195–2216.

[198] Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833 – 859.

[199] Slipher, V. M. (1917). Nebulae. *Proceedings of the American Philosophical Society*, 56:403–409.

[200] Smit, Merijn and Kuijken, Konrad (2018). Chasing the peak: optimal statistics for weak shear analyses. *Astronomy and Astrophysics*, 609:A103.

[201] Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., et al. (2003). Stable clustering, the halo model and non-linear cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, 341(4):1311–1332.

[202] Soo, J. Y. H., Joachimi, B., Eriksen, M., Siudek, M., Alarcon, A., et al. (2021). The PAU Survey: narrow-band photometric redshifts using Gaussian processes. *Monthly Notices of the Royal Astronomical Society*, 503(3):4118–4135.

[203] Springel, V. (2005). The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134.

[204] Stanford, S. A., Masters, D., Darvish, B., Stern, D., Cohen, J. G., et al. (2021). Euclid Preparation. XIV. The Complete Calibration of the Color-Redshift Relation (C3R2) Survey: Data Release 3. *The Astrophysical Journal Supplement Series*, 256(1):9.

[205] Stothert, L., Norberg, P., Baugh, C. M., Alarcon, A., Amara, A., et al. (2018). The PAU Survey: spectral features and galaxy clustering using simulated narrow-band photometry. *Monthly Notices of the Royal Astronomical Society*, 481(3):4221–4235.

[206] Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., and Oguri, M. (2012). Revising the Halofit Model for the Nonlinear Matter Power Spectrum. *The Astrophysical Journal*, 761(2):152.

[207] Tallada, P., Carretero, J., Casals, J., Acosta-Silva, C., Serrano, S., et al. (2020). Cosmohub: Interactive exploration and distribution of astronomical data on hadoop. *Astronomy and Computing*, 32:100391.

[208] Tenneti, A., Singh, S., Mandelbaum, R., di Matteo, T., Feng, Y., and Khandai, N. (2015). Intrinsic alignments of galaxies in the MassiveBlack-II simulation: analysis of two-point statistics. *Monthly Notices of the Royal Astronomical Society*, 448(4):3522–3544.

[209] Tonello, N., Tallada, P., Serrano, S., Carretero, J., Eriksen, M., et al. (2019). The pau survey: Operation and orchestration of multi-band survey data. *Astronomy and Computing*, 27:171–188.

[210] Torrado, J. and Lewis, A. (2019). Cobaya: Bayesian analysis in cosmology. Astrophysics Source Code Library, record ascl:1910.019.

[211] Tortorelli, L., Siudek, M., Moser, B., Kacprzak, T., Berner, P., et al. (2021). The PAU survey: measurement of narrow-band galaxy properties with approximate bayesian computation. *Journal of Cosmology and Astroparticle Physics*, 2021(12):013.

[212] Trujillo-Gomez, S., Klypin, A., Primack, J., and Romanowsky, A. J. (2011). Galaxies in ΛCDM with Halo Abundance Matching: Luminosity-Velocity Relation, Baryonic Mass-Velocity Relation, Velocity Function, and Clustering. *The Astrophysical Journal*, 742(1):16.

[213] Vale, A. and Ostriker, J. P. (2004). Linking halo mass to galaxy luminosity. *Monthly Notices of the Royal Astronomical Society*, 353(1):189–200.

[214] van Uitert, E. and Joachimi, B. (2017). Intrinsic alignment of redMaPPer clusters: cluster shape–matter density correlation. *Monthly Notices of the Royal Astronomical Society*, 468(4):4502–4512.

[215] Wadekar, D. and Scoccimarro, R. (2020). Galaxy power spectrum multipoles covariance in perturbation theory. *Phys. Rev. D*, 102:123517.

[216] Wang, Y. O., Lin, W. P., Kang, X., Dutton, A., Yu, Y., and Macciò, A. V. (2014). Satellite Alignment. I. Distribution of Substructures and their Dependence on Assembly History from N-body Simulations. *The Astrophysical Journal*, 786(1):8.

[217] Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., Hirata, C., Riess, A. G., and Rozo, E. (2013). Observational probes of cosmic acceleration. *Physics Reports*, 530(2):87–255.

[218] Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., and Labbé, I. (2009a). Detection of Quiescent Galaxies in a Bicolor Sequence from Z = 0-2. *The Astrophysical Journal*, 691(2):1879–1895.

[219] Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., and Labbé, I. (2009b). Detection of Quiescent Galaxies in a Bicolor Sequence from Z = 0-2. *The Astrophysical Journal*, 691(2):1879–1895.

[220] York, D. G., Adelman, J., Anderson, John E., J., Anderson, S. F., Annis, J., et al. (2000). The Sloan Digital Sky Survey: Technical Summary. *The Astronomical Journal*, 120(3):1579–1587.

[221] Zhou, C., Tong, A., Troxel, M. A., Blazek, J., Lin, C., et al. (2023). The intrinsic alignment of red galaxies in DES Y1 redMaPPer galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 526(1):323–336.

[222] Zuntz, J., Paterno, M., Jennings, E., Rudd, D., Manzotti, A., et al. (2015). CosmoSIS: Modular cosmological parameter estimation. *Astronomy and Computing*, 12:45–59.

Appendix A

Photo-z performance in individual fields

In this Appendix, we analyse the performance of the photo-z studied in Section 3.4.2, but stressing the comparison between each of the wide fields.

Table A.1 shows the number of photo-z (after applying the mask and the star flags), the area, the number density, the BCNZ σ_{68} and the BCNZw σ_{68} for the PAUS wide fields studied in this analysis. The G09 number density is slightly smaller than that of the W1 and W3 fields. Nonetheless, the number densities are comparable, indicating that a similar population was selected in the three PAUS wide fields, as shown in Fig. 3.2. As for both σ_{68} estimates, the values for the W1 and W3 fields are almost the same, while in the G09 field this value increases up to $\sigma_{68} = 0.030$, showing a lower performance.

Fig. A.1 shows the photo-z as a function of spec-z, as in the case of Fig. 3.8, but separated into the 3 PAUS wide fields, so that we can see the effect of applying the weight on the photo-z in each of the fields. A horizontal stripe of artificial photometric redshifts is seen

Table A.1 Number of photometric redshifts, area, number density, BCNZ σ_{68} and BCNZw σ_{68}, after applying the mask and rejecting stars, for the W1, W3 and G09 fields observed by PAUS.

Field	W1	G09	W3
# objects	308403	364592	568052
Area [deg^2]	10.2	12.54	18.25
Number density [deg^{-2}]	30236	29074	31126
BCNZ σ_{68}	0.018	0.029	0.019
BCNZw σ_{68}	0.019	0.030	0.018

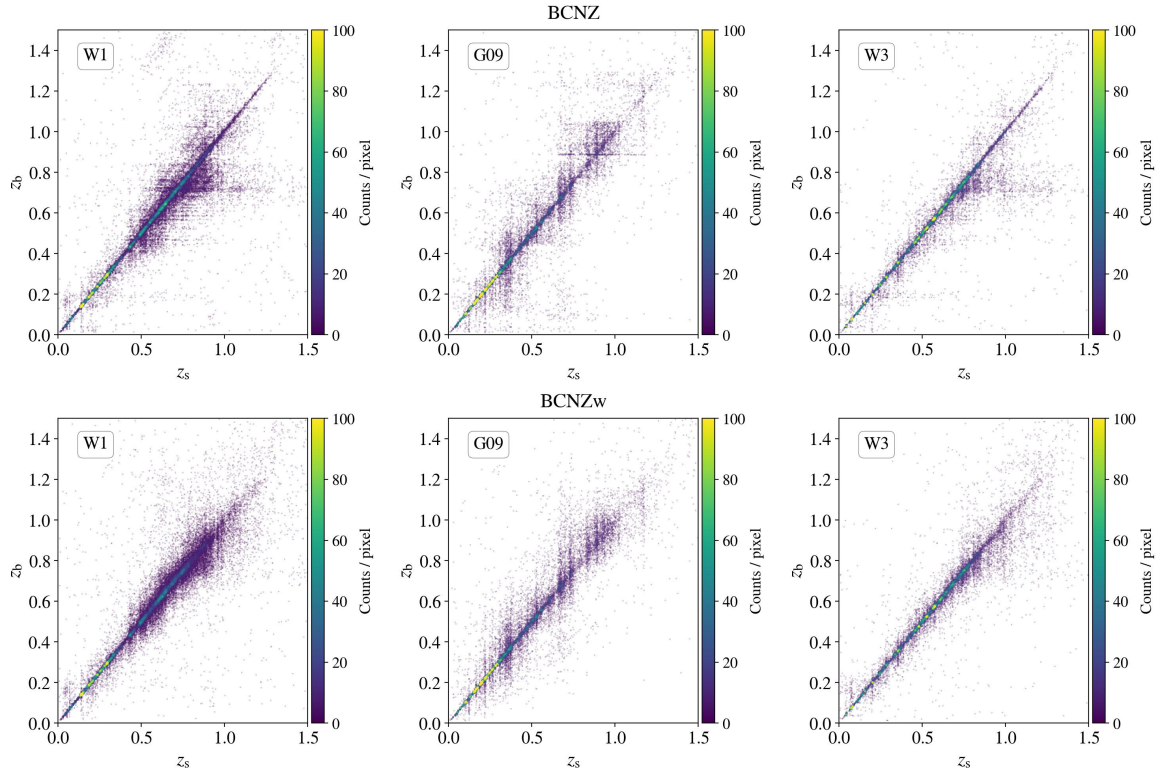

Fig. A.1 Photometric redshift vs spectroscopic redshift for the BCNZ photo-z (top) and for the BCNZw photo-z (bottom) for each of the 3 PAUS wide fields. The colour bar indicates the density of objects. The horizontal stripes at $z_b \approx 0.72$ and $z_b \approx 0.89$ are dissipated when weighting with the photo-z computed only with broad bands, which are obtained with another photometric redshift code.

at $z_b \sim$0.72 for the W1 and W3 fields and at $z_b \sim$0.89 for the G09 field, which is mainly corrected with the new weighted photo-z.

The weighted photometric redshift distributions are presented in Fig. A.2 for the W1, W3 and G09 fields. The W1 and W3 distributions are very similar, except for an increase in the W3 distribution from $z_b \sim 0.4-0.6$, not present in W1. In the case of the G09 field, there is an underdensity of objects at $z_b \sim 0.3-0.4$ and an overdensity at $z_b \sim 0.4-0.5$. However, it is not straightforward to assess if these are intrinsic differences coming from the fields themselves or if they are caused by the fact that we are comparing two photometric systems, with different star flags, different masks and different flux errors.

Fig. A.3 shows the performance of the W1, W3 and G09 fields as a function of i_{AB}, z_b and z_s for the weighted photo-z (solid lines) and the BCNZ photo-z (dashed lines), following the same structure as Fig. 3.13. We also include in this analysis the photo-z from the COSMOS field computed by Alarcon et al. [3] (solid black lines), so as to compare the photo-z with

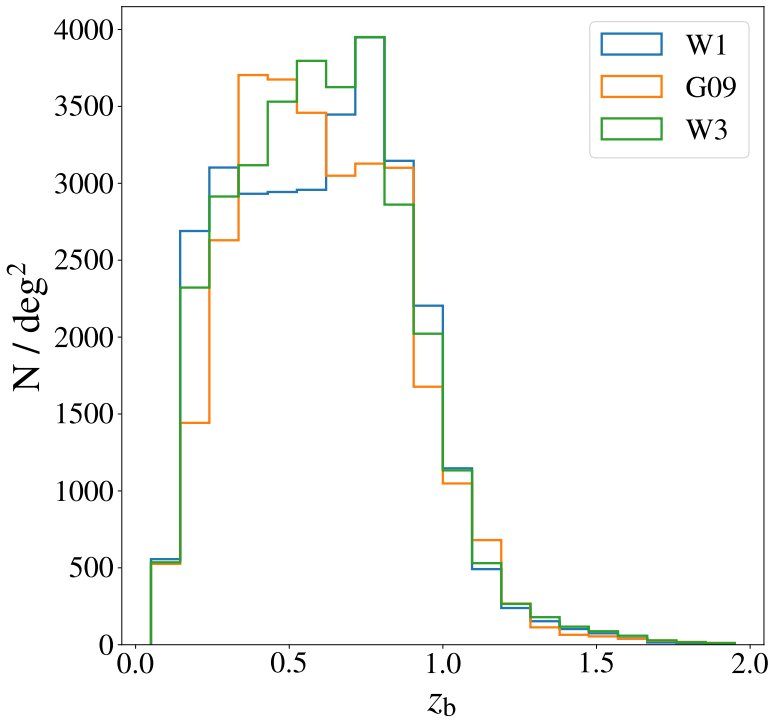

Fig. A.2 Weighted photometric redshift distributions for the W1, G09 and W3 fields, using the BCNZw photo-z. Under and overdensities are observed between the PAUS wide fields.

best accuracy that have been obtained by PAUS with the wide fields photo-z. Focusing on the PAUS wide fields, the σ_{68} values as a function of $i_{\rm AB}$ show that the performance of the G09 field is, in general, worse than the W1 and W3 fields, which are more comparable to one another. This is also observed when looking at the σ_{68} as a function of $z_{\rm b}$ and $z_{\rm s}$, except for the objects around $z_{\rm s} = 1$, where G09 performs slightly better. We note that the weighted photo-z, in general, present lower values of σ_{68}. For the BCNZw photo-z, the faintest objects ($i_{\rm AB} \sim 22.5 - 23$) have a $\sigma_{68} \sim 0.05 - 0.06$ for W1 and W3, while it arrives at $\sigma_{68} \sim 0.08$ for G09. When studying the outlier fractions, we note that G09 presents higher values than the W1 and W3 cases, which are very similar. G09 reaches an outlier fraction of $\sim$0.25 for the faintest bin, while W1 and W3 present a value of $\sim$0.15. As a function of $z_{\rm b}$ and $z_{\rm s}$, the outlier fraction is higher in the G09 field at intermediate redshifts ($z \sim 0.25 - 0.75$). As for the bias, we see that the behaviour is different for the G09 case as a function of $i_{\rm AB}$, showing positive values for the BCNZ photo-z, unlike the W1 and W3 fields, which show negative bias. G09 also presents higher bias as a function of $z_{\rm b}$. Focusing now in the COSMOS photo-z presented in Alarcon et al. [3], the σ_{68} and the outlier fraction values are lower than in the PAUS wide fields, while the bias is comparable between the combined wide fields shown in Fig. 3.13 and these COSMOS photo-z. It is important to highlight that the average

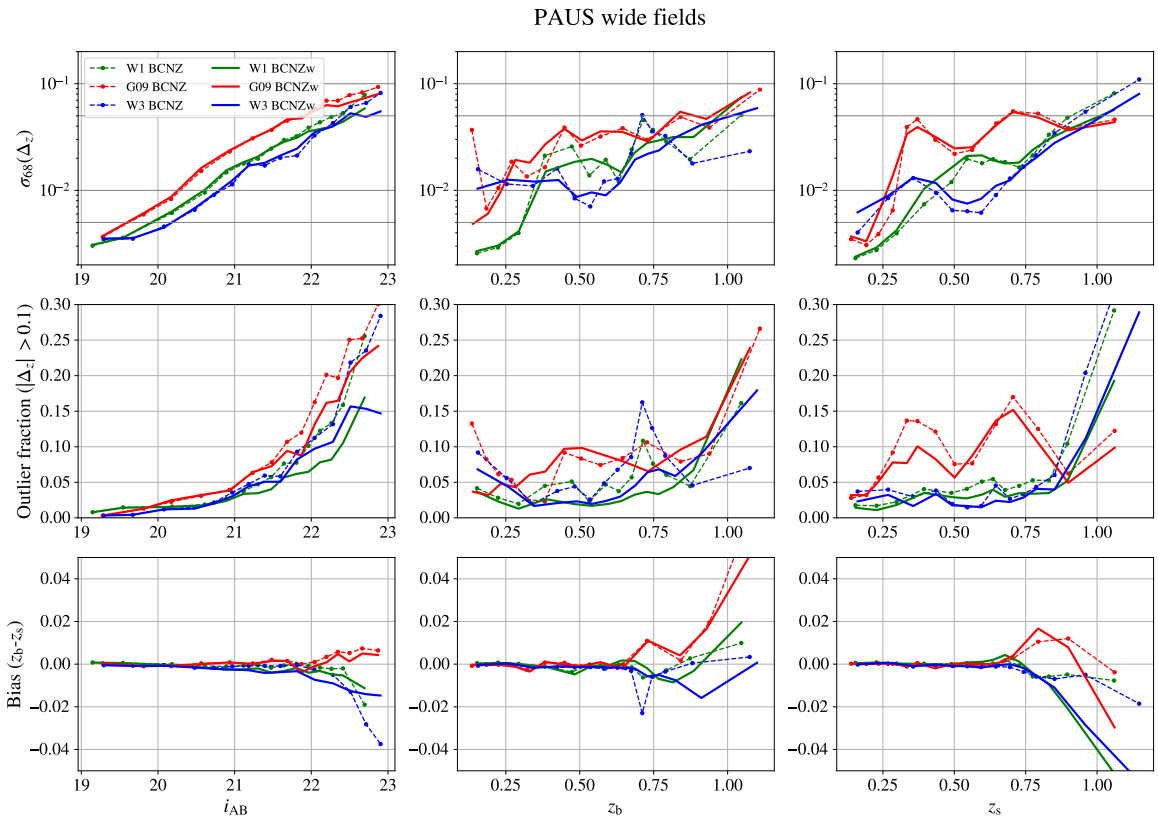

Fig. A.3 Performance of the W1, G09 and W3 fields (green, red and blue, respectively) for the BCNZw photo-z and for the COSMOS field photo-z in Alarcon et al. [3] (black). From top to bottom, the σ_{68}, the outlier fraction and the bias are shown as a function of i_{AB} magnitude, photometric and spectroscopic redshift. Each bin is defined to contain an equal number of objects.

number of PAUS observations in COSMOS is 5, while it is 3 in the PAUS wide fields. As a consequence, the SNR in COSMOS is higher than in the wide fields and the photo-z are more accurate. Additionally, for these COSMOS photo-z, the number of bands used was higher, with 26 broad, intermediate and narrow bands covering the UV, visible and near IR, besides the 40 PAUS NB. Finally, we would like to mention that the number of objects and the area of the COSMOS photo-z are much lower than the PAUS wide fields photo-z, with $\sim$40k versus $\sim$1.8 million objects and $\sim$1deg^2 versus $\sim$51deg^2, respectively. As a consequence, the PAUS wide fields photo-z have better statistics and are less affected by sample variance.

We now summarise the reasons why the G09 photo-z display a poorer performance than the other fields:

- Fig. A.4 shows the SNR, flux (ϕ) and flux error (ϕ_{err}) of the G09 field. Note that the flux errors of the BB decrease at a slower rate than in the W3 field case shown in

Fig. 3.6 and stop decreasing at $i_{\rm AB} \sim 22$. As a consequence, the SNR of both NB and BB decrease for almost all the $i_{\rm AB}$ range, with the SNR of the BB being still higher than the NB one. This lower BB SNR for the G09 field with respect to W1 and W3 is a key point that affects the photo-z estimation, given that the SNR drives the photo-z performance, as explained in Section 3.3.4. However, it is important to note that this difference in SNR between G09 and W1/W3 may be magnified, since we expect the CFHTLenS errors to be underestimated in comparison with the KiDS ones. The reason for this is that KiDS uses GAaP photometry [134], where the convolution of the image is lower than in CFTHLenS and is corrected for in the photometric error estimates. Nevertheless, we still expect lower SNR in KiDS measurements, since CFHTLenS is deeper than KiDS in most of the broad bands.

- Even though the KiDS photometric system has 9 broad bands instead of the 5 available in CFHTLenS, the VIKING near-infrared bands mostly help at $z > 1$, where they can better detect the Balmer and 4000Å break and where most of our objects are not located.

- The validation sample of the G09 field, which is obtained by merging the G09 and KiDZ-COSMOS spectroscopic redshifts, might not come from exactly the same processing, since the G09 objects are from KiDS DR4 and KiDZ-COSMOS come from KiDS DR5.

- The difference in the definition of the i-band between CFHTLenS and KiDS may also affect the comparison between both cases. On the one hand, even though we redefine $i_{\rm AB,\,CFHTLenS} = i_{\rm AB,\,KiDS} - 0.1$, we still have differences between both magnitudes, since this is only a first order correction, so that comparing the performance of KiDS and CFHTLenS as a function of $i_{\rm AB}$ is not straightforward. On the other hand, as we commented, the cut at $i_{\rm AB,\,KiDS} = 23.1$ allows to obtain very similar number counts compared with W1 and W3. However, there are still some differences that can be observed for bright objects in Fig. 3.2 that may affect the comparison.

- Finally, the fact that different outliers are found in the W1/W3 and the G09 fields, which correspond to artificial photo-z at $z_{\rm b} \sim 0.72$ and $z_{\rm b} \sim 0.89$, respectively, may affect the performance, even though we try to correct those outliers with the weighted photo-z.

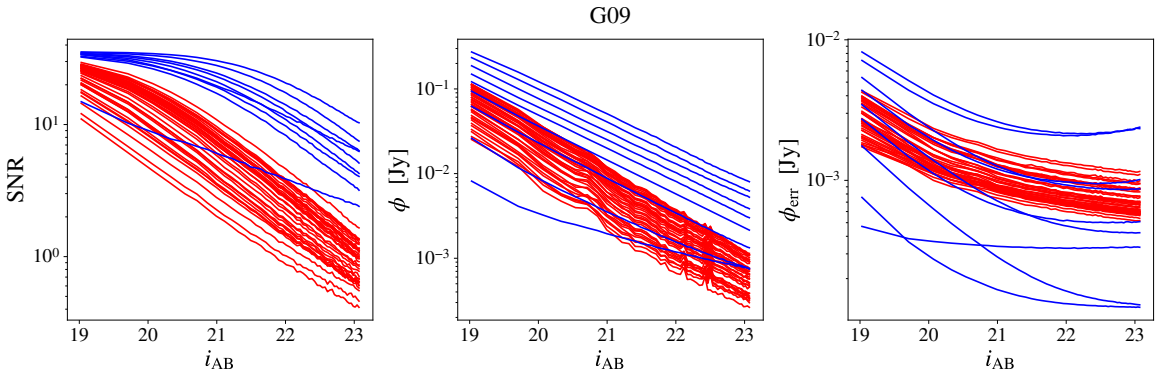

Fig. A.4 SNR, flux (ϕ) and flux error (ϕ_{err}) (from left to right) of the G09 field for the broad bands (blue lines) and the narrow bands (red lines). The SNR of the broad bands starts to decrease at $i_{\mathrm{AB}} \sim 21$, since the flux errors of the G09 broad bands stop decreasing around the same magnitude. In the case of the narrow bands, the SNR decreases as the objects become fainter, which is caused by the constant NB flux errors at faint magnitudes.

Appendix B

Scale cut analysis

In this appendix, we study the dependence on the scale cuts when fitting the $w_{gg} + w_{gp}$ data vector with the non-linear galaxy bias, for galaxy clustering, and the NLA model, for IA. As indicated in Section 4.4.6, we set $r_{p,\min} = 2.0\,h^{-1}$Mpc for both w_{gg} and w_{gp} in Chapter 4. However, in this appendix, we explore the scale cuts further by setting different values for $r_{p,\min}$ in the case of w_{gp}, while fixing $r_{p,\min} = 2.0\,h^{-1}$Mpc for w_{gg}. The justification of this approach is that perturbation theory, on which the non-linear galaxy bias is developed, breaks down below $r_{p,\min} = 2.0\,h^{-1}$Mpc. Nevertheless, in the case of the NLA model, previous studies set $r_{p,\min} = 6.0\,h^{-1}$Mpc based on observations, while Romain et al. (in prep.) shows that this mainly depends on the fact that linear galaxy bias is used. Given the high number density of PAUS, which allows measuring correlations down to smaller scales than other spectroscopic surveys, we want to exploit the range of validity of the NLA model by exploring different scale cuts. The scale cuts we test for w_{gp} range from the common $r_{p,\min} = 6.0\,h^{-1}$Mpc down to the uncommon $r_{p,\min} = 0.0\,h^{-1}$Mpc, which basically means setting no scale cuts. We analyse the variations in $\chi^2_{\nu,\mathrm{fit,SVD}}$ (the reduced χ^2 based in eq. 4.46), together with the change in the A_1 constraints, as a function of $r_{p,\min}$.

Fig. B.1 shows the dependence on the scale cuts when splitting the PAUS samples in red and blue galaxies. Both red and blue galaxies are evaluated at the same $r_{p,\min}$ values, but the points for blue galaxies in the figure are plotted at slightly larger values, to avoid overlapping. We highlight that we focus on the evolution of the $\chi^2_{\nu,\mathrm{fit,SVD}}$ value, rather than on its absolute value, given the remark made at the end of Section 4.6.3 about the values below and above 1 in the $\chi^2_{\nu,\mathrm{fit,SVD}}$. In the case of red objects, the $\chi^2_{\nu,\mathrm{fit,SVD}}$ starts increasing at $r_{p,\min} < 1.0\,h^{-1}$Mpc, from a stable value of $\chi^2_{\nu,\mathrm{fit,SVD}} \sim 0.4$ up to ~ 1.4. This impacts the value of A_1, which is similar for $r_{p,\min} = [1.0 - 2.0]\,h^{-1}$Mpc and varies more significantly for $r_{p,\min} < 1.0\,h^{-1}$Mpc. This value is also different for $r_{p,\min} = 6.0\,h^{-1}$Mpc, where also the

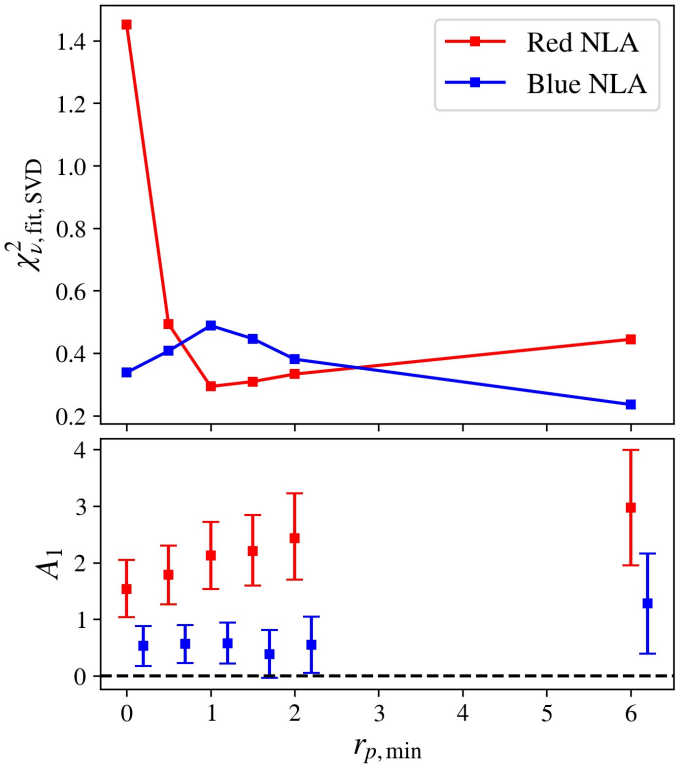

Fig. B.1 $\chi^2_{\nu,\mathrm{fit,SVD}}$ (top) and A_1 (bottom) as a function of the minimum separation considered in the modelling for red and blue galaxies.

error bars are larger than in the other cases. The reason for this may be the small amount of available data points to fit. In the case of blue objects, the $\chi^2_{\nu,\mathrm{fit,SVD}}$ is quite stable, but this is most likely due to the lack of signal for these types of galaxies.

Fig. B.2 shows the same configuration as in Fig. B.1 but for the redshift bins defined in Chapter 4. From left to right, the redshift bin increases. For the lower redshift bin, the $\chi^2_{\nu,\mathrm{fit,SVD}}$ is relatively stable for both galaxy colours, which could be explained by the fact that the signal is small at this low redshift. For the intermediate redshift bin, in the case of red galaxies, the NLA model is able to fit the data down to $r_{p,\mathrm{min}} = 0.0\,h^{-1}\mathrm{Mpc}$. This can be understood by observing Fig. 4.23, where the IA amplitude is consistent with 0 at most small scales. Nevertheless, if we observe the highest redshift bin for red galaxies, the situation is the opposite, with the $\chi^2_{\nu,\mathrm{fit,SVD}}$ value constantly increasing when moving to smaller scales. Again, this can be explained from Fig. 4.23, given that in this case the IA signal is not 0 at small scales and the NLA model is not able to fit this data. For the case of the intermediate and redshift bins in blue galaxies, the $\chi^2_{\nu,\mathrm{fit,SVD}}$ is rather stable, given that the IA signal is consistent with 0 at almost all scales. As an exception, the inclusion of the smallest scale

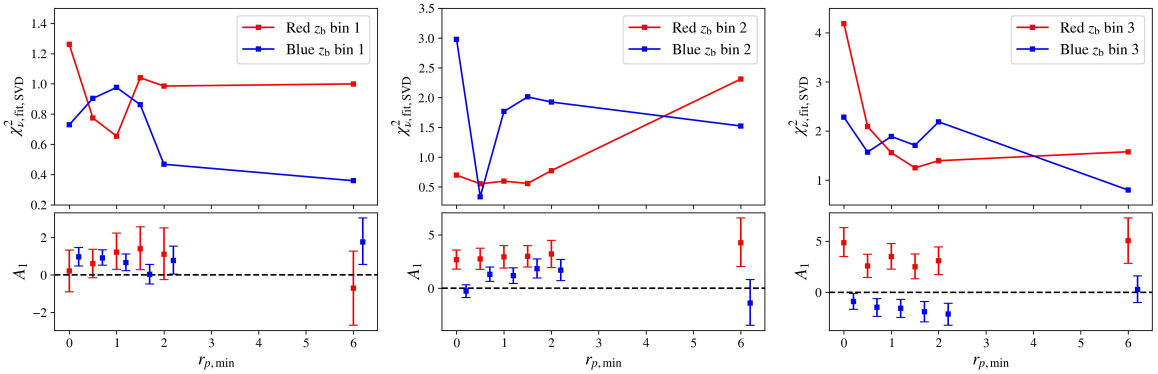

Fig. B.2 $\chi^2_{\nu,\mathrm{fit,SVD}}$ (top) and A_1 (bottom) as a function of the minimum separation considered in the modelling for red and blue galaxies as a function of redshift.

in the intermediate redshift bin, where the signal fluctuates away from 0 and the $\chi^2_{\nu,\mathrm{fit,SVD}}$ increases. In general, we note that, unless the $\chi^2_{\nu,\mathrm{fit,SVD}}$ increases drastically, the A_1 amplitude is stable even when including smaller scales. Nevertheless, this is not unexpected, since this happens when the IA signal at small scales is null.

Fig. B.3 shows the evolution of $\chi^2_{\nu,\mathrm{fit,SVD}}$ and A_1 as a function of luminosity. For the lowest luminosity bin, the $\chi^2_{\nu,\mathrm{fit,SVD}}$ constantly decreases down to $r_{p,\mathrm{min}} = 0.0\,h^{-1}$Mpc for red objects, since there is no clear IA signal at small scales, as shown in Fig. 4.27. The case for blue galaxies contrasts from that of red galaxies, with an increasing $\chi^2_{\nu,\mathrm{fit,SVD}}$. Nevertheless, we note that the A_1 amplitudes in this low luminosity bin are quite consistent for all $r_{p,\mathrm{min}}$, except for $r_{p,\mathrm{min}} = 6.0\,h^{-1}$Mpc. The case for the intermediate luminosity bin is similar for both red and blue galaxies, with the $\chi^2_{\nu,\mathrm{fit,SVD}}$ increasing at small scales, where the IA amplitude fluctuates. Finally, for the highest luminosity bin, both red and blue galaxies are fitted down to $r_{p,\mathrm{min}} = 0.5\,h^{-1}$Mpc, where the case of red galaxies is of special interest since there is a consistent IA amplitude at almost all scales.

As a concluding remark, in general we observe that the NLA model is able to fit scales with $r_{p,\mathrm{min}} < 2.0\,h^{-1}$Mpc if the IA amplitude is null, which is a reasonable result. Nevertheless, there is an exception for the most luminosity bin in red galaxies, where it is possible to fit it down to $r_{p,\mathrm{min}} = 0.5\,h^{-1}$Mpc.

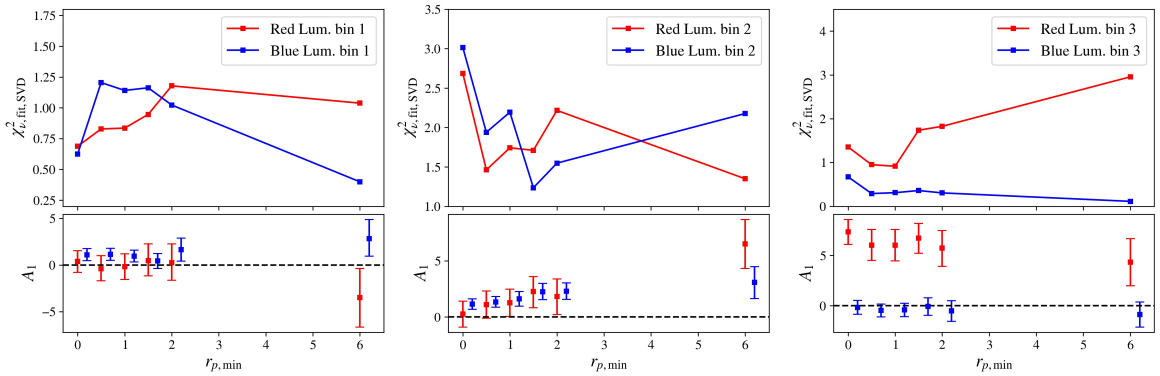

Fig. B.3 $\chi^2_{\nu,\mathrm{fit,SVD}}$ (top) and A_1 (bottom) as a function of the minimum separation considered in the modelling for red and blue galaxies as a function of luminosity.

Appendix C

Comparison of w_{gp} and w_{gx}

In this appendix, we analyse the signal of w_{gp} and w_{gx} for the cases that are deemed as *unclear*, in the sense that both w_{gp} and w_{gx} present high values of $\chi^2_{\nu,\mathrm{SNR}}$. As commented in Sec. 4.6, a high value of $\chi^2_{\nu,\mathrm{SNR}}$ may be caused by either correlated or uncorrelated points. In the case of correlated points in w_{gx}, this may indicate the presence of some systematic effect, while, in the case of uncorrelated points, this might be just noise. Based on Fig. 4.18, we focus on red galaxies, red galaxies for the z_b bin 2 and blue galaxies for the luminosity bin 2, since these cases show values of $\chi^2_{\nu,\mathrm{SNR}} \sim 2-3$ for both w_{gp} and w_{gx}.

The left-hand side plot in Fig. C.1 shows the w_{gp} (black) and w_{gx} (red) signal for red galaxies. While for w_{gp} we observe a clear trend of correlated points at scales of $r_p > 1\,h^{-1}$Mpc, this is not observed for w_{gx}, where there is no clear ascending or descending signal as a function of r_p. Nevertheless, we observe that most of the w_{gx} points have negative values, with upper bounds consistent with 0 in most cases. Thus, this may be a hint of the presence of some systematics in this measurement, although this is not entirely clear. The centre plot of the figure shows the case for the red z_b bin 2 configuration, where correlated points for w_{gp} are observed again at scales of $r_p > 1\,h^{-1}$Mpc, while there is no correlation for the w_{gx} points, just some noise at $r_p \sim 3\,h^{-1}$Mpc. Finally, on the right-hand side plot, the case of blue galaxies in the luminosity bin 2 shows signals of opposite sign for w_{gp} and w_{gx} at $r_p > 5\,h^{-1}$Mpc. This may indicate that this configuration is affected by some systematic, although further analyses should be done to disentangle its cause.

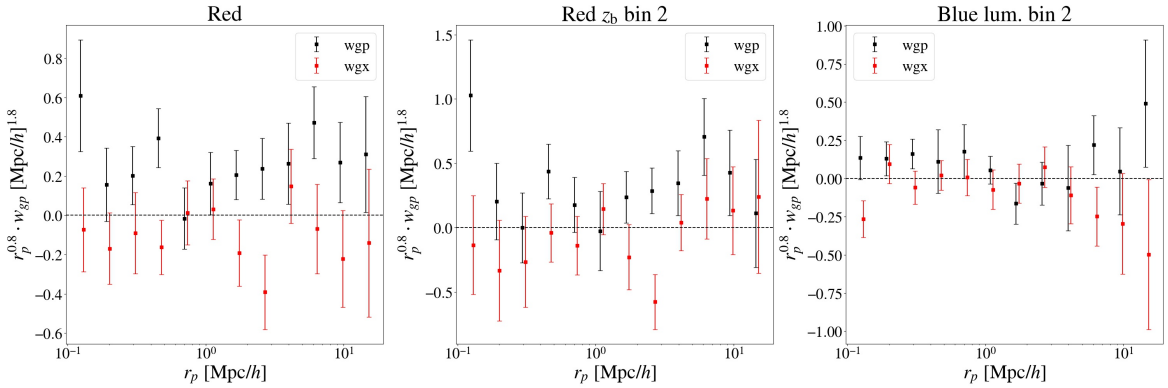

Fig. C.1 Comparison of the w_{gp} (black) and w_{gx} (red) signals for the unclear cases.