A Multi-probe Approach to Investigate Open Issues in Cosmology

Rishikesh Pandit

Supervisor: Prof. Enzo Branchini

Coordinator: Prof. Giuseppe Degrassi

Department of Mathematics and Physics
University of Roma Tre

This dissertation is submitted for the degree of
Doctor of Philosophy

PhD Cycle: XXXII

February 2020

To my beloved Universe…

Declaration

I, hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

The work presented in Chapter 5 was previously published in Dolfi et al. (2019) as "Clustering properties of TGSS radio sources" by Arianna Dolfi, Enzo Branchini, Maciej Bilicki, Andres Balaguera Antolinez, Isabella Prandoni and Rishikesh Pandit. This study was conceived by all of the authors. I carried out work related to 'Theoretical modelling of angular power spectrum using different halo bias models'.

Most of the work presented in Chapter 6 is my own, except where work which has formed part of collaboration is done as a part of *Euclid* Consortium and reference has been made to respective authors either by name of the person or his/her publication.

All the results presented in this thesis are computed using `INFN-ROMA3` cluster facilities at the Department of Physics in University of Roma Tre and `Science Machines` at Argelander Institute for Astronomy in University of Bonn. For programming and computational analysis I made use of: `C++`, `GSL`, `IPython`, `AstroPy`, `HealPy`, `NumPy/SciPy`, `Matplotlib` and `SuperMongo`.

Rishikesh Pandit
February 2020

Acknowledgements

First and Foremost, I would like to express my sincere gratitude to my supervisor Enzo Branchini for his exceptional guidance, clear and comprehensive mentoring at all times, and, for his admirable enthusiasm and passion in the field which constantly inspired me and helped me steer this research in the right direction. It is not possible to thank him enough for his patient and unending support during the impediments and hurdles in this thesis work.

Secondly, I am very grateful to Cristiano Porciani, who also supervised and mentored my master thesis as well as the research work carried out for the analysis of Chapter 4. His effortless ways of teaching and sparking insights, helped me to get a grip on my conceptual understanding of the subject.

A big thanks to the senior researchers in my group, especially to Alfonso Veropalumbo and Andres Balaguera Antolinez for their guidance, useful discussions and inputs for improving my work, both in writing and in analyses. I would commonly thank to all the colleagues and collaborators, with whom I interacted during most of this work: Lee Whittaker, Arthur Loureiro, Martin Kilbinger, Victoria Yankelevich, Joseph Kuruvilla, Emilio Romano Diaz, Jun Koda and Enrico Garaldi. Learning from their inputs that covered various aspects of Large Scale Structure Cosmology has helped me broaden my perspective in the field.

Furthermore, I am thankful to my friends spread out across different parts of the world, for adding interesting adventures to my life, without it I would have missed countless fun and memorable moments.

In the end, I am extremely grateful to my family, for constantly encouraging me to follow my ambitions and for standing by me in tough times. I cannot thank enough to my mother and father for always being supportive and kind to my choices in life; to my brother Yatin, who is my first inspiration and my mentor in life and whose sculpting hands can transform clay into life; to my sister-in-law Preshita, my second mother, whose unending motherly love and impeccable willpower in life has meant everything; and to my little niece Vangmayaa, who always grows a couple of inches taller when I visit home. And finally, to Vandita, who is my *Vita*, my Universe…

Table of contents

List of figures

List of tables

Introduction and Motivation

Today cosmology has emerged as a '*precision science*', thanks to the modern observational surveys and advanced theoretical models that, together, allow us to estimate the fundamental cosmological parameters with unprecedented precision. As a result the so-called concordance cosmological model, ΛCDM, has been remarkably successful. It is a simple theoretical model that, using Einstein's general relativity with a non-vanishing cosmological constant, is able to match with an impressively large and diverse amount of observational data: the cosmic microwave background; 3D distribution of large scale structure; supernovae distances; gravitational lensing; cosmic structure abundances and so on. This success is unparalleled by the physical understanding of this model, that requires the presence of dark components that would contribute to ~95% of the mass-energy budget of the Universe at the present epoch. The quest for the nature of these two elusive components, that we indicate as *Dark Matter* and *Dark Energy*, represent one of the most interesting open issues in cosmology and fundamental physics till date.

In the framework of the ΛCDM concordance model, the Universe would originate from an initial singularity, the so-called hot Big Bang, followed by an inflationary period, that would account for the flatness of the observable Universe. During its thermal history, a cosmological nucleosynthesis would occur, during which the light elements would form in a hot-dense, optically thick plasma from which a transparent matter-dominated Universe would emerge after the recombination epoch, i.e. when photons would decouple and freely stream away, allowing us to observe the last scattering surface of this process in the microwave range today: what we identify as the cosmic microwave background (CMB, Penzias and Wilson, 1965). CMB is nearly isotropic and uniform black-body radiation detected everywhere in the Universe with average background temperature of $T_{\text{CMB}} \sim 2.73$ K. However this uniform glow in the sky has tiny temperature variations of $\Delta T/T \sim 10^{-5}$ order of magnitude and have been observed by a number of successful satellite missions: *Cosmic Background Explorer* (COBE, Smoot et al., 1992), *Wilkinson Microwave Anisotropy Probe* (WMAP, Spergel et al., 2003) and finally, *Planck* (Planck Mission, Bersanelli et al., 2010). These temperature fluctuations are associated to fluctuations in the mass density and are believed to be originated from quantum fluctuations that enlarged to macroscopic scales during the inflationary era. CMB fluctuations are then further amplified, through the mechanism of gravitational instability, and eventually form the familiar cosmic structures that we observe in the nearby Universe.

The accelerated expansion of the Universe and the evolution of cosmic structures within, are driven by the aforementioned dark matter and dark energy. As a consequence, tracing the expansion history of the Universe and quantifying the growth of the large scale structures is an effective way of unveiling the nature of these dark components. This thesis work is motivated by the latter aspect, in particular, in measuring the rate, f, at which the structures grow. This quantity, that can be estimated from the observed large scale distribution of extragalactic objects at various epochs, is a powerful tool to investigate the nature of gravity. It could shed light on the nature of dark energy, as well as provide useful indications on how to build a new gravity model. It is to address these important questions that new, advanced observational campaigns have been or will be carried out. In the next years, large scale surveys like *Euclid* (Laureijs et al., 2011), *The Large Synoptic Survey Telescope* (LSST, Abate et al., 2012) *Dark Energy Spectroscopic Survey* (DESI, Aghamousa et al., 2016), *Wide-Field Infrared Survey Telescope* (WFIRST, Doré et al., 2018) will generate datasets containing billions of objects spanning over a significant fraction of the observable Universe. A major challenge to the scientific community is to build these experiments, collect and analyse the huge amount of data and elaborate accurate theoretical model to extract cosmological parameters, including $f(z)$, with unprecedented precision.

The main goal of this thesis is to study efficient ways to investigate the distribution of various mass density tracers. I will first consider the case of 3D distributed density tracers. As we know, deep spectroscopic redshift surveys, estimate the redshifts by scanning the spectrum of individual object, and accurately map the 3D distribution of matter around us. Tracers of the large scale matter density field, primarily galaxies, clusters of galaxies and ‘cosmic voids’, are a goldmine of scientific information that is typically extracted from this distribution by means of two-point statistics: in configuration-space (Hawkins et al., 2003; Zehavi et al., 2005; Eisenstein et al., 2005; Okumura, 2008; Cabré and Gaztañaga, 2009; Guo et al., 2013; Beutler et al., 2012; Veropalumbo et al., 2014; Ross et al., 2015; Jeong et al., 2015); in Fourier-space (Tegmark et al., 2004; Cole et al., 2005; Tegmark et al., 2006; Reid et al., 2010; Blake et al., 2011); or using both (Anderson et al., 2013, 2014; Alam et al., 2017). Over the past decade a number of spectroscopic redshift (hereafter spectro-z) surveys have been carried out, resulting in more than 3 million spectroscopic redshifts, largely of galaxies at low-to intermediate redshifts ($z \leq 0.3$) (Sloan Digital Sky Survey I-IV, York et al., 2000; Abazajian et al., 2009; Eisenstein et al., 2011; Blanton et al., 2017); (2-degree Field Galaxy Redshift Survey, Colless et al., 2001); (VIPERS, Guzzo et al., 2013). On the other hand, much larger are the photometric catalogues, that only list angular positions and magnitudes (in various bands) of objects (Dark Energy Survey (DES), Abbott et al., 2005); (Kilo Degree Survey (KiDS), de Jong et al., 2012); (2MASS Photometric Redshift Survey (2MPZ), Bilicki et al., 2013). Current photometric catalogues contain tens of millions of objects but can only be used to analyse the 2D-projected distribution of objects. If photometric redshifts (hereafter photo-z's) can be measured, the information of line-of-sight distribution of objects is insufficient to perform full 3D analysis, as the gain of number counts comes with a compromise of high uncertainty on redshift. Nonetheless, photo-z catalogues are ideal datasets to measure the gravitational lensing effect in the weak regime. However, by using tomographic clustering technique (Asorey et al., 2012; Alonso et al., 2015; Balaguera-Antolínez et al., 2018; Peacock and Bilicki, 2018) they can also be used to infer the 3D clustering properties of both the luminous tracers and the underlying matter. In this thesis I will focus on this second aspect.

The general framework of my thesis is the study of the large scale structure of the Universe through multiple aspects of clustering analysis. In the following subsections, I will sketch an outline of three different but complementary analyses that together build this thesis work.

Probing the Growth of Structure from Galaxy Clusters

Galaxy clusters, formed by gravitational clumping of thousands of galaxies, are the largest virialised structures in the Universe. In the ΛCDM bottom up scenario, they have formed relatively recently, when the Universe started accelerating its expansion. For this reason their abundance and clustering properties are sensitive probe to Dark Energy and Modified Gravity models (Rozo et al., 2007; Jain and Zhang, 2008; Ferraro et al., 2011; Ade et al., 2016b; Batista and Marra, 2017). Moreover, thanks to their intense Bremsstrahlung radiation, they're best detected and studied in the X-ray band. The recently launched *eROSITA* X-ray satellite, which has been designed to observe the largest cluster count to date, is expected to deliver an almost all-sky catalogue of about 10^5 clusters out to $z \sim 1$ (Merloni et al., 2012; Pillepich et al., 2012). We expect that 3D spectroscopic follow-up surveys (e.g. 4MOST Salvato and Merloni, 2015) after *eROSITA* (Reiprich, 2016, 2017) will provide reliable redshift estimates of *eROSITA* galaxy clusters at mean redshift $z = 0.46$.

Clusters are rare but bright. Therefore, they can be used to probe the underlying mass distribution on very large volumes, although in a rather sparse way. As a consequence, the study of their clustering properties and the so-called redshift-space distortions (RSD, Sargent and Turner, 1977; Davis and Peebles, 1983) provide a very effective way to measure the growth rate over a large interval of redshifts. Since the work of (Guzzo et al., 2008), studying RSD using 2-point statistics have become a standard, effective tool to constrain the growth rate $f(z)$. This has mainly been done using galaxies as mass tracers (Blake et al., 2011; Reid et al., 2012; Beutler et al., 2012; Samushia et al., 2014; de la Torre et al., 2013; Oka et al., 2014; de la Torre et al., 2016; Pezzotta et al., 2016; Qin et al., 2019; Hawken et al., 2019) although galaxy groups and more exotic objects like Active Galactic Nuclei (AGN) have been used as well (Okumura, 2008; Bianchi et al., 2012; Kwan et al., 2012; Marulli et al., 2017). A key ingredient for the measurement of the growth rate is an accurate model for redshift distortions. In the linear regime the model, originally proposed by Kaiser (Kaiser, 1987), has an analytic form which is simple to implement and allows one to estimate parameter combinations like $f\sigma_8$ (σ_8 being the *RMS* mass density fluctuation on 8 h^{-1} Mpc scale) or the amplitude of redshift distortions $\beta = f/b$ (where b is the linear bias parameter of the mass tracer). In fact, several past works have considered β as a benchmark to probe the growth of structure (Hatton and Cole, 1999; Berlind et al., 2001; Chuang and Wang, 2012; Tocchini-Valentini et al., 2012; Mohammad et al., 2016). A common issue when using galaxies as tracers of mass density field is that their RSD effects are poorly described by linear theory, and more sophisticated models need to be adopted instead.

The benefit of using clusters instead of galaxies is that, tracing cosmic structures and velocities over much larger scales, are intrinsically more linear or, more precisely, the RSD in their clustering properties are much more accurately described by the Kaiser-Hamilton model (Kaiser, 1987; Hamilton, 1992). This has triggered a widespread interest in clusters and prompted a number of theoretical works aimed at assessing their goodness as a RSD probe (Marulli et al., 2017) or for analyses related to Baryonic Acoustic Oscillations (BAO) (Veropalumbo et al., 2014).

In this thesis I will further explore this issue, keeping an eye on the upcoming *eROSITA* survey. My goal will be that of assessing how well the distortion parameter β can be estimated from the 2-point statistics of $\sim 10^5$ clusters in survey like *eROSITA*. I will address the problem in both configuration space, i.e. by measuring the anisotropic 2-point correlation function and its moments, and in Fourier space, by considering the cluster power spectrum. The rationale for considering both statistics is that, although they provide the same information, they are prone to different types of systematic errors, so that the comparison of the two results will clarify which statistics perform better in a particular range of scales or redshift. To do so, I will analyse the distribution of dark matter haloes extracted from 160 realisations of state-of-the art `zHORIZON` cosmological simulations

(Smith, 2009) considering only those objects with masses that match with those expected for the *eROSITA* clusters.

Extracting Clustering Signal from Radio Datasets

Radio-active extragalactic sources have long been regarded as interesting laboratories to study fundamental physical mechanisms and to highlight one of the most complex open problem in astrophysics: that of galaxy formation and evolution (Faint Images of Radio Sky at Twenty centimeters (FIRST), Becker et al., 1995; White et al., 1997, 1.4 GHz); (The Giant Metrewave Radio Telescope (GMRT), Ananthakrishnan, 1995); (Sub-millimeter Array (SMA) Blain et al., 2002); (The Phoenix Deep Field Survey (PDF) Hopkins et al., 2003). On top of this, radio sources are also precious mass tracers since they are little affected by galactic extinction, zodiacal light, starlight emission from our Galaxy and, more in general, by most of the processes that hamper our ability to observe extragalactic objects over all sky in the optical and infrared bands. The ability of observing radio sources over a large fraction of the sky has resulted in a number of large radio surveys, such as the TIFR GMRT Sky Survey at 150 MHz (TGSS, Swarup, 1991), the NRAO VLA Sky Survey at 1.4 GHz (NVSS, Condon et al., 1998), or the Sydney University Molonglo Sky Survey at 843 MHz (SUMSS, Bock et al., 1999), the Murchison Widefield Array (MWA, Tingay et al., 2013) that have been extensively used to study the large scale structure of the Universe both in 2D, when only the flux in some radio bands is measured, or in 3D, if the 21 cm line can be observed in the spectrum of the radio object. This unique characteristic has triggered ongoing international effort to carry out increasingly large radio surveys such as the Low-Frequency Array (LOFAR, van Haarlem et al., 2013) and, in the near future, Square Kilometre Array (SKA, Braun et al., 2015; Prandoni and Seymour, 2015).

While awaiting these future datasets to be collected and published, dedicated analyses can be carried out on the existing ones. In this thesis I have addressed the specific issue of analysing the two largest radio catalogues available to date, NVSS and TGSS, to investigate the clustering properties of these objects on very large angular scales that can only be probed by these wide radio surveys.

The study of large scale structure with radio objects presents its own characteristic problems that are different from those typically faced in the optical surveys. Detection and characterisation of radio sources requires a good understanding and removal of the foreground signal as well as the ability to detect the presence of multiple images of the same sources that may mimic the presence of different, nearby sources. Another characteristic of radio catalogues is their variety. Indeed a typical radio survey contains many different types of radio sources (radio galaxies, quasars, starburst galaxies, AGNs) that are characterised by different luminosity functions and, therefore, different redshift distributions. Furthermore, radio galaxies are typically located at high redshifts with very faint optical counterparts. As a result, only a small fraction of radio sources, especially those located in the local universe, have measurements of photometric and spectroscopic redshifts (e.g., Peacock and Nicholson, 1991; Magliocchetti et al., 2004). In general, the composite nature of the radio catalogue needs to be accurately modelled to extract cosmological information from the angular clustering properties of the radio sources.

As of today, the angular correlation properties of radio sources have been investigated using two-point statistics, both in real and harmonic space, in all available wide-angle radio samples (e.g., Cress et al., 1996; Loan et al., 1997; Blake and Wall, 2002b; Overzier et al., 2003; Blake et al., 2004b; Negrello et al., 2006; Chen and Schwarz, 2016; Blake et al., 2004a; Nusser and Tiwari, 2015). At small angular scales the results are in agreement with theoretical predictions of ΛCDM-based models. On the contrary, several authors have detected a significant excess of power on very large angular scales. A particular emphasis has been given to the dipole moment of the NVSS radio sources that was found to be higher than ΛCDM predictions, However, there is no agreement on the significance of this mismatch (Singal, 2011; Gibelyou and Huterer, 2012; Rubart and Schwarz, 2013; Fernández-Cobos et al., 2014; Tiwari and Jain, 2015; Tiwari et al., 2015). To clarify this issue, more recent studies have been focused on the newer TGSS sample (Rana and Singh Bagla, 2018). This catalogue significantly overlaps with NVSS and, because of its brighter flux threshold, is supposedly less affected by systematic errors that may mimic a spurious dipole. It turned out that not only the TGSS dipole is large, but that the discrepancy with theoretical predictions is even larger than in the NVSS case (Bengaly et al., 2018). The goal of this work, is to have a closer look to the issue and re-analyse the angular clustering properties of TGSS sources in harmonic rather than in real (configuration) space, considering all multipoles, instead of the dipole only. To do so, we have measured the angular power spectrum of different TGSS subsamples, compared these measurements with those obtained from the reference NVSS survey and, finally to compare with model theoretical predictions. My main contribution is on this latter aspect. To obtain the TGSS angular power spectrum I have modelled the redshift distribution and the bias of a multi-component population of radio sources and combined this information with the matter power spectrum predicted in the standard ΛCDM framework. The exercise has been repeated for a number of different TGSS sub-catalogues obtained using different flux cuts and angular masks, to assess the robustness of the results. The results of this analysis have been published in Dolfi et al. (2019).

Constructing a Pipeline to Simulate Next-Generation Wide Photo-z Survey Catalogues

Tomographic clustering analyses on photo-z catalogues have recently emerged as an alternative to that using 3D spectroscopic catalogues, to investigate clustering statistics over very large redshift ranges with much larger number of objects than with standard spectroscopic datasets (Asorey et al., 2012; Cai and Bernstein, 2012; Asorey et al., 2014; Balaguera-Antolínez et al., 2018). Forthcoming surveys will gather plethora of datasets in photo-z regime (factor of 100 more objects than spectroscopy) (Laureijs et al., 2011; Abate et al., 2012), for which tomographic methods will act as assets and therefore need to be fully tested with large simulated datasets which emulate the properties of the survey.

These next generation photometric datasets, will also require to be accurately modelled for the selection effects and statistical as well as systematic photo-z errors that are typically much larger than the spectroscopic ones. In addition, since photo-z samples are mainly designed for weak-lensing analysis, we need to assess whether photo-z catalogues are also suitable for clustering analysis of density tracers e.g. galaxies, if some additional selection need to be applied, and investigate further properties to find the optimal way of combining the two types of analyses. The best way to address these issues is by performing the planned clustering and/or weak lensing analysis on synthetic datasets that simulate the characteristics of a reference photo-z survey. The goal is that of assessing the magnitude of both statistical and systematic errors, their covariance, and design strategies to mitigate their impact.

However, this is a challenging task, since one needs to generate a very larger number of catalogues (of the order $(10^3 - 10^4)$) containing up to 10^9 objects with the possibility of changing the underlying cosmology and/or the modelled selection effects in the process. For this order of number of mock samples, the usual approach of performing N-body experiments is computationally too expensive and would not work in this case. Therefore, one needs to design approximated methods, that are both efficient enough to generate a large number of independent mock datasets and sufficiently accurate to mimic observational errors and selection effects (Agrawal et al., 2017; Balaguera-Antolínez et al., 2018). As a major part of the thesis work, I will develop full-pipeline to generate such mocks. The pipeline will be flexible enough to mimic any type of photo-z survey, although the one that is a part of this work will be specifically designed to imitate the upcoming photometric redshift survey that will be carried out by the *Euclid* satellite.

Briefly, the pipeline will consist of two stages: In the first one, I will focus on generation of ideal sets of mock catalogue containing discrete objects distributed along a light cone and divided into redshift slices. Central to this stage is the assumption that the number density field of discrete objects will obey a Lognormal statistics Coles and Jones (1991), a reliable assumption that is indeed approximately matched by real dataset Kayo et al. (2001). To assign the correct underlying statistical properties to the density field, one needs to generate the angular power spectrum of a population (or different populations) of extragalactic objects with specified bias. The corresponding spectrum is sensitive to the underlying cosmology. Together with the nonlinear model used to predict the mass power spectrum on small scales, the characteristics of the mass tracers such as their bias and redshift distribution also affects the behaviour of the spectrum. Especially for tomographic analysis, the angular auto- and cross- spectra for each redshift interval of the survey must be tuned accordingly. For this crucial stage, I will use specifically designed publicly available codes, one of which is `FLASK` (Xavier et al., 2016) that generate Lognormal maps and catalogues of discrete mass tracers. The code is mainly aimed and optimised for cosmic shear and weak-lensing convergence fields. My task will be to assess its usability as well as limit and optimise it for galaxy density fields. `FLASK` requires the input angular power spectrum, for which I will use `CLASS` (Lesgourgues, 2011a; Di Dio et al., 2013). The entire stage will be validated after the completion before moving to next stage.

The second stage of the pipeline will contain the framework that takes the ideal mock catalogue from first stage and turn it into a realistic mock catalogue. For this stage I will formulate original work, in which I will implement the properties of the *Euclid* photometric survey, starting by assigning the mean photometric redshift and the observational error. Then I will impose redshift distribution of the target *Euclid* sources along with the survey geometry. At last, using the ancillary codes which will be used for *Euclid* main pipeline, I will validate the final output.

Outline of Thesis

In Chapter 1 and 2, I will revise the main theoretical tools and concepts that are used through the thesis. I will begin with the basics of background cosmology and then move to structure formation using gravitational instability theory, both in the linear and nonlinear regime. In Chapter 3, I will present the main statistical tools to analyse both the 3D and the 2D clustering properties of the mass distribution in the Universe. In Chapter 4, I will present the first project concerning 3D clustering analysis of galaxy clusters using large number of simulated data. The second project, i.e. the angular clustering analysis of TGSS radio sources is presented in Chapter 5. Finally, the pipeline to produce simulated galaxy catalogues for tomographic clustering analyses is presented in Chapter 6 together with results from pipeline validation and optimisation.

Chapter 1

Background Cosmology

Cosmology has evolved as a data-driven field since past few decades, through back-to-back dedicated observational probes which have remarkably improved our understanding of the Cosmos. The key contribution has arrived from space-based missions such as CMB probes like COBE (Bennett et al., 1996), WMAP (Hinshaw et al., 2013), Planck (Ade et al., 2014, 2016a; Aghanim et al., 2018), Supernova Type Ia independent measurements (Riess et al., 1998; Schmidt et al., 1998; Perlmutter et al., 1999) and 3D Galaxy Redshift Surveys (SDSS and 2dFGRS, York et al., 2000; Colless et al., 2001). Observational data infer that we live inside a universe with flat geometry in space, homogeneous and isotropic on large cosmic scales and at present exponentially expanding with time. This inference underlines the foundation of modern cosmology. The idea is to start with a homogeneous and isotropic Universe to simplify the understanding of the physical structure of the Cosmos and follow its evolution with time. All its components are treated as ideal fluids in their theoretical description. I have referred to (Kolb and Turner, 1990; Padmanabhan, 2002; Mo et al., 2010; Jones, 2017; Carroll, 2019) for most of the topics described in this chapter.

In § 1.1, I discuss the cosmological principle that uniquely define the metrics and considerably simplify the solutions of Einstein's equations. The relevant equations are discussed in details in § 1.2. These equations, and its metric derived by Friedmann, also expose non-static nature of the Universe since its beginning, which I will discuss as a case of expanding Universe and distance measurements in § 1.2.1. In the end, a set of fundamental equations derived by Friedmann, are also used to obtain the dimensionless energy density parameters for individual components of matter and energy which I describe in last Section § 1.2.2.

We utilise following conventions for this chapter:

- Planck system of units (unless stated specifically), $\hbar = c = 1$;
- Metric sign convention is (-, +, +, +);
- Time derivative is represented as $\dot{x} = \frac{dx}{dt}$;
- Subscript ‘0’ corresponds to present time, e.g t_0 is cosmic time today.

1.1 Homogeneous-Isotropic Universe

The simplest assumption to start with is treating every spatial coordinate as same and indistinct from any other coordinate at random position in space. In other words, considering the Universe as homogeneous and isotropic. These two properties together are called as *cosmological principle*. The principle implies that there exists no special place anywhere in spacetime. Empirically, this assumption is valid only on very large spatial scales. We find our Universe to be homogeneous and isotropic on scales much larger than the average size of the largest virialised cosmic structures, i.e clusters of galaxies ($r \gg 100$Mpc). The *cosmological principle* lays the basis of modern cosmology.

1.1.1 Einstein’s Universe: General Relativity

The only fundamental force which governs the universe up to very large spatial scales is gravity. According to classical Newtonian gravity, the amount and the distribution of matter determines the gravitational field. The Poisson’s equation, $\nabla^2\Phi = 4\pi G\rho$, relates the gravitational potential Φ to its source: the mass density ρ, where G is the gravitational constant. In the framework of General Relativity theory (GR) (Einstein, 1916), the Poisson equation takes form of Einstein’s Field Equations (EFEs), i.e a set of non-linear differential equations that link the geometrical properties of the spacetime to its mass-energy content. EFEs read as follows:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G\, T_{\mu\nu} \,, \tag{1.1}$$

where $R_{\mu\nu}$ is *Ricci Tensor* and R is *Ricci Scalar*, both are contractions of the *Riemann Tensor* $R^{\sigma}_{\lambda\mu\nu}$. Left hand side (LHS) of the equations in (1.1) contains all information on spacetime structure and geometry. Whereas the *Energy-Momentum Tensor*, $T_{\mu\nu}$, which forms the right hand side (RHS), describes the energy density of the various components of the system, characterised by their pressure p and density ρ. On a brief note, the field equations (1.1) describe how the spacetime curvature acts on matter to manifest itself as gravity as well as how energy and momentum tell the spacetime to curve. For a single ideal fluid, the *Energy-Momentum Tensor* yields as:

$$T_{\mu\nu} = (p+\rho)u_\mu u_\nu + p g_{\mu\nu} \,, \tag{1.2}$$

where u_μ is a four-vector that represents velocity of the fluid. The equation of state (EoS) is usually expressed as:

$$w = \frac{p}{\rho} \,, \tag{1.3}$$

where w is called the Zel'dovich parameter and specifies the physical properties of various fluids: non-relativistic matter, photons etc. It also parametrises individual fluid components of matter and energy.

1.1.2 Metric of Spacetime

The topology of spacetime is characterised by its metric:

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu \,, \tag{1.4}$$

where $g_{\mu\nu}$ is a metric tensor. If the spacetime is homogeneous and isotropic i.e for a universe that obeys cosmological principle, it becomes *Friedmann-Lemaître-Robertson-Walker* (FLRW) metric:

$$ds^2 = -dt^2 + a(t)^2\, \gamma_{ij}\, dx^i dx^j \,, \tag{1.5}$$

where γ_{ij} is Euclidean three-metric, $a(t)$ is cosmic scale factor normalised such that $a(t_0)$ = 1 at present time t_0. The first part on the RHS of Equation (1.5) is temporal part of the metric and the latter part is spatial part in which the time-dependence is factorised by $a(t)$. When spatial component γ_{ij} is expressed in spherical coordinates then FLRW metric takes the form:

$$ds^2 = -dt^2 + \; a(t)^2 \left[\frac{dr^2}{1-Kr^2} + \; r^2\left(d\theta^2 + \sin^2\theta \; d\phi^2\right)\right], \tag{1.6}$$

where K embodies the curvature. The FLRW metric from Equation (1.6) describes the homogeneous, isotropic and expanding universe with $a(t)$ and curvature, K, which can take three different values that uniquely characterise the geometry of the Universe,

$$K = \begin{cases} -1, & \text{Negative Curvature, Open Universe.} \\ 0, & \text{Zero Curvature, Flat Universe.} \\ +1, & \text{Positive Curvature, Closed Universe.} \end{cases} \tag{1.7}$$

1.2 Cosmic Expansion

EFEs are in principle higher order non-linear differential equations. In 1922 Alexander Friedmann solved EFEs under assumption of FLRW metric; and derived equations that describe the dynamics of an expanding Universe (Friedmann, 1922). From the 00 component of the EFEs one obtains the first Friedmann equation:

$$H \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G \rho_i}{3} - \frac{K}{a^2}\,. \tag{1.8}$$

where i indicates the i-th fluid component, and $H = \dot{a}/a$ is Hubble parameter that characterises the expansion rate of the Universe. From the spatial component ii of the EFEs one obtains the second Friedmann equation:

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{8\pi G}{6}\,(\rho_i + 3p_i). \tag{1.9}$$

To solve the system of equations (1.8) and (1.9), one needs to specify the EoS that relates the pressure to density (p_i-ρ_i), where the boundary condition for ordinary fluid is $\rho_i + 3p_i \geq 0$ commonly known as '*strong energy condition*'.

1.2.1 Distances in Expanding Universe

In the late 1920's, Georges Lemaître and Edwin Hubble discovered independently, that Universe is expanding (Lemaître, 1927; Hubble, 1929) and, in a way, confirmed the Friedmann's theoretical work described above. It was observed that extragalactic objects recede with a velocity proportional to their distance from the observer:

$$v = H_0\,\mathbf{r}\,, \tag{1.10}$$

where $H_0 = 100\ h$ km s^{-1} Mpc^{-1} is *Hubble constant* which is parametrised in terms of dimensionless parameter h, and $\mathbf{r}$ is the physical separation vector that defines the actual distance between cosmic objects.

As the universe expands isotropically the space is stretched in all directions, causing the wavelengths of the emitted photons by a light source to be stretched as well. To understand this phenomenon better, let's elaborate more about expanding Universe and relative distances. The comoving separation $\mathbf{x}$ between two points, which stays same for space which is uniformly expanding with time, relates to physical separation $\mathbf{r}$ as:

$$\mathbf{r}(t) = a(t)\,\mathbf{x}\,. \tag{1.11}$$

In a uniform FLRW universe $\mathbf{x}$ remains unchanged in time, whereas the corresponding displacements in inhomogeneous universe will change $\mathbf{x}$ with time.

Light rays follow null geodesics i.e $ds^2 = 0$. If we consider a photon travelling on a radial null geodesic, from equation (1.6) and integrating along the photon's trajectory, we get:

$$D_{\rm PR} = \int_t^{t_0} \frac{dt}{a(t)} = \int_o^r \frac{dr}{(1-Kr^2)^{1/2}} \tag{1.12}$$

where LHS is the proper distance between two points o and r. Photons emitted at r at different times experience the stretching because of the expansion, $a(t)$. As a result, a monochromatic wavelength emitted from the source as $\lambda_{\rm em}$ is observed at o with a longer wavelength $\lambda_{\rm obs}$, which gives:

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = \frac{\nu_{\rm em}}{\nu_{\rm obs}} = (1+z) = \frac{1}{a(t)}\,, \tag{1.13}$$

which defines the cosmological redshift z. Because of the relation between the redshift z and the comoving position of the source located at r, redshift z is commonly used as a distance proxy.

Since proper distances cannot be measured directly in cosmology, we define other types of distances in a phenomenological way. The more relevant is the *Luminosity Distance*, which is defined from the observed flux F of a source with intrinsic luminosity L:

$$D_L = \left(\frac{L}{4\pi F}\right)^{1/2}\,; \tag{1.14}$$

and the *Angular Diameter Distance*, defined as the distance to a standard ruler[1] of size $2R$ with an angular separation θ:

$$D_A = \frac{2R}{\theta} \,. \tag{1.15}$$

These observable quantities are related to each other and give the cosmologically well defined proper distance as,

$$D_A = \frac{D_{\mathrm{PR}}}{(1+z)} = \frac{D_L}{(1+z)^2} \,. \tag{1.16}$$

1.2.2 Cosmological Parameters

By combining first and second Friedmann equations, (1.8) and (1.9), we get *continuity equation* which implies conservation of energy and momentum:

$$\dot{\rho}_i + 3H(\rho_i + p_i) = 0\,; \tag{1.17}$$

that when coupled with EoS of the fluid, specifies the evolution of the density,

$$\rho_i(t) = \rho_{0i}\; a(t)^{-3[1+w_i]} \tag{1.18}$$

where $\rho_{0i} = \rho_i(t_0)$ is energy density of i-th fluid today. The dimensionless density parameter, which is used to estimate the relative amount of i-th fluid, is classically expressed as a ratio of energy density of the fluid to its *critical density*, ρ_c :

$$\Omega_{0i} = \frac{\rho_{0i}}{\rho_{0c}}\;;\; \Omega = \sum_i \Omega_i \,. \tag{1.19}$$

The first Friedmann equation can be expressed using the dimensionless energy density parameter Ω:

$$H^2 = H_0^2 \left\{ \left[\sum_i \Omega_i \; a^{-3(1+w_i)} \right] - \frac{K}{a\,H_0^2} \right\} , \tag{1.20}$$

where H_0 is the expansion rate at present time. Current observations favour a flat Universe ($K = 0$), presently dominated by cosmological constant, and contributed by non-relativistic matter (dark and luminous). For this model Universe, total energy density parameter (as given above in (1.19)) is expressed as:

---

[1] A standard ruler is an object whose physical size $2R$ is known

$$\Omega_0 = \Omega_m + \Omega_\Lambda \simeq 1 \,, \tag{1.21}$$

where Ω_m is total matter energy density and Ω_Λ is cosmological constant energy density. In this model we can recast Equation (1.20) and obtain the normalised expansion history of the Universe:

$$E(z) = \frac{H(z)}{H_0(z)} = \left[\Omega_m (1+z)^3 + \Omega_\Lambda\right]^{1/2} . \tag{1.22}$$

Chapter 2

The Large Scale Structure

In Chapter 1, we have seen geometrical properties of homogeneous-isotropic Universe and its evolution with time. This approximation holds well on large spatial scales. However on small scales the *cosmological principle* is violated as we observe a cosmic foam of structures either clumped together in clusters or distributed in filaments and sheets, separated by underdense regions called *cosmic voids*. All these structures traced by galaxies are the building blocks of the Large Scale Structure (LSS). The amount of cosmological information encoded in the LSS is enormous. Studying the evolution of LSS is a way to comprehend the mechanism that assembles such gigantic structures.

This chapter is dedicated to cosmology of structures and their formation. It is divided into two main sections. The first section (§ 2.1) summarises the linear theory of density perturbations, valid as long as the amplitude of the perturbations is smaller than unity ($\delta \ll 1$). I start by defining density contrast and the statistical properties of a linear density field in § 2.1.2. The evolution of linear perturbations and the growing modes of matter perturbations that contribute to formation of structures are discussed in § 2.1.3. Section § 2.2 describes non-linear evolution of matter density perturbations, and the formation of Dark Matter haloes that host galaxies and galaxy clusters. Finally I discuss the halo mass function and the bias of density tracers. I have referred to (Kolb and Turner, 1990; Coles and Lucchin, 1995; Padmanabhan, 2002; Schneider, 2006; Mo et al., 2010; Jones, 2017) for writing this chapter.

Before we dive into specific details about how structure formation takes place, let's recall the main assumptions of the ΛCDM model:

1. Large scale Universe is homogeneous and isotropic with a flat geometry;
2. General Relativity is the theory that describes gravity;
3. The main components of our Universe are:
   (a) **Dark Energy**: dominating the energy density today and possibly driving the cosmic acceleration, which in its simplest form, the cosmological constant, fully satisfies all observational constraints;
   (b) **Cold Dark Matter**: collisionless, cold, massive particles that interact gravitationally with ordinary matter and act as a key driver of structure hierarchy;
   (c) **Baryons**: atomic matter which interacts to CDM through gravity;
   (d) **Photons**: relativistic massless particles that dominated the energy budget at early epoch;
   (e) **Neutrinos**: low mass particles that decouple during cosmological nucleosynthesis;
4. Perturbations in energy density were generated during the inflationary epoch in early Universe

2.1 Structure Formation: Linear Theory

2.1.1 Primordial Universe

At very early times, the Cosmos was extremely hot and denser than today. The ordinary matter and radiation were coupled through Thompson scattering. Small perturbations in the matter-radiation fluid existed then, and they are responsible for the tiny temperature fluctuations observed in the cosmic microwave background (CMB) of the order,

$$\frac{\Delta T}{T} \equiv \frac{\Delta \rho}{\bar{\rho}} \sim 10^{-5} \,. \tag{2.1}$$

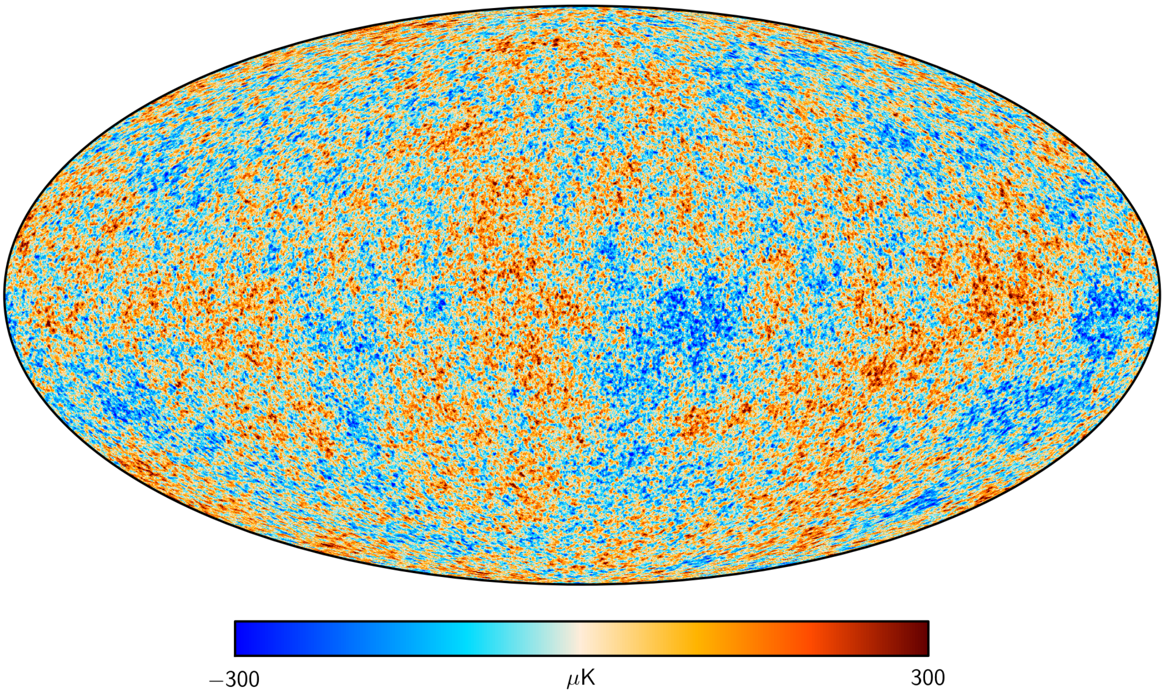


Fig. 2.1 Full-sky 2D projection of CMB temperature measurements, produced using SMICA which has been corrected for galactic plane, as observed by Planck satellite. The magnitude of temperature fluctuations is $\Delta T/T \sim 10^{-5}$ (Planck Collaboration et al., 2016).

When photons and matter decoupled, Universe became optically transparent, providing us with the possibility of mapping temperature fluctuations across the sky. This has been done for the first time by the COsmic microwave Background Explorer (COBE) satellite, and more recently by Planck. The Planck CMB map is shown in Figure 2.1. The gravitational instability model describes how these fluctuations grew to form the cosmic structures that we observe today.

The growth of the density perturbations mainly depend on their type, size in comparison with horizon size of the Universe and the dominant fluid component i.e the epoch at which they enter the horizon. This in principle takes into account the cosmology under consideration. We will discuss the evolution of density perturbations further in details in § 2.1.3. As long as the amplitude of these perturbations is considerably small, $\delta \ll 1$, the physics of their evolution can be described accurately with *linear theory*. The physical mechanism that drives the evolution of perturbations is *gravitational instability*. It predicts that only fluctuations with scale larger than the *Jeans length* can grow. While such overdensities evolve in time, they attract more matter and become denser leading to a gravitational collapse. The *Jeans length* is defined as:

$$\lambda_J = \left(\frac{\pi \, c_s^2}{G\rho} \right)^{1/2} , \tag{2.2}$$

where c_s is the speed of sound. For fluctuations with size larger than λ_J, self-gravity overcomes internal pressure and the overdensity collapses, growing indefinitely. This mechanism, originally adopted for star formation, has been applied to cosmology with considerable success. Before we learn about further implications of external parameters on growth of density perturbations, we must understand how the density perturbation is defined and its statistical properties.

2.1.2 Density Perturbations: Statistical Definition

The underlying statistical properties of density perturbations with extremely small amplitude are characterised by Gaussian probability distribution. A density perturbation $\delta(\mathbf{x})$ are defined as,

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\langle\rho\rangle} - 1 \, . \tag{2.3}$$

where $\rho(\mathbf{x})$ and $\langle\rho\rangle$ are matter density at point $\mathbf{x}$ and average matter density respectively. By definition a Gaussian random field is described with mean $\langle\delta\rangle = 0$. For a such a field field, the 1-point probability density of δ at a generic location $\mathbf{x}$ is,

$$\mathbb{P}[\delta] = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(- \frac{\delta}{2\sigma^2} \right) , \tag{2.4}$$

where σ^2 is *RMS* amplitude of fluctuation δ. It is convenient to expand $\delta(\mathbf{x})$ field in Fourier modes:

$$\delta(\mathbf{k}) = \frac{1}{(2\pi)^3} \int_{\mathcal{V}} \delta(\mathbf{x}) \; e^{-i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{x} \, , \tag{2.5}$$

where $\delta(\mathbf{k})$ are Fourier coefficients and $\mathbf{k}$ is the wave vector, $k \equiv |\mathbf{k}| = 2\pi/\lambda$. The inverse Fourier transformation, given $\delta(\mathbf{x})$ is a Fourier transform of $\delta(\mathbf{k})$, becomes:

$$\delta(\mathbf{x}) = \int_{\mathcal{V}} \delta(\mathbf{k}) \; e^{i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{k} \, . \tag{2.6}$$

As long as the perturbations are small, the Fourier modes $\delta(\mathbf{k})$ remain independent throughout the evolution of the perturbation $\delta(\mathbf{x})$. Besides defining 1-point statistical properties, Equation (2.4) is useful to characterise 2-point statistics of the overdensity field $\delta(\mathbf{x})$. We can define 2-point Correlation function (2PCF), $\xi(\mathbf{r})$, as the expected value of δ at distinct points $\mathbf{x}$ and $\mathbf{x}'$ separated by $\mathbf{r} = |\mathbf{x}-\mathbf{x}'|$, and it is expressed as:

$$\langle \delta(\mathbf{x})\delta(\mathbf{x}')\rangle = \xi(|\mathbf{x}-\mathbf{x}'|) \, . \tag{2.7}$$

If $\delta(\mathbf{x})$ is statistically isotropic then ξ depends only on the separation $|\mathbf{r}|$. Similarly, one can define the *Power Spectrum* as,

$$\langle \delta(\mathbf{k})\delta^*(\mathbf{k}')\rangle = P(\mathbf{k}) \, \delta^D(\mathbf{k}-\mathbf{k}') \, , \tag{2.8}$$

where δ^D is Dirac delta function. It is clear from the definition of $\delta(\mathbf{k})$ (Equation (2.5)) as well as Equations (2.7) and (2.8), that 2PCF $\xi(\mathbf{r})$ and power spectrum $P(\mathbf{k})$ are Fourier pairs,

$$P(\mathbf{k}) = \frac{1}{(2\pi)^3} \int \xi(r) \, e^{-i\mathbf{k}\cdot\mathbf{x}} \, d^3\mathbf{x} \, , \quad \xi(r) = \int P(\mathbf{k}) \, e^{i\mathbf{k}\cdot\mathbf{x}} \, d^3\mathbf{k} \, . \tag{2.9}$$

The power spectrum is a volume dependent quantity, $P(k) \equiv V^{-1}\langle |\delta(\mathbf{k})|^2\rangle$ and therefore has dimension of volume. It is conveniently defined as the *dimensionless power spectrum*:

$$\Delta^2(k) = \frac{k^3}{2\pi^2} \, P(k) \, . \tag{2.10}$$

But in practice, it is often convenient to consider a smoothed overdensity field, on a particular smoothing scale R as,

$$\delta(\mathbf{x}, R) = \int \delta(\mathbf{x}') \, W(|\mathbf{x}'-\mathbf{x}|, R) \, d^3\mathbf{x}' \, , \tag{2.11}$$

where $W(\mathbf{x}, R)$ is a window function which filters the density field on the scale R. One example is the Tophat window function:

$$W(\mathbf{x}, R) = \begin{cases} \text{const}, & |\mathbf{x}| \le R \\ 0, & \text{elsewhere} \end{cases} \, . \tag{2.12}$$

In Fourier space the variance of the smoothed overdensity is:

$$\sigma_R^2 = \frac{1}{2\pi^2} \int_0^\infty P(k) \, \tilde{W}^2(kR) \, k^2 dk \, , \tag{2.13}$$

where $\tilde{W}(kR)$ is a Fourier transform of $W(\mathbf{x}, R)$ and it has the form:

$$\tilde{W}(kR) = \frac{3[\sin(kR) - kR\cos(kR)]}{(kR)^3} \, . \tag{2.14}$$

From the smoothed variance, as expressed in Equation (2.13), on can set a characteristic mass scale assuming that a power spectrum behaves like a power law i.e $P(k) \propto k^{n_s}$ with n_s being scale invariant spectrum. The mass scale $M = \frac{4\pi}{3} R^3 \langle \rho \rangle$ contained in a sphere of radius R corresponds to a density field with variance that follows from (2.13):

$$\sigma_R^2 \propto \int_0^{1/R} k^{n_s+2} \propto R^{-n_s-3} \propto M^{-\frac{n_s}{3}-1} , \tag{2.15}$$

which implies that if $n_s > -3$, σ_R^2 is a decreasing function of M, which forms *hierarchical structure formation* scenario. These concepts along with the definition of σ_R^2, will be useful later in non-linear evolution of density perturbations, especially in § 2.2.3 and § 2.2.4. The above mentioned statistical measures that characterise cosmic density field, which also are essential tools to compare observed data with theoretical models through clustering analysis, are discussed in Chapter 3.

2.1.3 Evolution of Density Perturbations

Cosmic structures form due to gravitational collapse of matter density perturbations. The perturbation δ grows with the expanding Universe but is accelerated by mechanism of gravitational instability. Following the criteria of *Jeans length* as discussed before in § 2.1.1, different types of fluids (here radiation, DM and baryons) evolve according to their characteristic properties which dominate either gravity or pressure. If gravity dominates and pressure is negligible then overdensity grows further, in contrast if the pressure is on higher side the perturbation δ is suppressed. When two different types of fluids are coupled with interaction between fluid elements, as in the case of baryon-photon plasma in radiation energy dominated Universe before decoupling epoch t_{dec}[1], the interplay between gravitation and radiation pressure result in oscillated behaviour of baryon-photon $(\delta_b - \delta_r)$ perturbations. These oscillations, known as *Baryonic Acoustic Oscillations* (BAO), are frozen in the clustering information of the density tracers [see Figure 2.1]. BAO is also a powerful cosmological probe to constrain cosmological parameters and a useful tool to provide an insight in the nature of DE. In contrast to radiation and baryonic matter, DM perturbations δ_m remain unaffected by this because of their collisionless interaction with rest of the matter fluids, and continue to grow with the expansion, $a(t)$.

In previous chapter we saw in Equations (1.15) and (1.14), how physical distances relate to the expansion rate at that redshift z. In the same manner we asses the physical scale of the Hubble radius (horizon size) of the Universe, and it is expressed as:

---

[1] In many texts instead of epoch of decoupling t_{dec}, epoch of recombination t_{rec}, although both share the exact instant

Table 2.1 Summary of growth of perturbation as a function of the scale factor $a(t)$ at different epochs. The density perturbation δ in the radiation δ_r, baryons δ_b and *dark matter* δ_m components. All of them grow as the square of the scale factor when they are outside the horizon. After entering the horizon at radiation dominated epoch i.e $t < t_{\rm eq}$ baryons and radiation perturbations couple together and oscillate as acoustic waves. The growth of DM perturbations is frozen when they enter inside the horizon before matter-radiation equality because of Meszaros effect. After decoupling baryons follow the DM perturbations and grow with the scale factor $a(t)$

Epoch	δ_r	δ_m	δ_b
$t < t_{\rm enter} < t_{\rm eq}$; $\lambda > D_H$	$\propto a^2$	$\propto a^2$	$\propto a^2$
$t_{\rm enter} < t < t_{\rm eq}$; $\lambda < D_H$	oscillate	$\sim$ const	oscillate
$t_{\rm eq} < t < t_{\rm dec}$	oscillate	$\propto a$	oscillate
$t_{\rm dec} < t$	vanish	$\propto a$	$\propto a$

$$D_H = \int_{z'}^{\infty} \frac{dz'}{H(z')} \equiv a(t) \int_0^{t'} \frac{dt'}{a(t')} \,. \tag{2.16}$$

At early times, all the perturbations δ outside the Hubble radius (horizon) of the Universe grow as the square of the scale factor $\delta \propto a^2$, as there are no physical components affecting their growth. The perturbations outside the horizon are termed as *superhorizon* and those inside the horizon are *subhorizon* perturbations. Here the epochs that are important in growth of δ are following:

- $t_{\rm enter}$ or $a_{\rm enter}$: The epoch t or scale factor a at which the superhorizon perturbations become subhorizon perturbations;
- $t_{\rm eq}$ or $a_{\rm eq}$: The epoch at which Matter-Radiation energy density are equal;
- $t_{\rm dec}$ or $a_{\rm dec}$: The epoch at which photons decouple from radiation and free-stream.

In table below the growth of density perturbations according to their fluid type and the epoch at which they enter are summarised and in Figure 2.2 the evolution of δ with respect to the scale factor a is shown. It is clear from both that, in radiation era δ_m remain constant but in matter dominated era they grow as scale factor $\delta_m \propto a$ and baryons after being decoupled from radiation follow the dynamics of already grown δ_m. As long as the Universe is dominated with CDM content, the small density perturbations start growing earlier than Universe having only 'baryonic matter', resulting in boosted structure formation in CDM dominated Universe.

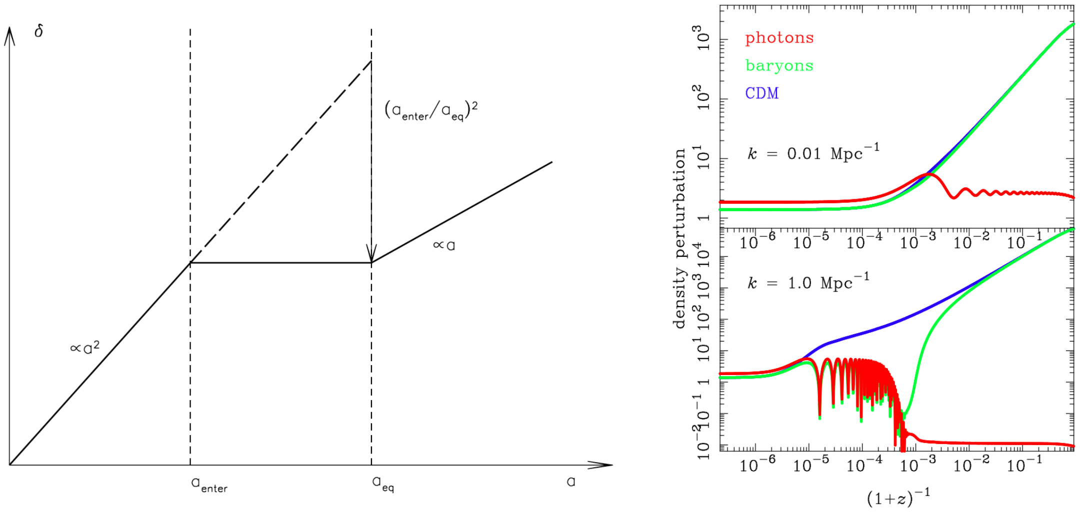


Fig. 2.2 **Left**: Evolution of a dark matter density perturbation with a scale larger than the Jean's length. Before entering the horizon the amplitude grows as a^2 ($a < a_{\rm enter}$). Then its amplitude remains constant as long as the Universe is dominated by radiation ($a_{\rm enter}<a<a_{\rm eq}$). Finally after matter-radiation equality, it grows as a. The growth of δ is suppressed by $(a_{\rm enter}/a_{\rm eq})^2$ for *superhorizon* perturbations at $a_{\rm eq}$ (Bartelmann and Schneider, 2001). **Right**: Evolution of density perturbations in various components (baryons + photons + CDM), for perturbation of different scales, around the epoch of decoupling $t_{\rm dec}$. Top panel shows the growth of *superhorizon* perturbations initiates as soon as the radiation perturbations decouple and oscillate, whereas in bottom panel subhorizon perturbations oscillate with photon-baryon coupling while CDM perturbations initiate their growth on the onset of matter domination. The baryons immediately interact gravitationally and attracted towards CDM potential wells after they become free from photons. [Plot taken from Challinor notes]

From their definitions expressed in (2.3) and (2.6), the density contrast which evolves in cosmic time t with comoving coordinate $\mathbf{x}$, can be expressed as:

$$\delta(\mathbf{x},t) = \frac{\rho(\mathbf{x},t) - \langle\rho(t)\rangle}{\langle\rho(t)\rangle}\,, \tag{2.17}$$

and its Fourier transform as,

$$\delta(\mathbf{x},t) = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3}\,\tilde{\delta}(\mathbf{k},t)\,e^{i\mathbf{k}\cdot\mathbf{x}}\,. \tag{2.18}$$

The *Fluid equations* govern the motion of fluid's particles. In case of pressureless matter, the comoving velocity of the particle is $\mathbf{u} = H\mathbf{x} + \mathbf{v}_{\text{pec}}$, where $\mathbf{v}_{\text{pec}} = a\dot{x}$ is its peculiar velocity with respect to the Hubble flow of expansion. The evolution of the fluid and its perturbation is described in an expanding Universe, by the following set of equations:

$$\begin{cases} \frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla_{\mathbf{x}}\cdot[(1+\delta)\mathbf{v}] = 0; & \text{Continuity Equation} \\ \frac{\partial}{\partial t}(a\mathbf{v}) + (\mathbf{v}\cdot\nabla_{\mathbf{x}})\mathbf{v} = -\frac{\partial\phi}{\partial x}; & \text{Euler Equation} \\ \frac{\partial^2\phi}{d\mathbf{x}^2} = 4\pi G\langle\rho\rangle a(t)^2\delta(\mathbf{x},t); & \text{Poisson Equation} \end{cases} \tag{2.19}$$

where δ is the mass density perturbation, $\mathbf{v}$ is the peculiar velocity of the fluid element, ϕ is the perturbation in the gravitational potential and $\langle\rho\rangle$ is the mean mass density.

If we assume small fluctuations, we can linearise (2.19) and find the solution to $\delta(t)$, for a flat cosmology ($K = 0$):

$$\ddot{\delta}(\mathbf{k},t) + 2H(t)\dot{\delta}(\mathbf{k},t) + \left(\frac{k^2c_s^2}{a(t)^2} - 4\pi G\langle\rho\rangle\right)\delta(\mathbf{k},t) = 0\,, \tag{2.20}$$

where the solutions of above Equation depends on the wave number $k \equiv |\mathbf{k}| = 2\pi/\lambda$ associated to the size of the perturbation λ. If $\lambda < \lambda_J$, then the solution oscillates and the perturbation propagates in the fluid as an acoustic wave, where as for $\lambda > \lambda_J$ then gravitational instability sets in and the overdensity grows.

In the limit $\lambda \ll \lambda_J$ the solution to (2.20) is the sum of two parts, one describing a growing solution, and the other a decaying one:

$$D_+ \propto a(t) \propto t^{2/3}\,, \tag{2.21}$$

$$D_- \propto a^{-3/2}(t) \propto t^{-1}\,, \tag{2.22}$$

where the decaying mode D_- is suppressed shortly after onset of gravitational instability. The general solution of Equation (2.20) is given by,

$$\delta(\mathbf{x},t) = D_+(t_i)\Delta_+(\mathbf{x}) + D_-(t_i)\Delta_-(\mathbf{x}) \,, \tag{2.23}$$

where t_i is initial time of perturbation's growth and $\Delta_+(\mathbf{x})$ is a space-dependent growing mode. In structure formation, only growing mode contributes to the growth of perturbation, therefore the linear evolution of $\delta(\mathbf{x},t) = D_+(t)\Delta_+(\mathbf{x})$.

In practice, the fiducial cosmological model defines the evolution of perturbations as given in Equation (2.23). The rate at which the perturbation grows is conveniently expressed by the growth rate:

$$f = \frac{d\ ln(D_+)}{d\ ln(a)} \,, \tag{2.24}$$

where f is a function of redshift z and it is sensitive to underlying background cosmology. An approximated, parametric expression for (2.24) has been given by (Peebles, 1980):

$$f(z) \simeq \Omega_m(z)^\gamma \,, \tag{2.25}$$

where $\gamma = 0.55$ for a ΛCDM model. The value of γ and its possible dependency on time and scale is sensitive to DE equation of state and deviations from GR.

2.2 Non-linear Evolution of Dark Matter Perturbations

Until now we have seen the evolution of density perturbations that are much smaller than unity. As soon as the density contrasts grow i.e $\delta \sim 1$ and increase further, the linear analytic description fails to describe their evolution accurately. Non-linear evolution couple the Fourier modes $\delta(\mathbf{k})$, making it more difficult to extract information from the analysis of the power spectrum. A full non-linear treatment for the evolution δ require the adaptation of numerical methods known as N-body simulations. These precious tools can also include hydrodynamical processes and dissipation physics, are however computationally expensive.

Alternatively, study of non-linear evolution, can be performed using simpler approaches, that use extrapolation linear theory under special assumptions. One of such ways is kinematic approach as Zel'dovich Approximation. Another is spherical collapse model. Though this approach serves limited purpose in understanding non-linear evolution of density perturbations, it provides a very powerful way to understand the collapse of matter overdensities and formation of Dark Matter haloes. The prediction about the expected number density of these haloes can be accurately estimated using such approximated mechanisms like spherical collapse model. We will explore in depth, the above mentioned approaches in following sections.

2.2.1 The Zel'dovich Approximation

Let us consider a massive particle at the physical position, $\mathbf{r}(t)$, or in comoving space, at $\mathbf{x}(t)$. The Zel'dovich approximation postulates that the physical coordinate of the particle at time t can be expressed as:

$$\mathbf{r}(t) = a(t)\chi(t) = a(t)[\mathbf{q} + b(t)\mathbf{f}(\mathbf{q})] \,, \tag{2.26}$$

where $\mathbf{q}$ is the comoving Lagrangian coordinate of a particle at $t = 0$, $b(t)$ is a time-dependent quantity which coincides with linear growth factor $D_+(t)$ and $\mathbf{f}(\mathbf{q})$ is a non-rotational velocity field that is proportional to the field of peculiar velocities $\nabla\psi(\mathbf{q})$. According to (2.26), particles move along straight lines on their velocity field which preserves their null vorticity as long as their orbits do not cross each other. The Zel'dovich approximation fails at orbit crossing. Before that, however, it is a powerful tool to follow the evolution of perturbations and the build-up of cosmic structures like filaments and voids.

2.2.2 Spherical Collapse Model

Evolution of a spherical mass overdensity with radius $r(t)$ enclosed in mass M can be treated analytically. The *spherical collapse model* is valid as long as the perturbations are spherical and the mass contained within each spherical shell is conserved. Under these conventions the equation of motion of the shell can be expressed as:

$$\ddot{r} = -\frac{GM}{r^2} \,; \tag{2.27}$$

where r is the radius of the shell and M is the mass contained within r. The equation of motion of the shell can be solved parametrically with $t(\theta) = B(\theta - \sin\theta)$ and $r(\theta) = A(1 - \cos\theta)$, where $A^3/B^2 = GM$. Following the expansion of the Universe, the shell starts expansion at $t = 0$ i.e. at $\theta = 0$. It reaches turnaround radius $r_{\rm max} = 2A$ at $\theta = \pi$ after $t_{\rm max} = \pi B$. Beyond this point the various shells cross each other, their mass is not conserved and the model is not valid anymore. However we can extrapolate the behaviour of the perturbation that will collapse at $\theta = \pi$ corresponding to $t_{\rm coll} = 2\pi B$. Before this happens, the system is virialised by holding following condition:

$$\mathrm{TE} = \frac{1}{2}\dot{r}^2 - \frac{GM}{R} \,, \tag{2.28}$$

Before shell crossing the evolution of matter density inside the shell can be expressed as:

$$\rho \propto \frac{M}{r^3} = \frac{3}{4\pi G B^2 (1-\cos\theta)^3}\,, \tag{2.29}$$

for early Universe dominated with matter perturbations. Time evolution of a spherical overdensity can be obtained by coping (2.29) with the evolution of the background density. For small t, when the perturbation is small, linear theory solution is recovered:

$$\frac{\Delta\rho}{\rho} = \delta \equiv \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}\,, \tag{2.30}$$

for $D_+ \propto t^{2/3}$. By the time of shell collapse $t_{\rm coll} = 2\pi B$, the overdensity is virialised until $\delta \sim 178$, the corresponding linearly extrapolated overdensity is:

$$\delta^{\rm lin} = \frac{3(12\pi)^{2/3}}{20} \simeq 1.686\,. \tag{2.31}$$

Equation (2.31) is very useful as it links the non-linear evolution of matter density perturbations δ_m, to linear theory predictions. This property allows ‘peaks’ of overdensities to be collapsed as mass haloes that will host luminous galaxies, to be related to the linearly extrapolated overdensity in the original density field. We stress that the spherical collapse model is ideal and breaks down for non-spherical fluctuations, that are stretched by external tidal fields generated by LSS. It is however a good approximation in describing the peaks of the density field that remain close to spherical, forming close-to-spherical virialised haloes.

2.2.3 The Press-Schechter Formalism

Spherical collapse model describes the non-linear evolution. Thanks to this model one can address the question on which fraction of matter in the Universe has formed collapsed structures at a given time and what is the mass distribution of these objects. Using spherical collapse model Bill Press and Paul Schechter in 1974 developed a "Mass Function" formalism (hereafter PS-formalism Press and Schechter, 1974) to estimate the number density of collapsed structures namely DM haloes, which also known as Press-Schechter theory.

In any realisation of a linear density field $\delta(\mathbf{x}, R)$, smoothed on the scale R by the window function defined in equation (2.12), there will be high density regions with $\delta > \delta_c$, where δ_c marks a threshold of critical density for triggering a collapse. The idea of the PS-formalism is to exploit the one-to-one relation between linear and non-linear overdensities and exploit the properties of a linearly evolved Gaussian field. Let us then use the statistical properties of a Gaussian random field as described previously in § 2.1.2. The goal is to predict the mass function of the collapsed haloes i.e. the abundance of DM haloes as a function of their mass. Let's recall the 1-point PDF of smoothed density field $\delta(\mathbf{x}, R)$:

$$\mathbb{P}[\delta, R]\, d\delta = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left(-\frac{1}{2\sigma_R^2}\right) d\delta\,, \tag{2.32}$$

where $\mathbb{P}[\delta, R]$ is probability that a smoothed density contrast exists between δ and $\delta + d\delta$. To obtain the mass function one assumes that collapsed objects form where the density contrast $\delta(\mathbf{x}, R)$ exceeds a certain threshold overdensity, which according to spherical collapse model is the critical density $\delta_c^{\mathrm{lin}} = 1.686$. The critical density δ_c is specified for the linearly evolved Gaussian random field. In an Einstein-de Sitter (EdS)[2] Universe its value is $\delta_c = 1.686$, which, as we have seen with respect to the virialisation threshold of a non-linear overdensity.

Let $f(M)dM$ be the number density of collapsed objects in the mass range $(M, M{+}dM)$ and let $F(M)$ account for those objects with masses greater than $M(R)$. The relationship between mass M and smoothing scale R varies with the window function, as the volume of the encompassing region depends on the type of window function. For Tophat window, mass $M = 4\pi\rho_M R^3/3$ and $M = (2\pi)^{3/2}\rho_M R^3$ for Gaussian window type. Following these relations, we use smoothing mass scale $\delta(\mathbf{x}, M)$ instead of $\delta(\mathbf{x}, R)$ that grows with time, and for any epoch we look for a region that exceeds δ_c on a scale $R(M)$ and are below δ_c on a larger scale $R' > R(M' < M)$. To retrieve the number of such isolated collapsed regions, a quantitative expression can be obtained by subtracting from the probability that the fraction δ_M exceeds δ_c:

$$F(>M) = \int_{\delta_c}^{\infty} \mathbb{P}[\delta_M]\; d\delta_M = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\nu}{\sqrt{2}}\right), \tag{2.33}$$

where $\mathrm{erfc}(x)$ is a complementary error function and $\nu \equiv \delta_c/\sigma_R = \delta_c/\sigma(M)$ is the overdensity threshold in units of the *RMS* density fluctuation. This formalism implies the presence of a typical scale for a collapse at any epoch, $M_\star$, such that the standard deviation is $\sigma_{R_\star} \equiv \sigma_{M_\star} = \delta_c$. However there is a limitation with this approach, as half of the collapsed objects are ignored following hierarchical structure formation for arbitrarily smaller smoothing scale $R \to 0$, and on contrary larger $\sigma_R \to \infty$ corresponds to structures with smaller mass $M \to 0$. The reason being $\mathrm{erfc}(0) = 1$ so that $F(0)$ in Equation (2.33) accounts for only half of the mass density is contained in collapsed objects. To bypass

---

[2] It is a flat matter only ($\Omega_m = 1$) FLRW Universe, where cosmological constant Λ is zero and it evolves with $a \propto t^{2/3}$.

the problem Press & Schechter multiplied (2.33) by factor of 2 to correct for the half number density, and defined the mass function of dark matter haloes, which accounts for distribution of virialised haloes between mass element M and $M + dM$ accordingly:

$$\begin{aligned} n(M)\, dM &= \sqrt{\frac{2}{\pi}}\, \frac{\langle \rho_M \rangle}{M^2}\, \frac{\delta_c}{\sigma} \left| \frac{d\, ln\, \sigma}{d\, ln\, M} \right| \exp\left(- \frac{\delta_c^2}{2\sigma^2} \right) dM\,, \\ &= \frac{\rho_M}{M^2}\, f_{\rm PS}(\nu) \left| \frac{d\, ln\, \nu}{d\, ln\, M} \right| dM\,, \end{aligned} \tag{2.34}$$

where $\langle \rho_M \rangle$ is the mean matter density and the quantity $f_{\rm PS}(\nu)$ is given by,

$$f_{\rm PS}(\nu) = \sqrt{\frac{2}{\pi}}\, \nu \exp\left(- \frac{\nu^2}{2} \right). \tag{2.35}$$

2.2.4 The Excursion Set Formalism

PS-formalism doesn't account for the smoothed overdensities $\delta(\mathbf{x}, R)$ at scale R which are underdense and have $\delta < \delta_c$. To assess this issue we use more stochastic approach with excursion set formalism, developed by Bond et al. (1991). The $1/2$ factor issue is not explained in the PS-formalism. To address the problem Bond et al. (1991) developed the extended PS-formalism (EPS).

In this approach, we use shorthand notation for mass variance, $S \equiv \sigma^2(M)$. Following hierarchical model, S is monotonically declining function of M, so that the larger the value of S, the smaller is the mass M. We can associate a trajectory $\delta(S)$ to each point in (S, δ_S) space. To satisfy the criteria for overdensity δ_S to follow Markovian Random walk, EPS adopts sharp k-space Tophat filter with $k_c = 1/R$, such that the smoothed density field is expressed as:

$$\delta_S(\mathbf{x}, k_c) = \int d^3\mathbf{k}\, W(kR)\, \delta_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{x}} = \int_{k<k_c} d^3\mathbf{k}\, \delta_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{x}}\,. \tag{2.36}$$

where $\delta_{\mathbf{k}}$ are Fourier coefficients of $\delta(\mathbf{x})$. The Gaussian properties of the random field are preserved, after a change $\Delta\delta(\mathbf{x}, k_c)$ is implemented due to an increase in the filter argument $k \rightarrow k_c + \Delta k_c$. The result is a Gaussian random variable with variance:

$$\begin{aligned}
\langle(\Delta\delta(\mathbf{x},k_c)^2)\rangle &= \langle(\delta(\mathbf{x},k_c+\Delta k_c)-\delta(\mathbf{x},k_c)^2)\rangle\ , \\
&= \int_{k_c}^{k_c+\Delta k_c}\Delta^2(k)dlnk\ , \\
&= \sigma^2(k_c+\Delta k_c)-\sigma^2(k_c)\ ,
\end{aligned} \tag{2.37}$$

which is independent of the original value $\delta(\mathbf{x},k_c)$. By definition, S is associated with $\sigma^2(M)$, where $M = 6\pi^2\langle\rho\rangle k_c^{-3}$. If $S \to 0$ then $M \to \infty$ and $\delta_S = 0$ for each $\mathbf{x}$. As S increases, the overdensity δ_S executes a Markovian random walk (See Figure 2.3).

Let us consider a mass scale S_1 corresponding to M_1 on x-axis in Figure 2.3, According to PS-formalism, the fraction of all trajectories having $\delta_S > \delta_c$ at S_1 is equal to the collapsed overdensities with mass $M > M_1$. There are two example trajectories A and B, in which an overdensity following B does not correspond to a collapsed overdensity at S_1 with mass $M > M_1$. However with mass scale $S_2 < S < S_1$ corresponding to $M > M_3 > M1$, is clearly a collapsed overdensity representing $\delta_S > \delta_c$. Thus it can be seen that PS-formalism fails to account for overdensities which follow B. Implementing stochastic approach with Markovian random walks, allows us to modify the PS-formalism ansatz, with a mirror-trajectory B' which is equally likely by following the change in $\Delta\delta(\mathbf{x},k_c)$ as explained in equation (2.37). The modified ansatz states that the probability that $\delta > \delta_c$ is the same as twice the fraction of mass contained in haloes with mass $M > M'$.

To obtain the halo mass function from new EPS ansatz, one needs to consider the fraction of trajectories that have their *first upcrossing* of the barrier $\delta_S = \delta_c$ at $S > S_1$, as given by trajectory A in Figure 2.3. Then according to new ansatz, these trajectories correspond to mass elements in collapsed objects of mass $M < M_1$. One can obtain this fraction of trajectories by formulating a conditional probability density function:

$$\begin{aligned}
F(<S_1) &= \int_{-\infty}^{\delta_c}(P(\delta,S_1)-P(2\delta_c-\delta,S_1))\ d\delta\ , \\
&= \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{\delta_c/\sqrt{S_1}} dx\ e^{-x^2/2} - \int_{-\delta_c/\sqrt{S_1}}^{\infty} dx\ e^{-x^2/2}\right]
\end{aligned} \tag{2.38}$$

where trajectories like B are subtracted and $P(B) = P(B')$ due to stochastic properties of trajectories. Equation (2.38) represents number density of masses included in haloes with $S > S_1$, i.e, $M < M_1$. Because the trajectory is a random walk, every trajectory will cross barrier as $S \to \infty$ and $M \to 0$, the number density will reach zero. Thus to obtain the final

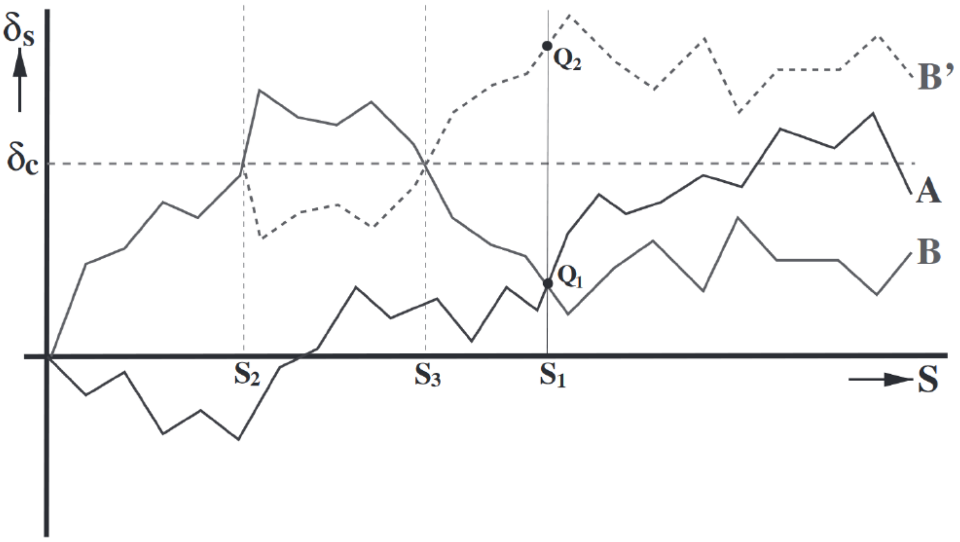


Fig. 2.3 Markovian Random walk of trajectories in (S, δ_S) space. Each trajectory corresponds to a mass element in initial (Gaussian) density field, and δ_S is the overdensity corresponding to smoothing mass scale $S \equiv \sigma^2(M)$, where δ_c is collapsing overdensity threshold (Mo et al., 2010).

EPS mass function, the fraction of haloes with masses $M > M_1$ gives $F(> M) = 1 - F(< S_1)$ and the mass function follows as:

$$\begin{aligned} n(M)\, dM &= \frac{\langle \rho_M \rangle}{M} \frac{\partial F(> M)}{\partial M}\, dM\,, \\ &= \sqrt{\frac{2}{\pi}} \frac{\langle \rho_M \rangle}{M^2} \frac{\delta_c}{\sigma} \left| \frac{d\, ln\, \sigma}{d\, ln\, M} \right| \exp\left(-\frac{\delta_c^2}{2\sigma^2} \right) dM\,. \end{aligned} \tag{2.39}$$

2.2.5 The Bias of Density Tracers

The Dark Matter haloes, of which we have formulated a theoretical framework to predict their abundance using EPS-formalism, host luminous structures such as galaxies and clusters of galaxies. In the general paradigm of galaxy formation, galaxies form through cooling of baryonic matter in virialised DM haloes. One of the challenges in the observational cosmology is how well the luminous objects within virialised DM haloes trace the underlying matter density field. This is the so-called galaxy bias. An effective way to tackle this problem can be explained in two steps. First one relates the virialised DM haloes to the mass density field i.e. addresses the 'halo bias'. Then one specifies the relation between luminous tracers and their host haloes. The excursion set formalism allows one to address the first part of the problem and quantify the halo bias. In this framework one can specify the fraction of collapsed haloes with the smoothing mass scale S_0 at overdensity δ_0. As one compares this quantity with that of all collapsed haloes (see equation (2.33)), the corresponding mass fraction contained in haloes with $M_h > M$ is given as:

$$F(> M) = \mathrm{erfc}\left(\frac{\delta_c - \delta_0}{2\Delta S} \right) \tag{2.40}$$

where $\Delta S = S - S_0$. As overdensity grows with time, F increases, and with $\delta_0 \to \delta_c$ $F \to 1$ i.e the whole density field corresponding to smoothed scale S will be considered as collapsed halo. The fraction of mass in haloes with mass range M and dM and overdensity δ_0 is given by:

$$\begin{aligned} n(M|\delta_0, S_0) dM &= \frac{1}{\sqrt{2\pi}} \frac{\delta_c - \delta_0}{\Delta S} \left| \frac{dS}{dM} \right| \exp\left[-\frac{(\delta_c - \delta_0)^2}{2\Delta S} \right] dM\,, \\ &= \frac{M_0}{M} f(M|\delta_0, S_0) \left| \frac{dS}{dM} \right| dM\,, \end{aligned} \tag{2.41}$$

where M_0 is a mass corresponding to uncollapsed spherical region with radius R_0. We further define the density contrast for haloes of mass M in Lagrangian space:

$$\delta_h^L = \frac{n(M|\delta_0, S_0)}{n(M)V_0} - 1 \tag{2.42}$$

where $V_0 = (4\pi/3)R_0^3$. If the mass contained in the larger region is much larger than the mass of haloes, $M \ll M_0$ and $\delta_c \gg \delta_0$, then expression simplifies to:

$$\delta_h^L = \frac{\nu^2 - 1}{\delta_c}\,\delta_0 \tag{2.43}$$

where ν is defined in (2.33). This relation was obtained by Mo and White (Mo and White, 1996) who also mapped the initial Lagrangian space to Eulerian one providing a way to define 'halo bias' b_h as,

$$b_h \equiv 1 + \frac{\nu^2 - 1}{\delta_c} \Longrightarrow \delta_h = b_h \delta_0 \,. \tag{2.44}$$

where halo overdensity is proportional to the matter overdensity. Since ν increases with M, the larger haloes are more biased with respect to the matter overdensity distribution than the small ones. In figure 2.4, the 'halo bias' $b_h(\nu)$ is shown. In the plot we compare different halo bias models specified below with empirical bias measured from an N-body simulation. The simple EPS model is shown with a solid curve.

Bias of density tracers is sensitive to cosmological assumptions, spatial scales, redshift as well as halo mass function under consideration. Various improvements to the original EPS model have been proposed to predict halo bias with higher precision, using N-body simulations as calibration. The most popular are:

- Sheth-Tormen 1999: (Sheth and Tormen, 1999) [ST99] The authors propose an improved approximation of the original Mo & White model by including the effect of stochasticity and is given as following:

  $$b_h^{\mathrm{ST}}(M_1, z_1) = 1 + \frac{a\nu_1^2 - 1}{\delta_1} + \frac{2p/\delta_1}{1 + (a\nu_1^2)^p}\,, \tag{2.45}$$

  where $a = 0.707$ and $p = 0.3$ were obtained by fitting the results of an N-body simulation. In Fig. 2.4 this model is represented the dashed line;

- Sheth, Mo & Tormen 2001:(Sheth et al., 2001) [SMT01] further improved ST99 model by considering Ellipsoidal instead of spherical collapse model. The ellipsoidal approach matches better with simulations, and the resulting bias parameter is :

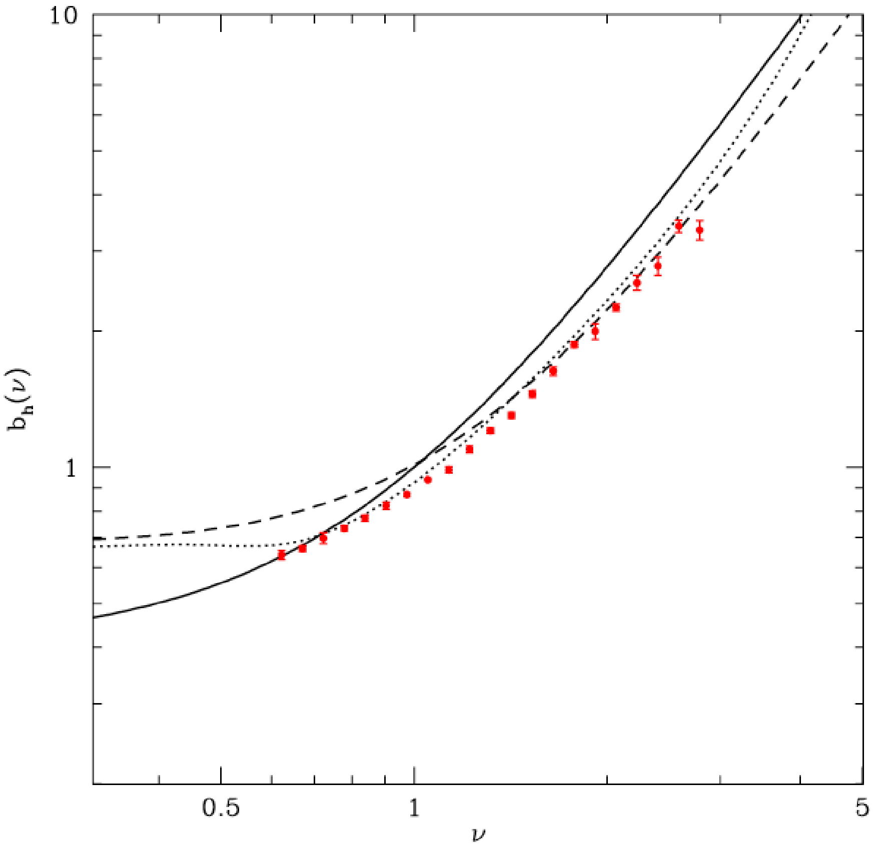


Fig. 2.4 The large scale halo bias as a function of ν. EPS-formalism prediction is depicted by solid line. Many approximated models, using fitting parameters obtained from N-body simulations, have been developed, of which the dashed line represents Sheth and Tormen (1999) and dotted line represents Seljak and Warren (2004) fits. The red datapoints are provided by N-body data. (Zentner, 2007)

$$b_h^{\mathrm{SMT}}(\nu_1) = 1 + \frac{1}{\sqrt{a}\delta_1}[\sqrt{a}(a\nu_1^2) + \sqrt{ab(a\nu_1^2)^{1-c}}] + \frac{-1}{\sqrt{a}\delta_1}\left[\frac{(a\nu_1^2)^c}{(a\nu_1^2)^c + b(1-c)(1-c/2)}\right], \tag{2.46}$$

where δ_1 is δ_c at redshift z_1, $\nu_1 = \delta_1/\sigma_1$, and $a = 0.707$, $b = 0.5$ and $c = 0.6$;

- Seljak & Warren 2004: (Seljak and Warren, 2004) SW04 proposed an improved analytical fit to the bias parameter measured from N-body simulations:

$$b_h^{\mathrm{SW}}(x) = 0.53 + 0.39x^{0.45} + \frac{0.13}{40x+1} + 5 \times 10^{-4}x^{1.5}, \tag{2.47}$$

where $x = M/M_{\mathrm{nl}}$ and M_{nl} is the mass within a shell with RMS amplitude is $\langle\delta^2\rangle = 1.69$. The x-range in which the approximation holds is $10^{-3} < x < 10^2$.

All the above expressions provide us with a recipe to relate haloes with the mass density field. All of them specify the bias of DM haloes with a mass M at redshift z. More often than not, haloes of different masses are used to trace the mass field. In this case one can define an effective bias by weighing the bias of a halo with a given mass by the fraction of haloes with the same mass. A typical case is that of a halo sample selected above a given minimum mass. In this case the effective bias parameter is:

$$b_h^{\mathrm{eff}} = \frac{\int_{M_{\mathrm{min}}}^{M_{\mathrm{max}}} b_h(M,z)\; n(M,z)}{\int_{M_{\mathrm{min}}}^{M_{\mathrm{max}}} n(M,z)}, \tag{2.48}$$

where $n(M,z)$ is the differential halo mass function at redshift z and the integral is evaluated within a range that encompasses all the halo in the sample selected above M_{min}. M_{max} is the largest halo of the sample.

2.2.6 Numerical Approach

So far we have seen extrapolation of linear theory through Zel'dovich Approximation and Spherical collapse model. Both of these approximations break down at certain stage of evolution of density perturbations. Therefore in order to assess the full non-linear evolution of density perturbations, one must rely on computational techniques such as N-body simulations.

A box, which is a fair sample of sufficiently large scale Universe i.e obeying cosmological principle, with periodic conditions, contains point mass particles representing Dark Matter (DM) particles. These particles then evolve by solving Newtonian equations of motion. The gravitational force acting on each particle is computed by summing over that of all particles. A particle is displaced using acceleration from their position at every time-interval Δt starting from their initial position. The finer the Δt, more accurate is the numerical solution of equation of motion.

The resultant force $\mathbf{F}_i$ acting on the particle is the time consuming part, where the computation time increases with number of particles and their mass. Some of the important techniques that are utilised for force computation, are introduced briefly as follows:

- Direct Summation (DS): The Newtonian gravitational pull acting on particle i from $N-1$ other particles is:

  $$\mathbf{F}_i = \sum_{j=1}^{N} \frac{m_j(\mathbf{x}_j - \mathbf{x}_i)}{(\epsilon^2 + |\mathbf{x}_i - \mathbf{x}_j|^2)^{3/2}}\,, \tag{2.49}$$

  where m is the mass of the particles situated at $\mathbf{x}_i$ and $\mathbf{x}_j$, and ϵ is softening length below which the acceleration force is coarse. This method was used heavily in seventies and eighties but alternative approaches were utilised later owing to its large CPU time which scales as $N(N-1)$ in this method.

- Particle Mesh (PM): A mesh-grid is formed and mass points are assigned to each grid point, which evolves with Poisson equation. Fast Fourier Transform (FFT) methods are used to recover gravitational potential on each particle. It is faster than DS as CPU time scales as $N \log N$.

- Particle^3-Mesh (P^{3}M): A fixed scale r_s is determined below which, the force computation is done by DS, and above which PM approach is used. This method as uses both approaches, can be used for high-resolution simulations.

- Tree codes: The mesh-grid is divided into hierarchical cells, where each cell exceeding one particle is divided into sub-cells and the division continues until a single particle is represented by each cell. This method is considered as effective as distant clumps of particles are treated as single massive particles and force computation takes place relatively faster in comparison with previous methods.

The output from simulations consists of particle positions and their velocities at a certain time snapshot. The initial conditions are set at some higher redshift z_{init}. As the system of particles evolves, one can take snapshots at different times, depending on the requirement. In Dark Matter only simulations, the numerical techniques described above are utilised. More complications occur with inclusion of gas dynamics of baryonic content, such simulations are referred as magneto-hydrodynamical (MHD) simulations.

In Chapter 4, I use simulated datasets from DM-only simulations, covering large cosmological volume. For such type of simulations, a large number of high-mass particles are used which ultimately provide coarse-grained cosmic structures, in contrary to very high resolution simulations that are aimed to study galaxy formation.

Chapter 3

Statistical Tools for Clustering Analysis

In the first two chapters we discussed the theoretical framework that leads to the formation of cosmic structures. In this chapter I focus on two-point statistics and their application to 2D and 3D cosmological density field. When observed, it can be seen that the density fields are traced by discrete objects. Using the distribution of these objects, clustering information can be estimated from Configuration space using reliable estimators of two-point correlation function, as well as from Fourier space using the Power spectrum estimates. However both the 3D statistical tools, for a 2D projected density field on the sky, are reformulated with spherical harmonic expansion of the density field. For observational datasets encoded with two-dimensional angular coordinates, angular clustering tools like angular power spectrum and angular correlation function are used. I discuss the estimators of two-point statistics, in relevance with our work involving 3D and 2D clustering analysis, along with their advantages and shortcomings in § 3.1 and § 3.2.

3.1 2-point Correlation Functions

3.1.1 Spatial 2-point Correlation Function (3D)

In the previous chapter, we have defined the two-point correlation function (2PCF) of a continuous density field $\delta(\mathbf{x})$ as its expectation value at two distinct positions separated by a distance vector $\mathbf{r}$ (Eq. (2.7)). However cosmological density fields are populated with discrete set of objects that trace, possibly with a biasing effect, the underlying matter density field. The interpretation of 2PCF in such case becomes as excess (or lack) of probability, compared to a uniform random distribution, of two objects separated by vector $\mathbf{r}$.

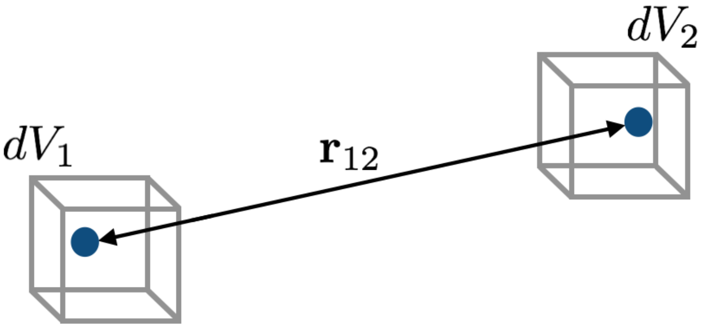


Fig. 3.1 The excess probability, with respect to a Poisson distribution, that a galaxy contained in a volume element dV_1 finds a neighbouring galaxy contained in a separate volume element dV_2 located at a distance $\mathbf{r}_{12}$ is determined by the two-point correlation function $\xi(r_{12})$.

This definition follows naturally when considering the joint differential probability dP of two objects in any two volume elements dV_1 and dV_2 as depicted in Fig. 3.1.

$$dP = \bar{n}_V^2 \,[1+\xi(r_{12})]\, dV_1 dV_2 \tag{3.1}$$

where $\bar{n}_V$ is the mean number density of the objects in sample volume V. For a statistically isotropic density field the two-point correlation function $\xi(r_{12})$ depends on the magnitude r_{12} of the separation vector $\mathbf{r}_{12}$. From Equation (3.6) it is clear that ξ represents a deviation from Poisson probability. If $\xi(r_{12}) > 0$ then it represents an excess probability of objects separated by r_{12}, i.e. objects are preferentially correlated (clustered) on the scale r_{12}. If $\xi(r_{12}) < 0$ (but > -1) then objects are preferentially uncorrelated, meaning that on these scales one is more likely to find underdense structures like voids. If the objects are chosen from the same datasets then $\xi(r_{12})$ interprets the auto-correlation, while if they are chosen from different datasets or of different types then a cross-correlation ξ^{ij} can be computed from pair of objects i and j from two distinct datasets.

In case of a specifically chosen direction, like observing objects along the line-of-sight by estimating their redshifts, the condition of isotropy as described above is violated and the measured 2PCF becomes anisotropic. In next chapter § 4.3 we use anisotropic 2PCF to perform redshift-space clustering analysis of clusters of galaxies.

It is not surprising that the estimate of the 2PCF of a discrete set of objects relies on pair counts, i.e. counting pair of objects separated by a distance r: $DD(r)$. Since real datasets are neither infinitely large nor ideally selected, one needs to correct for the combined effect of the geometry of the sample and its selection criteria that modulate the mean density of objects through the observed volume. This is typically done by comparing the counts of real objects pairs, with the pair counts of objects distributed in the same volume, with the same selection effects but without intrinsic clustering properties. Therefore all the proposed estimators of the 2PCF compare $DD(r)$ with the same counter performed in the so-called random sample, $RR(r)$.

One of the 2PCF estimators, that we shall use in this thesis, is proposed by Landy and Szalay (1993), that is often used for large scale clustering analysis. This estimator has the virtue of being both unbiased and minimum variance and expressed as:

$$\hat{\xi}_{\mathrm{LS}}(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} \,, \tag{3.2}$$

where DR corresponds to data-random galaxy pairs chosen simultaneously from data and random samples. To reduce shot noise error, which is driven by RR counts, the number of objects in the random catalogue is typically much higher as compared to the real data sample. In such case, the various pair counts in equation (3.2) need to be appropriately normalised by the total number of pair counts i.e. by multiplying RR with $(N_D/N_R)^2$ and DR with (N_D/N_R), where N_D and N_R are total number of galaxies in data and random samples, respectively. The selection function quantifies the combination of all observational effects and modulates the mean number density of objects. It is accounted for by the random catalogue, in which the same modulation as in the real data is induced, so that the selection effect is automatically corrected for, that cancels out in the ratio (3.2).

To quantify the error on the 2PCF estimate, one must have ensemble of independent density fields from which the clustering signal is measured. But we have only one Universe to observe, which is the so-called 'cosmic variance' problem, and this introduces a fundamental limit in the estimation of uncertainty. If the density field is Gaussian random field then the error can be modelled with Poisson distribution:

$$\sigma_{\xi} = \frac{1 + \xi(r)}{\sqrt{N_P(r)}} \,, \tag{3.3}$$

where $N_P(r)$ are number of independent pairs in a given bin of r. Since evolved cosmological fields are not Gaussian, the Poisson error estimate defined in Eq. (3.3) underestimates the actual error estimate. In such case, the errors and their covariance are best computed numerically by estimating the 2PCF from a large number of 'mock' catalogues that simulate the characteristics of the real density field. The shortcoming is that these estimates depend on the assumed fiducial cosmological model.

Mock catalogues can be either produced using numerical techniques (preferentially N-body simulations but we shall also describe in Chapter 6 alternative approximate methods based on assumed statistical properties of the underlying density field) or, if only the real dataset is available then one can obtain more realistic estimates on the error in $\xi(r)$, by dividing the data into a set of N similar subsamples and computing the covariance matrix:

$$\mathscr{C}_{ij} = \frac{1}{N-1} \sum_{k=1,}^{N} (\bar{\xi}_i - \xi_i^k)(\bar{\xi}_j - \xi_j^k) \,, \tag{3.4}$$

where ξ_i^k denotes the measurement of ξ at separation r_i from the k-th subsample. More sophisticated resampling approaches to obtain accurate error estimates can be utilised like bootstrap resampling and jackknife resampling.

3.1.2 Angular 2-Point Correlation Function (2D)

More often than not, cosmological fields are measured on the celestial sphere, rather than in the 3D space. This is the case, for example, of the cosmic microwave background but also of photometric surveys, for which the angular position and the flux of a set of discrete objects are measured but not their distance.

In these cases we can define the angular two-point correlation function (ACF) of the projected density field in analogy with (2.7):

$$w(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2) \equiv \langle \delta(\hat{\mathbf{r}}_1)\delta(\hat{\mathbf{r}}_2) \rangle = w(\theta) \,. \tag{3.5}$$

where $\delta(\hat{\mathbf{r}})$ is a realisation of 2D projected density field. In practice the observed density fields are populated with distinct set of objects, and the ACF can be reinterpreted using expression Eq. (3.6) as:

$$\delta P = \bar{n}_\Omega^2 \, [1 + w(\theta_{12})] \, \delta\Omega_1 \delta\Omega_2 \tag{3.6}$$

where θ is the angular separation between the two solid angle elements as shown in Figure 3.2, $\bar{n}_\Omega$ is mean number density of objects per unit solid angle element $d\Omega$.

Its unbiased, minimum variance estimator can also be defined in analogy with Eq. (3.2) as:

$$\hat{w}_{\mathrm{LS}}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)} \,, \tag{3.7}$$

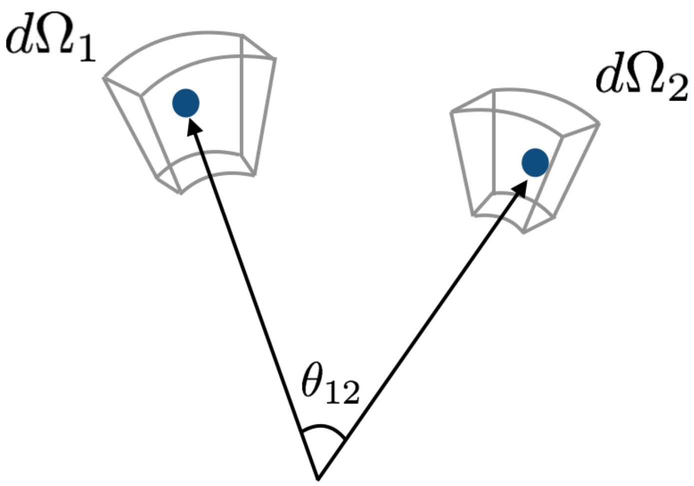


Fig. 3.2 Solid angle elements $d\Omega_1$ and $d\Omega_2$ are contained in 2D projected survey volume. The angular correlation function defines the excess probability that a pair of galaxies is enclosed in respective solid angle elements that is separated by angle θ_{12}.

with obvious meaning of the various symbols. The above discussed estimator is auto-angular correlation estimator, which we use while performing the clustering analysis of radio sources in Chapter 5. On the other hand in some cases, the estimated distances of objects have larger uncertainties, like in the case of galaxies in photometric redshift survey. Therefore it requires an optimal strategy to analyse two-point statistics, one of which is utilising a tomographic approach. In tomographic clustering analysis, the sample is sliced in a number of conveniently defined spherical shells. Then one measures both the auto-angular correlation function of the objects in each shell as well as the cross-angular correlation among objects in two different shells. I shall elaborate in detail on this approach in Chapter 6.

3.2 Power Spectra

3.2.1 Power Spectrum (3D)

As we have seen in the previous Chapter, the power spectrum is the Fourier transform of the 2PCF and, likewise, represents the expectation value of the transformed overdensity field in Eq. (2.5), at any two positions in Fourier space (See Eq. (2.9)). Unlike straightforward implementation as in the case of 2PCF, estimation of power spectrum requires more careful approach since it involves multiple correction and deconvolution steps before obtaining a 'true' estimate. To begin with, we shall consider a discrete density field, where N objects are sampled with a Poisson distribution, Eq. (2.3) takes the form:

$$\delta_N(\mathbf{x}) = \frac{n(\mathbf{x})}{\langle n \rangle} - 1 \,, \tag{3.8}$$

where the mean number density is specified by $\langle n \rangle$. The density contrast in discrete Fourier-space similar to Eq. (2.5) can be rewritten as:

$$\delta(\mathbf{k}) = \frac{1}{N} \sum \exp(i\mathbf{k}\cdot\mathbf{x}) \,. \tag{3.9}$$

The expression for power spectrum can then obtained by computing the variance of the $\delta(\mathbf{k})$ field. If the objects followed a random distribution, then Fourier coefficients $\delta(k)$ would originate from a random walk in a complex plane, with no intrinsic two-point correlation. The variance of $\delta(\mathbf{k})$ would be then:

$$\langle |\delta(\mathbf{k})|^2 \rangle = \frac{1}{N} \,. \tag{3.10}$$

However while dealing with the real data catalogues, one needs to account for the effects of the selection function which are absent in case of a uniform distribution of objects. The Fourier coefficients that are obtained from Eq. (3.9) are proportional to the transform of selection function $f(\mathbf{x})[1+\delta(\mathbf{x})]$ which needs to be subtracted from each coefficient:

$$\delta(\mathbf{k}) = \frac{1}{N} \sum \exp(i\mathbf{k}\cdot\mathbf{x}) - \frac{1}{N} \int \langle n_b \rangle \exp(i\mathbf{k}\cdot\mathbf{x}) \; d^3x \,, \tag{3.11}$$

where $\langle n_b \rangle$ is the number density of randomly distributed (background) objects. One can now subtract the spurious shot-noise from each **k**-mode to obtain the so-called FKP (Feldman et al., 1994) estimator of the Power spectrum:

$$\tilde{P}(\mathbf{k}) = \sum_k |\delta(\mathbf{k})|^2 - \frac{n_k}{N} \,, \tag{3.12}$$

where n_k is the number of **k**-modes associated with $|\mathbf{k}|$ involved in the sum. In the actual estimate of the power spectrum, one typically measures the power within a specified k-bin, i.e. a spherical shell in Fourier space with volume V_k. The angle averaged power spectrum in the bin is given by

$$\hat{P}(k) \equiv \frac{1}{V_k} \int_{V_k} d^3k' \tilde{P}(\mathbf{k}'). \tag{3.13}$$

where $\hat{P}(k)$ is the FKP estimator.

For the 3D Power spectrum the approach to the selection effect is the same as described above for 2PCF, i.e. one modulates the mean number density $\bar{n}_b(\mathbf{x})$ in (3.11) the same way as the real data. Geometry effects can be regarded as extreme selection effects, in which a fraction of the available volume does not contain objects because selection there has meant exclusion. In the estimation of the 2-point correlation function the geometry effects are accounted for by distributing random objects within the same volume as the real data. For the power spectrum, however, the effect of geometry, which is formally a product of the power spectrum by the volume window function in real space, becomes a convolution. So that if $F(\mathbf{x})$ is the window function that quantifies geometry effects, and it is defined as $F(\mathbf{x}) = 1$ within the observe volume and 0 outside, then the estimated power spectrum is:

$$\hat{P}(k) = P_{\mathrm{true}}(k) * |f(\mathbf{k})|^2 \,. \tag{3.14}$$

where $P_{\mathrm{true}}(k)$ is the power spectrum that would have been estimated in an infinite volume. $f(\mathbf{k})$ is the Fourier transform of $F(\mathbf{x})$ and * is the convolution product. In general $f(\mathbf{k})$ does not have an analytic form and needs to be estimated numerically. This is typically done by measuring the anisotropic power spectrum of a random population of objects distributed within the same volume as the real sample.

For mock catalogues extracted from N-body simulations, like those that will be used in this thesis, the cubic geometry and the use of periodic boundary conditions implies that no geometry effect will affect our power spectrum estimate and that the FKP estimator with method described as (3.11) can be safely used. In this case one can take advantage of the FFT technique to speed up the power spectrum measurement, subsequently making the estimation procedure less complicated for objects distributed in cubic volume. For such cases, using Fast Fourier Transform (FFT), one can estimate the power spectrum by distributing the particles in the box on a regular grid. I shall describe this procedure in details in § 4.3

3.2.2 Angular Power Spectrum (2D)

Theory

As discussed in § 3.1.2, many datasets, such as the photometric catalogues, specify the angular positions of objects along with their distance. In some other cases the distance can only be determined with very large errors, like in the photo-z catalogues. Since we shall use the angular power spectrum extensively in chapters 5 and 6, we describe in this section its properties in details. I will start with its definition and association with spatial power spectrum before defining the estimator. I also discuss the relevant observational effects related to partial sky coverage, geometry mask and slicing the redshift sky volume in spherical shells.

Let us define 3D spatial overdensity at any redshift, $\delta(\mathbf{r},z)$, when projected along the line-of-sight one obtains the surface overdensity[1] along that direction $\hat{r}$ as:

$$\delta(\hat{\mathbf{r}}) = \int_0^\infty \phi(\mathbf{r})\ \delta(\mathbf{r},z)\ r^2\ dr\ , \tag{3.15}$$

where $\phi(\mathbf{r})$ is a selection function and it accounts for all observational effects that modulate the number of observed objects. It is normalised such that $\int \phi(\mathbf{r}) r^2\ dr = 1$. The projected overdensity is defined over 2D spherical surface. It can therefore be expanded in spherical harmonics:

$$\delta(\hat{\mathbf{r}}) = \delta(\theta,\phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta,\phi)\ , \tag{3.16}$$

where θ and ϕ are angular coordinates that define the direction $\hat{\mathbf{r}}$ on the surface of the celestial sphere, $a_{\ell m}$ are spherical harmonic coefficients and $\mathrm{Y}_{\ell m}(\theta,\phi)$ are a complete orthonormal set of functions on the spherical surface similar to the plane-wave modes in flat 3D space. As their Fourier counter parts $\delta(\mathbf{k})$, the spherical harmonic coefficients are related to the surface overdensity field through:

$$a_{\ell m} = \int Y^*_{\ell m}(\hat{r})\delta(\hat{r}) d\Omega\ = \int Y^*_{\ell m}(\hat{r})\phi(\mathbf{r})\ \delta(\mathbf{r},z)\ d^3r \tag{3.17}$$

where for each multipole moment ℓ there are $2\ell+1$ degrees of freedom. The $a_{\ell m}$'s are complex numbers and satisfy:

$$\langle a^*_{\ell m} a_{\ell' m'} \rangle = C_\ell \delta^K_{\ell\ell'} \delta^K_{mm'}\ , \tag{3.18}$$

where δ^K_{ij} is the Kronecker symbol and the average is taken over ensemble of realisations of the density field. The quantity C_ℓ is the angular power spectrum (APS) and is defined as variance of harmonic coefficients analogous to the Fourier space definition expressed in Eq. (2.8):

$$C_\ell \equiv \langle |a_{\ell m}|^2 \rangle\ . \tag{3.19}$$

---

[1] In many references, the surface density is defined as $\sigma(\theta,\phi)$ in order to avoid confusion with the spatial overdensity. However I shall keep this notification as $\delta(\hat{\mathbf{r}})$ and $\delta(\theta,\phi)$ for spherical surface overdensity, $\delta(\mathbf{x})$ as spatial overdensity in real-space and $\delta(\mathbf{k})$ in Fourier space.

The quantity C_ℓ only depends on the multipole moment ℓ. However, since each realisation is identified by a different set of m-modes, one has to account for the intrinsic variance of C_ℓ, the so called '*cosmic variance*'. As in the 3D case, the cosmic variance can be analytically expressed if the field is Gaussian:

$$\left(\frac{\Delta C_\ell}{C_\ell}\right) \approx \sqrt{\frac{2}{2\ell+1}} \, . \tag{3.20}$$

Another way to express APS is also through Legendre transform of ACF, defined for a statistically homogeneous and isotropic density field by (see Eq. (3.5)),

$$C_\ell = 2\pi \int_{-1}^{1} d(\cos\theta)\, \mathscr{P}_\ell(\cos\theta)\; w(\theta)\,, \tag{3.21}$$

where $\mathscr{P}_\ell$ are Legendre polynomials. The Reyleigh expansion of plane waves in terms of spherical harmonics is given by:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{\ell m} i^\ell\, j_\ell(kr)\, Y_{\ell m}(\hat{\mathbf{r}})\, Y^*_{\ell m}(\hat{\mathbf{k}})\,, \tag{3.22}$$

where $\hat{\mathbf{r}}$ and $\hat{\mathbf{k}}$ are unit vectors in the direction $\mathbf{r}$ and $\mathbf{k}$, and j_ℓ is the spherical Bessel function. By redefining the Fourier transform of density contrast as described in Eq. (2.18) as a function of redshift $z(r)$ instead of cosmic time t as $\delta(\mathbf{k}, z(r))$, and from Eq. (3.17), the $a_{\ell m}$'s in Fourier space can be written as:

$$a_{\ell m} = 4\pi \int \mathrm{d}^3 r \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \delta(\mathbf{k}, z(r)) \sum_{\ell' m'} i^{\ell'}\, j_{\ell'}(kr) \phi(\mathbf{r})\, Y^*_{\ell m}(\hat{\mathbf{r}})\, Y_{\ell' m'}(\hat{\mathbf{r}})\, Y^*_{\ell' m'}(\hat{\mathbf{k}})\,, \tag{3.23}$$

which yields a direct relation with the 3D power spectrum using the definition from Eq. (3.21):

$$C_\ell = \frac{2}{\pi} \int dk\, k^2\, P(k) \left[\int_0^\infty dr\, r^2 \phi(\mathbf{r})\, j_\ell(kr) \right]^2 . \tag{3.24}$$

Above relation allows the comparison of measured quantity of APS with the theoretical prediction of 3D power spectrum.

Modelling APS for Observed Data

So far we have considered the case of a continuous field $\delta(\hat{\mathbf{r}})$. However, in our case the overdensity field is traced by discrete objects, and in the estimation of APS we will need to account for discreteness effects. One of which is Poisson noise, that adds another term to ΔC_ℓ in the (3.20):

$$\Delta C_\ell = \sqrt{\frac{2}{2\ell+1\ f_{\rm sky}}}\left(C_\ell + S_N\right)\ , \tag{3.25}$$

where the term $f_{\rm sky}$ is the fraction of observed sky and S_N is the added shot-noise term due to Poisson distribution of the objects and C_ℓ is the sample variance.

Moreover, photometric redshift surveys provide limited information on the radial position of the objects. In this case it is useful to divide the sample into radial shells and perform clustering analysis among objects within and across nearby shells. In this case we can rewrite the angular spectrum from Eq. (3.18) in more general form:

$$\tilde{C}_\ell^{ij} = \frac{1}{2\ell+1}\sum_m a^i_{\ell m}\, a^{j*}_{\ell m}, \tag{3.26}$$

where i and j indicates two shells, The case $i = j$ indicate the auto-angular spectrum of the shell and $i \neq j$ the cross-spectrum. In such case, one must account the Poisson noise for both shells:

$$\Delta C_\ell^{ij} = \sqrt{\frac{2}{2\ell+1\ f_{\rm sky}}}\left[\left(C_\ell^{ij}\right)^2 + \left(C_\ell^i + S_N^i\right)\left(C_\ell^j + S_N^j\right)\right]^{1/2}\ . \tag{3.27}$$

In most cases the selection function (3.15) can be approximated as the product of two functions: an angular mask, that accounts of the limited sky coverage of the sample, and a radial selection, that quantifies observational selection effects that depends on the objects' distances. In absence of angular dependent observational effects, the geometry mask can be expressed as a binary matrix with '0' values in those angular bins outside the survey area and '1' inside it. The selection function $\phi_i(\mathbf{r})$ then for i-th redshift shell can be expressed as:

$$\phi_i(\mathbf{r}) = \phi_i(r) M(\hat{r})\ . \tag{3.28}$$

where the normalisation of $\phi_i(r)$ to unity as described before is valid only for the redshift bin i. $M(\hat{r})$ is angular mask. Overall the selection function quantifies the probability whether a particular galaxy at $\mathbf{r}$ is included in the sample.

The radial selection is usually related to the luminosity of the objects in the observed band, L, and the fact that a survey is limited in flux, so that at any given radial distance r only objects brighter than $L_*(r)$ are included in the sample:

$$\phi(r) = \int_{L_*(r)}^{\infty} \phi(L)dL \Big/ \int_{0}^{\infty} \phi(L)dL \,. \tag{3.29}$$

As a consequence, only a fraction of objects enter in the survey. The probability of an objects at a distance r to be included in the survey can then be expressed as :

$$\phi(r) = \frac{1}{\mathcal{N}}\frac{dN}{dr} \,, \tag{3.30}$$

where $\mathcal{N}$ is the total number density of objects in the survey sample.

We have seen the impact of the selection function in determining the C_ℓ value which is quantified by Eq. (3.24). The impact of the geometry mask, on the other hand, is that of mixing of multipoles. If C_ℓ^{ij} is the angular auto- or cross- power spectrum measured over the all sky, the APS measured on a fraction of the 2-sphere characterised by an angular mask $M(r)$ is:

$$\langle a^i_{\ell m} a^{j*}_{\ell m} \rangle = \sum_{\ell'} \Gamma^m_{\ell\ell'} C^{ij}_{\ell} \,, \tag{3.31}$$

where the mixing of different ℓ-modes is specified by Γ and is related to the angular mask through:

$$\Gamma^m_{\ell\ell'} = \sum_{m'} \int \mathrm{d}\hat{\mathbf{r}}\, Y^*_{\ell m}(\hat{\mathbf{r}}) M(\hat{\mathbf{r}}) Y_{\ell' m'}(\hat{\mathbf{r}}) \int \mathrm{d}\hat{\mathbf{r}}'\, Y_{\ell m}\left(\hat{\mathbf{r}}'\right) M\left(\hat{\mathbf{r}}'\right) Y^*_{\ell' m'}\left(\hat{\mathbf{r}}'\right) \,, \tag{3.32}$$

which when averaged over m-modes, gives the angular mixing matrix, convolved with the theoretical angular power spectrum:

$$\tilde{C}^{ij}_{\ell} = \frac{1}{2\ell+1} \sum_m \langle a^i_{\ell m} a^{*j}_{\ell m} \rangle = \sum R_{\ell\ell'} C^{ij}_{\ell'} \,, \tag{3.33}$$

where $R_{\ell\ell'}$ is angular mixing matrix, which quantifies the effect of geometry mask on the true power spectrum C^{ij}_ℓ. The LHS of the above equation is the estimate of cross-APS, which is a convolution of the underlying theoretical angular power spectrum with angular mixing matrix $R_{\ell\ell'}$. The correlations among different ℓ-modes of C_ℓ that are induced by the partial sky coverage, are corrected by the mixing of orthonormal spherical harmonic coefficients as given in Eq. (3.32). The mixing matrix can be expressed in the form of $3j$-Wigner symbols as:

$$R_{\ell\ell'} = \frac{2\ell'+1}{4\pi} \sum_{\ell''} (2\ell''+1) W_{\ell''} \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2 \,, \tag{3.34}$$

where W_ℓ is the angular power spectrum of the mask:

$$W_\ell = \frac{1}{2\ell+1}\sum_m |M_{\ell m}|^2 \ . \tag{3.35}$$

As mentioned before, for a full-sky survey sample, the angular mask is $M(\hat{\mathbf{r}}) = 1$, $M_{\ell m} = \sqrt{4\pi}a_{\ell 0}a_{m0}$, mask power spectrum $W_\ell = 4\pi a_{\ell 0}$ and the mixing matrix $R_{\ell\ell'} = \delta^K_{\ell\ell'}$, and in that case:

$$\tilde{C}^{ij}_\ell = C^{ij}_\ell \ . \tag{3.36}$$

The measured auto or cross angular spectra need to be compared to theoretical models to extract cosmological information. In such model, that generalises (3.24), one needs to specify the power spectrum of the matter at different redshifts, its relation with that of galaxies, i.e. the bias function, the effect of the selection and that of the geometry mask. Let us ignore the geometry aspect that, as we have seen, can be included after the modelling of the all-sky angular spectrum, and assume that the radial selection is known. Also, let us use redshifts instead of radial distances. The relation between the two quantities depends on the cosmology, that is part of the C_ℓ model. All we need to specify is the 3D galaxy power spectrum at the generic epoch z. We assume that it is given by

$$P_{\rm gal}(k,z) = b^2(z)\ D^2(z)\ P_{\rm mat}(k,z=0)\ , \tag{3.37}$$

where $b(z)$ is the bias of the galaxy population at the redshift z and $D(z)$ is the linear growth factor of density perturbation introduced in (2.23). In the linear regime, one would then use linear theory to specify the mass power spectrum at the present epoch $P_{\rm mat}(k,z=0)$, and the model would be correct. In the non linear regime we can approximate $P_{\rm gal}(k,z)$ by modelling $P_{\rm mat}(k,z=0)$ as full non-linear matter power spectrum. The goodness of this approximation will need to be tested case by case, since non-linearities should be included also in the $b(z)$ model and in the growth factor $D(z)$. Also this expression ignores the correlation among the various k-modes.

That said, the cross angular power spectrum can be modelled as:

$$C^{ij}_\ell = \int_0^\infty k^2 P_{\rm mat}(k,z=0) F^i_\ell(k) F^j_\ell(k)\ dk\ , \tag{3.38}$$

where $F_\ell(k)$ is the radial power kernel in corresponding bin and quantified as:

$$F^i_\ell(k) = \sqrt{\frac{2}{\pi}}\int_0^\infty {\rm d}z \frac{{\rm d}N}{{\rm d}z} W_i(z) b(z) D(z) j_\ell(kz)\ , \tag{3.39}$$

which can be traced from Eq. (3.24). Here, I have substituted the radial selection function $\phi_i(z)$ as the product of the selection function of the full sample (see Eq. (3.30)) which is also a function of redshift z, and a window function $W_i(z)$ that characterises the distribution of the objects in the redshift bin.

When estimating Eq. (3.38), and depending on the characteristic properties of the sample under analysis, one can approximate the radial kernels under different limits of the selection function.

- *Thin Redshift Slicing:* In some cases, i.e. when dealing with spectroscopic rather than photometric surveys, slicing can be done using thin rather than thick shells. In this case one has,

  $$\frac{dN_i}{dz} \approx \frac{\delta^K(z-z')}{N^i_{\rm gal}} \,, \tag{3.40}$$

  where $N^i_{\rm gal}$ stands for number of galaxies in corresponding redshift slice i and δ^K is Kronecker delta. The kernel in this limit takes the form:

  $$F^i_\ell(k) = \sqrt{\frac{2}{\pi}}\frac{1}{N^i_{\rm gal}} W_i(z) b(z_i) D(z_i) j_\ell(kz_i)) \,, \tag{3.41}$$

  and the cross-angular power spectrum between two shells i and j can be defined as,

  $$C^{ij}_\ell = \frac{b(z_i)\,D(z_i)\,b\left(z_j\right)D\left(z_j\right)}{N^i_{\rm gal}N^j_{\rm gal}} \int_0^\infty k^2 P_{\rm mat}(k)\, j_\ell\left(kr\left(z_i\right)\right) j_\ell\left(kr\left(z_j\right)\right) {\rm d}k \,. \tag{3.42}$$

- *Limber's Approximation:* In general when one is interested in angular scales whose physical counterpart is smaller than the typical scale of the survey (the thickness of the shell), one can use the Limber approximation. In the limit of large multipole, i.e for $\ell \gg 1$, the double integral (3.38), becomes a 1D integral. The spherical Bessel function under asymptotic limit can be expressed as:

  $$\lim_{\ell\gg1} j_\ell(x) = \sqrt{\frac{\pi}{2\ell+1}}\delta^D(\ell + 1/2 - x) \,, \tag{3.43}$$

  where δ^D is Dirac delta. The C_ℓ under this approximation becomes:

$$C_\ell^{ij} = \int_0^\infty \mathrm{d}z \frac{dN_i}{dz}\frac{dN_j}{dz} b(z)^2 D(z)^2 \left(\frac{H(z)}{r^2(z)}\right) P_{\mathrm{mat}}\left(k = \frac{\ell+1/2}{r(z)}\right) , \tag{3.44}$$

and for thin slicing it further reduces to:

$$C_\ell^{ij} \approx \frac{\left(D\left(z_i\right) b\left(z_i\right)\right)^2}{N_{\mathrm{gal}}^i N_{\mathrm{gal}}^j} \left(\frac{H\left(z_i\right)}{r^2\left(z_i\right)}\right) P_{\mathrm{mat}}\left(k = \frac{\ell+1/2}{r\left(z_i\right)}\right) \delta^K\left(z_i - z_j\right) . \tag{3.45}$$

This means that for thin redshift shells, the approximation generates no correlation between different shells. The Limber's approximation is commonly used in weak-lensing analysis, when one is interested in small angular scales. In our case we are especially interested in large scales. For this reason in chapters 5 and 6 we shall use the complete expression (3.38) to model auto- and cross-angular power spectra. Also in chapter 6, we will explore the scenario when the cross-correlation in redshift shells is accounted for.

Pseudo-C_ℓ Estimator

To estimate the angular power spectra, i.e. the so-called Pseudo-C_ℓ of a distribution of objects in some patch of the sky, and within a given redshift range, we will make use of the `HEALPix` package (which is introduced in § 6.2.4), that offers the possibility, through the `MAP2ALM` routine to estimate the spherical harmonic coefficients, $a_{\ell m}$'s, of a distribution of discrete objects on a pixelised map as follows:

$$a_{\ell m}^{\hat{i}} = \Delta\Omega_p \sum_{k=1}^{N_{\mathrm{pix}}} \left(\frac{\mathcal{N}_{ik} - \bar{\mathcal{N}_i}}{\bar{\mathcal{N}_i}}\right) Y_{\ell m}^*(\hat{\Omega}) , \tag{3.46}$$

where $\Delta\Omega_p$ corresponds to pixel area in solid angle element, $\bar{\mathcal{N}_i}$ is the mean number density of objects i-th redshift shell and $\mathcal{N}_k$ is number of objects in k-th pixel. The cross-angular power spectrum can be estimated using a more generalised form of Eq. (3.26), after subtracting the shot-noise term and taking into the account the limited sky coverage:

$$\hat{K}_\ell^{ij} = \frac{1}{f_{\mathrm{sky}}(2\ell+1)} \sum_m |\hat{a}_{\ell m}^i \hat{a}_{\ell m}^{*j}| - \left(\frac{1}{\bar{\sigma}_i} \delta_{ij}^K\right) , \tag{3.47}$$

where f_{sky} is a fraction of sky in the survey under consideration and the last is shot noise. It is equal to the number density of galaxies per unit solid angle in i-th redshift bin,

$$\bar{\sigma}_i = \frac{N_{\mathrm{gal}}^i}{\Delta\Omega} , \tag{3.48}$$

The estimator described above in Eq. (3.47) is also called as pseudo-C_ℓ estimator. It was introduced by Peebles (1973) and used in clustering analyses performed by (Blake et al., 2011; Balaguera-Antolínez et al., 2018).

If the Distribution of discrete objects traces an underlying Gaussian overdensity field then the random error associated to the estimated auto-APS is (Dodelson, 2003):

$$\sigma_\ell^i = \sqrt{\frac{2}{(2\ell+1)\, f_{\rm sky}}}\,(C_\ell^i + S_N^i)\,, \tag{3.49}$$

where the quantities have the same meaning as defined in the beginning of this section in Eq. (3.25), also for the case of cross-APS estimator:

$$\sigma_\ell^{ij} = \sqrt{\frac{2}{(2\ell+1)\, f_{\rm sky}}}\left[(C_\ell^{ij})^2 + (C_\ell^i + S_N^i)(C_\ell^j + S_N^j)\right]^{1/2}. \tag{3.50}$$

However, the Gaussian hypothesis is not valid on small scales and at late evolutionary epochs. For this reason, when dealing with real catalogues, it is preferred to estimate errors and their covariance using realistic mock catalogues that mimic both the intrinsic clustering properties and the selection effects of the sample. In this case one simply measures the various angular spectra in each mock, $C_\ell^{(m,ij)}$, takes the average $\bar{\hat{C}}_\ell^{ij}$ and compute the covariance matrix as:

$$\mathbb{C}_{\ell\ell'}^{ij} = \frac{1}{N_m - 1}\sum_{m=1}^{N_m}\left(\hat{C}_\ell^{(m,ij)} - \bar{\hat{C}}_\ell^{ij}\right)\left(\hat{C}_{\ell'}^{(m,ij)} - \bar{\hat{C}}_{\ell'}^{ij}\right), \tag{3.51}$$

where N_m is number of mock catalogues and the summation is performed over total number of measured redshift bins.

Chapter 4

Probing Structure Growth from Clusters

Observing extragalactic objects through large redshift surveys is a very effective way to map the 3D mass distribution in the Universe at different epochs, and to probe the evolution of the large scale structure. Also we have seen in Chapter 2, the reliability of this mapping is intrinsically limited by the so-called biasing phenomenon, i.e. by the fact that luminous objects do not perfectly trace the mass. On top of galaxy bias, another effect, called redshift-space distortions [RSD] hampers this mapping. This distortion is introduced when measured redshifts are used as distance proxies without taking into account the presence of peculiar velocities that add up to the Hubble expansion. Unlike the bias, RSD pick up a preferential direction, the line-of-sight (LOS) to each object, and therefore destroy the statistical isotropy of the clustering properties. On the other hand RSD, being driven by peculiar velocities whose amplitude depends on the rate at which structures grow, contain precious cosmological information that can be retrieved once RSD are properly modelled and accounted for. For these reasons RSD are not anymore regarded as a nuisance effect. Instead, they have become one of the most powerful cosmological probe to investigate the origin of the accelerated expansion of the Universe (Guzzo et al., 2008).

Accurate theoretical model for RSD are required to extract cosmological information from observation. On those scales and at those epochs in which perturbations are small and linear theory is applicable, a simple RSD model has been proposed by Kaiser (1987) and Hamilton (1992). Unfortunately, the linear theory hardly applies to real samples for which nonlinear effects (reflecting both the complex dynamics and the nontrivial biasing phenomenon) as well as the ill-known impact of baryon physics cannot be neglected. Given these difficulties, the strategies to tackle RSD are typically of two types. The first one is to increase the complexity of the RSD model so that it can be safely applied to some existing dataset. The advantage of this approach is that it allows, in principle, to extract the maximum amount of information contained on those scales that are more affected by nonlinear effects. The price to pay is the introduction of a large number of parameters to characterise the model itself. The second approach, that I will follow in this chapter, is to select some special class of objects that are little affected by small scale dynamics and, instead, trace those large scale motions that can be described by the linear theory.

This second approach has been adopted by several authors to estimate the normalised growth rate of density fluctuations $f\sigma_8$, where f is the growth rate of structure as defined in Eq. (2.24) and σ_8 is the density variance Eq. (2.13) at 8 h^{-1} Mpc. Or, alternatively, to measure $\beta = f/b$, where b is the linear bias parameter of the tracers Eq. (2.44). The examples are Hatton and Cole (1999); Berlind et al. (2001); Chuang and Wang (2012); Tocchini-Valentini et al. (2012); Bianchi et al. (2012); Marulli et al. (2017).

In this chapter I will focus on galaxy clusters. The motivation for this choice is twofold. The first one is related to the clusters' scientific properties. Galaxy clusters, being the largest, virialised structures in the Universe are rare but very prominent and easily detectable out to very large distance. Because of these properties, they are almost ideal tracers of large-scale, coherent motions, that are well described by linear theory. Also, since they form in correspondence of the highest density peaks, they are also highly biased tracers of the underlying mass distribution. Because of these properties, they have been proposed as promising probes of RSD (Bianchi et al., 2012; Marulli et al., 2017) and of Baryonic Acoustic Oscillations (Veropalumbo et al., 2014). The second motivation is related to the data availability. New, dedicated surveys are being or will be soon carried out that will generate very large, homogeneous catalogues of clusters that will probe the large scale structure of the universe over very large volumes. This is the case of the upcoming *Euclid* survey (Laureijs et al., 2011), of the planned W-FIRST mission (Doré et al., 2018) and, which is more relevant for this thesis, the ongoing *eROSITA* survey (Merloni et al., 2012). The latter survey is designed to detect about 10^5 galaxy clusters with masses ranging from $5 \times 10^{13} h^{-1}$ $M_{\odot}$ and onward in the X-ray band out to redshift $z = 0.5$, most of which will have their spectroscopic redshift determined by a dedicated follow-up 4-MOST survey (Salvato and Merloni, 2015).

In this chapter I will investigate the possibility of determining the growth rate of density fluctuations using *eROSITA* clusters and assess the accuracy of this measurement. I will do that by analysing the RSD in the 2-point statistics measured on a set of mock galaxy catalogues extracted by the `zHORIZON` N-body experiment (Smith, 2009). The outline of the chapter is as follows: In § 4.1 I shall revise some of the theoretical concepts that are relevant to RSD. In § 4.2, the simulated datasets and the mass cuts for galaxy clusters are discussed. I provide the statistical tools required to analyse the RSD § 4.3 and finally I discuss the results of the RSD cluster analysis in § 4.4.

4.1 Theoretical Background

4.1.1 Redshift Space Distortions

Thanks to the expansion of the Universe, also known as the Hubble flow of matter, we can use the observed redshift as a distance proxy for any extragalactic objects emitting light. If the cosmic density tracers would follow pure Hubble flow, the observed redshift would provide a reliable measurement of their distances. However, as we have seen in chapter 2, the growth of density fluctuations generate peculiar velocities that add to the Hubble recession velocities. As a consequence, when we use redshift as distance estimator to map the three dimensional distribution of the mass in the Universe, we obtained a distorted representation of it. This is the so-called redshift-space distortion phenomenon. Let us lay out the theory that describe RSD and define the distance that we obtain from the redshift measurement, **s**. In comoving coordinates its relation with the real comoving position **x** is

$$\mathbf{s} = \mathbf{x} + (\mathbf{v}_{\rm pec} \cdot \hat{z})\hat{z}\,, \tag{4.1}$$

where $\mathbf{v}_{\rm pec}$ the is normalised peculiar velocity (divided by factor aH) and $\hat{z}$ is the LOS direction. Eq. (4.1) clarifies that the amplitude of the RSD is proportional to the typical magnitude of the velocity vector which, in linear theory, is proportional to the growth factor. If peculiar velocities were completely uncorrelated, then RSD would only introduce random error in the estimated velocities. This is the case for the so-called Fingers of God (FoG), caused by random motion of virialised satellites hosted in large haloes. However, since these velocities are instead highly correlated, they systematically modify the relative distances between objects and, as a consequence, their clustering statistics. To quantify the impact of RSD on clustering statistics it is convenient to decompose position vectors into parallel and perpendicular components along the line of sight. So that the parallel component of (4.1) is

$$s_{\parallel} = \mathbf{r} \cdot \hat{z} = r_{\parallel} + w_{\parallel}\,, \tag{4.2}$$

where w is the LOS component of the peculiar velocity, i.e $w_{\parallel} = |\mathbf{v}_{\rm pec} - \mathbf{v}'_{\rm pec}| \cdot \hat{z}$, and the perpendicular component which is unaffected by line-of-sight distortions,

$$s_\perp = r_\perp \, , \tag{4.3}$$

coincides with the real-space one.

It is convenient to follow the same approach to study two point statistics. This means expressing the separation vector **s** into parallel and perpendicular component to the LOS to the pair. So that, the 2-point correlation function (Eq 2.7) in redshift-space can be conveniently expressed as $\xi^S(\mathbf{s}_\perp, \mathbf{s}_\parallel)$, which is anisotropic between LOS-parallel and LOS-transverse directions unlike its real-space counterpart $\xi(\mathbf{r})$. It should be noticed that the LOS to the pair is not uniquely defined, as different choices are possible. This is why it is sometimes convenient to adopt, whenever possible, the distance observer approximation in which the LOS to each pair is the same and coincides with one of the Cartesian axis.

Figure 4.1 illustrates the collapsing overdensity shells as observed from a distance (left pane) and for an observer being a part of collapsing shell (right pane). Large scale inflows are coherent, where peculiar motions are extremely small as compared to the size of the overdensity shell, make it appear compressed along the direction of the observer. As the shell collapses further, the overdensity arrives at a turnaround phase where peculiar motions and Hubble flow cancel out. On small scales the random motions of objects becomes dominant and overdensity appears elongated along the direction of observer, creating non-linear effect of Fingers-of-God (FoG).

Since we observe luminous tracers of the underlying mass particles, one needs to account for the biasing relation to recover the cosmological information, i.e $\delta = b\,\delta_m$, where δ_m is mass overdensity. Since we are interested in galaxy bias, that traces the mass on large scales, we can adopt deterministic, linear bias relation in Eq. (2.44). With this assumption, the magnitude of the RSD depends on the so-called β parameter that is simply the ratio between linear growth rate and the linear bias parameter:

$$\beta = \frac{f(\Omega_m)}{b} \, . \tag{4.4}$$

The exact relation between f and Ω_m depends on the underlying cosmological model and gravity theory (refer to Eq. (2.25)). In the ΛCDM model, it is assumed that scale-independent exponent $\gamma = 0.545$ (Linder, 2005, 2017), which does not hold for other cosmological models. The approximated ratio described above in Equation (4.4) is valid only on large scales where non-linear effects are non relevant.

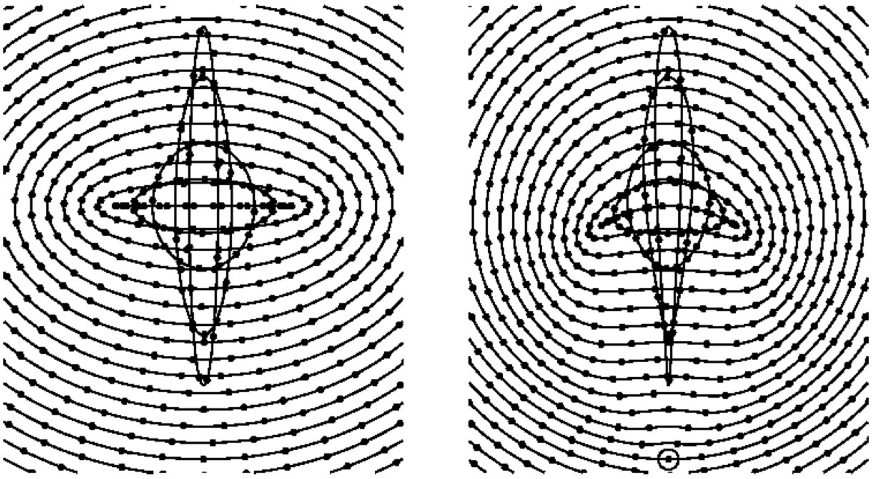


Fig. 4.1 The spherical overdensity shells observed in redshift space produce distorted patterns changing with respect to scales. The connected dots are the objects collapsing towards the centre positioned on the same overdensity shell. In the left pane, the observer is far away from the structure, i.e. somewhere below the figure, looking towards the overdensity centre. From large scales to small scales the geometry of shells changes from a squashed ellipsoid, collapses at the turnaround, to the elongated finger of god. Moreover, in the right pane the observer is a part of a collapsing shell (encircled dot). The elongated structures pointing towards him become sharp and at large separations kidney shaped distortions appear. (Kaiser, 1987; Hamilton, 1997)

4.1.2 Kaiser-Hamilton Formalism

In linear perturbation theory Kaiser (1987) has proposed an analytic model to describe the effect of RSD on 2-point statistics in Fourier space, i.e. on the power spectrum. In the distant observer limit the anisotropic power spectrum of a tracer can be expressed as:

$$P^S(k,\mu_k) = (1+\beta\,\mu_k^2)^2 P(k), \tag{4.5}$$

where k is the modulus of the wavenumber vector $\mathbf{k}$, μ_k is the cosine angle between $\mathbf{k}$ and the LOS direction, which is the same for all pairs in the sample, and β is the RSD amplitude defined in Eq. (4.4). Eq. (4.5) shows that the effect of RSDs (i.e. when $\beta \neq 0$) is that of distorting the iso-power contours, i.e. the curves at $P^S(k,\mu_k) = \text{constant}$, that would be circular if peculiar velocities could be ignored. Distortions in the iso-power contours are conveniently expressed by expanding $P^S(k,\mu_k)$ in Legendre polynomials:

$$P_\ell(k) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu_k\, \mathscr{P}_\ell(\mu_k)\, P^S(k,\mu_k). \tag{4.6}$$

where $\mathscr{P}_\ell(\mu)$ are Legendre polynomials. Because for the axial symmetry, for linear RSD, only the multipoles $\ell = 0, 2$ and 4 (i.e. monopole, quadrupole and hexadecapole) are non-zero, therefore the anisotropic power spectrum can be expressed as:

$$P^S(k,\mu_k) = P_0(k)\mathscr{P}_0(\mu_k) + P_2(k)\mathscr{P}_2(\mu_k) + P_4(k)\mathscr{P}_4(\mu_k), \tag{4.7}$$

where the spectral moments are related to the undistorted tracer real-space power spectrum $P^R(k)$ through:

$$P_0(k) = \left(1 + \frac{2\beta}{3} + \frac{\beta^2}{5}\right) P^R(k), \tag{4.8}$$

$$P_2(k) = \left(\frac{4\beta}{3} + \frac{4\beta^2}{7}\right) P^R(k), \tag{4.9}$$

$$P_4(k) = \frac{8\beta^2}{35} P^R(k). \tag{4.10}$$

This Legendre decomposition suggests a practical strategy to estimate β: measure $P^S(k,\mu_k)$ from the galaxy distribution, compute its first three moments and compare the results to theoretical predictions to obtain β from Eqs. (4.8), (4.9) and (4.10).

However, since real-space power spectrum is not directly measurable quantity, this approach requires assuming a cosmological model for $P^R(k)$. An alternative, more practical strategy is to consider multipole ratio, to eliminate the $P^R(k)$ dependence. A popular choice is to consider the so-called Q-ratio between quadrupole and monopole:

$$Q(k) \equiv \frac{P_2(k)}{P_0(k)} = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2}\,. \tag{4.11}$$

As long as the mass perturbations and bias are linear, this ratio is scale-independent and nearly model independent. At the present epoch, or at moderate redshift, this is true only on large scales. An effective way to define the scale above which linear theory can be safely applied is therefore that of measuring Q at increasing wavenumbers and search from deviations from its asymptotic value at small k.

Hamilton (1992) worked out an analogous formalism to characterise and quantify RSD in configuration space, i.e. using the anisotropic 2-point correlation function (2PCF) $\xi^S(s,\mu_s)$, where s is the modulus of the pair separation vector $\mathbf{s}$ and μ_s is the cosine angle between this vector and the LOS. As for Eq. 4.7, one can expand the anisotropic 2PCF in its first three moments:

$$\xi^S(s,\mu_s) = \xi_0(s)\mathscr{P}_0(\mu_s) + \xi_2(s)\mathscr{P}_2(\mu_s) + \xi_4(s)\mathscr{P}_4(\mu_s) \,, \tag{4.12}$$

where the modulus s is $\sqrt{s_\parallel^2 + s_\perp^2}$ and $s_\parallel, s_\perp$ are the components of the s vector along and across the LOS, and $\mu_s = s_\parallel / s$. The 2PCF moments are:

$$\xi_\ell(s) = \frac{2\ell+1}{2} \int_{-1}^{+1} \xi^S(s,\mu_s)\, \mathscr{P}_\ell(\mu_s)\, d\mu_s \,. \tag{4.13}$$

The relations between the moments of the 2PCF in redshift space and the real space 2PCF are analogous to those for the power spectrum monopole but involves integrals over scale r:

$$\xi_0(s) = \left(1 + \frac{2\beta}{3} + \frac{\beta^2}{5}\right)\xi^R(r) \,, \tag{4.14}$$

$$\xi_2(s) = \left(\frac{4\beta}{3} + \frac{4\beta^2}{7}\right)[\xi^R(r) - \bar{\xi}(r)] \,, \tag{4.15}$$

$$\xi_4(s) = \frac{8\beta^2}{35}\left[\xi^R(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)\right] \,, \tag{4.16}$$

where $\bar{\xi}(r)$ and $\bar{\bar{\xi}}(r)$ are spherically averaged correlation functions and are defined as,

$$\bar{\xi}(r) = \frac{3}{r^3}\int_0^r \xi^R(r')r'^2 dr' \,, \tag{4.17}$$

$$\bar{\bar{\xi}}(r) = \frac{5}{r^5}\int_0^r \xi^R(r')r'^4 dr' \,. \tag{4.18}$$

As in the Fourier case, one can work out the analogous of the Q-factor

$$Q(s) \equiv \frac{\xi_2(s)}{\xi_0(s) - (3/s^3)\int_0^s \xi_0(s')s'^2 ds'} = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2}. \tag{4.19}$$

which again, as in the previous case, depends neither on the true 2PCF nor on the separation s.

The information that can be extracted from a RSD analysis in Fourier space is the same as in configuration space, which basically consists of determining the growth rate in this case. However, there is a benefit in performing both analyses as these are affected by different types of random and systematic uncertainties. For example, if the geometry of the survey is simple and the typical scales involved in the analysis are close to linear, then the Fourier space covariance matrix is expected to be closer to diagonal and the uncertainties a combination of Gaussian errors and Poisson noise. This is not true in configuration space, due to the covariance of the correlation signal among different separation bins. On the contrary, when the geometry is not simple, the convolution by the geometry mask mixes k-modes resulting in a highly non diagonal covariance matrix.

Therefore, since power spectrum and 2PCF analyses are potentially affected by different types of systematic errors, we shall perform here both analysis, i.e. we will analyse the redshift space distortions in both $\xi^S(s,\mu_s)$ and $P^S(k,\mu_k)$, focusing on those scales that are large enough for linear theory to be applied safely. Indeed, as we have stated before, one of the goals of our analysis will be that of determining what are the correct scales to apply the linear RSD model in the upcoming cluster surveys like *eROSITA*.

4.1.3 Linear Bias Parameter

The reason to re-introduce the linear bias parameter for the analysis done in this chapter, is importantly to underline that the halo bias b_h introduced in Eq. (2.44) is different from linear bias parameter b. In principle, the bias parameter is scale dependent but for this work the parameter is treated equivalent to large scale galaxy bias (scale-independent). This assumption is justified as the analysis is only focused on very large spatial case. Also, throughout this work, DM haloes taken from the catalogues described in § 4.2 are considered to be unbiased representatives of *(eROSITA)* galaxy clusters. Now, according to the definition of two-point statistics in Fourier and configuration space (Equation (2.7) and (2.8)), the linear bias parameter b yields as square root of the ratio of halo distribution with that of dark matter:

$$b_h^P = \sqrt{\frac{P^R_{\rm hh}(k)}{P_{\rm mm}(k)}}; \qquad b_h^\xi = \sqrt{\frac{\xi^R_{\rm hh}(\mathbf{r})}{\xi_{\rm mm}(\mathbf{r})}}\,, \tag{4.20}$$

where $P^R_{\rm hh}(k)$ and $P_{\rm mm}(k)$ are real-space halo and matter auto power spectrum. On the other hand, $\xi^R_{\rm hh}(\mathbf{r})$ and $\xi_{\rm mm}(\mathbf{r})$ are real-space halo and matter auto correlation functions respectively.

4.2 Halo Catalogues

Galaxy clusters are less abundant tracers of matter density as compared to galaxies and groups of galaxies, since they form by clustering of galaxies and groups in late stages of structure evolution. Being rare objects (their total number in the currently available catalogues is a few thousands) distributed over very large volumes, the main source of uncertainties when studying 2-point statistics is sparse sampling. Also, the systematic errors in modelling their bias, that however, are expected to be rather small, being well described within the framework of EPS formalism in § 2.2.4. I will ignore the impact of observational systematics in selecting cluster samples since these are highly survey-dependent while our goal is to assess the potential of cluster as RSD probes.

In order to assess the impact of all these error sources we will perform our analysis using controlled N-body experiments from which mock galaxy cluster samples with well known properties can be extracted from. In particular we will use the Zürich Horizon simulations (`zHORIZON`, Smith, 2009). The initial conditions in these simulations are set to WMAP vanilla flat ΛCDM model; $\{\Omega_{\rm CDM} = 0.25$, $\Omega_b = 0.04$, $\Omega_\Lambda = 0.75$, $h = 0.7$, $n_s = 1.0$ and $\sigma_8 = 0.8\,\}$. DM haloes have been identified in all simulation outputs using a standard `FoF` algorithm with linking length $b = 0.2$. We ignore the presence of substructures in our analysis. Haloes with $\mathrm{M_{halo}} \gtrsim 10^{14}\ h^{-1}\mathrm{M}_\odot$ are grouped into catalogues in which each object is characterised by 3D position and velocity vector of its centre of mass. These are the 160 catalogues that I use in our analyses.

For each simulation run, I consider three different outputs: at redshifts $z = 0, 0.5$ and 1.0, with an average number of clusters equalling to 8.17, 3.46 and 1.05 (in $\times 10^4$ counts) respectively. As expected, the mean number density of mock clusters decreases with the redshift, from 2.421 to 0.311 (in $10^{-5}\ h^3\ \mathrm{Mpc}^{-3}$). The total number of clusters at all three snapshots is counts to ~127,000 (averaged over 160 realisations each with length $L = 1.5\ h^{-1}$ Gpc).

To quantify the number density of our clusters with the characteristics of the upcoming *eROSITA* cluster catalogue, I compare the statistics provided by *eROSITA* forecast paper: Pillepich et al. (2012). According to Fig. 3 and Table 4 of the referred paper, the survey is expected to observe 9.32×10^4 objects with masses above $5 \times 10^{13}\ h^{-1}\mathrm{M}_\odot$ and a median redshift $z \sim 0.35$. For objects with masses $\mathrm{M_{halo}} \gtrsim 10^{14}\ h^{-1}\mathrm{M}_\odot$, the corresponding expected number is 5.57×10^4 at median redshift of $z = 0.46$. This number in optimistic case (Deeper *eROSITA* in Table 4. of Pillepich et al. (2012)) will reach up to $\sim 9\times10^4$ which is comparable with our $z = 0$ dataset, whereas for lower estimate (Focused *eROSITA*) it decreases to $\sim 4\times10^4$, which of the order of $z = 0.5$ case. In both the cases number of clusters in our data at $z = 0$ and $z = 0.5$ are 10-12% underestimated than the survey.

The outputs at higher redshift i.e. at $z = 1.0$, even with very few massive haloes at hand, are analysed to compare results and look for a sweet spot where to best perform RSD analyses using clusters, keeping in mind that current and future cluster catalogues ,i.e. from the *Euclid* or the *LSST* surveys, will allow us to span the broader redshift range $z \sim [0.01, 2]$.

4.3 Methodology: Statistical Tools

4.3.1 Power Spectrum Estimator

Since this analysis is based on a mock galaxy catalogue extracted from a cubic box with periodic boundary conditions, the finite distance to the sample is less of a concern and wide angle effects can be ignored (that could be minimised using different power spectrum estimators), and the monopole, the quadrupole and the hexadecapole moments of the power spectrum can be measured. The hexadecapole moments are ignored since its measurement turned out to be very noisy.

The FKP estimator has been briefly described in § 3.2.1. More details can be found in the original work. What is relevant for this analysis is its application. First of all, as the interest here is the number density field, and not in the mass density field, I assume that all objects have the same mass. This field is interpolated from the objects distribution on a 512^3 cubic grid using a CIC scheme, over the same box as the of the `zHORIZON` simulations. This choice automatically sets the fundamental frequency $k_{\rm min} \equiv k_F = 2\pi/L = 0.0149\ h$ Mpc^{-1} whereas the maximum wavenumber is set by the Nyquist frequency $k_{\rm max} \equiv k_N = \pi/H = 1.072\, h$ Mpc^{-1}. The anisotropic power spectrum $P^S(k,\mu_k)$ is then estimated in 64×64 linearly spaced bin in the range $[k_{\rm min}, k_{\rm max}]$ and $\mu_k = [0,1]$. Since I am interested in sampling linear scales, I will restrict this analysis to $k < 0.1$, well below the wavenumbers potentially affected by aliasing effect that however, as we have checked, is already small at $k_{\rm max}$.

The detailed procedure to measure the anisotropic power spectrum of the 'clusters' from the `zHORIZON` halo catalogues is outlined below:

1. Clusters from the `zHORIZON` boxes are distributed onto a regular grid with added weights of the 'observable' quantity that describes the density field (in this case we assign unit mass weights). For distributing the particles one can use various schemes, one of them is a standard CIC scheme in which the linear interpolation is performed using a top hat window function (Eq. (2.12))

$$W(x) = \frac{1}{H}\begin{cases} 1 - \dfrac{|x|}{H} & \text{if } |x| < H; \\ 0 & \text{otherwise,} \end{cases} \tag{4.21}$$

   where H is the grid spacing;

2. I compute the halo number density contrast at each grid point as described in Eq. (3.8). The number density field $\delta_N(\mathbf{x})$ is then Fourier transformed and the Fourier coefficients $\delta(\mathbf{k})$ of the number density field are computed via FFT;

3. The power spectrum is computed using the scheme as described in Eq. (3.12) by subtracting the shot noise term. The obtained term $P(k)$ is however monopole of the power spectrum. The dipole and quadrupole can also be computed from the power spectrum by integration along the cosine angle, as indicated in Eq. 4.6. The numeric integration is performed in above defined linear bins between $k = [k_{\text{min}}, k_{\text{max}}]$ and $\mu_k = [0,1]$;

4. Since we are also interested in RSD anisotropic signature, when we estimate the power spectrum we split each **k** mode into its components parallel and perpendicular to the line-of-sight (that is the case coincides with the z-axis of the computational box), $k_\parallel$, $k_\perp$, and compute $P^S(k_\parallel, k_\perp)$.

4.3.2 2-point Correlation Function Estimator

To measure the anisotropic 2PCF of DM halos and its moments we use the unbiased, minimum variance LS estimator defined in Eq. (3.2). For this we need to create a random catalogue of objects, i.e. a set of fake objects distributed in the same cubic box that has periodic boundary conditions same as data box, with intrinsic clustering properties and affected by the same observational biases. Since we are not considering observational systematics in our analysis we simply generate a population of random objects with about 10^6 number of objects, to reduce the impact of shot noise error. However, since we are interested in measuring RSD, we decompose the separation vector into its component parallel and perpendicular to the LOS (i.e the z-axis) and count halo-halo (HH), random-random (RR) and mixed pairs (HR) accordingly. The LS estimator is then given by:

$$\xi^S(s_\perp, s_\parallel) = \frac{\mathrm{HH}(s_\perp, s_\parallel) - 2\mathrm{HR}(s_\perp, s_\parallel) + \mathrm{RR}(s_\perp, s_\parallel)}{\mathrm{RR}(s_\perp, s_\parallel)}, \tag{4.22}$$

where $s_\perp$ and $s_\parallel$ are separations in redshift-space along and across the LOS. Or, if we instead consider the cosine angle (μ_s) and the modulus of the separation vector:

$$\xi^S(s, \mu_s) = \frac{\mathrm{HH}(s, \mu_s) - 2\mathrm{HR}(s, \mu_s) + \mathrm{RR}(s, \mu_s)}{\mathrm{RR}(s, \mu_s)}. \tag{4.23}$$

Both $\xi^S(s,\mu_s)$ and $\xi^S(s_\perp,s_\parallel)$ are estimated in 50×50 linear bins in the ranges $s = [0,200]\ h^{-1}$ Mpc and $\mu=[0,1]$. Finally, we integrate numerically Eq. 4.13 to estimate the 2PCF monopole and quadrupole moments that we use to evaluate the $Q(s)$ value from Eq.4.19 in 50, equally spaced linear bins.

4.3.3 Covariance Matrices

Instead of relying on jackknife or bootstrap resampling methods that manipulate the available dataset, we take advantage of the 160 available mock catalogue and estimate the covariance numerically, from the scatter among the measurements in the various mocks. We point out that, in our analysis, the reference dataset will coincide with the mean 2-point statistics, i.e. our main goal it to assess the magnitude of the systematic errors induced when using linear theory to analyse 2-point cluster statistics.

The precision of our numerically-estimated covariance matrix is set by the number of available independent mock catalogues. As a rule of thumb, a % precision in the diagonal element of a N×N covariance matrix requires N^2 mocks. We sue this criterion to define upper limits to the number of bins in our covariance matrix, that we set equal to $12<\sqrt{160}$. With the additional consideration that we want to focus on linear, or quasi-linear scales, we compute the covariance matrices in the following ranges and binning: For the Quadrupole ratio in configuration space $Q(s)$ we consider 12 bins in the range $[50,200]\,h^{-1}$ Mpc. In Fourier space the covariance matrix for $Q(k)$ is estimated in 12 linear bins in the range $k=[0.02,0.12]$. In both cases the measured statistics, performed using a finer binning, are resampled to match the coarser binning.

With this in mind the $Q(s)$ and $Q(k)$ covariance matrices are computed using Eq. (3.4), in which now the estimated Q-ratio is used instead of ξ:

$$\mathscr{C}_{ij} = \frac{1}{N-1}\sum_{k=1,}^{N}(\bar{Q}_i - Q_i^k)(\bar{Q}_j - Q_j^k)\,, \tag{4.24}$$

were the i and j indicate the indexes of the 12 bins, k identifies one of the $N=160$ mock catalogues, Q_i^k represents the quadrupole value and $\bar(Q)_i$ its mean over the mocks.

In Figs. 4.2 and 4.3 we show the estimated covariance matrices of $Q(k)$ and $Q(s)$, respectively, normalised to their diagonal values:

$$C_{ij}^{\mathrm{Norm}} = \frac{C_{ij}}{\sqrt{C_{ii}\,C_{jj}}}\,, \tag{4.25}$$

where C_{ii} and C_{jj} are diagonal covariance matrix elements of indices ii and jj. The colour code in the plot is set according to the amplitude of the covariance matrix elements,

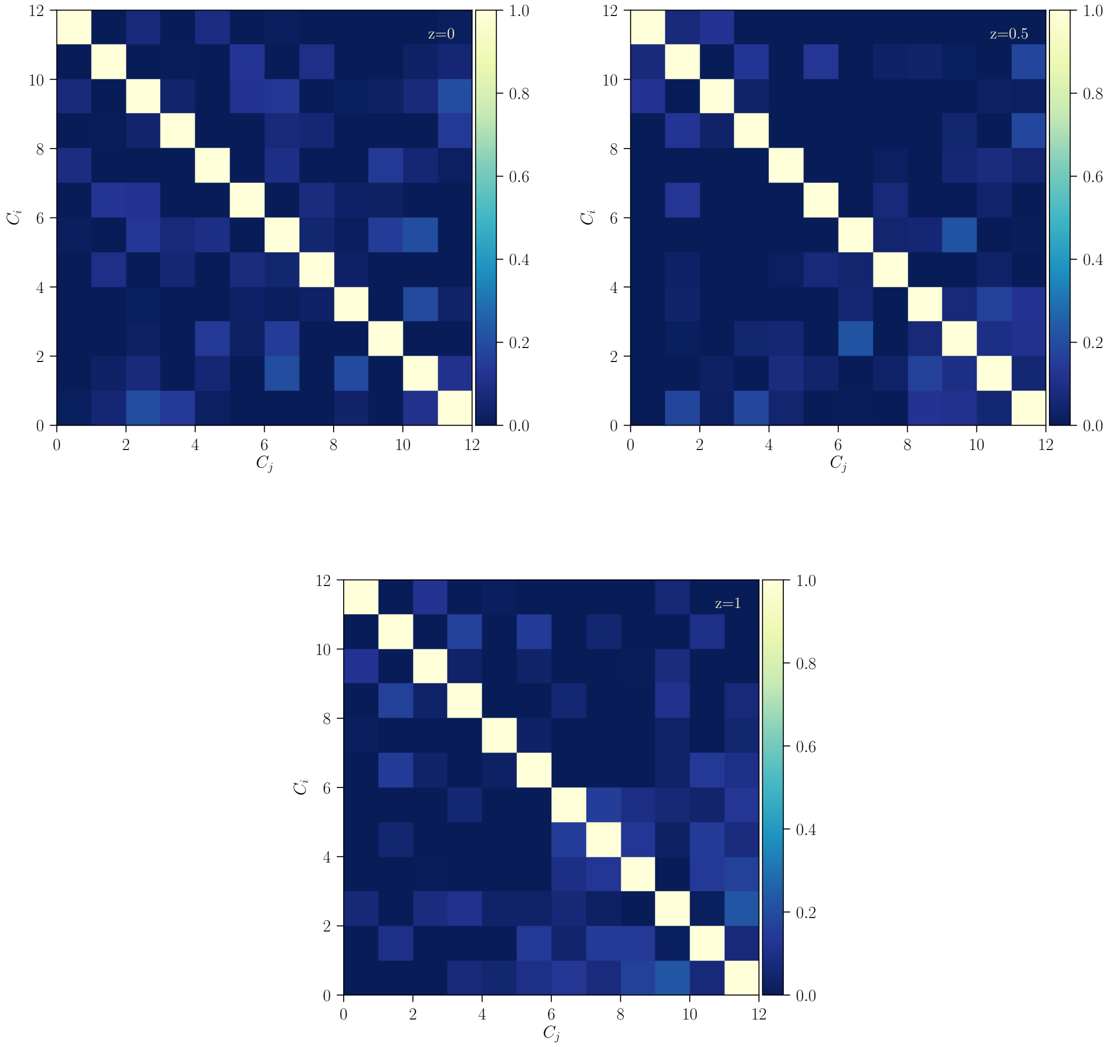


Fig. 4.2 Normalised covariance matrix for the quadrupole ratio $Q(k)$ at three different redshifts z=0 (top left), z=0.5 (top right) and z=1 (bottom) to right. The amplitude of the various terms are colour-coded from 0 to unity (the value of the diagonal elements, in white) as shown in the colour bar.

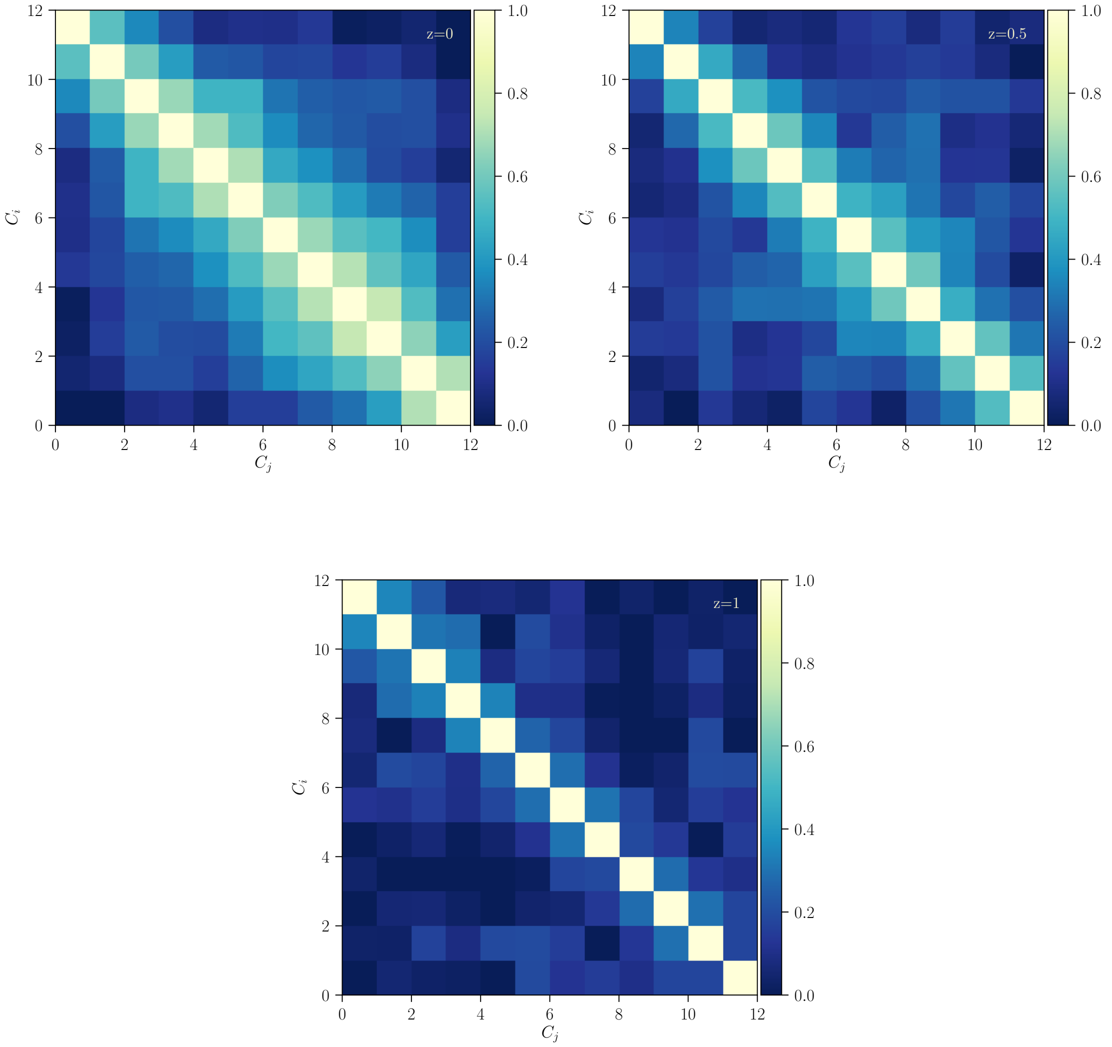


Fig. 4.3 Normalised covariance matrix for the quadrupole ratio $Q(s)$ at three different redshifts z=0 (top left), z=0.5 (top right) and z=1 (bottom). The amplitude of the various terms are colour-coded from 0 to unity (the value of the diagonal elements, in white) as shown in the colour bar.

as shown by the reference vertical bar. In Fourier space (Figs. 4.2) it is remarkably diagonal, as expected, up to the largest waveneumbers and at all redshifts considered in this analysis. This is quite reassuring since diagonal elements are estimated more precisely than the off-diagonal one. In configuration space (Fig. 4.3) the contribution from off-diagonal elements is comparatively larger as this is expected in configuration space (Grieb et al., 2016). This is particularly evident at $z = 0$ (left panel).

4.4 Results

4.4.1 Anisotropic Cluster Power Spectrum

As described in § 4.3.1, the anisotropic power spectrum separated in LOS parallel and transverse Fourier modes is obtained. As Fig. 4.4 shows the iso-power contours of the anisotropic halo power spectrum $P(k_{\parallel}, k_{\perp})$ at three different redshifts ($z = 0$, $z = 0.5$ and $z = 1$ from left to right). The plotted power spectrum is the mean among the 160 measured spectra in the mocks. In each panel the colour code represents the amplitude of the power spectrum, as indicated in the colour bar. The squashing of the contours is evident at all redshifts and reflects the impact of the redshift space distortions.

It can be seen that the distortion pattern is qualitatively similar to that expected if only linear RSD were present. Only in the $z = 0.5$ and $z = 1$ cases we see a negligible hint of FoG-like distortions (diverging contours) at large k values. Moving from the left to the middle panel, we notice that the colours grow darker, indicating that the power increases with the redshift. In fact, this is also true when moving from $z = 0.5$ to $z = 1$ and the reason why this is not evident from the contours is that the colour bar used in the right panel is different.

The models used in this section are computed using Kaiser approximation stated in Eqs. (4.8) and (4.9). The β in this case is obtained using Eq. (4.4) with linear growth rate and linear bias parameter b_h^P, which I have obtained using Eq. (4.20), i.e. the ratio of real-space power spectrum $P^R(k)$ to the matter power spectrum. In Table 4.1, different values of b and β used in the theoretical predictions are shown. The real-space power spectrum is measured from halo catalogues without considering the LOS axis. I use matter auto power spectrum at all 3 redshifts predicted with `CAMB` (Lewis et al., 2000) assuming the same cosmology as that of simulations used in this analysis.

The left and right panel of figure Fig. 4.5 compare the mean monopole $P_0(k)$ and quadrupole $P_2(k)$ of halo auto power spectrum measured from the catalogues at three different redshifts (shown with 1σ scatter in light shaded colours), with predictions from

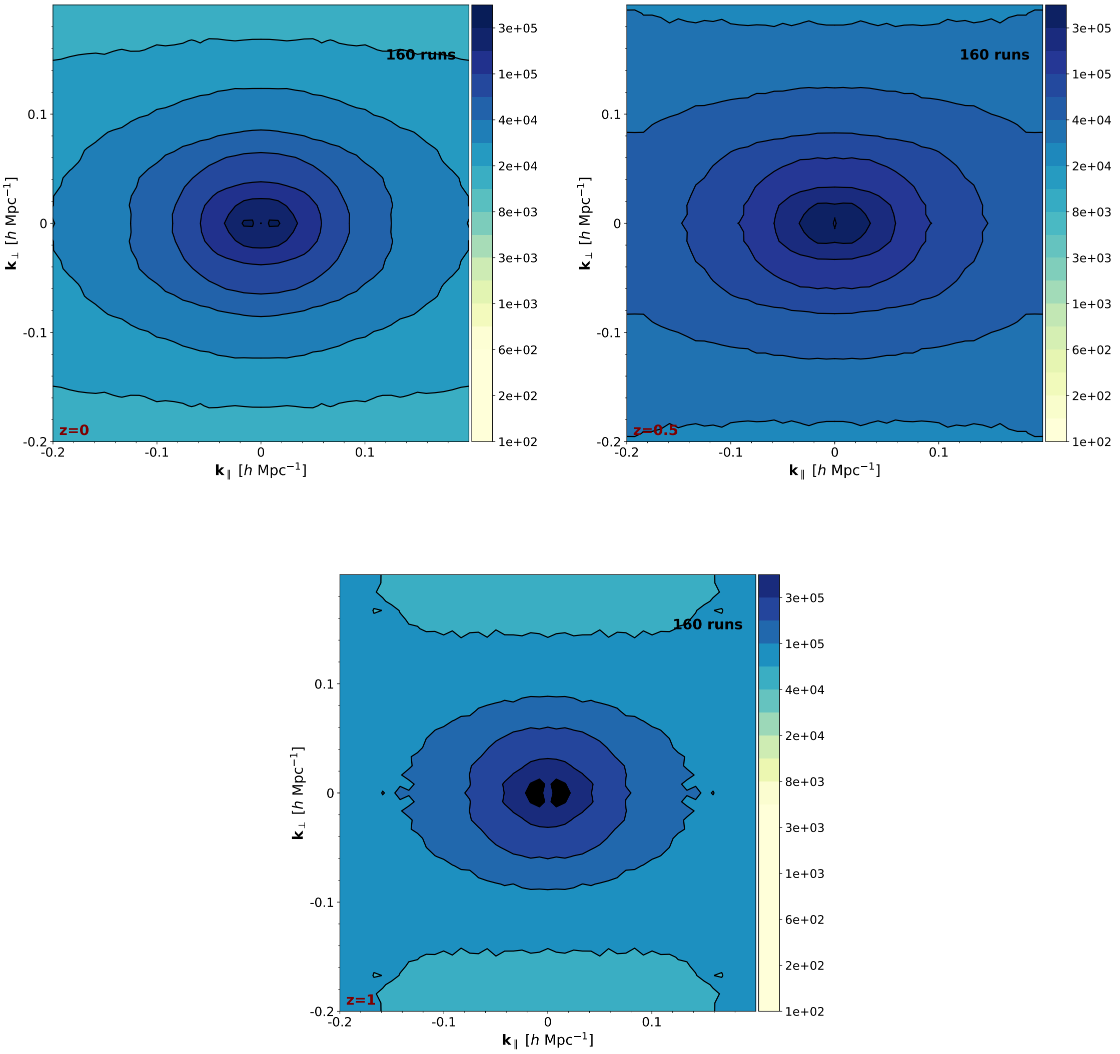


Fig. 4.4 Iso-power contours of the anisotropic halo power spectrum in the $(k_\parallel, k_\perp)$ plane at three different redshifts $z = 0$ (top left), $z = 0.5$ (top-right) and $z = 1$ (bottom). Colour code shown in the vertical bar. The power spectrum shown here is the average among the 160 mocks.

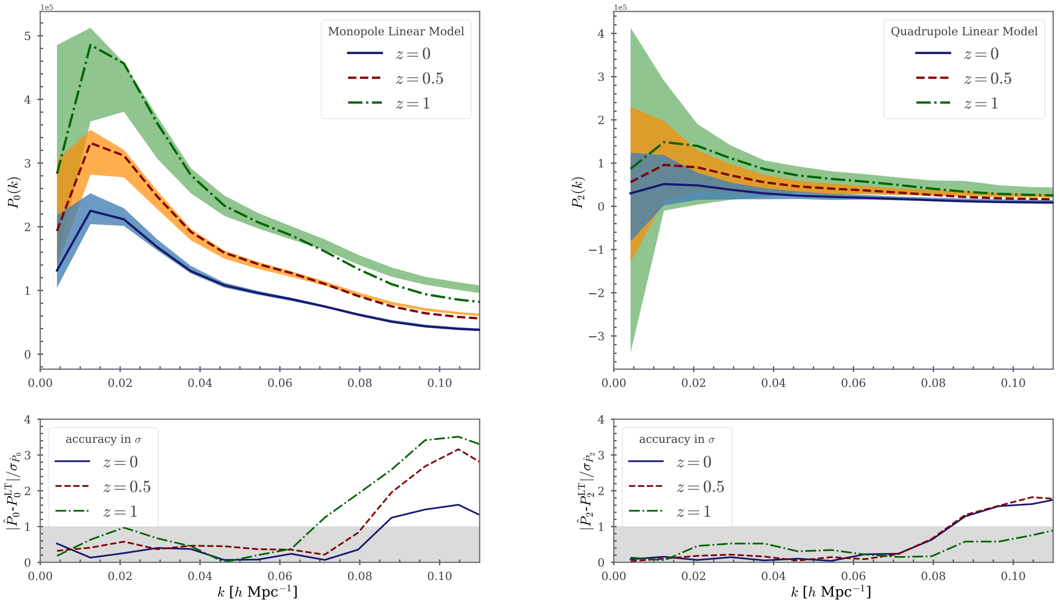


Fig. 4.5 Monopole (top left) and quadrupole (top right) halo power spectra at three redshifts. The shaded regions in light-colours are for measurements and errorbars, measured as standard deviation estimated from 160 catalogues. The light-blue, yellow and light-green represent $z = 0$, $z = 0.5$ and $z = 1$, where the dark coloured line curves are linear theory predictions at respective redshifts as indicated in the labels. In the bottom panels we show the mean systematic difference between measured and theoretical moments in units of 1σ random errors.

Table 4.1 Different halo biases and corresponding linear bias parameters used in our analysis as a function of the redshift. Linear bias in Eq. (4.20) used for obtaining theoretical prediction of β values shown in column 3 from Fourier space b_h^P and in column 5 from configuration space b_h^ξ. Both bias parameters are taken as average values only on very large spatial scales, i.e. for $s = [50, 200]$ h^{-1} Mpc and for $k = [0.01, 0.12]$ h Mpc^{-1}. In last two columns, linear theory prediction of b_h^{MW} using Mo & White (Mo and White, 1996) bias model (Eq. (2.44)) and corresponding β^{MW} is shown.

Redshift	b_h^P	β_{lin}^P	b_h^ξ	β_{lin}^ξ	b_h^{MW}	β^{MW}
0.0	2.639	0.177	2.796	0.168	2.623	0.179
0.5	3.980	0.229	4.162	0.219	3.968	0.230
1.0	6.130	0.238	6.455	0.226	5.970	0.244

linear theory (dark line curves) as described above. In bottom left panel of Fig. 4.5, 1σ accuracy between monopole model and measurement is shown. It can be noticed, the monopole at redshifts $z = 0, 0.5$ is well within 1σ accuracy of the linear prediction on scales $k = [0.01, 0.08]$ h Mpc^{-1}. The agreement is expected to increase with the redshift, but for $z = 1$ the model appears to deviate on smaller k's than other two cases, which is opposite to what is expected. One reason for this can be the linear bias parameter is not modelled appropriately at those scales, However in case of quadrupole (bottom right panel), this feature is not seen. In fact all three redshifts are in agreement with predicted quantities using linear theory. The larger spread seen on lower k-scales is cosmic variance.

These results indicate that, when considering cluster-sized objects, linear theory can be applied up to $\sim k = 0.08$ h Mpc^{-1} in the full redshift range $z = [0, 1]$. We shall then limit our χ^2 analysis to the interval $k = [0.01, 0.08]$ h Mpc^{-1}, where the lower k cut is introduced to discard spurious error-dominated bins. For estimating Q-ratio and, subsequently the parameter of interest β, we will use these scales where the measurements are in agreement with linear theory prediction.

4.4.2 Anisotropic Cluster 2-point Correlation Function

The anisotropic 2-point halo correlation function $\xi(s_{||}, s_\perp)$ is shown in Fig. 4.6. In analogy with Fig. 4.4 the three panels illustrate the iso-correlation contours at three different redshifts $z = 0$, $z = 0.5$ and $z = 1$ (from left to right). Contours squashing is clearly visible and qualitatively consistent with linear redshift distortions. A hint of non-linear effect is seen at small $s_\perp$ values at $z = 0.5$ and $z = 1$ in the form of contour elongation along the line of sight (the finger of god feature). Correlation amplitude, colour coded according to the colour bars, increases from $z = 0$ to $z = 0.5$ and, less so, to to $z = 1$, in agreement with the power spectrum results.

The measured monopole and quadrupole moments of the 2-point correlation function, both multiplied by s^2 are shown in the upper panels of Fig. 4.7. The light-coloured shaded region corresponds to measurements and 1σ error from 160 realisations. Here, only the selected range is shown $s = [50, 200]$ h^{-1} Mpc which is same as that considered for the covariance matrix analysis. In the monopole (top left panel), we clearly see the BAO peak feature at ~ 110 h^{-1} Mpc at all three redshifts. Below the BAO scale the amplitude of the correlation function increases with the redshifts, whereas on large scales we observe the opposite behaviour as ξ crosses 0 around 140 $h^{-1}Mpc$. For the normalised quadrupole (top right panel) the amplitude monotonically increases (decreases for the raw moment) the redshift at all separations. It should be noted that, the normalised quadrupole shown here is $(-\xi_2)$ which is already naturally negative. Once again, the significance of both the monopole and quadrupole trends with the redshift is larger than it appears in the plots in which we show the error on the single realisation (i.e. the 1σ scatter among the mocks) instead of the errors on the mean.

The dark coloured curves superimposed to the shaded measurements represent the Kaiser linear theory predictions for the monopole and quadrupole moments (Eqs. 4.14 and 4.15,respectively). For the theoretical prediction, same method as Fourier space analysis is followed, only by calculating the right side term of Eq. 4.20. Similar to the power spectrum case, to appreciate the adequacy of linear theory when estimating β from the anisotropic correlation function of a cluster sample, we plot in the bottom panels the systematic error, quantified by the difference between the measured and the expected correlation, in units of the random error, estimated as the 1σ scatter σ among the 160 realisations. The results show that systematic errors are typically within 1σ except that in correspondence of the BAO peak (for the monopole) and at separations smaller than ~ 100 h^{-1} Mpc for the quadrupole. Finally, systematic errors decrease with the redshift, showing that, as expected, linear approximation becomes progressively more precise when moving to higher redshifts.

4.4.3 Measurements of β

From the measured quadrupole over monopole ratios in Fourier and configuration space $Q(k)$ and $Q(s)$ respectively, the corresponding β^P and β^ξ values are estimated by minimising the $\chi^2(\beta)$ function:

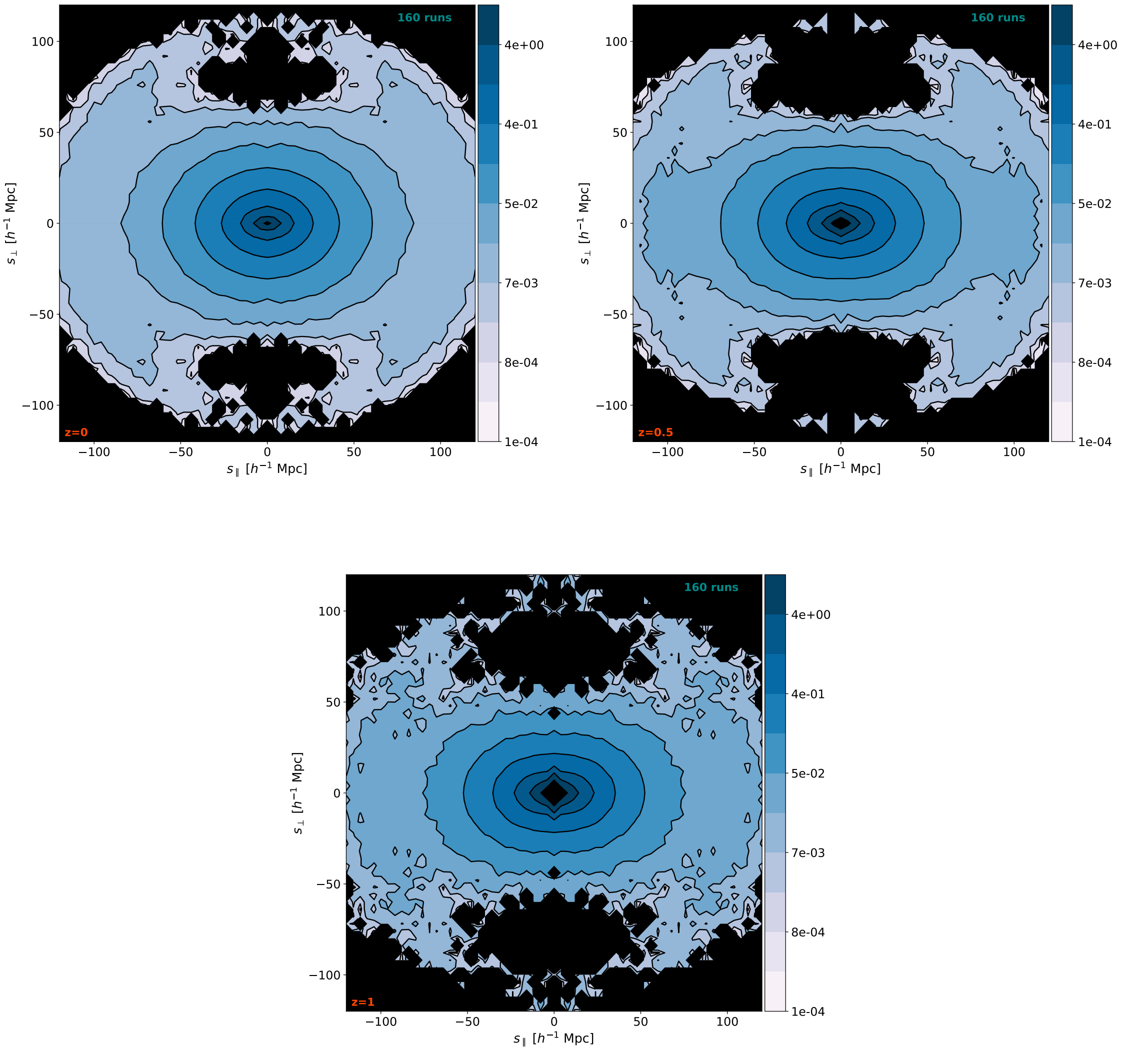


Fig. 4.6 Iso-correlation contours of the two-point correlation function $\xi(s_{||}, s_\perp)$ measured at $z = 0$ (top left), $z = 0.5$ (top right) and $z = 1$ (bottom). The correlation is the mean among those measured in the 160 realisations. The colour-bars on the right set amplitude of the correlation contours.

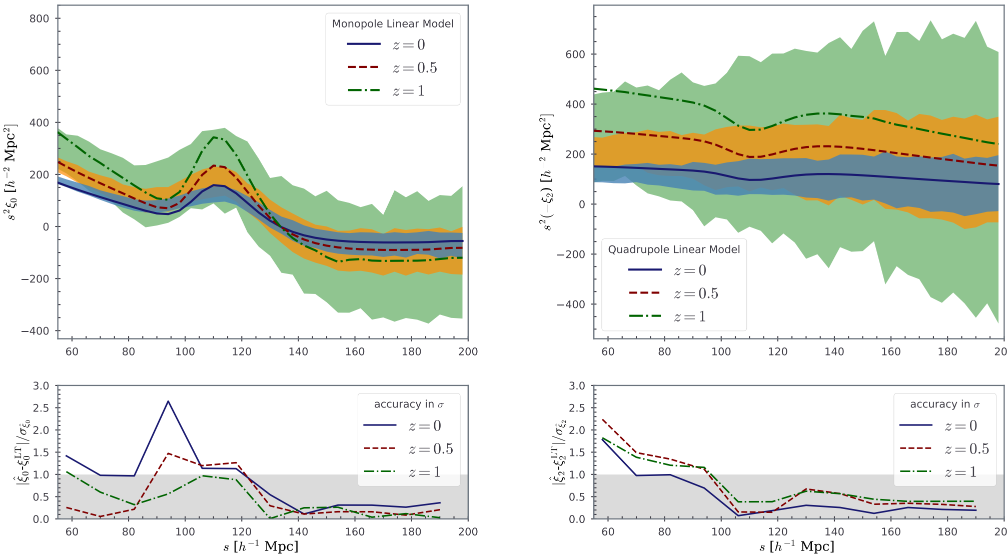


Fig. 4.7 Monopole (Top left) and quadrupole (top right) of the 2PCF in the separation range $s = [50, 200]$ h^{-1} Mpc. Different light-coloured shaded regions are for measured values plus the 1σ scatter at different redshift, starting from light-blue, yellow and light-green for redshifts $z = 0, 0.5$ and 1 respectively. The darker line curves represent linear theory predictions. In the bottom plots we show the mean systematic difference between measured and theoretical moments in units of 1σ random errors.

Table 4.2 Estimated β values and their 1 σ errors in Fourier and configuration space at three redshifts. The mean number density (per box) of halos with masses above $10^{14}\ h^{-1}\mathrm{M}_\odot$ in the 160 `zHORIZON` simulations is also provided

Redshift	$\hat{\beta}^P \pm \sigma_\beta$	$\hat{\beta}^\xi \pm \sigma_\beta$	$\bar{n}_h$ per $(1.5)^3$ Gpc^3
0.0	0.188 ± 0.021	0.171 ± 0.037	$81,700$
0.5	0.236 ± 0.023	0.217 ± 0.043	$34,800$
1.0	0.211 ± 0.038	0.197 ± 0.072	$10,500$

$$\chi^2 \equiv \sum_{i,j=1}^{N} [\hat{Q} - Q(\beta)]_i\, \Psi_{i,j}\, [\hat{Q} - Q(\beta)]_j\,, \tag{4.26}$$

where i and j run over the $N = 12$ bins, $\hat{Q}$ indicates the estimated value of either $Q(k)$ or $Q(s)$ and $Q(\beta)$ the corresponding theoretical Kaiser linear models Eqs. 4.11 or 4.19. $\Psi_{i,j}$ is the precision matrix obtained by inverting the covariance matrix described previously.

The results, listed in Table 4.2, are obtained by considering the quantities measured in the ranges $k = [0.01, 0.12]\, h\ \mathrm{Mpc}^{-1}$ and $s = [50, 200]\, h^{-1}$ Mpc binned in 12 linear intervals. The results of the two analyses are in agreement with linear prediction, as expected. This indicates that systematic errors in either the measurement or in the theoretical model do not dominate the random ones, neither in configuration nor in redshift space. A second, interesting, result is that uncertainties are about 50% smaller when the analysis is performed in Fourier space. This difference, which is mostly analysis dependent (Sanchez et al., 2017), reflects in this case, the larger error covariance in configuration space, evident in Fig. 4.3.

It is important not to over emphasise the importance of this result since the analysis is performed using a periodic data cube, whereas real surveys have complex geometry that substantially modify the structure of the error covariance as well as their amplitude.

It is nevertheless reassuring that β estimates in Fourier space agree with those in configuration space within 1σ uncertainties.

To further investigate the presence of systematic errors it is instructive to compare the values of $\hat{\beta}^P$ and $\hat{\beta}^\xi$ estimated using theoretical linear bias models as well with the expected from the ratio Eq. (4.20) as listed in Table 4.1. In the last column of the table, Mo & White (MW, Mo and White, 1996) model bias is shown, which I will use here for comparison test. For this, ideally one would use the weighted average of bias, by summing over mass functions of the haloes present in the data, i.e. Eq. (5.2). As the goal of this analysis is to estimate β as a precursor of linear growth rate as well as validate the linear theory assumption for large scale clustering analyses using galaxy clusters, instead of the effective β, I will use here the prediction from i.e. Eq. (2.44), using cluster mass cut-off ($10^{14}h^{-1}$ $\mathrm{M}_{\odot}$) that is used in this analysis, which is predicts the β^{MW} at fairly comparable values to those measured in this analysis. I shall leave the case with β^{eff} obtained using effective bias b_{eff} for future analysis.

Top panels of Fig. 4.8 (Fourier space) and Fig.4.9 (configuration space) show the results from the comparison tests. In the plots the datapoints and errorbars show the estimated $\hat{\beta}^{P}(k)$ and $\hat{\beta}^{\xi}(s)$ values in 12 bins from 160 realisations. The horizontal line shows the MW bias model β^{MW} value from Table 4.2. The coloured regions indicate linear theory prediction β^{P}_{lin} and $\beta^{\xi}_{\mathrm{lin}}$ that are shown in Table 4.2, with the scatter among the realisations and quantifies the theoretical uncertainties induced by the assumption of the linear halo model and perturbation theory. In configuration space the sparse sampling at $z=1$ with low number density of clusters has resulted in enlarged errorbars, which can be seen in the figure.

In the bottom panel of Figs. 4.8 and 4.9, the accuracy of compared models with measured estimates can be seen. The results in both configuration and Fourier space show that, measured beta values agree well with theoretical expectations, with the possible exception of the interval $k \geq 0.8$ h Mpc^{-1} in Fourier space. The mismatch is more significant at low redshifts, suggesting that this is a manifestation of nonlinear effects at small scales. To assess the impact of this effect we have repeated the χ^2 minimisation using Eq. 4.26, considering the smaller interval $k=[0.01, 0.08)h$ Mpc^{-1}. The resulting β values are fully consistent with those in Table 4.2, indicating that our results are robust to nonlinear effects over the full $k-$ range explored. Also in all 6 cases shown in Figs. 4.8 and 4.9, the MW model gives a fair prediction for the measured estimates, as well as reassures our comparison tests are correct.

4.5 Summary and Conclusions

In this chapter I analysed a set of 160 realisations of cluster-sized haloes, with masses $M_{\mathrm{halo}} \gtrsim 10^{14} M_{\odot}$, extracted from `zHORIZON` simulations. The analysis was mainly performed to interpret the accuracy at which we can estimate β, as a precursor of linear growth rate, using Kaiser (linear) model from galaxy cluster samples, like that of currently ongoing *eROSITA* survey. As discussed in 4.2, the cluster samples used at redshift z=0 and z=0.5 are comparable in number density and minimum mass cut-off for Deeper and Focused *eROSITA* samples, after suitable follow up procedures by spectroscopic as well as photometric datasets (Pillepich et al., 2012; Salvato and Merloni, 2015; Reiprich, 2017).

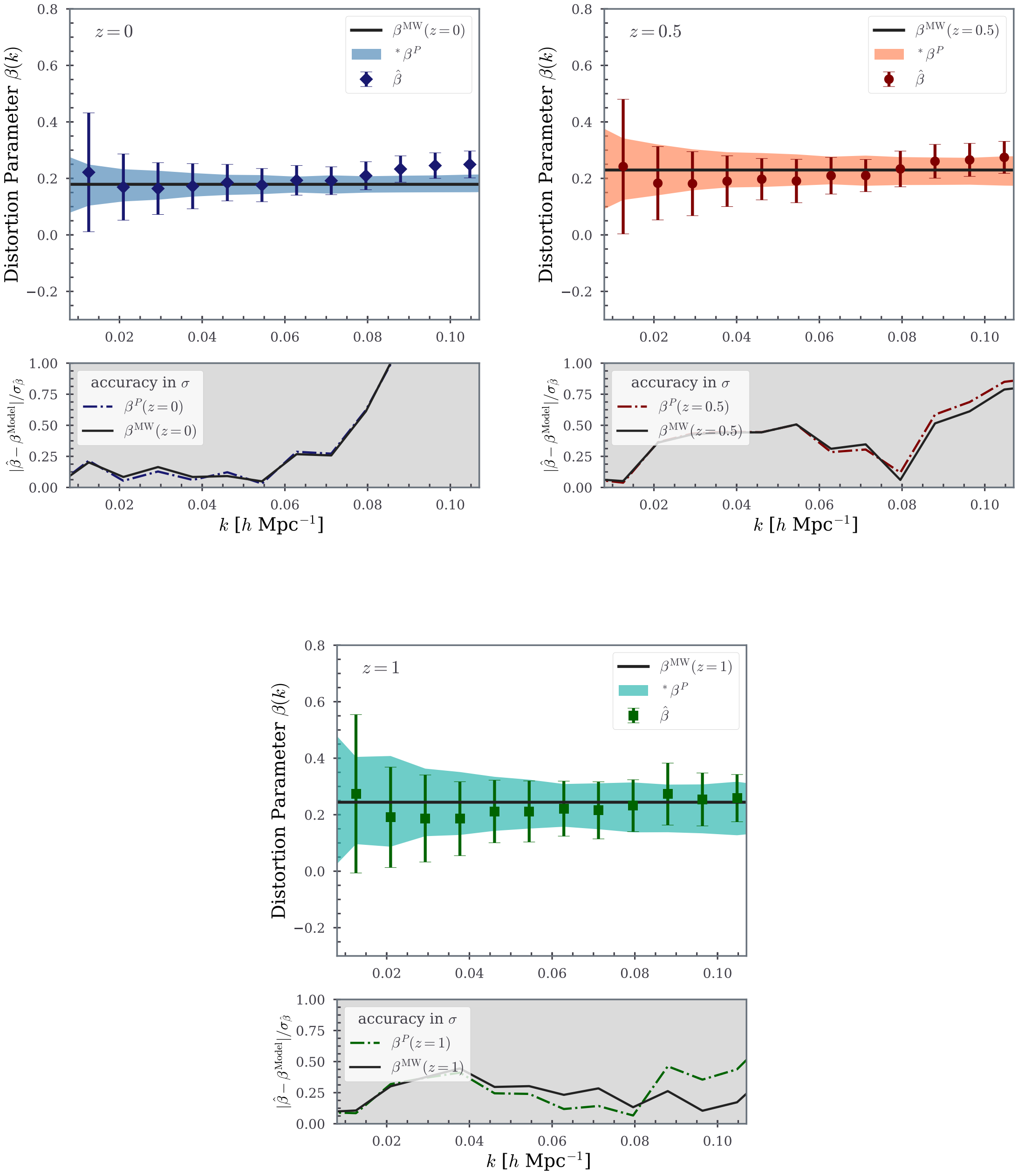


Fig. 4.8 $\hat{\beta}^P(k)$ in 12 bins at redshifts z=0 (top left), z=0.5 (top right) and z=1 (bottom). Datapoints with errorbars indicate the mean and the 1σ scatter from 160 mock halo catalogues. Colour shaded regions show the expected $\beta^P_{\mathrm{lin}}(k)$ mean and standard error on the mean from the mocks. The horizontal lines show the β^{MW} values from Table 4.2. The $*$ before $\beta^P_{\mathrm{lin}}(k)$ indicates adjusted errors (for the plot) from 1σ to standard error on the mean $(1\sigma/\sqrt{N})$ for better visualisation of the shaded portion.

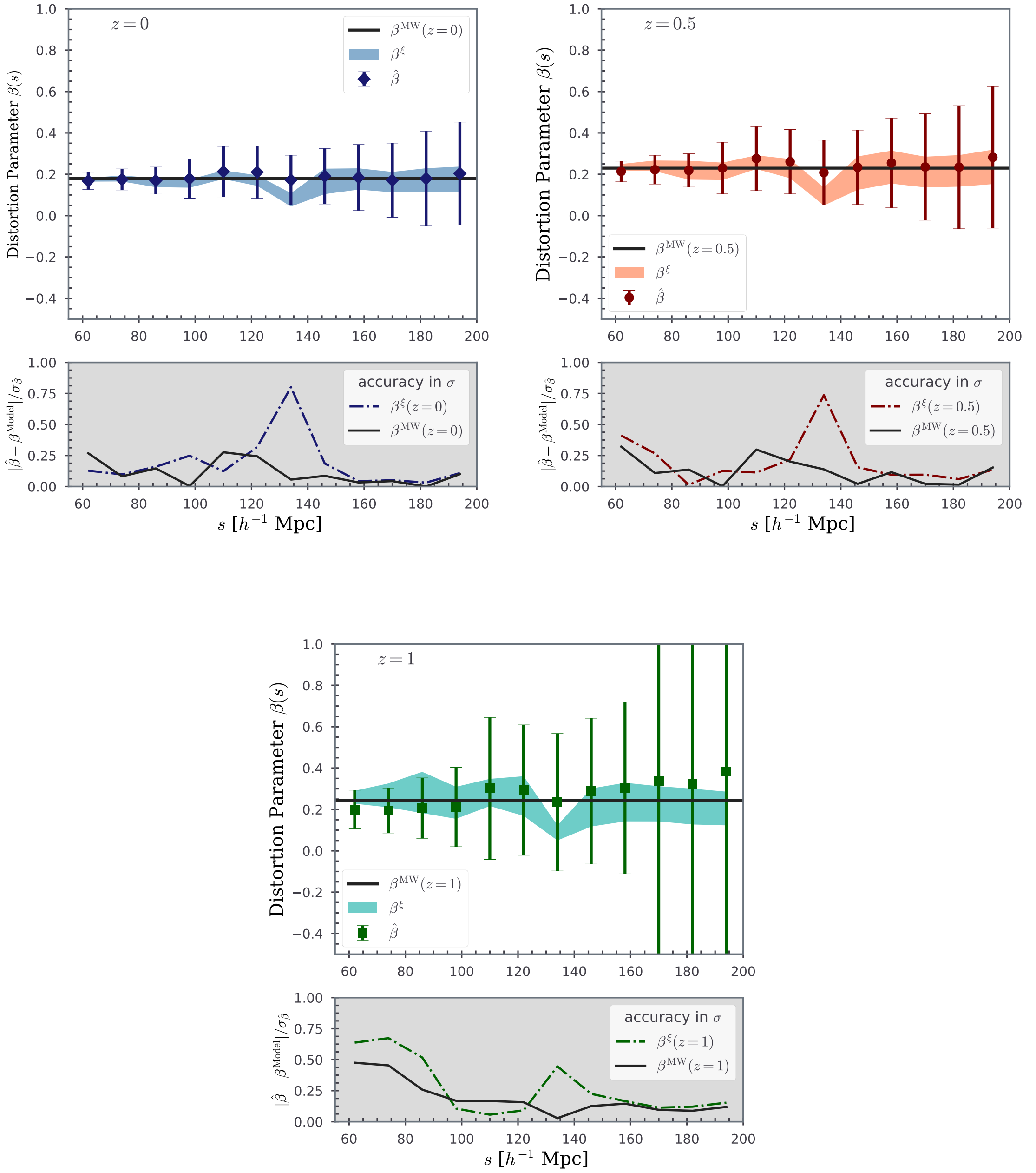


Fig. 4.9 Analogous to Fig. 4.8, $\beta^{\xi}(s)$ in 12 bins at redshifts z=0 (top left), z=0.5 (top right) and z=1 (bottom). Datapoints with errorbars indicate the mean and the 1σ scatter from 160 mock halo catalogues. Colour shaded regions show the expected $\beta^{\xi}_{\rm lin}(s)$ mean and scatter from the mocks. The horizontal lines show the $\beta^{\rm MW}$ values from Table 4.2.

For the clustering analysis, that was performed in Fourier space and configuration space, the measurements were compared with linear theory prediction at redshifts z=0, z=0.5 and z=1.

The analysis confirms that the Fourier space has advantage over configuration space and produces more reliable estimates of β. When β is measured using power spectrum, there is a 50% observed reduction of statistical uncertainties as compared to that of correlation function on similar scales (Grieb et al., 2016; Sanchez et al., 2017). In addition to this, Fourier space provides a wider window of scales in the linear regime but it is also noticed that the non-linearity sets earlier ($\sim 80\ h^{-1}$ Mpc) when going from large to small separations, i.e. towards larger values of k in comparison with configuration space ($\sim 45\ h^{-1}$ Mpc). Two-dimensional clustering results in both formats $\xi(s_\parallel, s_\perp)$ and $P(k_\parallel, k_\perp)$ shows salient features of Kaiser effect and no significant signs of FoG as expected from massive haloes with almost no substructure with random motions.

To testify the accuracy of the model and investigate the contribution of systematic effects in the data, the estimates $\hat{\beta}$ and the errors from 160 realisations, are compared with the linear theory prediction $\beta_{\rm lin}$ values in both Fourier and configuration space (See values in Tables 4.1 and 4.2). It is found that the estimates match with predicted values within 1σ limits on selected range of scales. The analysis also confirms the predictions by Bianchi et al. (2012) and Marulli et al. (2017) that the model is adequate for galaxy clusters, when the bias of the objects is appropriately modelled at higher redshifts, and can be used, without caring for complex non-linear effects, to extract β and possibly the linear growth rate $f\sigma_8$ with 10% accuracy. As thought by Okumura and Jing (2011), I consider more conservative approach on selection of scales in s-space and state scales above 50 h^{-1} Mpc appropriate to use linear approximation.

Finally, on scales $\gtrsim 50\ h^{-1}$ Mpc and $0.01 < k < 0.08\ h\ \mathrm{Mpc}^{-1}$, the estimated $\hat{\beta}^P_{z=0} = 0.188 \pm 0.021$ and $\hat{\beta}^P_{z=0.5} = 0.236 \pm 0.023$ from Fourier space analysis. Using correlation function the statistical errors add up significantly and yield $\hat{\beta}^\xi_{z=0} = 0.171 \pm 0.037$ and $\hat{\beta}^\xi_{z=0.5} = 0.217 \pm 0.043$. Also we notice that in configuration space the measured β is underestimated by 10%. The analysis confirms that future cluster samples from full-*eROSITA* survey (provided with redshift estimates by follow-up surveys) can be analysed with simple Kaiser-Hamilton model to estimate structure growth up to redshift $z \sim 0.5$, provided the systematic errors are minimised.

Chapter 5

2D Clustering with Radio Datasets

Unlike 3D distributed cosmic structures, the information extracted from a ‘sky-map’ data containing the projected 2D angular distribution is limited. In such cases, the datasets lack the radial information to the objects which makes more difficult to appreciate the time evolution of the relevant physical quantities. Nonetheless, there is still plenty of cosmological study that can be done using 2D information only, i.e. simply the angular coordinates on the projected map. A very good example of such analysis is the Cosmic Microwave Background (CMB) (Planck Collaboration et al., 2016; Aghanim et al., 2018; Nørgaard-Nielsen, 2018). As of today, a large number of observational datasets, contains angular coordinates and the flux of the observed sources, one such example is datasets obtained from observations in the radio band. Radio frequencies allow us to probe distant, high-redshift region of the Universe, and observe a large number of extra-galactic objects, potentially probing the large scale structure of the Universe over a wide redshift range. Outside the plane of our Galaxy, most of the sources detected at centimetre and metre wavelengths are extragalactic and often at very high redshifts. This is related to the emission mechanisms at such frequencies, which are nonthermal and occur in specific environments where electrons are accelerated to relativistic velocities and produce synchrotron radiation. The observed extragalactic radio sources are therefore hosts of powerful engines such as active galactic nuclei or sites of intensive star formation, and can be detected from very large cosmological distances. This, together with the fact that they are unaffected by dust extinction, makes extragalactic radio sources very useful to probe large cosmological volumes.

Amongst existing radio catalogues, a few wide-angle, sub-arc-minute-resolution catalogues cover areas up to thousands of square degrees that can be suitably used for LSS studies. Some notable examples include the Green Bank survey at 4.85 GHz (87GB, Gregory and Condon, 1991), the Parkes-MIT-NRAO survey also at 4.85 GHz (PMN, Wright et al., 1994), the Faint Images of the Radio Sky at Twenty centimeters (FIRST, Becker et al., 1995), the Westerbork Northern Sky Survey at 325 MHz (WENSS, Rengelink et al., 1997), the NRAO VLA Sky Survey at 1.4 GHz (NVSS, Condon et al., 1998), or the Sydney University Molonglo Sky Survey at 843 MHz (SUMSS, Bock et al., 1999). More recently large swaths of sky have been mapped by the Giant Metrewave Radio Telescope (GMRT, Ananthakrishnan, 1995), the Low-Frequency Array (LOFAR, van Haarlem et al., 2013), and the Murchison Widefield Array (MWA, Tingay et al., 2013). In the near future, such efforts are expected to accelerate the growth of wide-angle radio datasets by orders of magnitude thanks to forthcoming surveys such as the VLA Sky Survey[1] (VLASS, Myers and VLASS Survey Team, 2018) or those that will be undertaken by the Square Kilometre Array (SKA, Braun et al., 2015; Prandoni and Seymour, 2015) and its precursors (see e.g., Norris et al., 2011).

Studying LSS with radio imaging raises some specific challenges. The nonthermal character of radio emission means that the observed intensity of radio sources is only very weakly related to their distances, unlike in the optical where the bulk of the flux is black-body-like and hence readily provides information about luminosity distance. Another issue is related to the often complicated morphology of the radio sources. While usually point-like or at least concentrated to a small ellipse at short wavelengths, in the radio domain galaxies often present double or multiple structure with very extended lobes that generates clustering signal on small scales, with optical/IR counterparts that are difficult to identify. Furthermore, radio galaxies are typically located at high redshift with very faint optical counterparts. As a result, only a small fraction of radio sources, typically located in the local universe, have measurements of photometric and spectroscopic redshifts (e.g., Peacock and Nicholson, 1991; Magliocchetti et al., 2004). The only viable approach towards studying the LSS with radio continuum data is therefore via angular clustering. Despite its limitations, such two-dimensional (2D) clustering analyses can be very useful to identify the nature of radio sources, probe their evolution, and reveal subtle observational systematic errors.

The angular correlation properties of wide-angle radio catalogues have long been detected and analysed both in configuration and in harmonic space, usually using two-point statistics. The two-point angular correlation function (ACF) of the radio sources has been studied in several of the above-mentioned wide-angle radio samples (e.g., Cress et al., 1996; Loan et al., 1997; Blake and Wall, 2002b; Overzier et al., 2003; Blake et al., 2004b; Negrello et al., 2006; Chen and Schwarz, 2016). As for the harmonics analysis, the main catalogue used to measure the angular power spectrum (APS) so far is the NVSS (e.g., Blake et al., 2004a; Nusser and Tiwari, 2015).

---

[1] https://science.nrao.edu/science/surveys/vlass

The results of these analyses have shown that the clustering properties of radio sources can be accounted for in the framework of Λ cold dark matter (ΛCDM) and halo models, in which radio sources are located in massive dark matter halos, typically associated to large elliptical galaxies and active galactic nucleus (AGN) activity, sharing a common cosmological evolution. One exception to this success is represented by the dipole moment in the distribution of the radio sources in the NVSS and other radio surveys. After its first detection (Blake and Wall, 2002a), it was clear that the dipole direction agrees with that of the cosmic microwave background (CMB) dipole. However, several subsequent analyses indicated that its amplitude is larger than expected. The tension with the CMB dipole and theoretical predictions has been quantified by Singal (2011); Gibelyou and Huterer (2012); Rubart and Schwarz (2013); Fernández-Cobos et al. (2014); Tiwari and Jain (2015); Tiwari et al. (2015) to name a few examples. All of these studies agree that the observed dipole is difficult to reconcile with the predictions of the standard cosmological model (Ellis and Baldwin, 1984), although the significance of the mismatch depends on the analysis and can be partially reduced by taking into account the intrinsic dipole in the local LSS (Fernández-Cobos et al., 2014; Tiwari and Nusser, 2016; Colin et al., 2017) or by pushing the analysis to the quadrupole and octupole moments (Tiwari and Aluri, 2018) .

New, large, homogeneous datasets at different radio frequencies are clearly welcome to investigate the clustering properties of the radio objects in more depth. This is one of the reasons why the TIFR GMRT Sky Survey (TGSS) at 150 MHz, carried out at the GMRT[2] has received so much attention. Rana and Singh Bagla (2018) studied the clustering properties of this sample in configuration space by measuring its ACF. Their analysis, which is focused on angular scales smaller than $\theta = 0.1^\circ$, confirms that in this range the ACF is well described by a single power law with a slope comparable with that of NVSS but a larger amplitude. In another work, Bengaly et al. (2018) investigated the TGSS clustering properties in the harmonic space, focusing on the much-debated dipole moment. Quite surprisingly, they showed that the TGSS dipole is also well aligned with that of the CMB, but that its amplitude is large, much larger in fact that the one observed in NVSS.

The main goal of this chapter is to expand the analysis of Bengaly et al. (2018) by considering the full TGSS angular spectrum and compare it with theoretical expectations, focusing on the large-scale behaviour. As previous APS models have adopted simplifying hypotheses and neglected theoretical uncertainties, we shall emphasise the modelling aspects by including all the effects that contribute to the clustering signal and by propagating the uncertainties on the nature, redshift distribution, and bias of the radio sources

---

[2] http://www.gmrt.ncra.tifr.res.in/

into the APS model. We aim at quantifying possible departures from ΛCDM on all scales, using all multipoles $\ell > 1$.

The results of the analysis presented in this chapter have been published in (Dolfi et al., 2019)[3] The outline of this chapter is as follows: In § 5.1, I briefly describe the datasets we used in this work. These include the TGSS survey that constitutes the focus of our research, the NVSS survey that we mainly use as a control sample, a catalogue of radio sources obtained by cross-matching TGSS with NVSS object which we use to identify systematics and, in addition, a sample of quasars extracted from the Sloan Digital Sky Survey (SDSS) spectroscopic catalogue, to trace the distribution of TGSS objects at large redshifts. In § 5.2 I include the result of our analysis in configuration (i.e., the ACF) and harmonics (i.e., the APS) space. The motivation for considering the ACF is to assess its behaviour on angular scales smaller than those explored by Rana and Singh Bagla (2018) in order to isolate and characterise the clustering signal generated by multiple-component radio sources. The section also features the various tests performed to assess the robustness of the results. The model APS is presented in § 5.3. In this section I include my main contribution to this analysis i.e. test results related to redshift-dependent halo bias models that I have computed using minimum halo mass $\mathrm{M}_{\mathrm{min}}$ that characterises different radio sources following the work from (Ferramacho et al., 2014) and (Nusser and Tiwari, 2015). Finally, the results of its comparison with the measured TGSS power spectrum are presented in § 5.4. The conclusions of our work are discussed in § 5.5.

In this chapter, we assume a flat ΛCDM cosmological model with parameters taken from Planck Collaboration et al. (2016): Hubble constant $H_0 = 67.8$ km s^{-1} Mpc^{-1}, a total matter density parameter $\Omega_m = 0.308$, baryonic density parameter $\Omega_b = 0.048$, the rms of mass fluctuations at a scale of 8 h^{-1}Mpc $\sigma_8 = 0.815$, and a primordial spectral index $\mathrm{n}_s = 0.9677$.

5.1 Observational Datasets

The main dataset used in this chapter is the TIFR GMRT Sky Survey (TGSS) of radio objects detected at 150 MHz. A large fraction of them are in common with those in the NRAO VLA Sky Survey (NVSS). We analyse both the NVSS as well as the catalogue of common objects (dubbed TGSS×NVSS). These radio datasets are employed for angular clustering measurements. We also use the quasar catalogue from SDSS Data Release 14, but only to probe the redshift distribution of the TGSS sample.

---

[3] In the rest of this chapter, I shall refer to this collective work of authors with 'we' pronoun.

5.1.1 TGSS Catalogue

TGSS[4] is a wide-angle continuum radio survey at the frequency of 150 MHz, performed with the GMRT radio telescope (Swarup, 1991) between April 2010 and March 2012. The survey covers 36,900 deg^2 above $\delta > -53°$ (i.e., ∼ 90% of sky). In this analysis we use the TGSS Alternative Data Release 1 (ADR1, Intema et al., 2017)[5], which is the result of an independent re-processing of archival TGSS data using the SPAM package (Intema et al., 2009).

TGSS ADR1 contains 623,604 objects for which different quantities are specified. For this analysis we use angular positions, as well as the integrated flux density at 150 MHz and its uncertainty. The overall astrometric accuracy is better than $2''$ in right ascension and declination, and the flux density accuracy is estimated to be ∼ 10%. We shall consider only objects with integrated flux density above $S_{150} = 100\,\mathrm{mJy}$, where the ADR1 catalogue is ∼ 100% complete and more than 99.9% reliability being defined as (fraction of detection corresponding to real sources, Intema et al. 2017). The resolution of the survey depends on the declination: it is $25'' \times 25''$ north of $\delta \sim 19°$ and $25'' \times 25''/\cos(\delta - 19°)$ south of $\delta \sim 19°$.

The red histogram in Fig. 5.1 shows the TGSS source counts $N(S)$ per logarithmic flux bin ($\Delta \log(S) = 0.114$) per solid angle. The turnover at $S_{150} \sim 70\,\mathrm{mJy}$ reveals the completeness limit of the survey and justifies our conservative choice of considering only objects that are brighter than 100 mJy. Beyond this flux the $N(S)$ is well fitted by a power law that, as pointed out by Bengaly et al. (2018), has a slope $S^{-0.955}$ in the range $100\,\mathrm{mJy} < S_{150} < 500\,\mathrm{mJy}$. At brighter fluxes the $N(S)$ becomes steeper.

For our analysis we extract a subsample of the TGSS objects. The main selection criterion is the noise level, which is not constant across the survey (the median *RMS* value is 3.5 mJy/beam); it increases towards the Galactic plane and near bright radio sources. The TGSS subsample used in this analysis was selected as follows:

- We exclude all objects with declination $\delta < -45°$, where the *RMS* noise is higher than ∼ 5 mJy/beam, which is the value below which 80% of all measurements lie (see, Intema et al., 2017, Fig. 7).

- We discard objects with Galactic latitude $|b_{\mathrm{Gal}}| < 10°$, where the *RMS* noise is also large due to bright diffuse synchrotron emission of the Galaxy and to the presence of Galactic radio sources.

---

[4] http://tgss.ncra.tifr.res.in

[5] http://tgssadr.strw.leidenuniv.nl/doku.php

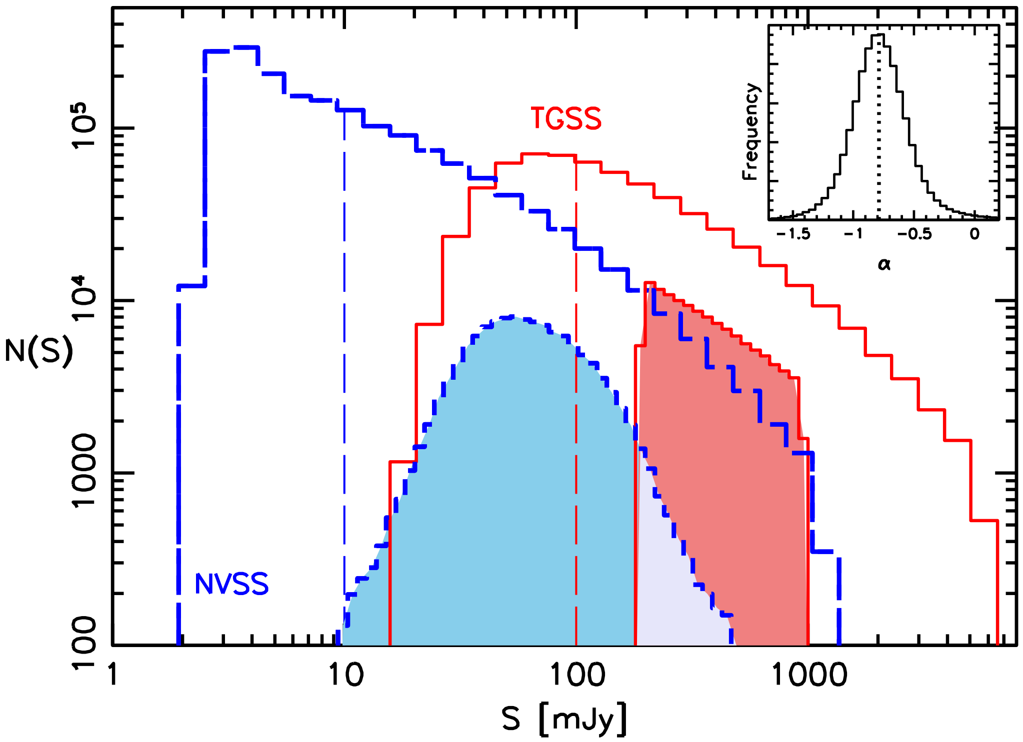


Fig. 5.1 Source counts of the TGSS (red, continuous) and NVSS (blue, long-dashed) catalogues. The red shaded histogram on the right shows the number counts (in S_{150} flux unit) of the objects in the TGSS×NVSS catalogue. The blue shaded area on the left shows the number counts (in $S_{1.4}$ flux unit) of the same TGSS×NVSS objects. Dashed vertical lines indicate the lower flux thresholds assumed for the analysis presented in this chapter. The histogram in the insert shows the distribution of the 150 MHz - 1.4 GHz spectral index of the sources in the TGSS×NVSS catalogue. The vertical dotted line indicates the peak of the distribution at $\alpha = -0.77$.

- We discard the sky patch of coordinates $97.5^\circ < \alpha < 142.5^\circ$ and $25^\circ < \delta < 39^\circ$, corresponding to the problematic observing session on January 28, 2011 characterised by bad ionospheric conditions (Intema et al., 2017).

- Following visual inspection using the Aladin Desktop tool, we mask out 34 of the brightest extended radio sources that appear as a cluster of many points in the catalogue which could produce anomalous large counts in small regions, mimicking spurious small-scale clustering (Nusser and Tiwari, 2015).

The areas of the sky identified by these constraints are represented by a binary `HEALPix` (Górski et al., 2005) mask with resolution $N_{\rm side} = 512$, which corresponds to a pixel size of 0.114° ($\sim 7'$) or a pixel area of $0.013\,{\rm deg}^2$. The maximum multipole corresponding to this angular resolution is $\ell_{max} \simeq 1024$. Nevertheless, in our analysis we only consider modes $\ell < 100$, to minimise nonlinear effects, as detailed in Sect. 5.2. After applying this mask, the fraction of the sky covered by the TGSS catalogue is $f_{\rm sky} \simeq 0.7$. Both very bright and very faint sources have also been excluded. Since different subsamples are considered for the clustering analyses, here we only specify the least restrictive flux cuts, that define the largest sample considered, and those used to extract the TGSS sample that we use as a reference; referred to as the *Reference* sample. The other flux cuts will be specified in Section 5.2.2, where they are used.

- We exclude all objects brighter than 5000 mJy since they increase the *RMS* noise in localised regions and produce spurious clustering signal. This threshold corresponds to the flux cut of about 1000 mJy in the 1.4-GHz band that we have adopted for the NVSS sample (see following section). For the *Reference* TGSS sample we set a more conservative conservative flux cut $S_{150} = 1000\,{\rm mJy}$ to minimise the chance of systematic effects that, as shown below, have a more significant impact than the random sampling noise. However, we demonstrate in Sect. 5.2.2 that the results of our analysis are very robust to the choice of the upper flux limit, in particular when this is set equal to 5000 mJy.

- Similarly, as already mentioned, we exclude all objects fainter than the completeness limit of $S_{150} = 100\,{\rm mJy}$, but in our *Reference* sample we use a stricter lower cut $S_{150} = 200\,{\rm mJy}$.

To summarise, we have defined a TGSS *Reference* catalogue of 109,941 radio sources with fluxes in the range $S_{150} = [200, 1000]\,{\rm mJy}$ located outside the masked area defined above. The main properties of this sample, together with two others used in the analysis (see below), are provided in Table 5.1.

Sample	N. objects	f_{sky}	Shot noise	$\Delta C_\ell \times 10^{-6}$
Ref. TGSS	109,941	0.7	8.01×10^{-5}	8.51
Ref. NVSS	518,894	0.75	1.82×10^{-5}	2.55
TGSS×NVSS	103,047	0.67	8.23×10^{-5}	8.51

Table 5.1 Main datasets used in this analysis and their characteristics. Col. 1: Dataset name. Col. 2: Number of objects. Col. 3: Fraction of the unmasked sky. Col. 3: Shot Noise. Col. 4: APS correction for multiple sources in units 10^{-6}.

The top panel of Fig. 5.2 shows a Mollweide projection of the observed TGSS counts in equatorial coordinates. The colour code indicates the number counts per squared degree, N/deg^2, of TGSS objects with flux in the range $200\,\text{mJy} < S_{150} < 1000\,\text{mJy}$. The masked areas are plotted in a uniform white colour.

5.1.2 NVSS Catalogue

The 2D clustering properties of the NVSS sources, especially their dipole moment, have been investigated in a number of works. The reason for repeating such an analysis here is twofold. First of all, it constitutes a useful cross-check for the analogous analysis of the TGSS catalogue. The second, more compelling reason is that, as we see below, the large majority of TGSS sources are also listed in the NVSS catalogue. Comparing the clustering properties of this population with those of their parent catalogues is a useful tool to spot systematic effects and to check the robustness of the results to the selection criteria.

The NVSS survey (Condon et al., 1998) at 1.4 GHz contains ~ 1.8 million sources over an area similar to that of TGSS and is 99% complete above $S_{1.4} = 3.4\,\text{mJy}$. Previous works have used various selection criteria and, consequently, analysed slightly different NVSS samples. Our data cleaning is similar to that of Blake et al. (2004a):

- We ignore the low signal-to-noise-ratio (S/N) region with declination $\delta < -40°$.
- We exclude objects near the Galactic plane $|b_{\text{Gal}}| < 5°$, to minimise spurious contribution of Galactic foreground and radio sources.
- We mask out 22 square regions around bright extended radio sources that can be fitted by multiple elliptical Gaussians and would generate spurious clustering signal (Blake et al., 2004a).

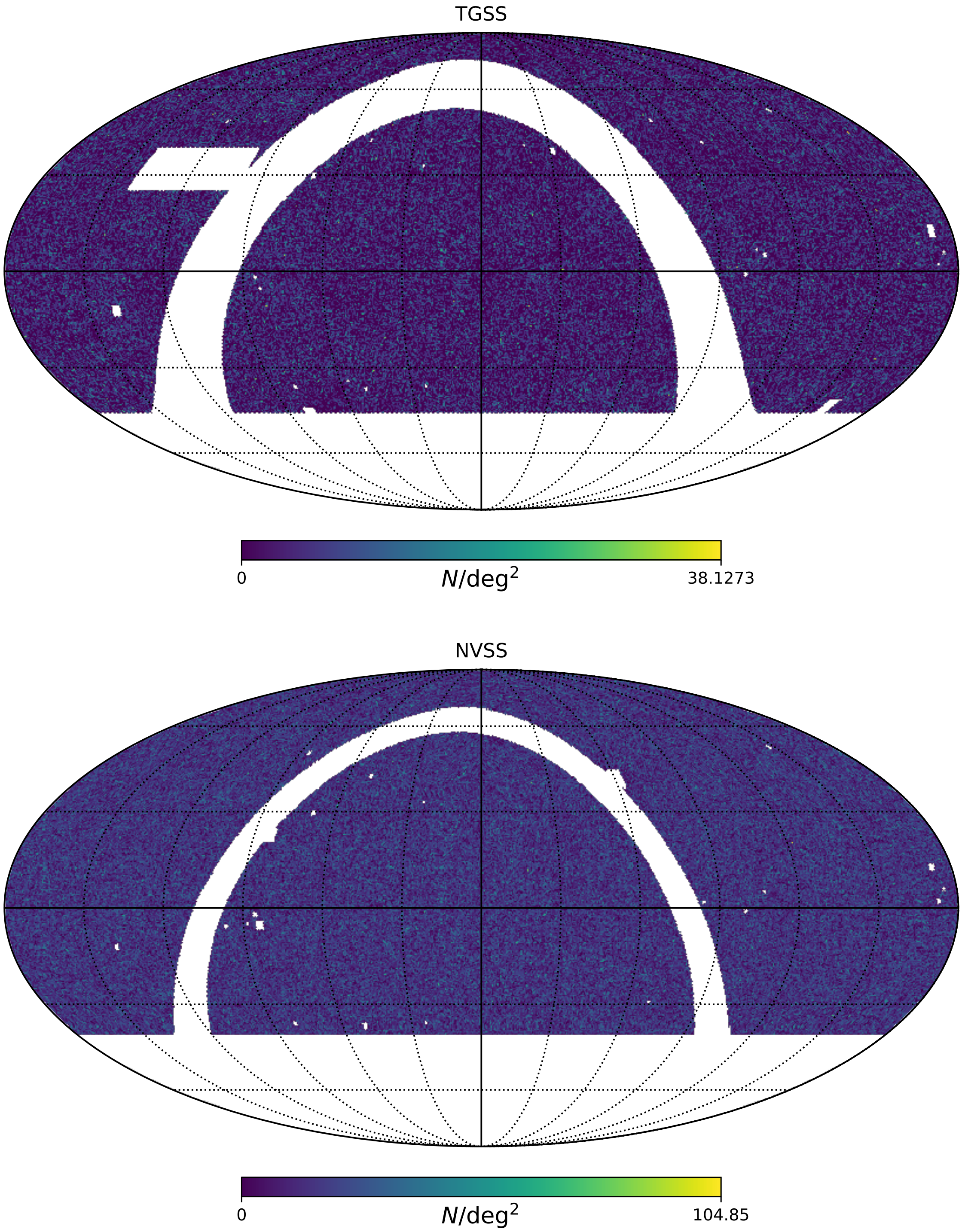


Fig. 5.2 Mollweide projection of TGSS (top) and NVSS (bottom) samples in equatorial coordinates. The plots show reference catalogues with selection criteria described in the text. The colour code in the bottom bar refers to N/deg^2, denoting the number counts per deg^2 in the pixel. The resolution of the map is $N_{\mathrm{side}} = 128$.

We create a binary `HEALPix` map to quantify the masked region. After masking, the sky fraction covered by NVSS is $f_{\rm sky} \simeq 0.75$. Similarly to the TGSS case, we define a *Reference* NVSS catalogue using the additional flux cuts:

- A lower cut at $S_{1.4} = 10\,$mJy, since below this limit the surface density of NVSS sources suffers from systematic fluctuations (Blake et al., 2004a).
- An upper cut at $S_{1.4} = 1000\,$mJy, since brighter sources may be associated to extended emission.

Our reference NVSS catalogue then consists of 518,894 radio sources with fluxes in the range $S_{1.4} = [10, 1000]$ mJy outside the masked area. Its source counts are represented by the blue histogram in Fig. 5.1. Above the 10-mJy threshold (vertical dashed line) the shape of the distribution is similar to that of the TGSS and can be superimposed to it by assuming a TGSS versus NVSS flux ratio $S_{150}/S_{1.4} \simeq 5$ (Bengaly et al., 2018).

The bottom panel of Fig. 5.2 shows the surface density of NVSS sources outside the masked areas. It is worth noting that the footprints of the two surveys are very much alike. This means that the effects of the two masks are very similar and the APS measured in the two samples can be compared directly.

5.1.3 TGSS × NVSS Catalogue

A detailed analysis of the properties of the objects in common between the TGSS and the NVSS was performed by Tiwari (2016) and de Gasperin et al. (2018). Our goal here is simply to build a matched catalogue to estimate the spectral index $\alpha_\nu = -1.03\log(S_{150}/S_{1.4})$ of the sources in common and to investigate their clustering properties in comparison to those of the parent catalogues.

Our procedure of matching the two datasets is as follows:

- we consider a TGSS source;
- we search for NVSS sources within $45''$ radius, corresponding to the NVSS survey resolution;
- if a single NVSS source is found, we accept the NVSS object as the cross-match with TGSS;
- if more than one NVSS source is found, we take the closest one as the cross-match.

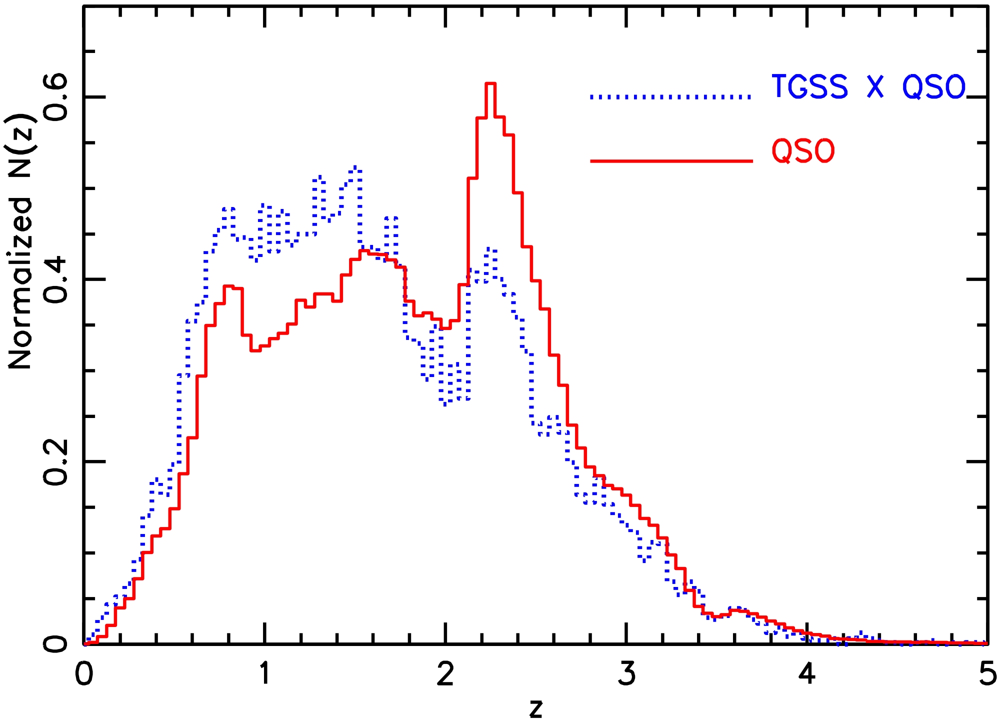


Fig. 5.3 normalised redshift distribution of the cross-matched TGSS×QSO catalogue (blue, dotted histogram) and of the parent SDSS-DR14 QSO catalogue (red, continuous).

The resulting cross-matched TGSS×NVSS catalogue contains 103,047 sources within the reference TGSS and NVSS flux limits, corresponding to ∼ 94% of the TGSS parent sample. The typical separation between the NVSS and TGSS sources is $1.2''$ and less than 10 % of them are separated by more than $8''$, comparable to the astrometric accuracy, as expected for genuine matches.

The number counts of the TGSS × NVSS objects are shown in Fig. 5.1 in both S_{150} (shaded red histogram on the right) and $S_{1.4}$ (shaded blue histogram on the left) flux units. The counts distribution, characterised by sharp cuts in S_{150} flux, has a distribution close to log-normal in units of $S_{1.4}$ flux.

The distribution of the spectral index α_ν is shown in the upper insert of Fig. 5.1 and is close to a Gaussian, with a peak at $\alpha_\nu \simeq -0.77$, in agreement with previous results (Tiwari, 2016; de Gasperin et al., 2018; Rana and Singh Bagla, 2018).

5.1.4 TGSS×SDSS-QSO Sample

The last catalogue considered here was obtained by cross-matching TGSS sources with the quasar (QSO) sample of the Sloan Digital Sky Survey Data Release 14 (SDSS DR14, Pâris et al., 2018). We point out that, unlike for the other catalogues described above, we do not expect the TGSS × SDSS-QSO sample to be statistically representative and, for this reason, we do not use it to perform any clustering analyses. Instead, it is only employed to show that the redshift distribution, $N(z)$, of TGSS sources extends out to large redshifts. The motivation is that so far the $N(z)$ of TGSS objects has been estimated directly only at relatively low redshifts by cross-matching them with the galaxies of the SDSS spectroscopic sample (Rana and Singh Bagla, 2018), which do not reach beyond $z = 1$.

The observed and model luminosity function of the radio sources (Willott et al., 2001) suggests however that the distribution of TGSS objects should extend to much higher redshifts than SDSS galaxies, so it is worth checking directly that this is indeed the case. Indirect verification of this prediction already exists: Nusser and Tiwari (2015) cross matched the NVSS catalogue with two small spectroscopic surveys (CENSORS and Hercules, Best et al. 2003; Waddington et al. 2001) and found that the distributions of NVSS sources extends out to $z \simeq 3$.

To prove that this is also the case for the TGSS sources, we performed a similar matching procedure as described above to build a TGSS×SDSS-QSO cross-matched catalogue. In the process we ignored astrometric errors in the QSO positions, which are negligible, and searched for TGSS - QSO matches within the angular resolution of TGSS ($25''$) and in the area common to the two surveys. We found $9,645$ matches corresponding to $\sim 1.5\%$ of the TGSS sources, most of them within $8''$ of the target object. The fraction of matched objects is small but still considerably larger than that of objects with an optical counterpart in the SDSS galaxy catalogue ($\simeq 0.6\%$, Rana and Singh Bagla 2018). The distribution of the TGSS×SDSS-QSO sample extends to $z \sim 4$ (blue, dotted histogram in Fig. 5.3), which is far beyond the redshift probed by Rana and Singh Bagla (2018), and is characterised by a double peak like that of the parent DR14 QSO sample (red, continuous histogram in Fig. 5.3), suggesting that this small cross-matched catalogue traces the redshift distribution of the optically selected QSO population.

The fact that this redshift distribution is so different from the one found by Rana and Singh Bagla (2018) strongly suggests that the TGSS catalogue contains various types of radio sources. We shall take into account this fact to model their correlation properties.

5.2 Angular Clustering Analysis

In this section we describe the statistical tools and the main results of the TGSS clustering analysis. We mainly use the angular power spectrum. However, the auto correlation function is also considered.

5.2.1 2-point Angular Correlation Function

We measure the angular two-point correlation function using the `TreeCorr` package (Jarvis et al., 2004). The estimator for this measurement is defined in § 3.1.2. We use ten times more sources in random catalogues in comparison with the real catalogue, by accounting for the complex geometry of the sample. The `TreeCorr` package generates ACF in bins of width $\Delta\log(\theta^\circ) = 0.1$, along with estimated errors obtained from propagating the Poisson noise.

The cosmic variance contribution could be estimated under the assumption of Gaussian errors from Eq. 20 of Eisenstein and Zaldarriaga (2001). Here we prefer to ignore this term since at the angular separations considered in our analysis ($\theta \leq 0.1^\circ$) the Gaussian approximation is expected to break down and the error budget to be dominated by Poisson noise rather than cosmic variance. For the same reason we ignore the effect of the "integral constraint", that is, the fact that the mean surface density of the sources is computed over a fraction of the sky (e.g., Roche and Eales, 1999). Given the large areas covered by the radio samples and the small angular scales considered here, the integral constraint is small and can be neglected.

Figure 5.4 shows the measured ACF of the *Reference* TGSS catalogue(red dots with error bars) and of the *Reference* NVSS catalogue(small cyan asterisks). Both ACFs exhibit a characteristic double power-law shape (Blake and Wall, 2002b) which reflects the fact that while on scales larger than $\theta \simeq 0.1^\circ$ the signal is dominated by the correlation among sources in different dark matter halos (i.e., the two-halo term), at smaller scales it is dominated by correlation of multiple sources within the same halo (the one-halo term). This second term depends on the density profile of the source, the typical number of radio components per source and the fraction of sources with multiple radio components. In the harmonic space, this one-halo term generates an almost constant, shot-noise-like signal that needs to be accounted for to compare the measured APS with theoretical predictions. The magnitude of this term depends on the characteristics of the sample: in brighter samples with a larger number of extended sources (and thus with a larger fraction of objects with multiple radio sources) this term is large, which explains why in Fig. 5.4 the amplitude of the TGSS ACF increases with the flux threshold.

Following Blake and Wall (2002b) we compute this term by fitting a power law to the measured ACF below $\theta = 0.1°$, under the hypothesis that the number of radio components per TGSS source is the same as in NVSS. In Fig. 5.4 we show the best-fitting power law to the reference TGSS sample (dashed line vertically offset to avoid confusion) together with the values of the best-fit amplitude A and slope γ. As these are different for the different TGSS subsamples, the best-fitting procedure has been repeated for all TGSS subsamples considered in our analysis.

For the *Reference* TGSS sample we estimate that the fraction of TGSS sources with multiple components is $e = 0.09 \pm 0.009$, where errors on e are propagated from the uncertainties of the measured ACF parameters A and γ. The corresponding shot-noise-like correction that we apply to the measured angular spectrum is $\Delta C_\ell \simeq 2e\sigma_N/(1+e) = (8.51 \pm 0.66) \times 10^{-6}$, where σ_N is the mean surface density. For the *Reference NVSS* case the corresponding values are $e = 0.07 \pm 0.005$ and $\Delta C_\ell = (2.55 \pm 0.18) \times 10^{-6}$ (see Table 5.1).

5.2.2 Angular Power Spectrum

To measure the auto- and cross- APS we use the Pseudo-C_ℓ estimator defined in Chapter 3 in § 3.2.2. We focus our analysis on the multipole range $2 \leq \ell \leq 100$ and consider the angular power in bins $\Delta\ell = 5$. We neglect the mode $\ell = 1$ because the dipole of TGSS sources has already been studied by Bengaly et al. (2018). The reason for setting $\ell \leq 100$ is to reduce the impact of nonlinear effects that correlate modes with large multipoles ℓ. Mode coupling is also induced by the incomplete sky coverage, although the effect is not expected to be large, given the wide areas of the NVSS and TGSS catalogues. The $\Delta\ell = 5$ bin is introduced to further reduce the effect of mode coupling because the effect of binning is to decorrelate measurements, resulting in a more Gaussian likelihood (Thomas et al., 2011). For all these reasons we assume Gaussian-independent random errors that, for the individual ℓ mode are computed using Eq. (3.49) where the values of $f_{\rm sky}$ corresponding to the unmasked sky covered by the sample, are shown in Table 5.1.

Figure 5.5 compares the measured APS of TGSS (red dots) and NVSS (blue squares) samples, as well as that of the TGSS×NVSS cross-matched sample (green triangles). All spectra are corrected for the multiple source contributions ΔC_ℓ listed in Table 5.1. Error-bars represent the 1σ Gaussian uncertainties (Eq. 3.49).

The NVSS and TGSS samples considered in the plot are slightly different from the *Reference* ones since we applied the same angular mask obtained by multiplying the TGSS and NVSS masks pixel by pixel. The sky fraction covered by both samples is $f_{\rm sky} \simeq 0.67$, the same as that covered by TGSS×NVSS. The rationale behind this choice is to eliminate

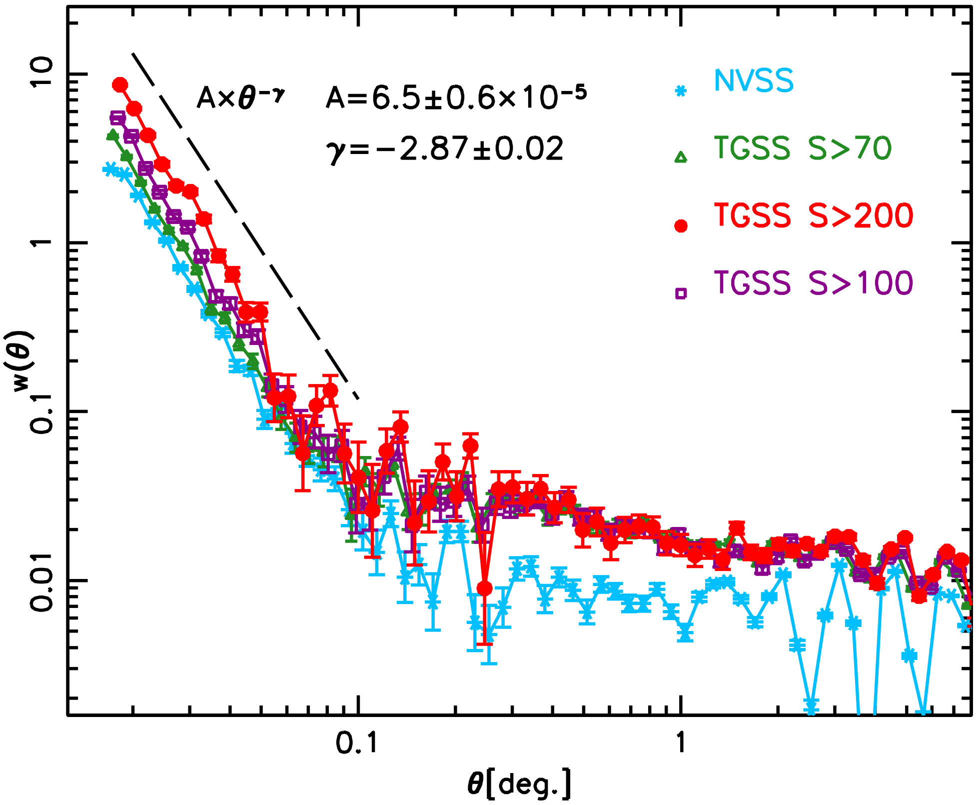


Fig. 5.4 Angular two-point correlation function for the *Reference* TGSS (red dots) and NVSS (light blue asterisks) samples. Green triangles and purple squares represent the ACF of two additional TGSS subsamples selected at different flux cuts $S_{150} > 70$ mJy and $S_{150} > 100$ mJy, respectively. Error bars represent Poisson uncertainties. The black dashed line shows the best-fit power law to the ACF of the *reference* sample at $\theta < 0.1°$. A vertical offset has been applied to avoid overcrowding. The best fitting parameters are indicated in the plot.

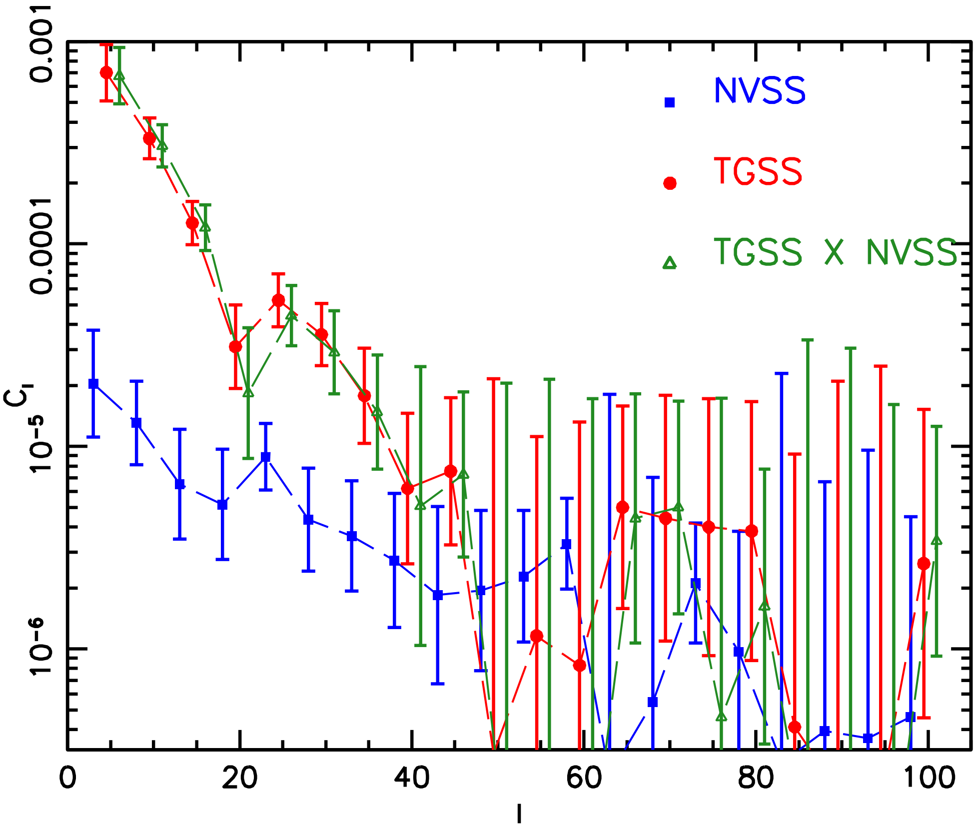


Fig. 5.5 Angular power spectrum of the NVSS (blue squares) and TGSS (red dots) samples with Gaussian error bars. Small green triangles show the APS of the TGSS×NVSS matched catalogue. All spectra are corrected for shot noise and multiple source contributions ΔC_ℓ.

all the differences that may result from sampling different regions (cosmic variance) and geometries (convolution effects).

There is a striking difference between the TGSS and NVSS angular spectra below $\ell \simeq 30$, where the amplitude of the former is significantly larger than that of the latter. At larger multipoles the two spectra agree with each other within the errors. The angular spectrum of the matched TGSS×NVSS catalogue is similar to that of TGSS-only, which should be expected considering that almost 95% of the TGSS reference sample have counterparts in NVSS. Taking into account the lack of a one-to-one relation between multipoles ℓ and angular separations θ, we identify the amplitude mismatch between the TGSS and NVSS power spectra at $\ell \leq 30$, with the amplitude difference of the angular correlation functions seen at $\theta \geq 0.3°$ (Fig. 5.4).

A useful sanity check to assess the reliability of this result is to compare our NVSS APS with the ones measured by Blake et al. (2004a) and Nusser and Tiwari (2015). The test is successful in the sense that it shows a good qualitative agreement with the measured spectra in both cases. Analogously, it is useful to compare our TGSS angular spectrum with the one computed by Bengaly et al. (2018). Although their analysis focused on the dipole moment, figure 4 in their paper shows a significant mismatch between the TGSS and the NVSS APSs at $\ell < 30$ which is analogous to the one detected in our analysis.

This discrepancy between the two spectra is quite unexpected, considering the similarities between the two samples both in terms of surveyed areas and the likely nature of the sampled sources. It may either reflect a genuine physical origin, related to the intrinsic clustering properties and redshift distribution of the two samples, or it could be an artefact produced by some observational systematic errors that have not been properly identified and accounted for. In the remainder of this section we explore the latter possibility by performing a number of tests aimed at testing the robustness of the APS measurements to different observational quantities that are expected to correlate with the measured radio flux and Galactic emission.

Robustness to Flux Cuts

Spurious clustering features on large angular scales can be generated by errors in the flux calibration that are coherent across large areas. This type of systematic uncertainty is indeed present and can be significant for low-frequency radio observations (Schwarz et al., 2015) reaching up to $10-20\,\%$ in amplitude for the case of the TGSS survey (Hurley-Walker, 2017). The angular scale of coherence is related, in the TGSS case, to the size of the area covered during the observing session which is typically of the order of $\sim 10^\circ$ (Bengaly et al., 2018). The impact of this effect was simulated by Bengaly et al. (2018) who focused on the dipole moment, and it turned out to be quite small ($\sim 1\,\%$ on the dipole amplitude). This is much smaller than the TGSS versus NVSS power mismatch and can hardly explain it, even taking into account that its amplitude may increase at $\ell > 1$, on the angular scales corresponding to those of the typical observational session. For this reason we exclude this possibility and neglect the effect of flux calibration errors in this analysis.

Other possible systematic errors that are not related to flux calibration can be induced by the flux threshold used to select the sample. For example, random uncertainties in the flux measurements, that in the TGSS case are of the order of 10% (Intema et al., 2017), can scatter objects fainter than the completeness limit of the survey into the catalogue. Their impact in the APS can be appreciated by changing the value of the lower flux threshold $\bar{S}_{150}$. Analogously, including bright, extended objects associated with multiple sources may artificially increase the clustering signal. In this case, an effective robustness test would be to change the upper flux cut $\bar{S}_{150}$ of the TGSS survey.

To quantify the impact of the systematic errors related to the flux cuts we ran a set of tests in which the TGSS APS was measured by varying the values of $\bar{S}_{150}$ and $\bar{S}_{150}$, keeping the geometry mask fixed. The upper panel of Fig. 5.6 illustrates the sensitivity to $\bar{S}_{150}$. The curves drawn with different line styles indicate the difference between the APS of the TGSS sample selected at a given cut $\bar{S}_{150}$ with respect to the *Reference* sample, for which $\bar{S}_{150} = 200\,\mathrm{mJy}$. The difference ΔC_ℓ is expressed in units of the Gaussian error, σ_{C_ℓ}, of the reference APS. The results are remarkably independent of the choice of the lower flux cut. Selecting objects with $\bar{S}_{150} = 100\,\mathrm{mJy}$, that is, brighter than the formal completeness limit of the TGSS sample, does not significantly modify the results. Similarly, when we use more conservative flux cuts of $\bar{S}_{150} = 300$ and $400\,\mathrm{mJy}$ (the second one not shown in the plot to avoid overcrowding) we also find results that are consistent with the *Reference* ones within the 1-σ Gaussian errors.

We also tried forcing the lower cut below the TGSS completeness limit, by setting $\bar{S}_{150} = 50\,\mathrm{mJy}$. The rationale behind this choice is to identify possible systematic effects that may be present also in the complete sample. We find that using this cut significantly enhances the power at low multipoles, especially at $\ell \simeq 20$. This is a sizeable effect that interestingly occurs on the angular scale ($5^\circ \times 5^\circ$) of the mosaics that constitute the building blocks of the TGSS survey. Since the overall TGSS source catalogue is obtained by summing up mosaic-based data, this effect is likely to be attributed to sensitivity variations in adjacent mosaics, or even to the fact that the sensitivity pattern in these mosaics is replicated in adjacent mosaics. As a consequence, the surface density of faint objects with fluxes below the completeness threshold coherently varies across each mosaic, generating a spurious clustering signal on the angular scale of the mosaic itself. A small excess of power is also seen at $\ell \simeq 20$ if larger $\bar{S}_{150}$ cuts are applied. However, its statistical significance is much less than in the $\bar{S}_{150} = 50\,\mathrm{mJy}$ case. This fact corroborates the hypothesis that this excess power reflects an observational systematic effects that are corrected for by selecting objects above the completeness limit of $S_{150} = 100\,\mathrm{mJy}$. Our results are also robust to the choice of the upper threshold $\bar{S}_{150}$, as shown in the middle panel of Fig. 5.6, which compares two more permissive upper flux cuts at 3000 and 5000 mJy against the *Reference* of $\bar{S}_{150} = 1000\,\mathrm{mJy}$.

Finally, we performed analogous robustness tests on the TGSS × NVSS catalogue by similarly modifying the upper (lower) flux cuts in both samples below (above) the completeness limits. As for the TGSS sample, we find no significant departures from the *Reference* angular power spectrum. Further tests aimed at detecting possible systematic effects in the TGSS sample that may generate spurious clustering signal can be found in the appendix of (Dolfi et al., 2019).

Robustness of Mask Selection

To quantify possible systematic effects induced by Galactic foregrounds or by any other effect related to the presence of the Galaxy, we tested the impact of using different geometry masks characterised by more conservative cuts in the Galactic latitude. We explored two cases. In the first one we excluded all objects with $|b| < 15°$ and in the second one we discard the region $|b| < 20°$. The unmasked sky fraction is consequently reduced to $f_{\rm sky} \sim 0.61$ and $f_{\rm sky} \sim 0.56$, respectively. In both cases we considered the same flux cuts as the reference TGSS sample.

We then computed the residuals of the corresponding angular power spectra with respect to the TGSS reference case in units of Gaussian error. The results, displayed in the lower panel of Fig. 5.6, appear to be robust to the inclusion of objects near the Galactic plane. It is worth noticing that some difference in the various spectra is to be expected because different geometry masks are used here. They are obviously small, since they contribute to the plotted residuals.

Relatively nearby sources can generate high-amplitude clustering signal that is not fully accounted for in the modelling. As we do not have information on the distance of the sources, an effective strategy to minimise the impact of the nearest ones is to exclude objects near the Supergalactic plane. Tiwari and Jain (2015) adopted this approach in measuring the NVSS dipole and found that this cut has a negligible impact on the dipole signal. They conclude that the dipole is largely generated by distant objects. In Section 5.3.1, we show that the model redshift distribution of TGSS sources (in Fig. 5.7) does not feature the prominent local ($z < 0.1$) peak that, instead, characterise the NVSS one. Given the lack of a prominent local population of TGSS objects, we conclude that removing TGSS objects near the Supergalactic plane will likely only increase the shot noise error and, therefore, we decided not to apply additional cuts to the geometry mask.

5.3 Modelling the Angular Power Spectrum

The analyses performed in the previous sections indicate that the APS of the TGSS sources is significantly larger than that of the NVSS at $\ell \simeq 30$ and that the mismatch cannot be attributed to known potential sources of observational systematic errors.

In this section we consider the alternative hypothesis that the large-scale TGSS power is genuine and reflects the intrinsic clustering properties of the TGSS radio sources. To test

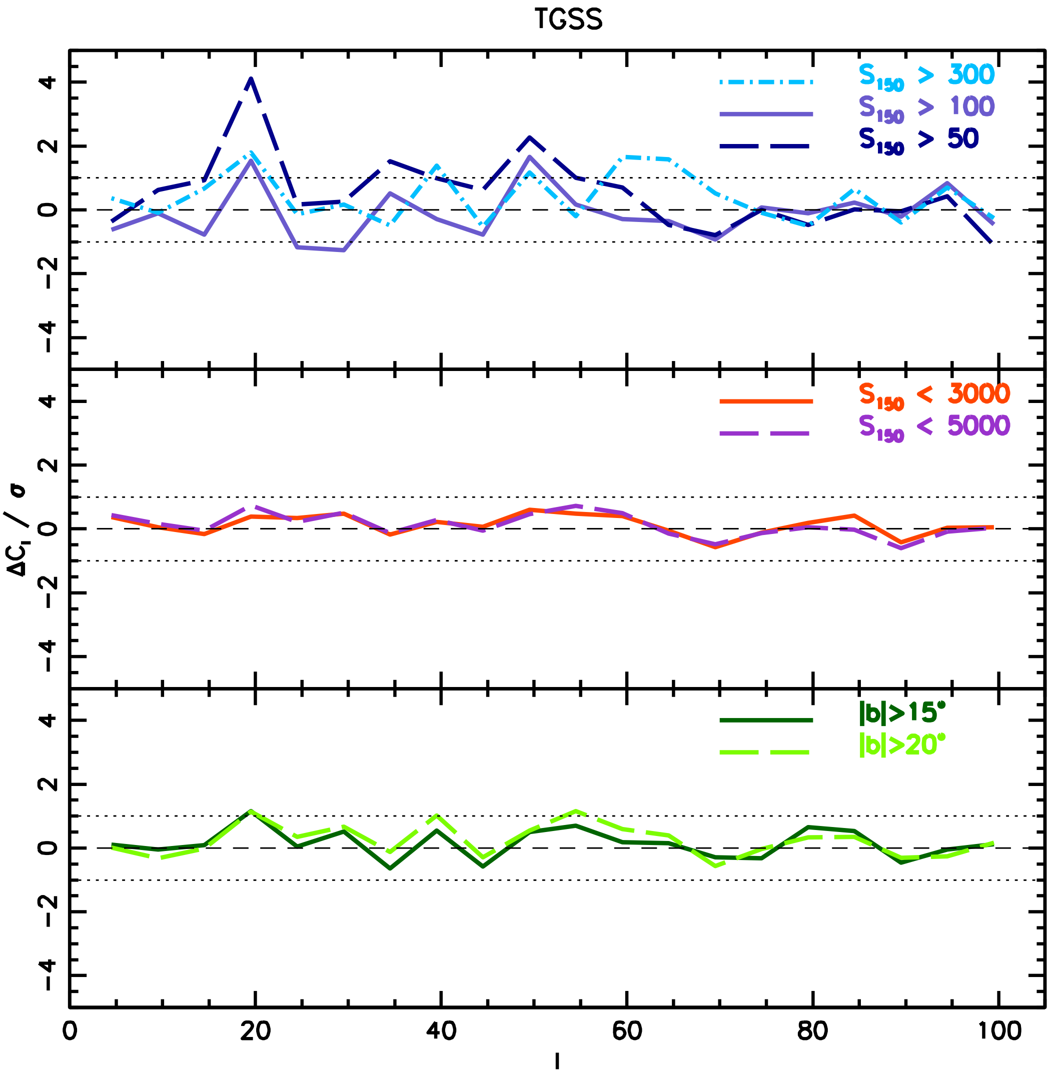


Fig. 5.6 Angular power spectrum residuals of different TGSS samples with respect to the *Reference* case, expressed in units of Gaussian errors. The upper panel shows the normalised residuals of the TGSS samples selected at different values of the minimum flux cut, $\bar{S}_{150}$, indicated in the plot, compared to the *Reference* case of $\bar{S}_{150}$ = 200 mJy. In the middle panel we consider samples selected at different values of the maximum flux cut, $\bar{S}_{150}$; the *Reference* is $\bar{S}_{150}$ = 1000 mJy. The bottom panel shows the residuals for samples with different geometry masks, cut at different values of the Galactic latitude, also indicated in the plot, referred to the baseline case of $|b| > 10°$. The dotted horizontal lines in all panels indicate the 1 σ Gaussian error of the *Reference* sample. The dashed horizontal line indicates the zero residual level.

this hypothesis we compare the measured APS with the theoretical predictions obtained assuming a Planck ΛCDM cosmology (Planck Collaboration et al., 2016) and physically motivated models for the redshift distribution, $N(z)$, and bias, $b(z)$, of TGSS sources. Since we are interested in large scales, we limit our comparison to the range $\ell \leq 100$. In this comparison we do not try to infer cosmological parameters, as we assume that the background cosmological model is well known. Instead, we consider various realistic $N(z)$ and $b(z)$ models to investigate whether the large-scale power of TGSS can be accounted for within the known observational and theoretical errors. To assess the validity of this approach we perform the same comparison for the NVSS sample. Only the *Reference* TGSS and NVSS samples are employed here.

To model the APS of TGSS sources we use the code `CLASSgal` (Lesgourgues, 2011b; Di Dio et al., 2013) which accounts for nonlinear evolution of matter density fluctuations and offers the possibility to include physical effects such as redshift space distortions, gravitational lensing, and general relativistic effects. Required inputs are the parameters of the underlying cosmological model (given by our fiducial set of parameters), the redshift distributions of the sources and their linear bias.

All our APS models share the same treatment of the mass power spectrum and differ in the choice of $N(z)$ and $b(z)$. The characteristics of the model mass power spectrum are described below. We also quantify the impact of the various physical effects that contribute to the clustering signal by considering the $N(z) + b(z)$ model S^3-HB described in the following section.

- *Nonlinear effects.* The nonlinear evolution of mass density fluctuations is modelled within the so-called `HALOFIT` framework (Smith et al., 2003; Takahashi et al., 2012). On the scales of interest ($\ell \leq 100$) nonlinear effects are expected to be small, and for this reason have been ignored altogether in some of the previous APS analyses (e.g., Nusser and Tiwari, 2015). To quantify the impact of nonlinear effects we compared the APS predicted with `HALOFIT` with the one obtained using linear perturbation theory, using the redshift distribution and bias of the model S^3-HB. We found that at $\ell = 100$ the nonlinear evolution enhances the angular power by just $\sim 0.5\%$.

- *Redshift space distortions* (RSDs). Peculiar velocities amplify the clustering signal on large angular scales. We have compared the APSs obtained with and without including RSD and found that RSDs amplify the clustering signal by $\sim 3.5\ \%$ at $\ell = 2$. The amplitude of the effect decreases at larger multipoles; it is $\sim 2\ \%$ at $\ell = 20$ and $\sim 1\ \%$ at $\ell = 40$.

- *Magnification lensing.* Gravitational lensing modulates the observed flux of objects and therefore reduces or increases the number counts above a given flux threshold. This effect generates an additional correlation (or anti-correlation) signal that can be described in terms of magnification-magnification and magnification-density correlations (Joachimi and Bridle, 2010). The magnitude of the effect depends on the slope of the cumulative luminosity function at the limiting flux of the sample (Joachimi and Bridle, 2010; Di Dio et al., 2013). Because of the composite nature of TGSS and NVSS, which contain different types of objects with different luminosity functions (see e.g. below), one needs to account for their individual contributions to the magnification signal. We do that by considering an effective luminosity function slope that we computed by considering the luminosity function of each object type at different redshifts (from Willott et al. 2001), estimating their slope in correspondence of their limiting flux and computing the effective slope as $\tilde{\alpha} = \sum_i \sum_j \alpha(i,j) N_i(z_j) / \sum_i \sum_j N_i(z_j) \simeq 0.3$, where i runs over all object types, j runs over the redshift values, $N_i(z)$ is the redshift distribution of object type i and $\alpha(i,j)$ the slope at the redshift j. For this we have assumed the S^3-HB model. We find that in the TGSS case the magnification lensing provides a small but significant, negative contribution to the clustering signal. On the scales of interest ($\ell < 40$) the amplitude of the effect is $\sim -6\%$, increasing to $\sim -9\%$ at $\ell = 2$ and decreasing to $\sim -3\%$ at $\ell = 100$.

- *General Relativistic effects.* `CLASSGal` provides the opportunity to include general relativistic contributions to the APS. Their impact, however, is small and limited to very large angular scales. It is of the order of 1% at $\ell \sim 4$, sharply decreasing to 0.1% at $\ell = 30$.

In addition to these physical effects that are included in all our models, there are some approximations and corrections that we need to make explicit before considering different model predictions and their comparison with data.

- *Limber approximation.* Several APS models in the literature have adopted the Limber (1953) approximation (refer to Eq. (3.44) and the related text) to speed up the APS numerical integration. As anticipated in Chapter 3, we do not adopt Limber approximation in this analysis. However, it is useful to quantify its impact when comparing our results with those of other analyses. `CLASSGal` allows one to switch the Limber approximation option on and off and to select the ℓ value above which the approximation is adopted. The Limber approximation boosts the modelled angular power at small ℓ values. In the S^3-HB model the effect is as large as $\sim 15\,\%$

at $\ell < 5$ but then its amplitude rapidly decreases to ~ 7 % at $\ell = 10$ and to ~ -1 % at $\ell = 20$.

- *Geometry mask.* The effect of the geometry mask is to modulate the signal and to mix power at different APS multipoles. This effect can be expressed as a convolution of the form $\tilde{C}_\ell = \sum_{\ell'} R_{\ell\ell'} C_{\ell'}$, where C_ℓ is the model APS predicted by `CLASSGal` and $R_{\ell\ell'}$ is the mixing matrix, evaluated APS of the survey mask (see e.g. equation 6 of Balaguera-Antolínez et al. (2018)). The main effect of this mask is to modulate power at small multipoles. As we are interested in the range $2 \leq \ell \leq 100$, we do not account for the survey beam, which instead would modulate power at large multipoles.

To finalise our APS models of the TGSS and NVSS catalogues we need to specify the redshift distribution and the bias of the sources. A specific $N(z) + b(z)$ model was adopted to quantify the impact of the various effects that contribute to the APS. Now we describe and justify the adoption of that model and explore its uncertainties by considering a number of physically motivated models of both $N(z)$ and $b(z)$ that have been proposed in the literature. We quantify the related theoretical uncertainties by taking into account the scatter in the corresponding APS predictions.

5.3.1 Redshift Distribution Models

The analysis of the cross-matched TGSS×SDSS QSO catalogue has confirmed that the distribution of TGSS sources extends to much larger redshifts than those probed by cross correlating them with galaxy redshift catalogues (Rana and Singh Bagla, 2018). As a consequence, although the majority of the TGSS APS signal at low multipoles is probably built up at $z \leq 0.1$ as in the NVSS case (Blake et al., 2004a; Nusser and Tiwari, 2015), a non-negligible contribution could also be provided by highly biased objects at higher redshifts. To test this hypothesis we need to identify the nature of the TGSS sources and to probe their distribution along the line of sight.

As we discussed in the introduction, the difficulty in finding IR/optical counterparts to the objects identified in low-frequency radio surveys makes it difficult to measure their $N(z)$ directly. Only Nusser and Tiwari (2015) have adopted such an approach by cross correlating the NVSS catalogue with a deep but small sample of objects with measured spectroscopic redshifts. With about 300 matches they were able to trace the redshift distribution of NVSS objects out to $z \sim 3$. Unfortunately we cannot repeat this procedure with TGSS because of the small number of TGSS matched objects. Therefore we need to change the approach and instead model the TGSS redshift distribution.

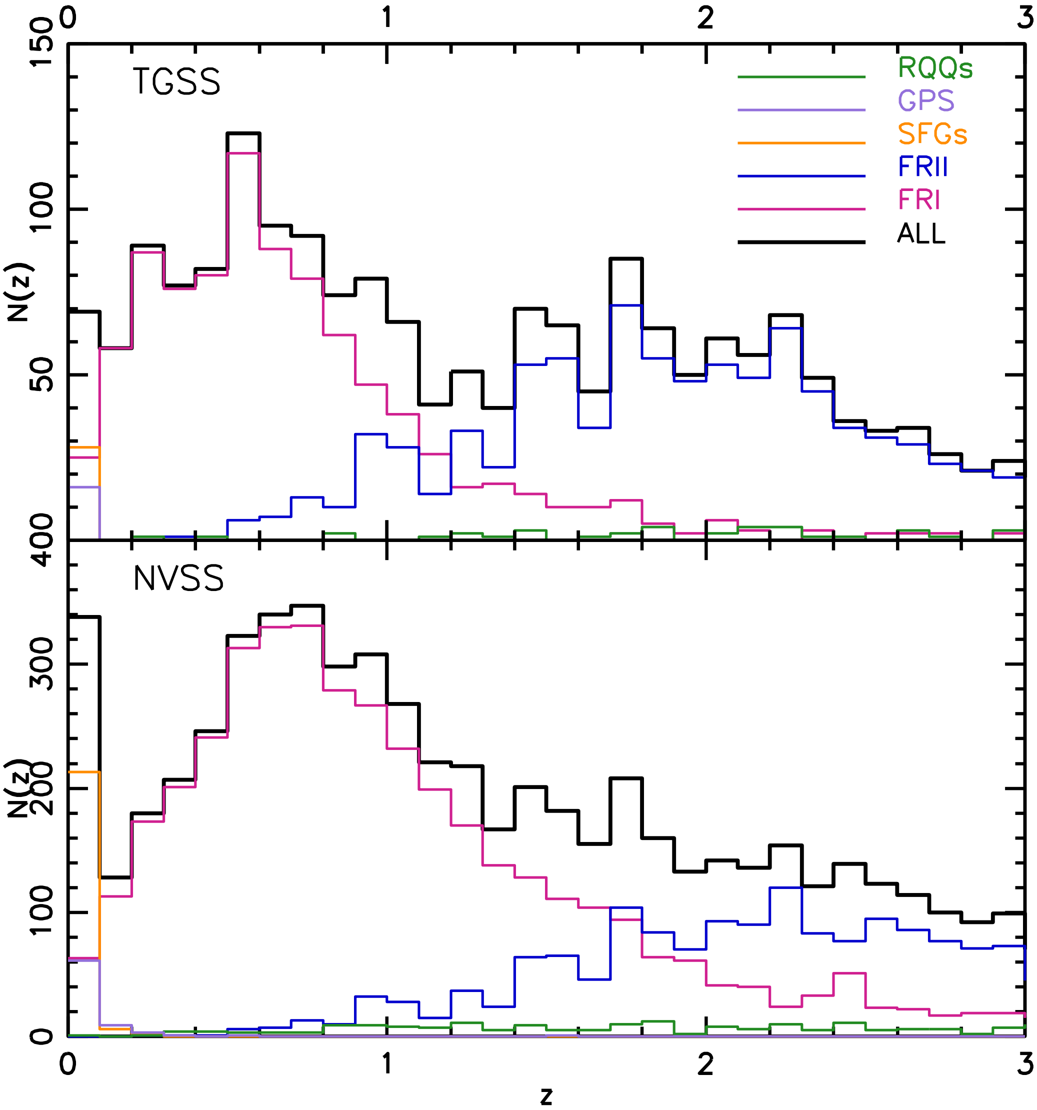


Fig. 5.7 S^3 model redshift distributions $N(z)$ of the various types of sources in the TGSS (top) and NVSS (bottom) samples. The redshift distribution of each source type is represented by a different colour, as specified in the upper panel. The thick, black histogram shows the redshift distribution of all types of sources combined.

For the redshift distribution modelling we use the SKA Simulated Skies (S^3) database[6]. This tool, described in detail in Wilman et al. (2008), is meant to model radio observations in a given band within a sky patch. It is a phenomenological model in the sense that it uses constraints on the available, observed luminosity functions at different redshifts. This simulator also mimics the clustering properties of radio sources by assuming a model for their bias. This latter aspect, however, is quite uncertain, as shown by recent clustering analyses of radio sources (Magliocchetti et al., 2017; Hale et al., 2018). In principle, we could have used the newer simulator, T-RECS (Bonaldi et al. 2019) that, in addition to predicting more realistic clustering properties than S^3, also implements more recent evolutionary models for SFGs, and treats RQ AGNs as part of the SFG class, under the assumption that their radio emission is dominated by star formation. However, considering that: *i)* we use the simulator to model the redshift distribution of the radio sources and not their clustering properties and *ii)* the number of SFGs and RQ AGNs expected in our TGSS sample is negligible, using T-RECS instead of S^3 would have little or no impact on our results. Therefore we decided to stick to S^3 instead of using T-RECS that only became available when our analysis was in a very advanced stage of completion.

In our application we simulated two radio surveys over the same sky patch of 400 deg^2 at 150 MHz and at 1.4 GHz, and considered objects with fluxes above the flux limits of our *Reference* samples, that is, $S_{1.4} > 10$ mJy and $S_{150} > 200$ mJy, respectively. No upper flux cuts were considered since, as we have seen, results are very robust to the upper flux cut. As a result, we obtained two samples of ~ 2000 TGSS-like and ~ 5000 NVSS-like sources, respectively.

The simulator generates five types of radio sources: *i)* star forming galaxies (SFGs); *ii)* radio quiet quasars (RQQs); *iii)* Fanaroff-Riley class I sources (FRI); *iv)* Fanaroff-Riley class II sources (FRII) and *v)* GHz-peaked radio sources (GPSs). Their redshift distributions in the simulated TGSS and NVSS catalogues are shown respectively in the upper and lower panels of Fig. 5.7 together with the cumulative $N(z)$ (thick line). In both catalogues the counts are dominated by FRI and FRII-type radio sources. The distribution of FRI objects peaks at $z \sim 0.6$ and dominates the counts at $z < 1$. The distribution of FRII objects is much broader and dominates the counts at higher redshifts. The number of SFGs and GPS objects is much smaller. However, being concentrated in the local universe, they represent a significant fraction of the counts at $z \leq 0.1$. Radio quiet quasars are also comparatively rare and have a very broad distribution, being a sub-dominant population at all redshifts.

---

[6] http://s-cubed.physics.ox.ac.uk/

This $N(z)$ model, which we refer to as S^3, is the one adopted to predict the APS of both the NVSS and TGSS samples. It is implemented in the form of a step function with the same bin size $\Delta z = 0.1$ used in Fig. 5.7.

5.3.2 Linear Bias Models

The linear halo bias parameter $b(z)$ (defined in Ch. 2, Eq. (2.44)) has a significant impact on the APS model since it amplifies or decreases the clustering signal of the various populations at different redshifts. Direct estimates of linear bias based on cross-matches with CMB lensing convergence maps (Allison et al., 2015), spectroscopic/photometric redshift catalogues (Lindsay et al., 2014) or by joining the lensing and the clustering information (Mandelbaum et al., 2009) are few, limited to small samples and, therefore, coarsely trace the bias evolution. In this analysis in order to appreciate the impact of bias model uncertainties on the APS prediction we decided to explore four different, physically motivated, bias models taken from the literature. They all assume a deterministic, linear bias that evolves with time (redshift).

- *Halo Bias model [HB]*. This bias prescription relies on the halo model and assumes that radio sources are hosted in dark matter halos of different masses (and biases). Because of the rarity of radio sources, we assume that halos can host at most one radio source, located at their centre. For consistency, we also assume that the radio sample contains the same classes of sources as in the $N(z)$ model. We make some hypotheses on the halo host: we adopt the Sheth-Mo-Tormen(SMT) halo bias model, $b_h(M,z)$, defined in Eq. (2.46). and assume that the masses of the halos that host a given source type are Gaussian distributed around a typical mass $\hat{M}$ with a standard deviation $0.2\,\hat{M}$ (Ferramacho et al., 2014). Indicating the Gaussian distribution as $G(M,\hat{M})$, we estimate the bias of each type i of radio source as

  $$b_i(z) = \int_0^\infty G_i(M,\hat{M}_i)\, b_h(M,z)\, \mathrm{d}M\,. \tag{5.1}$$

  The values of $\hat{M}_i$ are also taken from Ferramacho et al. (2014): $\hat{M}_{\rm SFG} = 1\times10^{11}M_\odot$, $\hat{M}_{\rm RQQs} = 3\times10^{12}M_\odot$, $\hat{M}_{\rm GPS} = \hat{M}_{\rm FRI} = 1\times10^{13}M_\odot$ and $\hat{M}_{\rm FRII} = 1\times10^{14}M_\odot$.

  The current implementation of `CLASSGal` does not allow one to specify different analytic bias functions $b_i(z)$ for the different source types. To circumvent this problem we approximate each $b_i(z)$ with a step function with the same binning as

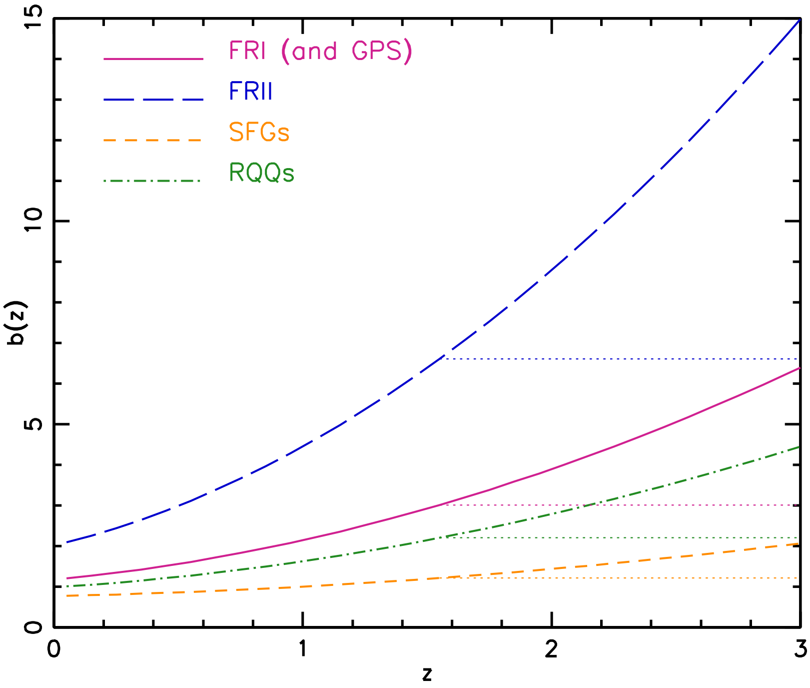


Fig. 5.8 Model linear bias evolution for different type of Radio sources. FR II (blue dashed curve) sources are most biased amongst others. The horizontal lines are $b_i(z > 1.5) = b_i(z = 1.5)$ cut-off margin for truncated bias models THB and TPB.

$N_i(z)$, and compute the effective *bias function* of the catalogue:

$$b_{\text{eff}}(z) = \frac{\sum_i N_i(z) b_i(z)}{\sum_i N_i(z)}. \tag{5.2}$$

We then feed the `CLASS` code with an effective redshift distribution of the objects $\tilde{N}(z) = b_{\text{eff}}(z) \times N(z)$. Next we estimate the effective *bias parameter* of the sample $b_{\text{eff}} = \sum_i \sum_j b_i(z_j) N_i(z_j) / \sum_i N_i(z_j)$ and feed this single parameter to the code as the linear bias of the whole sample. In Figure 5.9 we show the effective *bias function* $b_{\text{eff}}(z)$ of the NVSS (top) and TGSS (bottom) catalogues for all the models explored and, in particular, for the HB model (purple, continuous curve).

This somewhat cumbersome procedure is analogous to using a normalised redshift distribution $\hat{N}(z_j) = \sum_i N_i(z_j)/\sum_i \sum_j N_i(z_j)$ and a normalised biasing function $\hat{b}(z_j) = \sum_i b_i(z_j)N_i(z_j)/\sum_i \sum_j N_i(z_j)$ as input parameters to `CLASSGal`.

- *Truncated Halo Bias model [THB]*. Some previous analyses (e.g., Tiwari and Nusser 2016) have assumed a truncated bias evolution in which the halo bias does not increase indefinitely with the redshift but remains constant beyond $z = 1.5$, that is, $b_i(z > 1.5) = b_i(z = 1.5)$. Although this is clearly a rough approximation and there is no compelling theoretical reason to justify an abrupt cut on the bias at high redshift, we also consider this possibility for the sake of completeness and as a robustness test. In Fig. 5.9 this model is represented by the blue short-dashed curve. For different radio sources, the HB (dashed and dotted curves) and THB (horizontal dotted lines) are represented in Fig. 5.10.

- *Parametric Bias model [PB]*. Tiwari and Nusser (2016) proposed a parametric bias model for the NVSS sources also used by Bengaly et al. (2018) to model the TGSS bias. The parameters of the parametric models, specified in these works, have been determined by best-fitting the number counts and angular spectra of the radio sources.

  This parametric model relies on the physical model proposed by Nusser and Tiwari (2015). Here we prefer to avoid using their parametric expression since the parameters were derived by constraining the correlation properties of the sources that they were investigating. Instead, we follow the original Nusser and Tiwari (2015) approach and: *i)* assume that the radio activity is a strong function of the stellar mass; *ii)* adopt the expression provided by Nusser and Tiwari (2015) to quantify the fraction of radio sources as a function of stellar mass and redshift: $\mathscr{F}(M_*, z)$, where M_* is stellar mass; and *iii)* assume that the stellar mass is related to the halo mass as proposed by Moster et al. (2013). We use the analytical expression presented in Eqs. 14 and 15 of Nusser and Tiwari (2015), and assume that radio sources can be accounted for the function of stellar mass. The value of $b(z)$ is quantified by minimum stellar mass that is the lower limit of the integral in Eq. (5.2) To do this we use the SMT bias model (Eq. (2.46) halo bias model and the halo mass function of Jenkins et al. (2001). In this framework the difference between the TGSS and NVSS bias is determined by the choice of the minimum halo mass that can host a radio source, which sets the lower limit of the integration. For the NVSS case we adopt $1.4 \times 10^{11} M_\odot$, as in Nusser and Tiwari (2015), whereas for TGSS, which contains brighter objects, we use $10^{12} M_\odot$. However, as we have verified, this bias model is

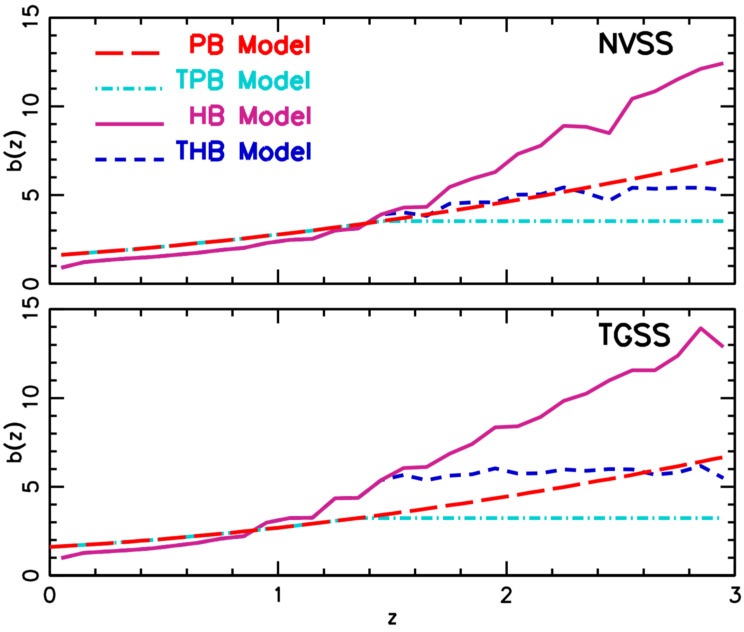


Fig. 5.9 Effective bias function (Eq. 5.2) for all the models listed in Table 5.2. The different bias models have different line styles, as indicated in the label. Top panel: Bias function of the *Reference* NVSS catalogue. Bottom panel: Bias function of the *Reference* TGSS catalogue.

not very sensitive to the choice of this minimum mass. The effective halo bias of the PB model is represented by the red, long-dashed curves in Fig. 5.9.

- *Truncated Parametric Bias model [TPB]*. It is the same as the PB model but, like in the THB case, we assume no bias evolution beyond $z = 1.5$ The corresponding bias function is shown as the light blue, dot-dashed curve in Fig. 5.9. We show the impact on minimum stellar mass on the α_0 and α_1 parameters defined by (Nusser and Tiwari, 2015) in Fig. 5.10. The red and black lines represent minimum masses used for TGSS and NVSS and are coinciding for redshifts $z > 0.5$. Instead, for more massive objects, the bias is significantly different (blue and green lines).

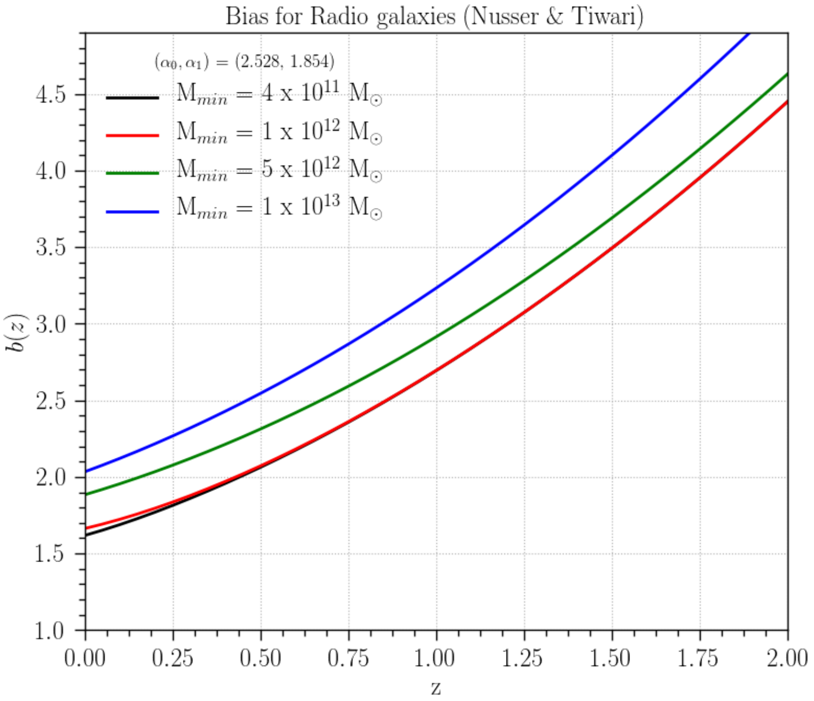


Fig. 5.10 Model bias evolution as a function of minimum mass. The minimum masses used for TGSS and NVSS are represented by red and black curves. The parametric values are taken from (Nusser and Tiwari, 2015).

5.4 Models vs. Measurements: χ^2 Analysis

In Figures 5.11 and 5.12 we compare the measured NVSS and TGSS angular power spectra shown earlier in Fig. 5.5, with the APS models described in Sect. 5.3.

Already the visual inspection reveals that none of the APS models succeed in reproducing the angular power of TGSS sources at $\ell \leq 30$. The magnitude of the mismatch is remarkable indeed. To quantify the discrepancy we have computed the reduced χ^2 in two intervals: $\ell = [2,30]$ (χ^2_{30}) to focus on the range in which the mismatch is larger and $\ell = [2,100]$ corresponding to the full multipole range considered in our analysis ($\chi^2_{\rm TOT}$). The χ^2 was evaluated as follows (e.g. Dodelson 2003):

$$\chi^2 = \sum_{\ell_1,\ell_2} (C_{\ell_1} - C^M_{\ell_1}) \mathscr{C}^{-1}_{\ell_1,\ell_2} (C_{\ell_2} - C^M_{\ell_2}) \tag{5.3}$$

where C_ℓ corresponds to the measured APS and C^M_ℓ to the APS model. We assume that the covariance matrix is diagonal, that is, $\mathscr{C}_{\ell_1,\ell_2} = \sigma_{C_{\ell_1}} \delta_{\ell_1,\ell_2}$, where σ_{C_ℓ} is the Gaussian error

Sample	$b(z)$	$\chi^2_{30}/$d.o.f. ($Q = P(> \chi^2)$)	$\chi^2_{\rm TOT}/$d.o.f.
NVSS	HB	1.34 (0.25)	1.83
	THB	1.64 (0.16)	1.21
	PB	0.61 (0.65)	1.62
	TPB	0.66 (0.62)	1.30
TGSS	HB	9.40 (4.5×10^{-7})	3.18
	THB	9.62 (2.9×10^{-7})	3.18
	PB	9.36 (4.7×10^{-7})	3.09
	TPB	9.42 (4.3×10^{-7})	3.10

Table 5.2 Angular power spectrum model parameters used in the χ^2 analysis and results. Col. 1: Type of catalogue. Col. 2: Bias model (see text for the meaning of the acronyms). Col. 3: reduced χ^2 value obtained when considering the multipole range $\ell = [2,30]$ and the probability $Q = P(> \chi^2)$. Col. 4: reduced χ^2 value obtained when considering the full multipole range $\ell = [2,100]$. In all the cases the redshift distribution $N(z)$ is based on the S^3 simulations as detailed in the text.

in Eq. (3.49). The sum runs over all Δ_ℓ bins from $\ell = 2$ to either $\ell = 30$ (χ^2_{30}) or $\ell = 100$ ($\chi^2_{\rm TOT}$). The number of degrees of freedom $N_{\rm d.o.f.}$ is set equal to the number of Δ_ℓ bins. The values of the reduced χ^2 are listed in Table 5.2 together with, for χ^2_{30} only, the probability $Q = P(> \chi^2)$.

We stress that here we are using the χ^2 statistics to quantify the goodness of the fit, assuming no free parameters in the model APS. The mismatch between prediction and measurement is so spectacular and the corresponding χ^2 value is so large that it is not worth performing a more rigorous maximum likelihood analysis that accounts for error covariance which, as we have argued, is expected to be small. This result clearly shows that none of the physically motivated APS models built within the ΛCDM framework can account for the excess TGSS power on large scales, also when one takes into account theoretical uncertainties, quantified by the scatter in model predictions.

The only possibility to match the measured large-scale power would be to advocate a population of relatively local and highly biased radio sources that, however, is neither supported by direct observational evidence nor by the results of the NVSS clustering analyses, which instead show that theoretical predictions match the measured APS, as visible in Fig. 5.11. The value of the reduced χ^2 for NVSS in Table 5.2 is close to unity for all models explored and quantifies the agreement in all the cases.

It is interesting to look at the differences between the APS models. At low redshifts the effective bias of the PB and TPB models is larger than that of the HB and THB ones. This, and the fact that the angular power on large scales is largely generated locally (see e.g., Fig.

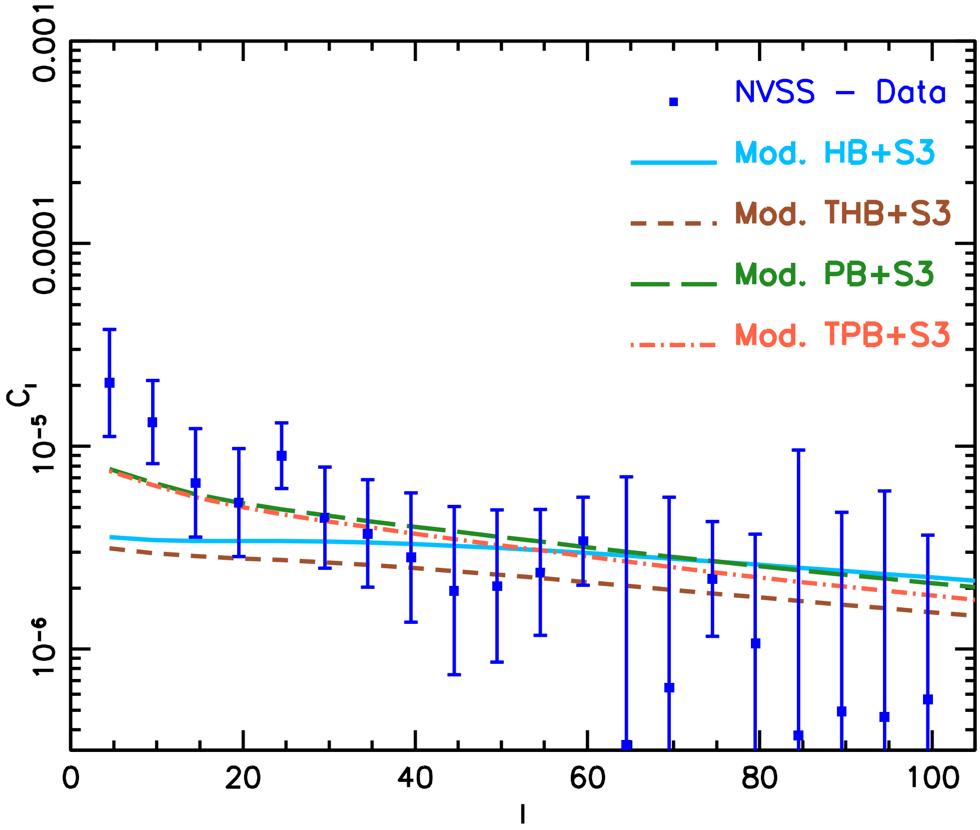


Fig. 5.11 Measured NVSS APS (blue squares from Fig. 5.5) vs. model predictions. The different models are listed in Table 5.2 and described in the text, and represented with different line styles, as indicated in the plot.

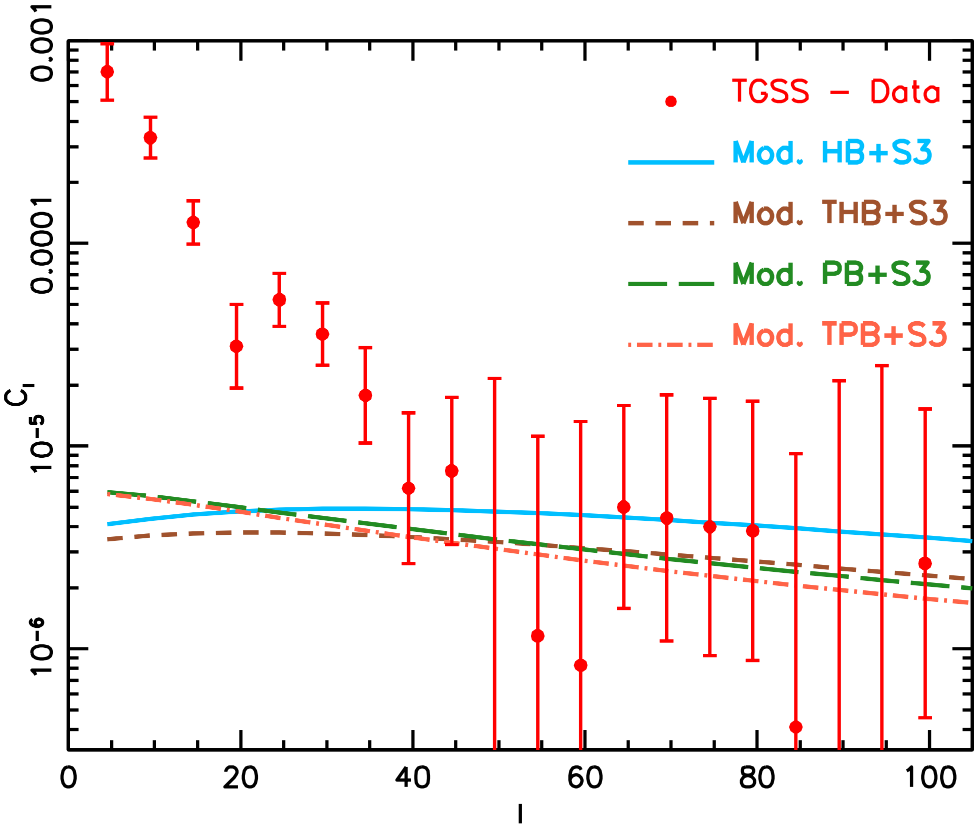


Fig. 5.12 As in Fig. 5.11 but for the TGSS sample. The measured APS (red dots) is compared to model predictions (continuous curves with different line styles).

7 in (Nusser and Tiwari, 2015)) , explains why the APS predicted by the PB models is larger than that predicted by the HB models at low multipoles, and why the former provide a better fit to the NVSS data. Also, truncating the bias evolution at $z = 1.5$ has very little impact on our results since distant objects, even if highly biased, are quite sparse and provide a shot-noise-like signal rather than producing coherent power on large angular scales.

It is worth pointing out that in this analysis we are considering the power within relatively large ℓ bins. Therefore, our result has no implication on the NVSS and TGSS dipole whose anomaly has been analysed in a number of previous works. In this respect, all we can infer is that if indeed the NVSS dipole is anomalously large, then our analysis implies that the TGSS dipole is even larger, in qualitative agreement with the conclusions of the Bengaly et al. (2018) analysis.

5.5 Discussion and Conclusions

In this chapter we have analysed the angular clustering properties of the radio sources in the First Alternative Data Release of the TGSS survey. Our analysis was performed in the harmonic space, to minimise error covariance and to facilitate the comparison with theoretical predictions, and focused on relatively large angular scales. This choice was motivated by the results of recent clustering analyses that revealed a large clustering signal (compared to that of the NVSS sources) at angular separations larger than $\Delta\theta \simeq 0.1°$ (Rana and Singh Bagla, 2018) and an anomalously large dipole amplitude, in clear tension with ΛCDM expectations (Bengaly et al., 2018). Our aim was to investigate the behaviour of the TGSS angular power spectrum at multipoles $\ell > 1$ and compare it with theoretical predictions, taking into account known observational and theoretical uncertainties. The clustering analysis of the TGSS sample was repeated on the NVSS catalogue and on a sample of TGSS objects with a NVSS counterpart. The rationale behind this choice was to compare our results with those of a well-studied sample that contains most of the TGSS sources distributed over a similar sky area.

The main results of our analysis are as follows.

- The vast majority of TGSS sources have a counterpart in the NVSS catalogue (about 94% when we consider our *Reference* samples) and are characterised by a spectral index Gaussian distributed around the value $\alpha_\nu \simeq -0.77$, similar to that of the NVSS sources and suggesting that the two catalogues contain similar classes of radio sources.

- The redshift distribution of TGSS sources extends well beyond $z = 0.1$, that is, the typical scale probed by galactic counterparts with measured redshifts (Rana and Singh Bagla, 2018). We proved this point by cross-matching TGSS sources with optically identified QSOs in the SDSS-DR14 catalogue. The fraction of cross-matched objects is small (~ 1.5 %) but sufficient to show that the distribution of TGSS sources extends beyond $z = 3$, like the NVSS sources (Nusser and Tiwari, 2015).

- The angular two-point correlation function of TGSS sources exhibits a double power-law behaviour, qualitatively similar to that of the NVSS sources. Although not surprising, this result was not discussed by Rana and Singh Bagla (2018) since they focused on angular scales larger than $0.1°$. In that range the amplitude of the TGSS ACF is larger than that of the NVSS. At small angles the behaviour of the ACF is determined by the presence of radio sources with multiple components. We analysed the behaviour of the ACF on these small scales to quantify the clustering signal produced by multiple components and subtracted it from the measured angular power spectrum.

- The angular spectrum of TGSS sources has significantly more power than that of the NVSS in the multipole range $2 \leq \ell \leq 30$. Beyond $\ell = 30$ the two spectra agree with each other within the errors. This mismatch is also seen when the TGSS×NVSS cross matched catalogue is considered instead of the TGSS one.

To check the robustness of this result to the known observational systematic errors we considered different TGSS samples obtained by varying the lower and upper flux selection thresholds and by using different geometry masks that exclude progressively larger regions of the sky near the Galactic plane. The measured APS is remarkably robust to these changes and the TGSS versus NVSS power mismatch remains significant even when going beyond the completeness limit of the TGSS catalogue. We did not explore the impact of errors in the flux calibration since these were found by Bengaly et al. (2018) to be small with respect to the magnitude of the mismatch.

Altogether these results excluded the hypothesis that the observed power mismatch could be attributed to known systematic errors related to the treatment of the data or to the observational strategy, and opened up the possibility that it may reflect genuine differences in the clustering properties of radio sources in the two catalogues.

To investigate this possibility we performed an absolute rather than a relative comparison between the measured TGSS angular spectrum and the one predicted in the framework of the ΛCDM model. In doing this we took special care in modelling all the physical effects that contribute to the clustering signal and in propagating model uncertainties. Among the physical effects, the ones that contribute the most to the large-scale clustering amplitude are the redshift space distortions, which can boost the correlation signal by $\sim 3\%$, and magnification lensing, which reduces the amplitude by $3-9\%$, depending on the multipole considered. These effects were generally ignored in previous analyses. Although not negligible, their amplitude is far too small to explain the anomalous TGSS power. Finally, we find that the use of the Limber approximation, which has been adopted in many of the previous APS analyses, would spuriously enhance the predicted APS amplitude by $7-15\%$, again depending on the multipole considered and being largest at $\ell < 5$.

The physical effects described above are well known and their contribution can be modelled with small errors. The largest uncertainties in modelling the TGSS spectrum are related to the composition of the catalogue, the redshift distribution of its sources and, most of all, their bias. To model the composition of the catalogue and the redshift distribution of each source type we used the SKA Simulated Skies tool and found that our *Reference* TGSS catalogue is mainly composed of FRII and FRI sources. Fainter radio objects like SFGs and GPS are comparatively fewer but very local, and therefore they represent a sizeable fraction of the TGSS population at $z < 0.1$. These objects are characterised by different redshift distributions and trace the underlying mass distribution with different biases.

The biases of these sources and their evolution is the single most uncertain ingredient of our APS model. To account for these uncertainties we considered four bias models. All of them assume a linear, deterministic biasing process and were conceived in the widely accepted framework of the "halo bias" model. All of them are physically plausible, as they were designed to match the observed radio luminosity functions and number-count statistics. They differ from each other in the evolution of the bias beyond $z = 1.5$ and in the relation between the radio sources and the mass of the host halo.

With all the ingredients and hypotheses previously described we generated four models for the APS of TGSS sources, and none of them are able to match the observed power at low multipoles. In fact the tension between models and data at $\ell \leq 30$ is so large that it makes a sophisticated error analysis unnecessary. Our simple χ^2 estimate is sufficient to reveal that the observed TGSS angular power spectrum cannot be generated within the framework of a ΛCDM model: none of the physical effects described are large enough to generate such a signal and none of the hypotheses on the nature, distribution and bias of TGSS radio sources can be stretched enough to simultaneously satisfy the luminosity function and the clustering properties of these sources.

It is remarkable that, instead, our models match the angular spectrum of NVSS galaxies in the same range $2 \leq \ell \leq 30$ once the observational errors and theoretical uncertainties are taken into account. This result confirms that our APS models are indeed physically viable.

We are left with the uncomfortable evidence of an excess large-scale clustering in the angular distribution of the TGSS ADR1 sources. The excess is seen both in the comparison with similar analyses carried out on the NVSS dataset which shares many similarities with TGSS, and in the comparison with theoretical predictions. In Section 5.2.2 we searched for possible observational effects that may generate a spurious clustering signal large enough to explain the tension detected by our study but failed to identify an obvious candidate.

There is an obvious continuity between our results and those of Bengaly et al. (2018) who detected an anomalous large amplitude in the $\ell = 1$ dipole moment of the TGSS angular spectrum. For this reason we agree with their conclusion that the observed mismatch indicates the presence of unidentified systematics in the data not captured by the ones that we have explicitly searched for in this study. It may be that this issue can only be clarified with future TGSS data releases or thanks to other forthcoming wide-angle radio surveys carried out at similar frequencies like the ongoing LOFAR Two-Metre Sky Survey (LOTSS; Shimwell et al. (2017, 2018)) and/or future, deeper releases of the GLEAM catalog (Hurley-Walker et al., 2017; White et al., 2018).

Chapter 6

Constructing Mock Catalogue Pipeline

In chapters 4 and 5, we have carried out, two, rather typical types of clustering analyses: a 3D one, involving the use of spectroscopic samples (of clusters) and a 2D one using catalogues of objects for which only the flux is measured but not their distance. As anticipated, there is a third, somewhat intermediate case, in which the distance to the objects is known with large uncertainty: i.e. the photometric redshift (hereafter photo-z) catalogues. In this case, the limited amount of information along the radial direction advises against performing a full, time consuming 3D clustering analysis. On the other hand a plain 2D study would miss a significant amount of precious information about the rate at which cosmic structures evolve. The ’tomographic’ approach, in which 2D auto- and cross- correlation analyses of the objects’ distribution within and across spherical shells, provides a good compromise, that allows one to retrieve interesting, linear information, along the radial direction together with 2D angular clustering at different epochs. The interest in this type of analyses has been triggered by the recent availability of large datasets from Dark Energy Survey (DES, Abbott et al., 2005); Kilo Degree Survey (KiDS, de Jong et al., 2012) or the 2-MASS Photometric Redshift Catalogue (2MPZ, Bilicki et al., 2013) that have been analysed in the recent years (Asorey et al., 2012; Alonso et al., 2015; Balaguera-Antolínez et al., 2018) also in combination with spectroscopic data (Cai and Bernstein, 2012; Salazar-Albornoz et al., 2014). The importance of tomographic analysis is bound to increase with the advent of the much larger datasets that will be released by from projects such as *Euclid* and *LSST*.

In principle, photo-z surveys, in combination with high resolution imaging, are mainly designed for performing weak-lensing analyses, since they will allow us to measure the properties of cosmic shear fields and their two-point statistics at various redshifts, and hence to constrain the total mass distribution over very large volumes. However, since shear comes with the estimate of angular position and photo-z’s of individual objects, these catalogues can also be exploited for pure position-based clustering analyses as well as for combined weak-lensing and spatial clustering correlation studies. In this chapter I will focus on the spatial clustering aspect, and will only mention the potential advantages of combining the two analyses at the end of the chapter.

The main advantage of photo-z samples over the spectroscopic one is the statistics. Next generation photo-z catalogues will contain billions of objects, i.e. a factor of ~ 100 more objects than in future spectroscopic samples. This advantage poses, however, a significant number of technical problems to be faced in order to fully exploit these samples. The most important of which is the need to quantify systematic and statistical errors and their covariance for such large datasets. Procedures that use the datasets itself, like *Jackknife* or *Bootstrap*, are not suitable for investigating the origin and the impact of systematic errors and, more important, are not good enough to provide an accurate assessment of errors (Friedrich et al., 2016; Shirasaki et al., 2017). This is why most of the recent studies rely on mock, realistic catalogues that mimic the characteristics of the datasets to estimate the errors and their budget. Ideally, one would like to use full N-body numerical experiments to build such mocks, since in this case all relevant nonlinear physics can, in principle, be accounted for in the error budget. However, the sheer number of different mock catalogues required to fully account for the cosmic variance and build a reliable covariance matrix is simply too large to make this approach feasible. The next best approach is to use simplified methods which are computationally faster than N-body simulations. An approximated mechanism is required to generate independent catalogues of synthetic objects, later in which, different observational selection effects (and their uncertainties) can be included. In this case, the mock catalogues are built to carry out statistical analyses, construct the covariance matrix and assess the precision and accuracy with which the relevant cosmological parameters can be estimated.

In this chapter I will describe a pipeline to build a large set of mock catalogues mimicking the next generation photo-z samples suitable for tomographic analysis. I will consider the case of the upcoming Euclid photometric survey as a case study. The pipeline I have assembled combines, in an efficient way, different elements obtained from various numerical cosmological codes that are publicly available. The main element of the pipeline is the so-called `FLASK`[1] (*Full-sky Lognormal Astro-fields Simulation Kit*, Xavier et al., 2016) code to generate correlated Lognormal random fields on a set 2D-spheres sharing a common origin. The use of such code obviously relies on the assumption that the probability distribution function of galaxy counts is well approximated by a Lognormal distribution (Coles and Jones, 1991). The goodness of this assumption will be discussed in the following sections. Other elements of the pipeline are: the Boltzmann solver `CLASS` to generate the

---

[1] http://www.astro.iag.usp.br/~flask

theoretical angular power spectra of the fields in the spherical shells; the `HEALPix` package to generate Gaussian realisations and maps in spherical shells; an original package that I have programmed and optimised to include survey-specific, observational effects like photo-z errors and angular masks etc. and, finally, the angular power spectrum estimator [known as the Pseudo-C_ℓ (`PCL`) estimator] that is currently being developed and has been selected to become the official code for the measurement of the angular spectrum of the Euclid galaxies.

This chapter is organised as follows: In § 6.1, I discuss the Lognormal hypothesis and provide the theoretical background to generate correlated Lognormal field. I also introduce the basic structure and the flowchart of the pipeline in this section. In § 6.2 I describe the first stage of the pipeline, i.e. the framework to generate Lognormal mock catalogues. I start by providing an overview of the `Stage-I` computational framework and then describe in detail its main elements. To validate the elements of the first stage I have performed a number of "unit" tests, that I briefly describe in the end of the section. Later in § 6.3, the second stage of the pipeline, primarily consisting of the original package 'survey mock generator'. In this code, I apply properties that are most dataset-dependent, i.e. photo-z information and corresponding observational errors. At this stage, I also append the survey properties like redshift distribution, mask etc. to the datasets, in order to make them "realistic" ones, as those expected from the survey. At the end of this section I include the results from `Stage-II` validation using `PCL` estimator. The computational budget, pipeline 'run' performance and optimisation techniques are presented in § 6.4 together with a discussion on future features to be added to the pipeline.

6.1 Introduction

6.1.1 Lognormal Fields

To outline the properties of a Lognormal density field and its relation with the Gaussian one I follow the work from Kayo et al. (2001). The reader is advised to refer the paper for detailed description and derivation.

It is well known that a Lognormal 1-point probability distribution function (PDF) given by,

$$\mathbb{P}^{(1)}_{LN}(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left[\ln(1+\delta)-\mu\right]^2}{2\sigma^2}\right\} \frac{1}{1+\delta}\,, \tag{6.1}$$

provides a very good description of the mass overdensity field, δ, measured in N-body simulations well into the nonlinear regime (Kayo et al., 2001; Lahav and Suto, 2004; Joachimi et al., 2011; Neyrinck, 2011; Seo et al., 2012). The field δ is assumed to be filtered on scale R. This PDF only depends on the variance of the field $\ln(1+\delta)$ since $\mu = -\sigma^2/2$ to ensure that expected value $\langle\delta\rangle = 0$. The variance of the nonlinear field δ can be obtained by filtering the full nonlinear power spectrum $P_{\rm nl}(k)$ (very similar to linear one in Eq. 2.13):

$$\sigma^2_{\rm nl}(R) \equiv \frac{1}{2\pi^2}\int_0^\infty P_{\rm nl}(k)\, \tilde{W}^2(kR)\, k^2 dk\,, \tag{6.2}$$

where the subscript 'nl' indicates nonlinear quantities and $\tilde{W}^2(kR)$ represents the smoothing filter. The variance σ in Eq. 6.1 is related to the variance of δ field through:

$$\sigma^2(R) = \ln\left[1+\sigma^2_{\rm nl}(R)\right]. \tag{6.3}$$

Much of the theoretical interest in the lognormal field stems from the fact that there is a one-to-one mapping between the linear Gaussian and the nonlinear Lognormal field. As we have seen, a linear overdensity field $\delta_{\rm lin}$ smoothed on the (same) scale R obeys a 1-point PDF Gaussian statistics (same as Eq. (2.4)):

$$\mathbb{P}^{(1)}_{G}(\delta_{\rm lin}) = \frac{1}{\sqrt{2\pi\sigma^2_{\rm lin}}} \exp\left\{-\frac{\delta^2_{\rm lin}}{2\sigma^2_{\rm lin}}\right\}, \tag{6.4}$$

where 'lin' indicates linear quantities and the variance of the linear density field is defined in analogy with 6.2, i.e. as given in Eq. 2.13. The overdensity field defined from $\delta_{\rm lin}$

$$1+\delta = \frac{1}{\sqrt{1+\sigma_{\rm nl}^2}} \exp\left[\frac{\delta_{\rm lin}}{\sigma_{\rm lin}}\sqrt{\ln\left(1+\sigma_{\rm nl}^2\right)}\right] , \tag{6.5}$$

has a PDF that can be computed with $(d\delta_{\rm lin}/d\delta)\mathbb{P}_G^{(1)}(\delta_{\rm lin})$ and corresponds to the Lognormal differential PDF in Eq. 6.1. The Lognormal cumulative PDF is given by,

$$\mathbb{C}_{LN}^{(1)}(\delta) = {\rm erf}\left(\frac{\ln(1+\delta)-\mu}{\sigma}\right) , \tag{6.6}$$

where ${\rm erf}(x)$ indicates the error function

$${\rm erf}(x) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-t^2/2}\, dt\, . \tag{6.7}$$

The fact that, the Lognormal model provides a good fit to the PDF mass overdensity field measured in numerical experiments, and that its variance can be related to that of the Gaussian fields (Chiang et al., 2013; Alonso et al., 2015), makes the use of the Lognormal approximation very useful for quickly generating large number independent realisations of cosmological fields. Also consequently to create the mock datasets that imitate large, already existing as well as next generation catalogues of objects' positions and lensing shear measurements.

So far we have considered a single Gaussian variable $\delta_{\rm lin}$ and its Lognormal counterpart but the same procedure can be used to define a multivariate Lognormal random variable X_i from a multivariate Gaussian variable Z_i with known mean vector and covariance elements related through:

$$X_i = e^{Z_i} - \lambda_i\, , \tag{6.8}$$

where the quantity λ 'shifts' the distribution to ensure zero mean of the underlying Gaussian field. For the case of cosmological overdensity fields, which is of interest here, the variable $X_i = 1+\delta_i$ represents the Lognormal field, $\lambda_i = 0$ and the constraint that $\langle X_i\rangle = 0$ implies that $Z_i = \delta_{{\rm lin},i} - \sigma_{{\rm lin},i}^2/2$ is Gaussian distributed with zero mean and variance $\sigma_{{\rm lin},i}^2 = \xi_{\rm g}^{ii}$. The quantity $\xi_{\rm g}^{ij}$, which represents the covariance matrix of the Gaussian density field $\delta_{{\rm lin},i}$ fields, is related to the covariance matrix of the Lognormal field δ_i through

$$\xi_{\rm g}^{ij} = \ln\left(\frac{\xi_{\rm ln}^{ij}}{\alpha_i\alpha_j}+1\right) \tag{6.9}$$

where $\alpha_i = \langle Z_i \rangle + \lambda_i > 0$ and $\langle Z_i \rangle$ being the ensemble average of Lognormal variables. For more specific and detailed derivation, referring to Xavier et al. (2016) is recommended. Eq. (6.9) constitutes one of the most important relations, that is used to generate correlated density fields across different spherical shells.

6.1.2 Pipeline Overview

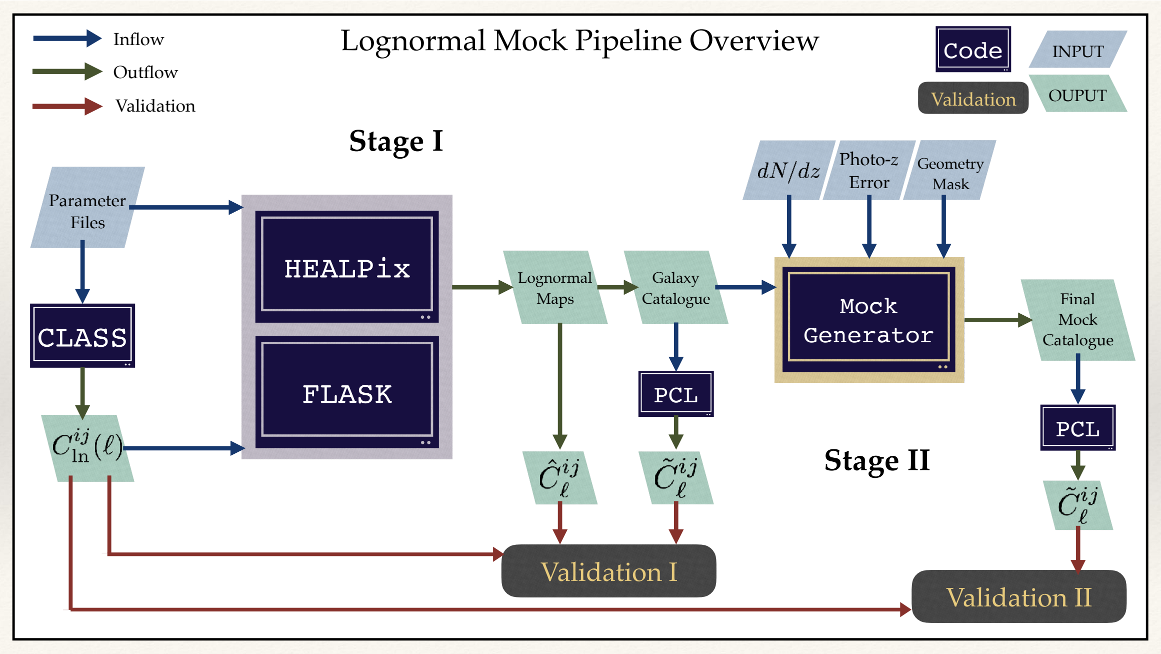


Fig. 6.1 Flowchart of the photo-z mock pipeline showing the flow of input information (light blue parallelograms) through the computational infrastructure, which is represented with dark boxes with names of the associated routines, producing the output files (light green parallelograms). The auto- and cross- angular power spectra, that are validated at the bottom of the flowchart are represented with: *i)* input Lognormal spectra as $C_{\rm ln}^{ij}(\ell)$'s; *ii)* Map recovered spectra from `FLASK` as $\hat{C}_\ell^{ij}$'s and *iii)* `PCL` estimated spectra from output catalogues as $\tilde{C}_\ell^{ij}$'s. dN/dz is the radial survey selection function.

Since this thesis work is a part of *Euclid* collaboration, the pipeline and its individual elements are selected in tandem with the collaboration. The general overview of the pipeline's flow and the important computational elements are highlighted in Fig. 6.1. As introduced earlier, the pipeline's first stage consists of three publicly available repositories, `FLASK`, `CLASS` and the `HEALPix`, which I shall elaborate in context of their usage for this pipeline in the following sections. The first stage creates archetype Lognormal mock catalogues populated with galaxies, which are validated using `PCL` at the end of the stage, and passed further to second stage in order to impose characteristic properties of the target photo-z survey.

The second stage of the pipeline involves assigning the photo-z and its uncertainty to the galaxies, applying the radial survey selection function, and finally the footprint of the survey that is composed in the form of a binary mask. As it can be noticed in the figure, similar to `Stage-I`, the output from `Stage-II` is also passed through the validation procedure, to assess the accuracy and quality of the final output catalogue. In rest of the chapter, I shall expand these stages in details and present the relevant results.

6.2 Stage-I: Generating Ideal Mocks

6.2.1 FLASK: Code Overview

The core and the backbone of this pipeline is the publicly available `FLASK` code developed by Xavier et al. (2016). This multi-purpose code is already well structured as a pipeline. So the task here is to channelise `FLASK` for this pipeline and construct further by adding new elements to it, check their functioning and validate the final results. The entire code is structured in `C++` for rapid creation of Lognormal mocks. `FLASK` is designed to generate correlated 2D as well as 3D scalar fields of cosmological interest like the mass density and the shear lensing field. Once the underlying cosmological model is specified, `FLASK` generates a realisation of these fields that obey either a Gaussian or a Lognormal statistics. These fields can be `Monte Carlo` sampled to generate catalogues of discrete objects like galaxies. Here, I focus on the generation of correlated 2D cosmological density fields with lognormal 1-point PDF.

Fig. 6.2 shows the complex structure of the `FLASK` code. The various boxes indicate the processing elements of the code (blue boxes), input data (yellow boxes) to be provided by the user and that, in this case, are generated by additional pipeline elements, and finally output data (green boxes) that I use to run validation and unit tests to check the quality of the results. The light-blue boxes are specific to the creation of the Lognormal mocks. The data flow normally follows the black arrows. It must be noted that the flowchart shown in Fig. 6.2, differs from the original flowchart of entire package[2], and displays only the `FLASK` processes that are used by our pipeline, since we focus on the creation of density fields and position-only galaxy catalogues and all other processes specific to the shear fields are omitted.

---

[2] http://www.astro.iag.usp.br/~flask/documentation

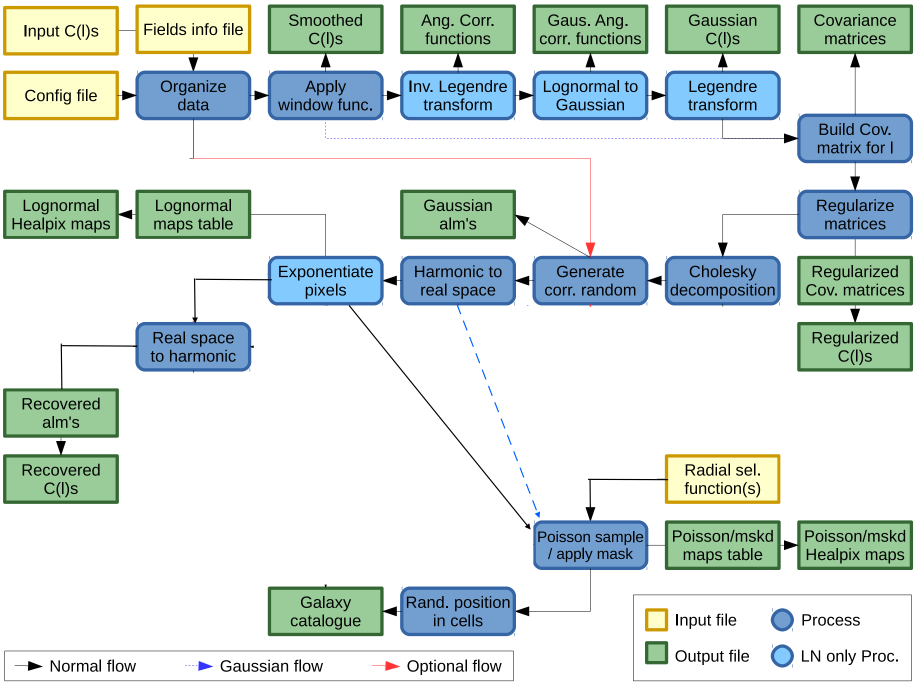


Fig. 6.2 `FLASK` flowchart describing the important `I/O` processes and functions for this pipeline. The detailed description is explained in the text.

Before describing in detail the individual steps of the `FLASK` processing functions, we briefly summarise the steps for the creation of a Lognormal mock catalogues.

1. Prior to running the main module of the code, following parameters are specified: (i) the type of the density field; (ii) the number of redshift slices with edges of each slice i.e z_{min} and z_{max}; and (iii) the mean value and shift parameter (Eq. (6.8)), which are tabulated in an external file '`XXX-fields-info.dat`';

2. The input model auto- and cross- angular power spectra $C^{ij}_{\mathrm{ln}}(\ell)$ are generated using external code `CLASS`, that defines the characteristic properties of the Lognormal density field;

3. The above information is provided to the configuration file '`XXX.config`', and other parameters that are essential for the output maps are specified in the configuration file: (i) A random `int` type variable via `RNDSEED` which results in an independent realisation; (ii) An integer with a base of 2 defines the resolution of output maps through `NSIDE`; (iii) Regularisation options for processing the covariance matrix. At the end the output options (green boxes in Fig.6.2) are specified by enabling or disabling output files. After setting all the essential parameters, the code is

initiated with input command ‘`flask XXX.config`’, here the code organises the input information and proceeds to next process function;

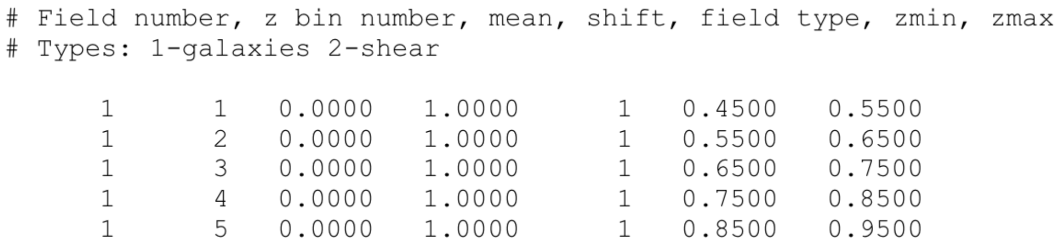
4. `HEALPix` pixel window function is applied to the input angular power spectra, for smoothing the power to account for the size and the shape of the pixel corresponding to the resolution set by `NSIDE` parameter;
5. The initial goal is to obtain the Gaussian covariance matrix. The smoothed ‘Lognormal’ spectra $C^{ij}_{\rm ln}(\ell)$’s are then transformed to Lognormal correlation function $\xi^{ij}_{\rm ln}(\theta)$ using the inverse Legendre transform;
6. Using the relation (6.9)), the Gaussian correlation function $\xi^{ij}_{\rm g}(\theta)$ is obtained. By performing Legendre transformations on $\xi^{ij}_{\rm g}(\theta)$, give the Gaussian angular power spectra $C^{ij}_{\rm g}(\ell)$, which builds the Gaussian covariance matrix;
7. The covariance matrix is regularised, i.e. checked for being ‘positive-definite’, using the regularisation method provided in configuration file. This step is meant to restore the small scale power that is lost in the Lognormal transform. The effect is significant for the shear analysis, that uses very high multipoles, but much less so for clustering analysis, i.e. that focuses on larger angular scales;
8. Regularised covariance matrix undergoes Cholesky decomposition to generate correlated Gaussian harmonic coefficients $A^i_{\ell m}$’s (with zero-mean and unit variance). The Cholesky decomposition gives out correlated Gaussian fields, to account for the correlation among shells, i.e. to correctly mimic the cross-angular power spectra between the neighbouring shells;
9. The obtained correlated Gaussian spherical harmonics coefficients, $A^i_{\ell m}$’s, are used to generate maps, and are stored in separate output files. Pixelised density maps are generated using `ALM2MAP` function of the external package `HEALPix` with the map resolution set by `NSIDE`;
10. The pixelised density field is exponentiated using (6.8), to obtain a Lognormal density field. The corresponding maps are saved to output file;
11. At this step, the Lognormal angular spectra are measured by recovering $A^i_{\ell m}$’s from the realisation maps. The angular auto- and cross- spectra are computed from these coefficients by calling `MAP2ALM` function of `HEALPix` and stored;
12. The Lognormal maps are Poisson sampled to generate a discrete catalogue of objects.

13. A mock catalogue is produced that specifies the angular positions and the redshift of the objects in the shell. Because of the map's finite resolution, and since no information is provided on the radial distribution of objects within the shell, the angular position of the objects is randomly assigned within each pixel and its redshift is drawn from a uniform distribution within the redshift interval of each shell.

We elaborate the above described points in following subsections.

6.2.2 Input 1: Parameter Files

The very first step that is required by `FLASK` is to set the density field type and the number of redshift slices of the target Lognormal field. The file '`XXX-fields-info.dat`' is a text file which sets these properties. Fig. 6.3 shows the example of the input parameters list through this file. One is also required to provide the values for mean and shift of the Lognormal field as described in Eq. (6.8). For the choice of Lognormal fields, these parameters are set to their default values 0 and 1 respectively. Last but not the least, the bin-edges of the redshift slice i.e. $z_{\rm min}$ and $z_{\rm max}$, determine the width, $\Delta z = |z_{\rm max} - z_{\rm min}|$, of the slice. As it can be seen from the figure, the file will be used to generate galaxy density field consisting of 5 redshift slices of width $\Delta z = 0.1$ with mean redshift z ranging from $[0.5, 0.9]$.



```
# Field number, z bin number, mean, shift, field type, zmin, zmax
# Types: 1-galaxies 2-shear

      1       1    0.0000    1.0000       1    0.4500    0.5500
      1       2    0.0000    1.0000       1    0.5500    0.6500
      1       3    0.0000    1.0000       1    0.6500    0.7500
      1       4    0.0000    1.0000       1    0.7500    0.8500
      1       5    0.0000    1.0000       1    0.8500    0.9500
```

Fig. 6.3 A screenshot of an example fields-information file. As it can be noticed from the first line of the file, which is commented with '#', the parameters that go in the columns are: field number, serial number of the redshift shell, mean of Lognormal distribution, shift of Lognormal distribution, type of the field (numbers specified in the second commented line of the file), and last two columns correspond to redshift interval of the shell with $z_{\rm min}$ and $z_{\rm max}$ values.

The `FLASK` configuration file '`XXX.config`' is the main text file which contains all the information about the processing algorithm and its parameters. It controls the entire `I/O` of the code, which can be tuned according to user's choice. First important parameter is the random seed, an integer type variable, provided via '`RNDSEED`' option in `config` file, to generate independent realisation. The next is the set of input files containing the above described fields information file, all the auto- and cross- angular power spectra that we shall discuss next in § 6.2.3. The resolution of pixelised maps that will be produced by the code is fixed with `NSIDE` parameter, which goes in the `config` file.

6.2.3 Input 2: Angular Power Spectra

The second input to `FLASK`, i.e. the model angular power spectra must be produced by an external "Boltzmann solver" a shorthand for those numerical codes that, for a given cosmological model, predict angular 2D and 3D power spectra for a number of different fields (CMB temperature, density, cosmic shear etc.) also accounting for a number of cosmological effects (lensing, peculiar motions, presence of massive neutrinos etc.) that modify or modulate the fluctuation spectrum. A few of such computational routines which are publicly available: `CMBFAST` (Seljak and Zaldarriaga, 1996; Zaldarriaga and Seljak, 1998), `CMBEASY` (Doran, 2005) and very popular one `CAMB` (Lewis et al., 2000; Howlett et al., 2012) that are written in `FORTRAN90`. Recently developed `python` based routine `AngPow` (Campagne et al., 2017) is particularly aimed at computing auto- and cross- tomographic angular power spectra using Limber's approximation (refer to Eq. (3.44) in § 3.2.2 for details). However, in this work we shall utilise `CLASS` (*Cosmic Linear Anisotropy Solving System*) which is originally written in `C++` by Lesgourgues (2011a); Blas et al. (2011), and subsequently modified by various contributors [e.g. CLASSgal (Di Dio et al., 2013)], to account for all General relativistic effects including gravitational lensing and peculiar velocities. The fact that `CLASS` is also written in C++ has been the main motivation for its adoption.

For this pipeline, `CLASS` is used to generate, for a given cosmological model, angular auto- and cross- spectra in various redshift shells, for a set of objects with a given redshift distribution. To compute the power spectra, relevant parameters are specified in an input parameter file '`XXX.ini`' that contains:

- All relevant cosmological parameters that uniquely define the mass power spectrum of density fluctuation;
- Option to choose among the linear regime, i.e. generate a linear mass power spectrum, or a nonlinear one. In this second case, which is used for this task, `CLASS` relies on the semi-analytical '`HALOFIT`' model (Smith et al., 2003; Takahashi et al., 2012);

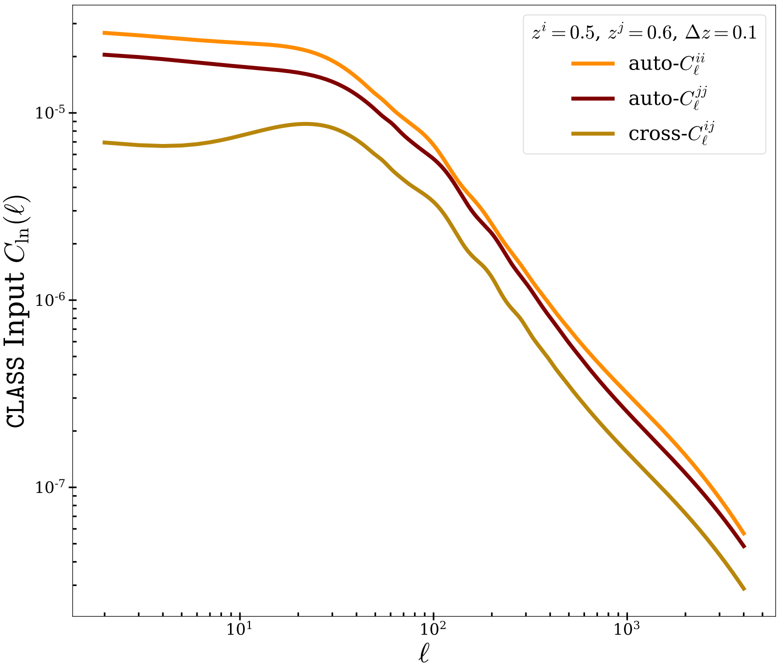


Fig. 6.4 Sample input power spectra at two redshifts $z^i = 0.5$ and $z^j = 0.6$ with orange and maroon curves showing auto-angular power spectra, generated using `CLASS`. Whereas the light-green curve is the cross-spectrum between i and j. The width of the Top-Hat redshift slice is set to $\Delta z = 0.1$.

- An option that specifies if and what types of general relativistic effects need to be accounted for. I only include peculiar velocities.
- The Limber approximation can be enabled if required. I will not activate this option in our applications.
- For specifying the properties of the redshift shells, their total number by providing mean redshifts, shell-width, and shapes are specified. Included options for defining the shape are Dirac Delta, Top-Hat and Gaussian. The shapes of the redshift shells correspond to the window function W_i as given in Eq. (3.39). In this pipeline (for now), a Top-Hat selection is used to specify shells in real space. Gaussian selection is used in photo-z space, to account for the impact of photo-z errors. In this application I ignore comparably smaller spectroscopic redshift (hereafter spectro-z) errors, i.e. I identify real space with spectro-z space.

- A radial selection function, dN/dz to specify the redshift distribution of objects in the target survey.
- The linear galaxy bias as a function of the redshift, $b(z)$, to generate maps of objects counts rather than mass density. The function $b(z)$ is included in the integrand of Eq. (3.39).

Fig 6.4 shows a typical output of `CLASS`. It displays the auto-angular spectra in two top-hat slices at redshift $z = 0.5$ and $z = 0.6$, respectively, together with their cross spectrum, characterised by a significantly lower amplitude. Since in this application, a single (density) field is considered, I generate, together with the N_z auto-spectra (N_z being he number for the shells) also $N_z(N_z - 1)/2$ cross spectra. All generated auto- and cross- input angular power spectra: $C^{ii}_{\rm ln}(\ell)$ and $C^{ij}_{\rm ln}(\ell)$ where i and j represent two different redshift slices.

6.2.4 Pixel Window Function

The pipeline also features another software package, already integrated with `FLASK`, which is mainly used for map generation. These maps are in fact, number of pixels distributed on the sphere, each with equal surface area. The NASA Jet Propulsion Lab developed tool, `HEALPix`[3] (acronym for Hierarchical Equal Area isoLatitude Pixelisation of a sphere) (Górski et al., 2005), is designed for pixelising spherical surfaces by subdividing it with pixels covering the same surface area as every other pixel.

In this pipeline, `HEALPix` is used for three different tasks. To create pixelised maps from a given set of spherical harmonics coefficients (`ALM2MAP` routine); to estimate spherical harmonic coefficients and their angular power spectrum from a given pixelised map (`MAP2ALM`) and to provide the window function associated to the pixels.

Pixelisation sets the angular resolution of the map. As anticipated, in the `HEALPix` notation the map resolution quantified by the `NSIDE` parameter (same as $N_{\rm side}$) which takes the numbers in binary form i.e 2^n. The number of pixels on the sphere are multiples of base level 12, and can be obtained from $12 \times N^2_{\rm side}$. For higher resolution i.e. $N_{\rm side} > 128$, the higher limit of maximum multipole is $\ell_{\rm max} = 4N_{\rm side}$. In this work, maps with resolution $N_{\rm side} = 2048$ are produced. Therefore, pixelising is analogous to smoothing and removes power on angular scales similar to and smaller than the size of the pixel. The effect can be seen in Fig. 6.5 that shows model C_ℓ before and after the application of pixel window function.

Since in configuration space the smoothing effect of the pixel is quantified by a convolution integral, it is more convenient to quantify its effect in harmonic space, where

---

[3] https://healpix.jpl.nasa.gov/

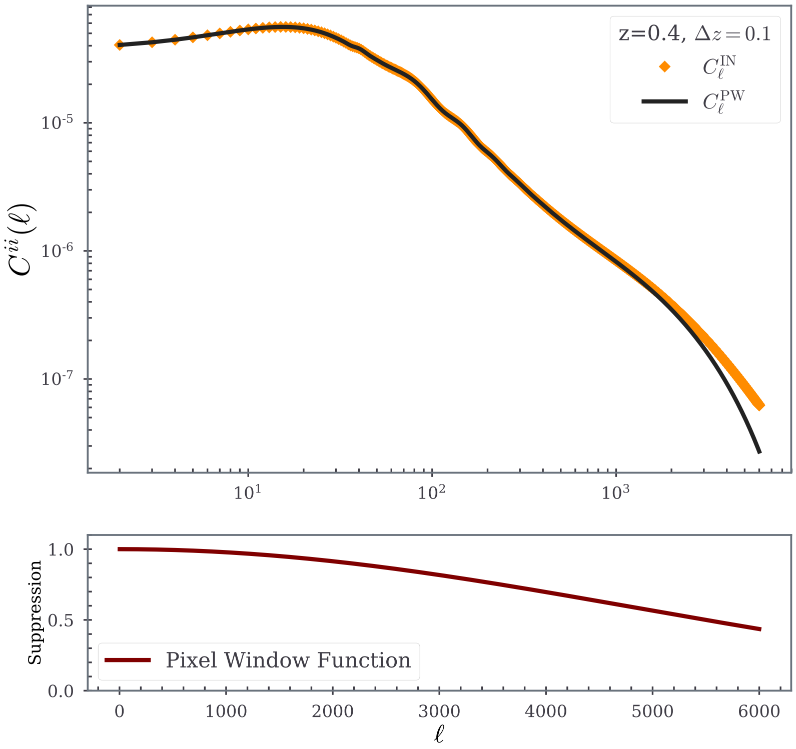


Fig. 6.5 **Upper panel:** Smoothing of the input angular power spectrum by applying the `HEALPix` pixel window function. The black curve is the smoothed spectrum while the orange datapoints are input C_ℓ's from `CLASS`. **Lower panel:** Pixel window function suppression with the corresponding multipole range (Note that the multipole-scale and angular power-scale in upper panel is Log-scale).

convolution becomes a simple multiplication. For this reason `FLASK` calls `HEALPix` to obtain, for a given map resolution, the window function ω_ℓ for a pixel p defined as:

$$\omega_\ell = \left(\frac{1}{N_{\rm pix}} \sum_{p=0}^{N_{\rm pix}-1} \omega_\ell^2(p) \right) . \tag{6.10}$$

This is applied to the angular input power spectrum from `CLASS` to obtain the convolved (smoothed) power spectrum which can be used as a reference for validation, i.e. to assess the quality of the mock catalogues and of the measured angular spectra as from the mock pixelised maps.

6.2.5 Obtaining the Gaussian Spectra

The multipoles of the unsmoothed angular spectra from `CLASS`, $C_{\text{ln}}^{ij}(\ell)$, assumed to be that of the Lognormal density field in shells, are manipulated to obtain the multipoles of the associated Gaussian field defined by $C_{\text{ln}}^{ij}(\ell)$.
This is done in three steps:

- At first the angular power spectra are inverse Legendre-transformed into the angular correlation function:

$$\xi_{\text{ln}}^{ij}(\theta) = \frac{1}{4\pi}\sum_{\ell=0}^{\infty}(2\ell+1)C_{\text{ln}}^{ij}(\ell)\mathscr{P}_{\ell}(\cos\theta)\,, \tag{6.11}$$

  where to define the angular two-point correlation function, here $\xi(\theta)$ is used instead of $w(\theta)$ to be consistent with Xavier et al. (2016) notations. At this step, the auto- and cross- angular correlation functions are saved and stored;

- Using the transformation from Eq. (6.9) the angular correlation function of the Gaussian (linear) field $\xi_{\text{g}}^{ij}(\theta)$ is obtained;

- At last, $\xi_{\text{g}}^{ij}(\theta)$ is Legendre-transformed back into the angular power spectra of the Gaussian field $C_{\text{g}}^{ij}(\ell)$ using:

$$C_{\text{g}}^{ij} = 2\pi\int_{0}^{\pi}\xi_{\text{g}}^{ij}(\theta)\mathscr{P}_{\ell}(\cos\theta)\sin\theta\, d\theta\,. \tag{6.12}$$

  All of the obtained spectra are stored on the disk.

Using a direct derived relation between the angular spectra of the Lognormal and Gaussian density fields,

$$C_{\text{g}}^{ij}(\ell) = 2\pi\int_{-1}^{1}\ln\Big[\sum_{\ell'=0}^{\infty}\frac{2\ell'+1}{4\pi}\frac{C_{\text{ln}}^{ij}(\ell')}{\alpha_i\alpha_j}\mathscr{P}_{\ell'}(\mu)+1\Big]\mathscr{P}(\mu)\,d\mu\,, \tag{6.13}$$

the code estimates the covariance that, for each multipole, is represented by the matrix $\mathbb{C}_{\text{g}}(\ell)$ with elements $C_{\text{g}}^{ij}(\ell)$.

From Eq. 6.13, it is clear that the relation between the Gaussian and Lognormal multipoles is non-local. Each Gaussian C_ℓ is obtained by summing over all Lognormal C_ℓ's. As a consequence, the resolution effects, that suppress the contribution from higher multipoles, would systematically bias the result of the integration in (6.13) and underestimate the Gaussian C_ℓ at large multipoles. A practical way to reduce this bias, that is adopted in this work, is to specify the input angular spectrum up to a maximum multipole much larger than that actually needed in the catalogue analysis. Following the Xavier et al. (2016) recommendation, I've set $\ell_{\max} = 6000$ in this pipeline (which is roughly of the order $3 \times N_{\text{side}}$). I also have verified that using this $\ell_{\max}$ value completely resolves this limitation.

6.2.6 Generating Correlated Gaussian Fields

Since the generated density fields are isotropic in 2D, the Gaussian multipoles of C_ℓ's in each shell are independent and so are the spherical harmonics coefficients. However, one expects that the same multipoles, (ℓ, m), in different shells should be correlated. Therefore, one needs to generate correlated Gaussian random variables from uncorrelated ones. This problem is solved by the so-called Cholesky decomposition in which each individual covariance matrix $\mathbb{C}_{\text{g}}(\ell)$ defined for each ℓ, is uniquely decomposed into lower triangular covariance matrices $\mathcal{T}(\ell)$:

$$C_{\text{g}}^{ij}(\ell) = \sum_k \mathcal{T}_k^i(\ell)\mathcal{T}_k^j(\ell) \,, \tag{6.14}$$

`FLASK` does that by calling the external routine from `GSL` library. Once the triangular matrices are generated, then the correlated coefficients can be obtained from a set of uncorrelated ones:

$$A_{\ell m}^i = \sum_k \mathcal{T}_k^i(\ell) A_{k,\ell m}^0 \,, \tag{6.15}$$

where $A_{\ell m}^0$'s are Gaussian random variables generated from a random Normal distribution using the seed from `RNDSEED` provided in the `config` file. The generated coefficients $A_{\ell m}^i$ satisfy the property given in Eq. (3.18) for C_{g}^{ij}.

Cholesky decomposition is only doable if the initial covariance matrix is positive definite. This condition is not guaranteed in the current case. As it was discussed before, the fact that the relation between Lognormal and Gaussian variables is non-local does not guarantee that given a set of Lognormal variable with positive definite covariance, the associated set of Gaussian variables has a positive definite covariance too. This is a fundamental problem that has only a phenomenological fix if deviations from the definite positiveness is not too large. `FLASK` estimates the amplitude of these deviations and numerically regularises the covariance matrix accordingly. Deviations from the definite positiveness of the matrix can be significant for the shear field in weak-lensing analyses. However, it turns out to be very small for the density field, i.e. the correction factor of the order 10^{-4} is sufficient to fix this. This correction is the so-called 'fractional change'

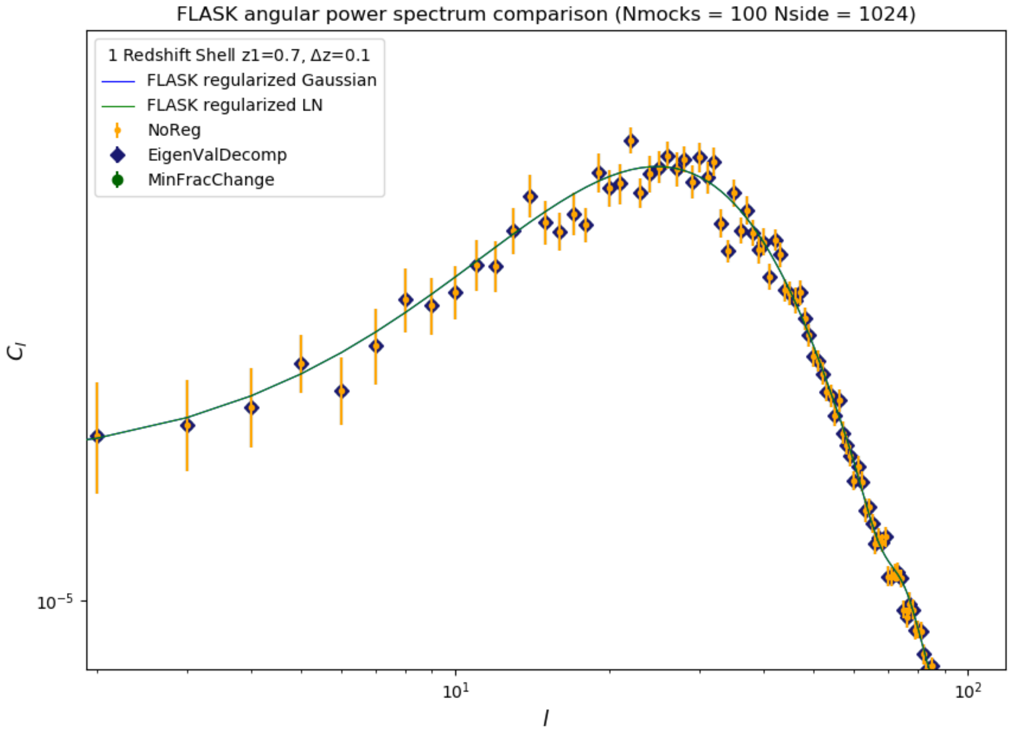


Fig. 6.6 The comparison with different methods of obtaining the regularised matrix. It can be noticed that density fields have no significant impact of regularisation, where the detailed methods of regularisation are described in the text.

(blue datapoints in Fig. 6.6) in non-positive matrix elements by manually tuning them to be positive-definite. Moreover, `FLASK` enables another way to resolve this problem is via direct ‘fixing’ the non-positive Eigenvalues through Eigenvalue decomposition (blue datapoints in Fig. 6.6), which is rather a time consuming approach. This is one of the peculiar features of `FLASK`, and therefore provides a reliable framework for generating correlated Lognormal fields, which is the main criterion for choosing this framework for the pipeline. Fig. 6.6 illustrates the results from a series of tests I have performed to assess the impact of this regularisation step.

6.2.7 Generating Lognormal Maps

Using the `ALM2MAP` routine from `HEALPix` module, FLASK generates a set of correlated density maps from the correlated Gaussian coefficients $A^i_{\ell m}$’s obtained previously. These maps, whose resolution has been specified by `NSIDE`, provide the value of the Gaussian density field in at each pixels of each shell. These correspond to a set Gaussian Z_i variables, that, using Eq. (6.8), are transformed into Lognormal variables that represent the target object density at the pixel location. To avoid aliasing effect, the resolution of the Lognormal map has been set higher than that of the Gaussian one, i.e $N_{\rm side} \sim \ell_{\rm max}/3$.

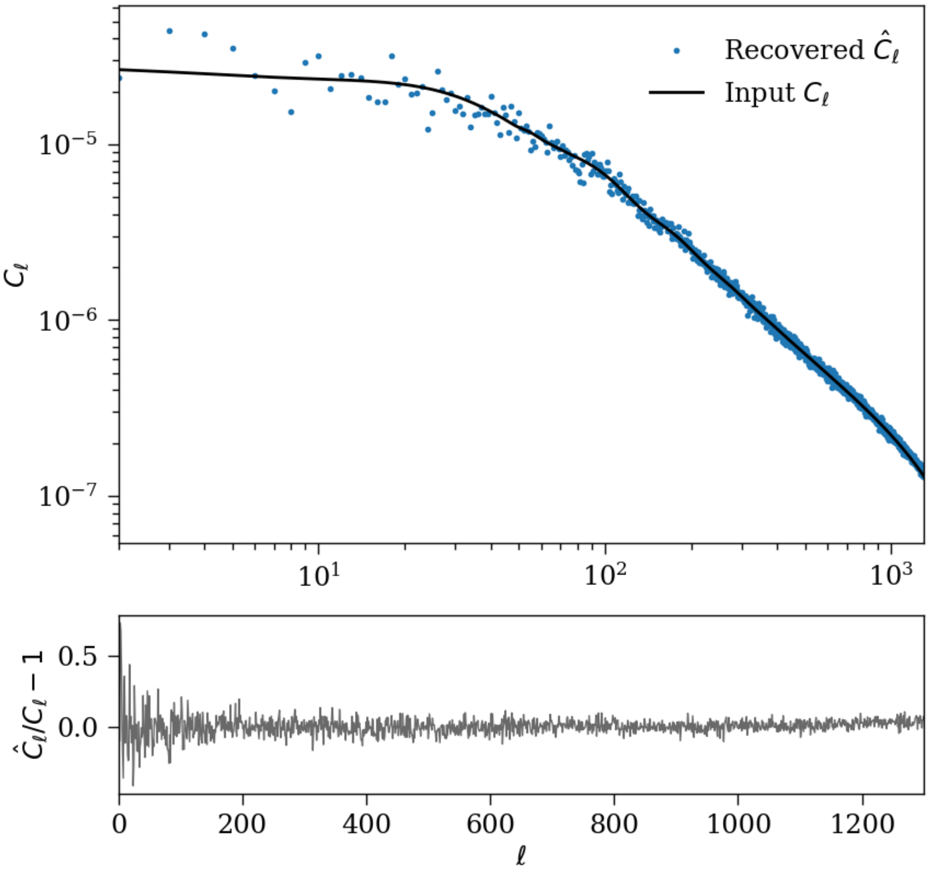


Fig. 6.7 The input angular power spectrum (solid black) versus the recovered C_ℓ (blue datapoints) from the generated multipoles as described in Eq. (6.16), which is measured from one realisation of LN density field. The top panel shows the spectra, whereas the bottom panel shows the fractional difference between the two.

Another routine `MAP2ALM` is then used to estimate the spherical harmonic coefficients of the Lognormal map and the angular auto- and cross- spectra are obtained by:

$$C_\ell^{ij} = \frac{1}{\ell+1} \sum_{m=0}^{\ell} A^i_{\ell m} A^{j*}_{\ell m} , \tag{6.16}$$

where the realised multipoles are not summed over negative m's as the fields are real and satisfy $(-1)^m a^i_{\ell -m} = a^{i*}_{\ell m}$.

The quality of the generated map can be assessed by comparing the angular spectra measured from the Lognormal map with the reference, smoothed input spectra generated at the beginning of the procedure. The result of one such tests is shown in Fig. 6.7. The continuous curve represents the reference angular spectrum of a density field in a shell at $z = 0.5$ with thickness $\Delta z = 0.1$. The points show the measure C_ℓ from one map realisation. The scatter is consistent with the expected cosmic variance and the residuals (bottom panel) show no sign of systematic errors. It can be noticed that, being the spectrum of a continuous field, no shot noise contribution has been added to the spectrum. The map from which this spectrum has been measured is also shown in Fig. 6.8 (left panel) together with a zoomed in square portion of 200×200 pixels, to better appreciate the density contrast range.

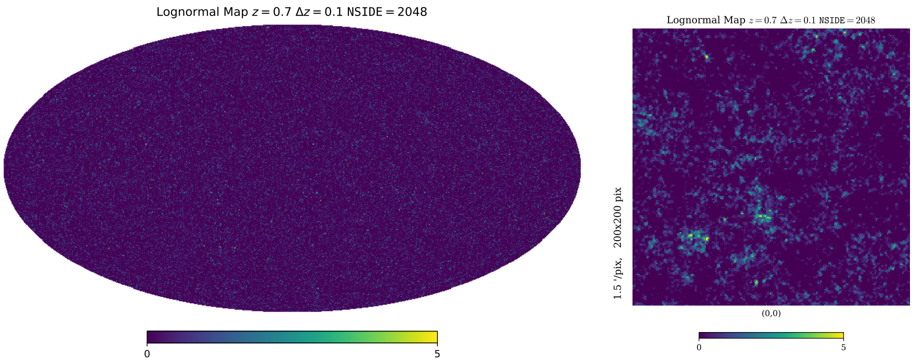


Fig. 6.8 **Left panel** The Lognormal overdensity map produced using `HEALPix`, with the full-sky Mollweide view output by `FLASK` at redshift $z_s = 0.7$ with a Top-Hat bin width $\Delta z_s = 0.1$. **Right panel** Zoomed in portion of 200×200 pixels of the same resolution as the map i.e. $N_{\rm side} = 2048$

6.2.8 Generating Catalogue of Discrete Objects

As a final step of `Stage-I`, the pixelised, Lognormal density maps are `MonteCarlo` sampled into catalogues of discrete objects (in this case galaxies)[4]. `FLASK` offers two sampling options (in addition to the no-sampling one): Poisson and Gaussian sampling. For this work, I have opted for Poisson sampling.

---

[4] So far, I have referred them as objects because only after assigning (astro-) physical properties such as flux and mass, they can be called 'galaxies'. It should be noted that, throughout this work, catalogues contain point objects and their position although sometimes they are referred as galaxies only in colloquial sense.

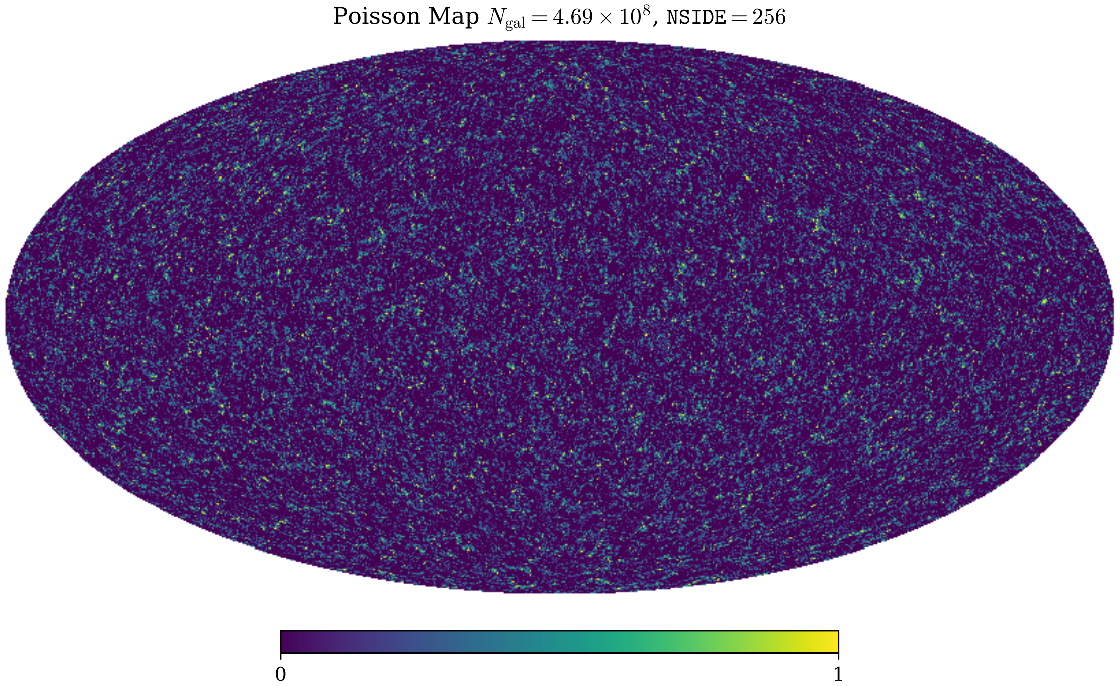


Fig. 6.9 Example of a Poisson sampled map with about 470 million galaxies at $z_s = 0.7$. Though for the analysis we use a high resolution of $N_{\text{side}} = 2048$, we have descaled this map to a lower resolution ($N_{\text{side}} = 256$) in order to visualise the structure of galaxy distribution

The mean number of galaxies in a generic pixel, p, of shell i is set by the selection function of the survey ϕ_i. And the number of galaxies in that pixel is drawn from a Poisson distribution with mean:

$$\langle N_{\text{gal}}(i,p)\rangle = \bar{n}_{\text{gal}}(i,p)[1+\delta_{\text{gal}}(i,p)]\Delta\Omega \,, \tag{6.17}$$

where $\Delta\Omega$ is angular area traced by a pixel in arcmin^2, δ_{gal} is the (Lognormal) galaxy overdensity that is defined according to the input power spectrum. The quantity $\bar{n}_{\text{gal}}(i,p)$ is the expected number of galaxies per arcmin^2 per pixel inside a homogeneous Universe. This is a mean value of the Poisson distribution and contains following components:

$$\bar{n}_{\text{gal}}(i,p) = a_s \times M(p) \times \phi_i(z) \,, \tag{6.18}$$

where a_s is scaling factor (adjustable through the parameter file); $M(p)$ is binary geometry mask that describes the footprint of the survey; $\phi_i = (dN/dz_i)$ is the input radial selection function, both of them can be provided as external input to the code. I will focus on these two quantities later in § 6.3.

Fig. 6.9 shows the final outcome of the sampling procedure. It represents the Aitoff projection of the mock galaxy distribution in an full-sky Lognormal catalogue of 4.69×10^8 in a shell at $z = 0.7$ with thickness $\Delta z = 0.1$.

6.2.9 Stage-I Validation: Model vs. Measured Spectra

To test the quality of the synthetic catalogues and validate the first stage of the pipeline, I measure the angular auto- and cross- spectra of the mock galaxies in real space (i.e. spectro-z shells) and compare the results with the input power spectra, by taking the shot noise contribution into account.

To estimate the angular spectra from the catalogue I use the `PCL` estimator. The general procedure to estimate angular power spectra is described in Chapter 3. However, the `PCL` code, which I have included in the pipeline to perform on-the-fly quality checks, is the one developed for the *Euclid* collaboration. Therefore, as a standard procedure to measure the power spectrum from a survey catalogue, the code generates overdensity maps from the catalogue by distributing the galaxies on `HEALPix` pixelised map, and draws fluctuation coefficients $a^i_{\ell m}$'s using `MAP2ALM` routine. The auto- ($\tilde{C}^{ii}_\ell$) and cross- ($\tilde{C}^{ij}_\ell$) power spectrum is estimated from Eq. (3.47). The shot noise power spectrum is measured separately by randomising the angular position of the mock galaxies as:

$$N^{ij}_\ell = \delta^K_{ij} \frac{\Delta\Omega}{N^i_{\rm gal}} , \tag{6.19}$$

where δ^K_{ij} is Kronecker delta function, $N^i_{\rm gal}$ is the number density of galaxies in i-th redshift shell and $\Delta\Omega$ is full-sky solid angle. The shot noise is subtracted from the measured spectra before comparing it with the analytic one.

The result for two redshift shells from validation tests is shown in Fig. 6.10. It compares the shot noise subtracted, angular auto-spectra measured from two mock catalogues created in two different shells centred at $z = 0.5$ and $z = 0.8$ with thickness $\Delta z = 0.1$, with orange and green datapoints, respectively). I overplot the input angular power spectra produced by `CLASS` (thick solid curves). Errorbars represent the 1σ scatter among 100 mock realisations. The residuals from the measured versus the input spectra can be seen at both redshifts in the bottom plot.

For quantitative assessment, the quality of the output catalogue is verified by computing the χ^2 statistics:

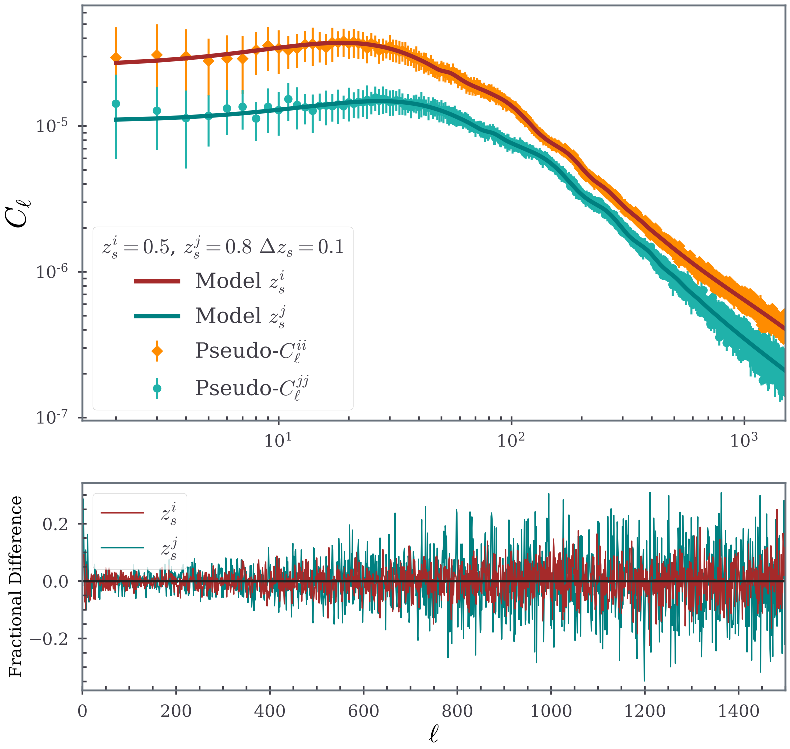


Fig. 6.10 **Upper panel:** Measured auto-pseudo-C_ℓ's for $z^i = 0.5$ and $z^j = 0.8$, represented by orange and green datapoints (mean) respectively, are fitted with model spectra which are input spectra from `CLASS` in this case with Top-Hat bin width $\Delta z = 0.1$. Errorbars are the scatter over 100 realisations. **Lower panel:** Residuals of input vs. measured C_ℓ's for both redshifts.

$$\chi^2 = \sum_i^{N_m} \frac{\left(\hat{C}_\ell^i - C_\ell^i\right)^2}{(\sigma^2/N_m)}\,, \tag{6.20}$$

where N_m is number of mock realisations, the error term in the denominator of the χ^2 formula is error on the mean (since in this case we are comparing the model spectra with mean from many realisations). The obtained values of χ^2 analysis for `Stage-I` validation are shown with `Stage-II` in Table 6.1. At this step, according to the schematic of pipeline introduced in the beginning of the chapter (Fig. 6.1), the set of ideal mocks are validated and ready to be passed on to '`Survey Mock Generator`' of the pipeline.

6.3 Stage-II: From Ideal to Realistic Mocks

The goal of the second stage of the pipeline is to transform the archetypal, full-sky Lognormal catalogue of objects distributed in different redshift slices into a realistic mock catalogue resembling the characteristics of a well defined photo-z survey. The addition of observational properties is done through different, independent processing elements. In the current implementation of the pipeline 3 of such properties are considered: *i)* the assignment of photo-z and its error; *ii)* the enforcement of the expected redshift distribution from the survey; *iii)* and finally the masking of full-sky region with the survey-specific footprint to account for incomplete sky coverage. At later stages, more effects can be added (i.e. Stellar foreground, Galactic absorption, Zodiacal light etc.) if they are relevant to characterise the survey selection criteria or observational biases. In fact, I have constructed this pipeline in a way, to simulate the characteristics of any photo-z survey sample. For this thesis I focus on the above three specific properties for *Euclid* photo-z survey. This stage includes the original computational framework, that I have programmed independent to `Stage-I` codes, i.e. the '`Survey Mock Generator`' (hereafter `SMG`) in Fig. 6.1, that takes in the input parameters which tune-in the catalogue to emerge exactly similar to the survey observed one. In following subsections I expand the processing elements of `Stage-II` in details.

6.3.1 Adding Photo-z Information

As mentioned in the last step of § 6.2.1, the output catalogue from `Stage-I` contains the angular coordinates of the galaxies along with their redshifts. The redshifts are drawn from uniform random distribution within the input redshift shell interval, which I assume as 'true' (cosmological) redshifts, are the spectroscopic ones and possess very small errors with respect to the photometric counterparts. For this reason, the spectro-z errors $\sigma_z(z_s)$ (subscript s denotes spectroscopic) can be ignored. As a result the spectro-z distribution dN/dz_s resembles a '*Top-Hat*' (flat) histogram within the shells (redshift bins) of size Δz_s. Ideally, the number of galaxies within each redshift shell can be modulated by a survey selection function $\phi_i(z)$. It is equivalent to the probability distribution $P(z)$ and is expressed as (recall Eq. (3.39)):

$$\phi_i(z) \equiv P(z) = W_i(z)\,\frac{dN_{\rm gal}}{dz_i}\,, \tag{6.21}$$

where $N_{\rm gal}$ is number of galaxies in a given redshift shell and W_i is the window function.

To assign the photo-z and its error for each galaxy in the catalogue, it is essential to understand the relation between the spectroscopic and photometric distribution. For more detailed description on this relation, I recommend referring to the work in Sheth and Rossi (2010); Asorey et al. (2012) and Balaguera-Antolínez et al. (2018). Because of the higher uncertainty on photo-z estimates, z_p (the subscript p denotes photometric)[5], the 'observed' distribution dN/dz_p in photo-z space is different from the true one dN/dz_s. The probability distribution (PDF), $P(z_p)$, for photo-z is given as:

$$P(z_p) \equiv \left(\frac{dN}{dz_p}\right) = \int_0^{\infty} P(z_s)\; P(z_p|z_s)\; dz_s\,, \tag{6.22}$$

where $P(z_s)$ can be expanded in the form Eq. (6.21) and the quantity $P(z_p|z_s)$ are photo-z errors, i.e. the conditional probability of a photo-z (z_p) given the spectro-z (z_s). Ideally, photo-z errors [$\sigma_z(z_p)$] can be drawn from any type of distribution. Once assigned, these errors are used to displace objects along the line-of-sight. In case of tomographic analyses for photo-z surveys, the sample slicing is done using photo-z rather than the cosmological ones. Consequently the true distribution, i.e. spectro-z's are scattered in and out of the slice, modifying the correlation properties of the objects in the photo-z slice with respect to those in the original (true) redshift slice. This effect needs to be properly quantified and accounted for, to compare model prediction with observations.

In `Stage-II` module of the pipeline, two different procedures are implemented to assign a photo-z information for each object, given the spectroscopic one. First is object based assignment, i.e. radially perturbing individual galaxy with an assumed distribution type. Alternatively, since the photo-z error displacement conserves the number of the objects, the cumulative probability distribution in both spectroscopic and photo-z space is also conserved and z_p can be assigned by one-to-one mapping from the sorted spectro-z distribution. In principle, both procedures are flexible enough to permit the adoption of a generic photo-z errors, that can obey any specified statistics and whose magnitude can depend on the redshift of the shell. In this thesis, I use the case of redshift-independent Gaussian error, which provides reasonably good approximation to realistic photo-z errors, once the catastrophic redshift errors are ignored. Keeping the *Euclid* survey in perspective, I assume that these errors are Gaussian distributed with the *RMS* $\sigma_z(z_p) = 0.02$. In this scenario, I summarise the two procedures used in the pipeline.

---

[5] Convention Note: In the thesis, I use photo-z and spectro-z as general convention to refer photometric and spectroscopic redshifts, while z_p and z_s correspond to mean redshift estimates in the sample with their errors represented by $\sigma_z(z_p)$ and $\sigma_z(z_s)$ respectively.

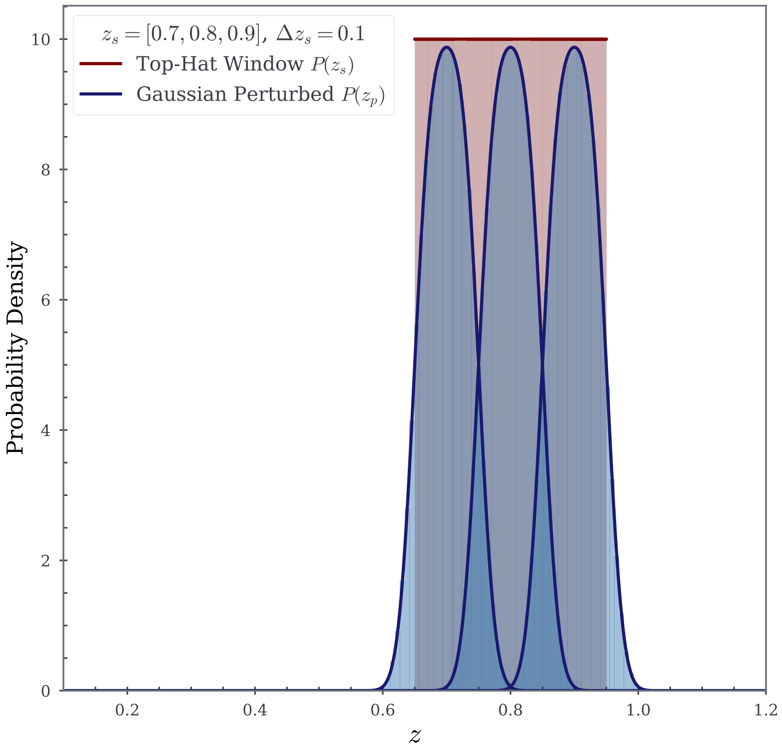


Fig. 6.11 PDF of example ideal catalogue with three adjacent Top-Hat sliced spectro-z shells with their galaxy distribution showed in light-red histograms. The light-blue histograms are 'observed' distribution of photo-z of objects in spectro-z shells. The dark curves are model PDF computed using Eqs. (6.21) and (6.22) for spectro-z and photo-z.

1. **Single Object:** The first method consists of using approximated $P(z_p|z_s)$ by radially displacing each individual object centred at z_s. Here I generate a Gaussian deviate with mean z_s and variance σ_z:

$$P(z_p|z_s) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{(z_p - z_s)^2}{2\sigma_z^2}\right], \tag{6.23}$$

where σ_z is the photo-z error. For spectroscopic distribution $P(z_s)$, the redshift shell has the *Top-Hat* form within the redshift interval $z_{s,\mathrm{min}}$ and $z_{s,\mathrm{max}}$. The integral in Eq. (6.22), in that case, can be evaluated analytically:

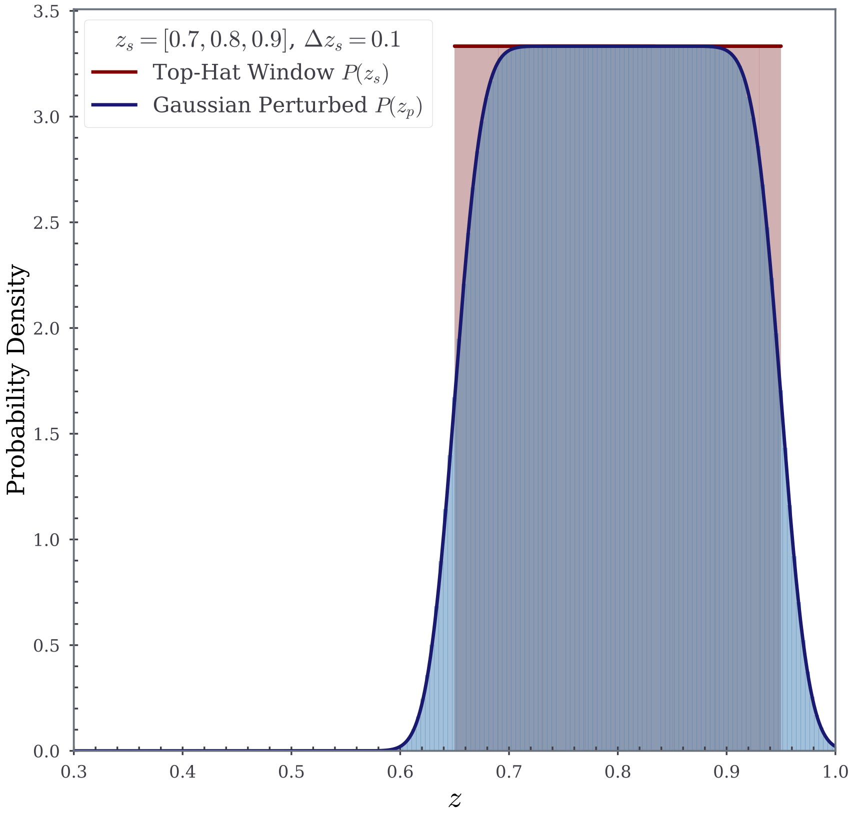


Fig. 6.12 PDF of same example ideal catalogue as in Fig. 6.11. Spectro-z (observed in light-red histogram vs. model in dark-red curve) and photo-z PDF (observed in light-blue vs. model in dark-blue) for full sample.

$$P(z_p) = \left(\frac{1}{2} \text{erf}\left[\frac{z_{s,\max} - z_p}{\sqrt{2}\sigma_z} \right] - \frac{1}{2} \text{erf}\left[\frac{z_{s,\min} - z_p}{\sqrt{2}\sigma_z} \right] \right) , \tag{6.24}$$

where erf(x) is error function. In Fig. 6.11 and 6.12, the example of this method is presented for objects in individual shells and for the full sample respectively. In both the figures, the original catalogue has 3 *Top-Hat* shaped shells in spectroscopic space, centred at z_s=[0.7, 0.9] with shell width $\Delta z = 0.1$ and distribution $P(z_s)$ shown in light-red histograms. The light-blue histograms are the $P(z_p)$ distributions measured in the catalogues after objects displacements. The thick dark-blue curves show the expected $P(z_p)$ distribution according to Eq. (6.22). The agreement between the curves and the blue histograms quantifies the quality of the z_p-assignement. Fig. 6.12 shows full $P(z_p)$ distribution of all objects in the sample.

`SMG` computes Eq. (6.22) for each object to assign z_p and the error $\sigma_z(z_p)$. In terms of computational budget, since the method performs a loop over total objects in the catalogue, the computational time scales linearly with $N_{\rm gal}$. In this pipeline, this option is set to default and is made efficient by distributing the computation Eq. (6.22) on multiple threads of the process i.e. using parallelisation on multi-core machine. In § 6.4, I shall present the relevant results when I describe the pipeline architecture.

2. **Sorting Objects:** In case of a single core machine, the use of second procedure is recommended, as it is significantly faster than the individual z_p-assignment. This is achieved by drawing z_p values from the model photo-z cumulative distribution (CDF), $\Phi(z_p)$:

$$\Phi(z_p) = \frac{1}{2}\mathrm{erfc}\left(-\frac{1}{\sqrt{2}}\frac{z_p - z_s}{\sigma_z}\right), \tag{6.25}$$

where erfc(x) is complementary error function. It must be noted that, the number of object in $\Phi(z_p)$ and $\Phi(z_s)$ are conserved, i.e. $\Phi(z_p) = P(> z_p) \equiv P(> z_s) = \Phi(z_s)$ In practice, since both $\Phi(z_p)$ and $\Phi(z_s)$ are monotonic functions and preserve the ranking of sorted objects, the photo-z can be obtained from the spectroscopic one through inversion:

$$\Phi^{-1}(z_p) = \sqrt{2}\sigma_z\, \mathrm{erf}^{-1}(2z_p - 1) + z_s\,. \tag{6.26}$$

Above relations are valid only for Gaussian error case (see Eq. (6.23)). `SMG` module follows a simple flow of steps:

- The $N^i_{\rm gal}$ objects in i-th shell are ordered according to their spectro-z's;
- A list of z_p values equal to $N^i_{\rm gal}$ are generated according to the expected $P(z_p)$ and ordered according to z_p;
- z_p are assigned to each object according to their ranking, i.e. the largest z_s value will be associated to the largest z_p value, the second largest to the second largest and so on.

  The results of this method are shown in Fig. 6.13. Again in this example for two redshift shells, the quality of z_p-assignment procedure using sorting method is verified by comparing the model $P(z_p)$ distribution (thick dark-blue curves) with the 'observed' histograms in light-blue. The corresponding spectro-z distribution $P(z_s)$ is shown in red.

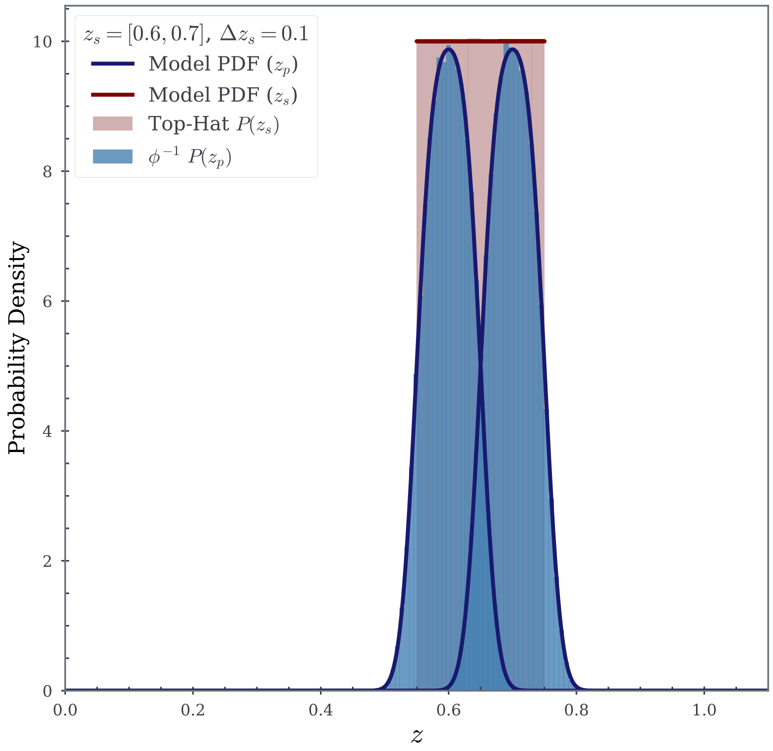


Fig. 6.13 Photo-z PDF, shown in blue, that is obtained using sorting method i.e. Eq. (6.26) for two redshift shells, shown in red, Top-Hat sliced in spectro-z-space. The mean redshifts are $z_s = 0.6$ and 0.7 with shell-width $\Delta z_s = 0.1$. The light colour represents observed distribution and dark curves are model PDF of the respective distribution.

For tomographic clustering, the photo-z catalogue will be sliced into *Top-Hat* shells. `SMG` takes the user input information of number of photo-z shells to be sliced for the analysis, along with their width Δz_p. This information goes into the pipeline via a single file which is represented under '*Photo-z error*' title in pipeline schematic, Fig. 6.1.

An example of photo-z slicing is shown in Fig. 6.14. The light-blue histogram shows the photo-z distribution of the full sample originally selected in the spectro-z space centred at mean redshifts ranging from z_s=[0.5,0.9] and divided in 5 equal-sized shells. Because

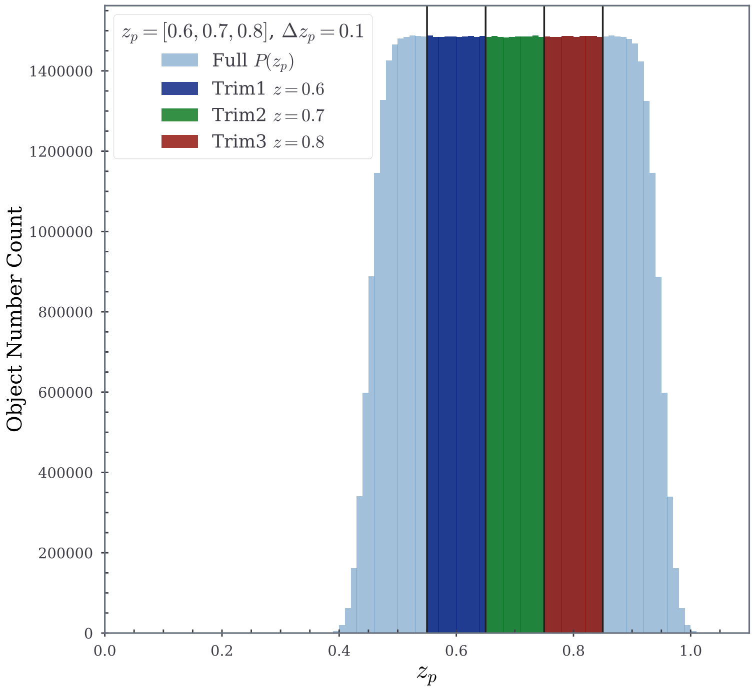


Fig. 6.14 Trimming of photo-z slices in Top-Hat with shell-width $\Delta z_p = 0.1$ for redshifts $z_p = 0.6, 0.7$ and 0.8 in dark blue, green and red respectively. The faint blue distribution is full photo-z corresponding to 5-shells in spectro-z-space with mean redshifts $z_s = 0.5, 0.6, 0.7, 0.8$ and 0.9.

of the photo-z displacement, the number of shells in parent catalogue is higher than the final photo-z one. The central 3 photo-z centred at z_p=[0.6, 0.8] are 'trimmed' as photo-z catalogue after discarding the fraction of objects residing in the tails of the distribution. At this step, the associated 'true' $P(z_s)$ distribution of objects in each photo-z shell must be assessed by evaluating the 'model' prediction. An accurate evaluation of this quantity is fundamental in order to accurately predict the angular power spectrum to compare with the measured one in photo-z space. This will be clear at the end of this section, where I shall validate `Stage-II`.

The relevant model $P(z_s)$ can be obtained, analogously to Eq. (6.22), by integrating the following expression:

$$P(z_s) = \int_0^\infty P(z_p)\, P(z_s|z_p)\, dz_p\,, \tag{6.27}$$

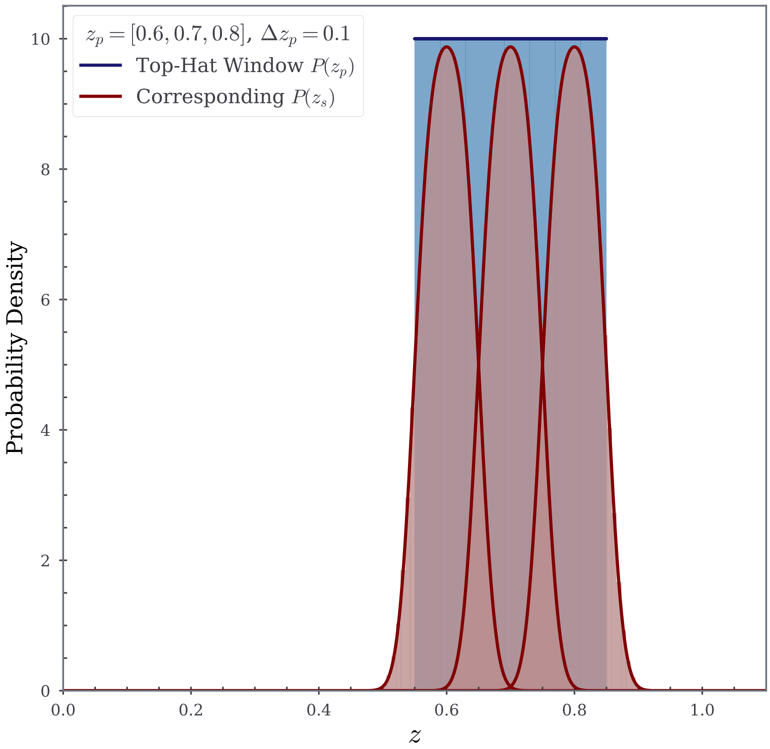


Fig. 6.15 PDF with *Top-Hat* sliced shells in photo-z-space (in blue) and associated spectro-z (in red). Lighter shades correspond to measured distribution and dark curves represent model PDFs for respective redshifts. The mean redshifts and the shell-width for three photo-z shells are identical to Fig. 6.14.

where $P(z_p)$ is now a *Top-Hat* distribution that defines the photo-z shell and $P(z_s|z_p)$ can be either obtained from the joint distribution $P(z_s|z_p)$ of the objects in the mock catalogues. In the case of Gaussian photo-z errors that are considered here, it turns out that $P(z_s|z_p)$ is well approximated by a Gaussian distribution as in analogous case, that I have adopted in this application, as analogous terms in photo-z and spectro-z spaces specified in Eq. (6.22) and Eq. (6.27) related through Bayes' theorem:

$$P(z_s|z_p) = \frac{P(z_s)P(z_p|z_s)}{P(z_p)} \,. \tag{6.28}$$

The result of this analysis is shown in Fig. 6.15, which is, as expected, analogous to Fig. 6.11 except that now the *Top-Hat* light-blue histograms show the distribution of the

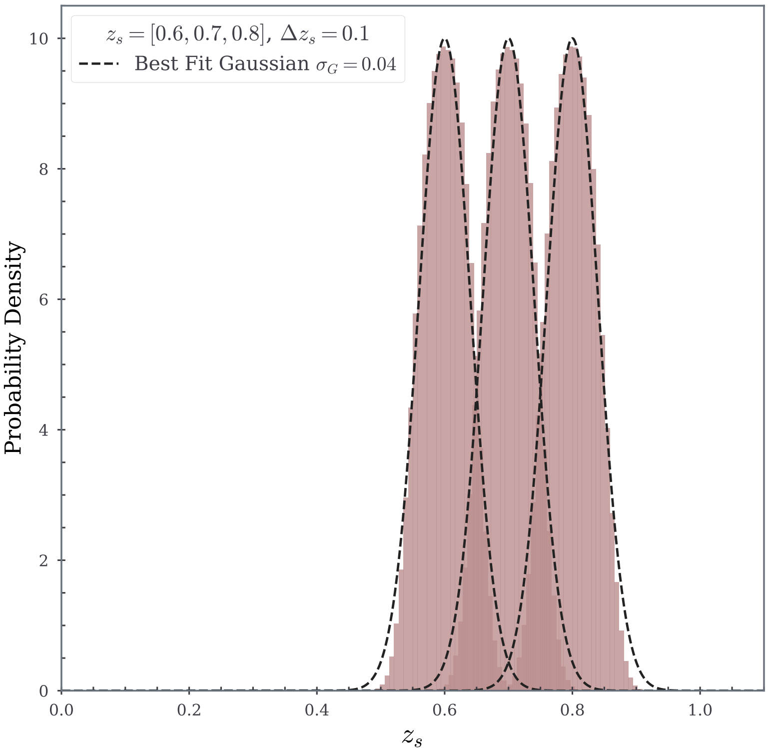


Fig. 6.16 Observed spectro-z PDF corresponding to Top-Hat shell in photo-z space for three redshift shells trimmed in case of Fig. 6.14. Black dashed lines are Gaussian fit with best fit value of σ_G.

photo-z sample. On the other hand, the 3 light-red histograms show the corresponding $P(z_s)$ of the objects in each shell. The same three $P(z_s)$ distributions are also shown in Fig. 6.16 as light-red histograms. The corresponding dashed black curves are the best-fit Gaussian approximations that better model the measured quantities. For the input Gaussian photo-z error with *RMS* $\sigma_z(z_p) = 0.02$, the obtained best fit value of $\sigma_G = 0.04$ is stored in log-file, which will be used to obtain the theoretical angular power spectrum from `CLASS` to compare it with the measured one from photo-z shell. The approximation, as it can be seen in figure, is good except at the tails of the distribution. As we shall see in § 6.3.4 this mismatch does not hamper the ability to predict the auto- and cross- angular spectra accurately for the photo-z samples.

6.3.2 Photo-z Survey Radial Selection Function

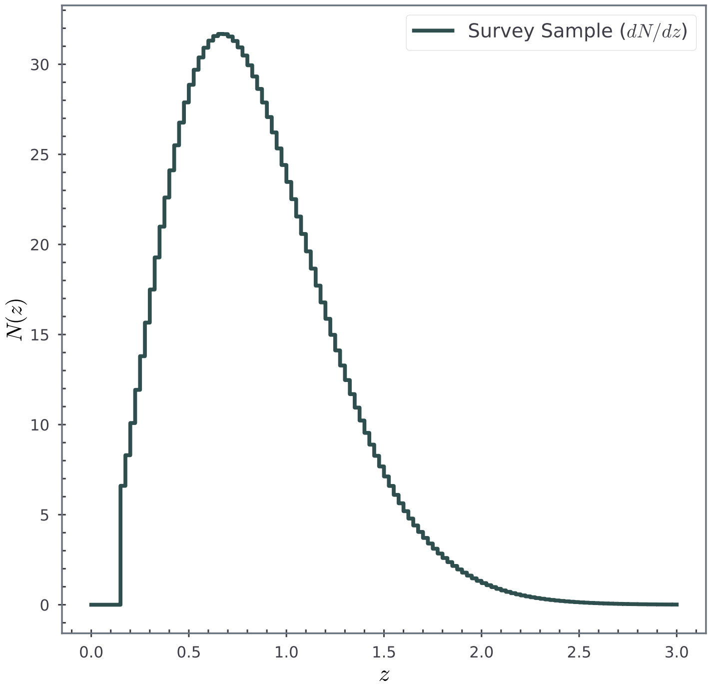


Fig. 6.17 Euclid photo-z radial selection function, the dN/dz_p, showing the number density in arcmin^2 per redshift interval dz.

The catalogue at this step has the photo-z information and, more importantly, modelled true distribution, $P(z_s)$, in photo-z shells, that provides the input information for `CLASS` to model underlying statistical properties (through the theoretical auto- and cross-spectra) of the catalogue. But the number of objects in each photo-z shell is yet to be matched with the expected one, i.e. the radial selection function, more specifically the redshift distribution dN/dz_p of the survey. Fig. 6.17 shows the sample dN/dz_p for *Euclid* photo-z survey which is the target distribution for the final catalogue. To devise a method that guarantees the number of object in each redshift shell to be a Poisson deviate of the target one, it is important to assess the loss of objects due to displacement of photo-z errors. As we have seen in § 6.3.1, the number of objects in spectro-z shell within Δz_s in ideal `Stage-I` catalogue, is different from the photo-z catalogue sliced at the same redshifts in photo-z space Δz_p. In fact Fig. 6.14 shows that number of objects ending up in the photo-z sample will be smaller because of the trimming of objects at the edges of the redshift ranges.

Since the difference in population of objects in each shell varies with the amplitude of the photo-z error and depends on the expected dN/dz_p of the *Euclid* survey, I use a brute-force approach for this case. I populate 20% more objects in each shell of the initial (ideal) catalogue than those expected in the survey by adjusting the scaling factor a_s in Eq. 6.18, after the photo-z assignment, the excess objects are rejected using `Monte-Carlo` resampling to match with the expected number in each shell.

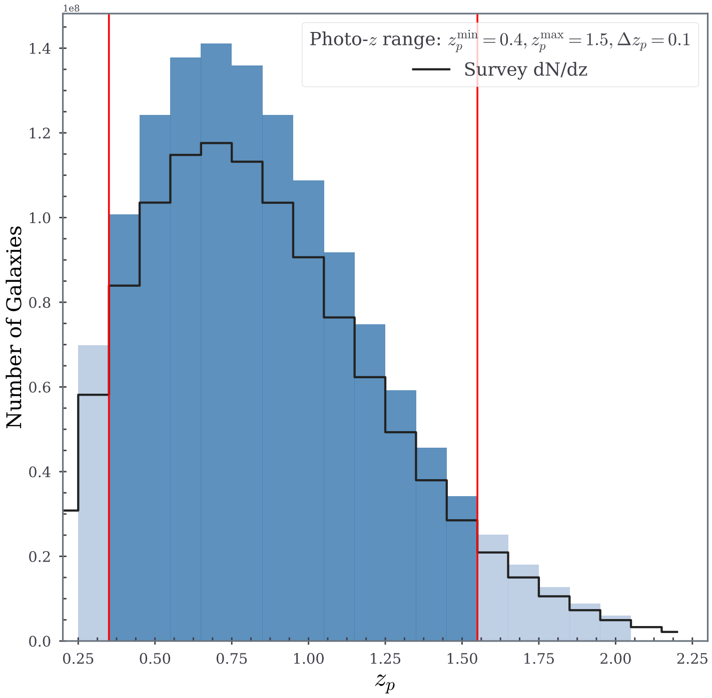


Fig. 6.18 Euclid survey dN/dz_p (black lines) is used to systematically overpopulate parent distribution $P'(z_p)$ (histograms), which is 1.2 times higher than dN/dz_p. The vertical red lines mark the cut-off margin for selecting the photo-z redshift range, where darker histograms represent the selected range from $z_p = 0.4$ to 1.5 with shell-width of Δz_p. The light-coloured histograms are refrained from the final distribution, as the number of galaxies significantly low to overcome the average shot-noise from all shells. (Note: The shell corresponding to $z_p = 0.3$, even though has significant objects, is trimmed from the spectro-z catalogue, and therefore will not be considered.)

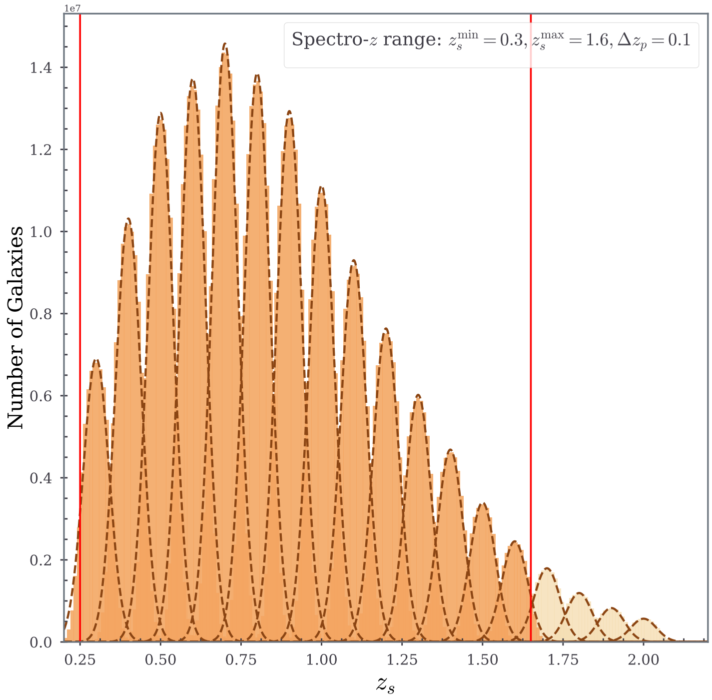


Fig. 6.19 Associated distribution of spectro-z shells (in coloured histograms) to *Top-Hat* photo-z shells shown in Fig. 6.18. The dashed lines are Gaussian window function fits to the spectro-z shells centred at $z_s = 0.3$ to 2.0, where the cut-off for final catalogues is shown with vertical red lines.

Fig. 6.18 illustrates the outcome of the procedure. The blue-shaded histogram shows the photo-z distribution of the objects after the photo-z assignment step. It is systematically larger than the target dN/dz_p of the survey (black histogram). The `Monte Carlo` resampling step guarantees that the final sample matches the expected redshift distribution. It can be noticed that the parent sample extends beyond the redshift interval of the final one, indicated by the two vertical red lines. This is done to account for those objects outside the interval in spectro-z space are scattered in by the photo-z displacement. The effect can be appreciated in Fig. 6.19. It shows the spectro-z distribution of the objects that reside in the photo-z shells of 6.18 The red vertical lines now indicate a larger redshift range that include 2 more redshift shells (an inner one centred z_s=0.3 and an outer one centred at z_s=1.6) The objects in the two corresponding shells are scattered and therefore trimmed in the photo-z sample. The bulk of mock catalogues that I have generated by using the *Euclid* survey redshift distribution contains ~ 1.3 billion galaxies in each catalogue. For handling this number density of objects, it is very crucial for the pipeline to be fully optimised and parallelised to efficiently produce large number of mock catalogues. I shall discuss this in details in § 6.4.

6.3.3 Photo-z Survey Footprint (Geometry Mask)

Real surveys do not cover the entire sky. Even when performed under most favourable circumstances and in the most 'clear' electromagnetic window, an extragalactic survey can only cover a fraction of the whole sky. The actual footprint of a survey is specific to its observing characteristics and possibly also time-dependent for long lasting surveys.

`SMG` offers the possibility to account for the survey footprint as long as it is quantified by a pixelised binary map, with 0 value in the unobserved areas. The mask is generated using `HEALPix` with a resolution appropriate to describe the survey geometry. As seen in Eq. (6.18), the geometry mask has the form of $M(p)$ vector, where p is the pixel. For more details about the application of geometry (angular) mask, Eq. 3.28 and its description can be referred in Chapter 3.

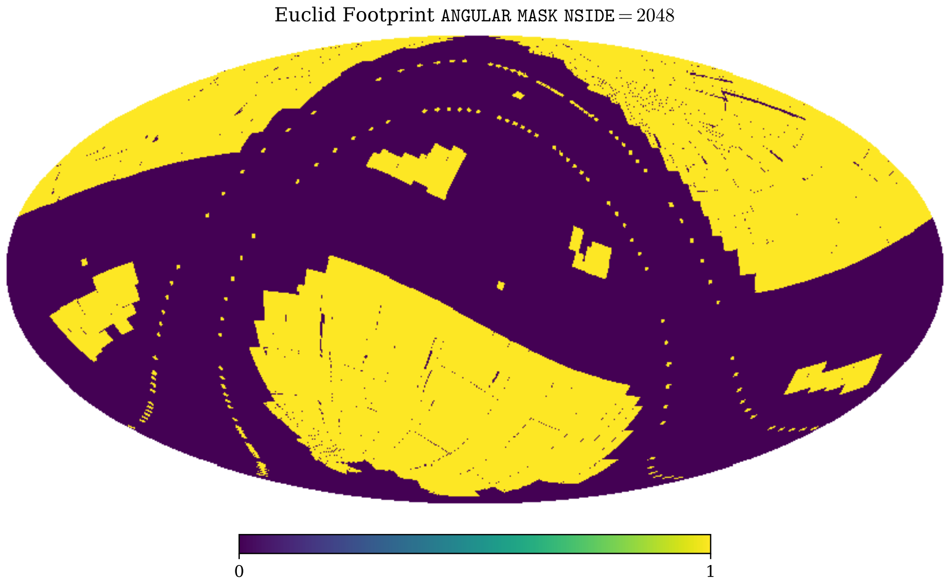


Fig. 6.20 Euclid survey mask, with yellow area representing the observed portion of the sky by the survey (pixels masked with value = 1) and the purple colour corresponds to unobserved area, that contains the regions of the sky involving galactic plane and zodiacal light from the ecliptic (pixels masked with 0). The mask is developed by *Dr. Martin Kilbinger* and his team. The fraction of observed sky with this mask is $f_{\rm sky} = 0.364$.

In the current implementation of the geometry mask, I use the one developed in the collaboration for *Euclid* photo-z survey. The *Euclid* survey[6] is expected to cover $\sim 15,000$ deg^2 in three near infra-red and one optical band. Unobserved areas will be a strip around the Galactic plane, the area above and below the Zodiacal plane and some other smaller regions, which are characterised by anomalously large absorption or the presence of very bright objects. The expected survey footprint is represented in Equatorial coordinates in Fig. 6.20. The coverage will not be uniform, in the sense that some fields may be observed longer than others, reaching a deeper or shallower magnitude limit. I have not considered this aspect that must be addressed using a non-binary mask in which each pixels is assigned an appropriate weight to account for local depth. Instead I consider the *Wide-Euclid* survey to have the same magnitude limits in all bands throughout the full survey area. The mask used in this analysis is set to a resolution of $N_{\rm side}$=2048.

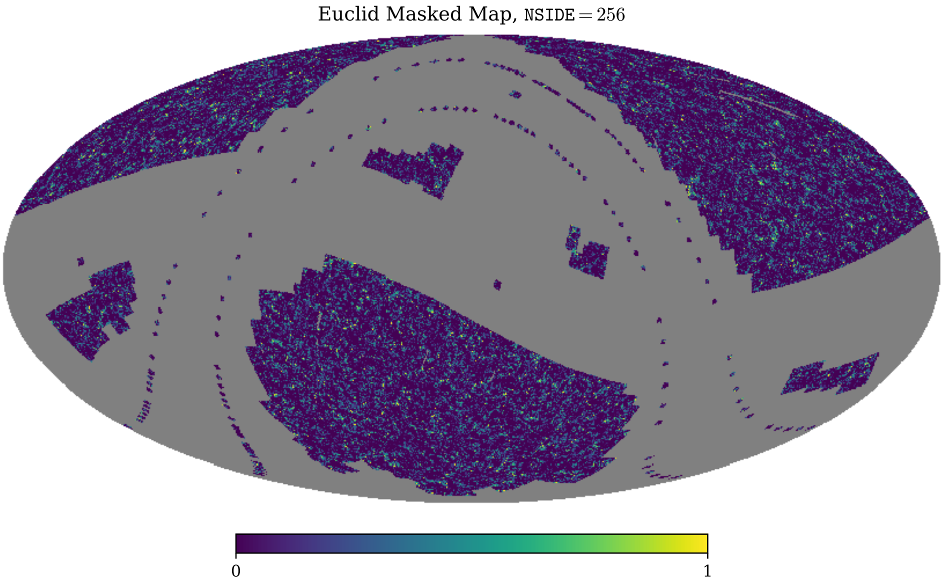


Fig. 6.21 Euclid masked catalogue, showing the survey observable sky. For this case, $N_{\rm side}$ of the mask shown in Fig. 6.20 is reduced to 256 to highlight the structure.

Fig. 6.21 shows masked photo-z catalogue generated by `SMG`, which is post-processed with survey selection function in § 6.3.2. It is clear that the grey region is masked, while the purple region shows the 'galaxy' population that is expected to be observed by the survey. Since the current version of the pipeline is designed for full-sky maps, the number of objects in the full sample has to be 3× higher ($\sim 4 \times 10^9$), as *Euclid* footprint cuts down

---

[6] More details about the survey can be accessed at https://www.euclid-ec.org/.

population by factor of `1/3`. This further stretches the computational requirements of the pipeline. However in the final version, the mask will be integrated from `Stage-I`, i.e. before Poisson sampling the Lognormal density field, subsequently reducing the computational time.

6.3.4 Stage-II Validation: Model vs. Measured Spectra

The final and most important validation step of the pipeline, is done by measuring the angular power spectrum of the '`SMG` post-processed' mock photo-z catalogues from `Stage-II` and compare the results with theoretical predictions. From the catalogue, auto- and cross- angular power spectrum is estimated using `PCL`. As said before, the code is developed in *Euclid* collaboration by *Dr. Lee Whittaker* and collaborators. Since the code is still in development stage, I have contributed to the parallelisation of the code, so that the different estimates can be computed efficiently allowing a fast validation of the pipeline. The estimated output from `PCL` is verified by comparing the results with those obtained by other estimators, like the one used to measure the angular power spectrum of TGSS sources described in Chapter 5 and also by measuring the signal from pixelised maps using `ANAFAST` routine from `HEALPix` package.

Until now it is understood that objects in a *Top-Hat* photo-z shells have a different 'true' redshift distribution in spectro-z space. The result of analysis performed in § 6.3.1 makes it clear that this distribution is well approximated by a Gaussian function. Also, as I have described at the end of § 6.3.1, the Gaussian function fitting *RMS* value of $P(z_s)$ distribution, which is stored in `SMG` log-file, is used to model the appropriate shape of the window function of `CLASS` angular power spectrum (see Eqs. (3.38) and (3.39)).

Fig. 6.22 shows the comparison of measured auto- and cross- angular power spectra with theoretical predictions obtained from `CLASS`. The results displayed here are the `I/O` spectra from two adjacent photo-z shells of equal width and centred at z_p=[0.6,0.7]. It can be noticed that, as expected, cross- spectra (in green) has lower amplitude than two auto- spectra (in red and blue), where light-coloured datapoints with errorbars are measurements and thick dark curves are models. The results shown here in Fig. 6.22 are only representative of two shells from the full sample. As anticipated before (Fig. 6.18), validation test is performed on the *Euclid* catalogue sample, I have used the 12 shells with range z_p=[0.4,1.6]. All `I/O` spectra are validated using χ^2 statistics defined in Eq. (6.20). In Table 6.1, a sample result from complete validation test is shown with reduced-χ^2 values, i.e. χ^2 divided by number of degrees of freedom. I have also used the 'recovered' spectra in `Stage-I` measured from Lognormal overdensity maps by `FLASK` as a reference for comparison tests.

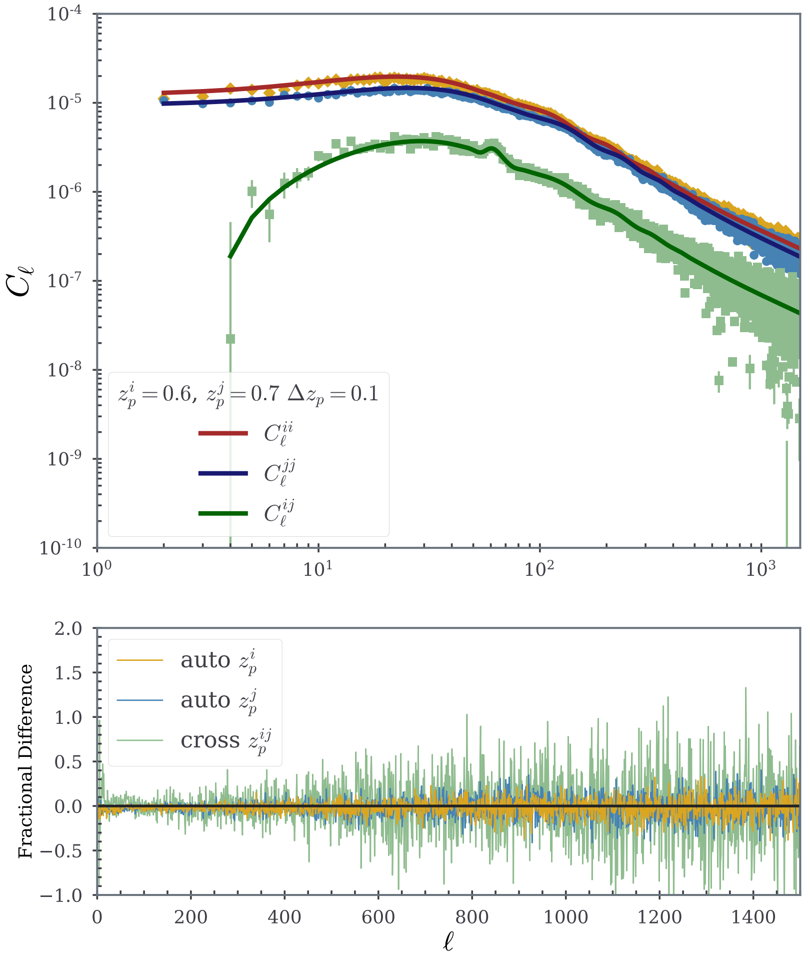


Fig. 6.22 Measured vs. Input (model) auto- and cross- C_ℓ's from sample of 100 photo-z mock catalogues. **Upper panel:** Measured Pseudo-C_ℓ from `PCL` are shown in light-orange, light-blue and light-green data points with standard error on the mean. Dark curves are model fits. The corresponding redshifts of two shells is shown in plot legend. **Lower panel:** Fractional difference of measured Pseudo-C_ℓ's and model C_ℓ's. For auto- spectra at $z_p = 0.6$ and $z_p = 0.7$ in light-orange and light-blue colour while for cross- spectra it is shown in light-green colour.

On top of this result, the use of a Geometry mask would require an additional manipulation to account for the mixing of the various multipoles (has been discussed in details in § 3.2.2) which is not accounted in this case. Instead, the results shown in Fig. 6.22 is the

Auto/Cross C_ℓ	z_s or z_p	`Stage-I` (z_s)	`Stage-II` (z_p)	Recovered C_ℓ
Auto	0.5	0.99	1.20	0.98
Cross	0.5-0.6	0.96	1.00	1.06
Cross	0.5-0.7	1.03	1.10	1.05
Cross	0.5-0.8	0.99	1.02	1.05
Auto	0.6	1.05	1.19	1.08
Cross	0.6-0.7	1.03	1.10	1.03
Cross	0.6-0.8	1.02	1.13	1.03
Auto	0.7	1.02	1.11	1.10
Cross	0.7-0.8	1.02	1.07	1.00
Auto	0.8	1.02	1.19	1.07

Table 6.1 Validation χ^2-test results shown for 4 (subsampled) shells ranging in z=[0.5,0.8] (column 2). In column 3, 4 and 5 the values are χ^2/ν from comparison of model C_ℓ with the measured ones. For `Stage-I` spectro-z (column 3) and `Stage-II` photo-z (column 4) catalogues, the spectra are estimated using `PCL`. Instead, in the last column the model is compared to reference spectra measured from Lognormal overdensity maps.

validation of the photo-z catalogues by considering full-sky samples only. Evaluating the effect of the mask is one of the next features to be added to this pipeline.

6.4 Pipeline Architecture and Computational Budget

As shown in the schematic of the pipeline in Fig. 6.1, it currently consists of two independent computational units i.e. `Stage-I` (hereafter SI): with `HEALPix+FLASK` as core codes which are written in `C++`; and `Stage-II` (hereafter SII) with `SMG` as the main module which is written in `Python`. Both units use `CPU`-based computation and not `Memory`-based. For this reason I will only limit this section to discussion of pipeline parallelisation on a multi-core architecture, and dominant factors in the processing elements of the pipeline.

For the first unit (SI), I have performed some quick tests, shown in Fig. 6.23, that summarise the pipeline's performance and expose the dominant factors in computational time budget. For these tests, I have used the full mock catalogue sample (same as used in Fig. 6.19), that has been provided to *Euclid* collaboration. SI as an external package, is already parallelised using `OpenMP`. The computational time (run-time) of the `FLASK` routine of SI is dominated by two parameters. The first one is $N_{\rm side}$, shown for two different core architectures in the left panel of Fig. 6.23, that fixes the output map resolution. It can be seen that, larger the $N_{\rm side}$ value, the higher the run-time for the code. As it is proportional to maximum multipole, $\ell_{\rm max}$, of the angular power spectrum, larger

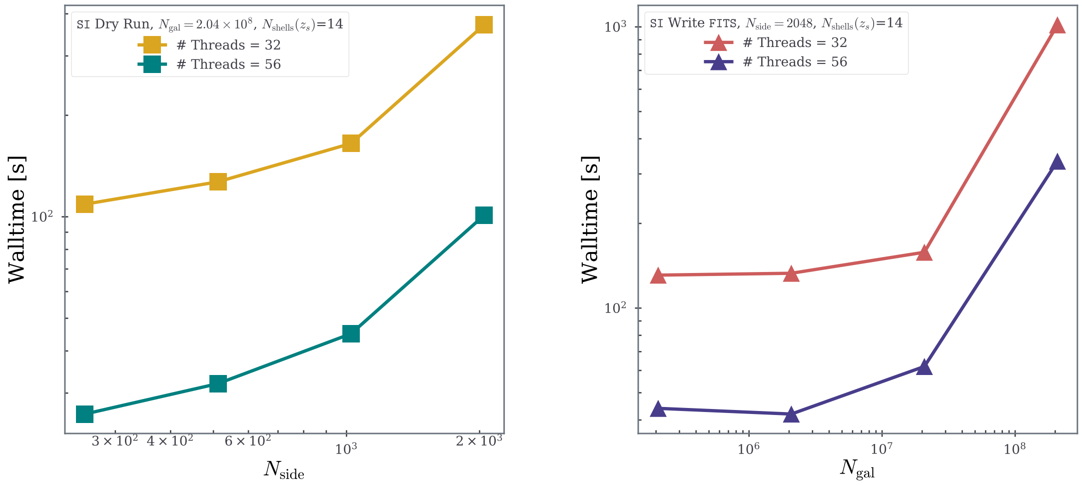


Fig. 6.23 **Left panel:** `Stage-I` code run-time for 14 z_s-shells, as a function of the map resolution $N_{\rm side}$. The output files are not written for this test (Dry Run). Golden datapoints represent code runs on 32 Core architecture and the teal-coloured are performed on 56 Core architecture using all threads. $N_{\rm gal}$ is kept constant for all cases, i.e. 2×10^8. **Right panel:** Writing time for `Stage-I` catalogues with $N_{\rm side}$ fixed to 2048 and $N_{\rm gal}$ varied from roughly 200K galaxies to 200M galaxies. Pink and purple datapoints correspond to 32 and 56 Cores respectively.

covariance matrix is used to generate correlated Lognormal fields. Due to this factor, a significant deviation from almost linear trend until $N_{\rm side} = 1024$ can be noticed, i.e. for $N_{\rm side} = 2048$ which corresponds to $\ell_{\rm max} = 6000$.

The second dominant parameter in computational budget is the total number of objects $N_{\rm gal}$. The processes where $N_{\rm gal}$ plays the role in SI include Poisson sampling Lognormal maps, generation of mock catalogue and writing the catalogue on the disk. In this case the writing part is the bottleneck, where the code does not use multi-threading option, i.e. owing to conventional `FITS` file format that is hard-coded in `FLASK`. The code uses `ASCII` tables for catalogue `FITS` files where the code consumes maximum amount of run-time. Therefore, I have subtracted the other processes where $N_{\rm gal}$ is parallelised and considered only writing time, which is shown in the right panel of Fig. 6.23. It is visible in the plot that for a catalogue containing ~ 200 million objects, the run-time 'explodes' significantly. Also, it is obvious that for faster computation, I prefer the 56-core machine on the `INFN-Roma-3` cluster[7].

---

[7] http://web-cluster.fis.uniroma3.it/

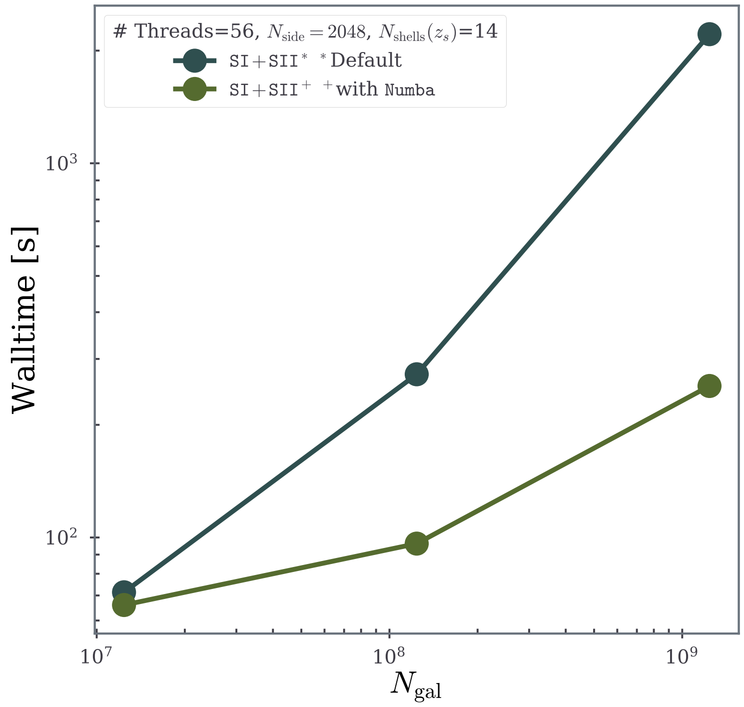


Fig. 6.24 Full pipeline run-time for 1 catalogue containing 12 photo-z shells (i.e. 14 spectro-z shells). The darker and lighter green datapoints represent without and with parallelised implementation of `Stage-II` respectively.

For SII, I have implemented parallelisation to radially displace redshifts in ideal catalogues, using various `Numba`[8] routines. Also, the `SMG` performance is improved significantly with targeted improvements using task-specific and optimised `python-numpy` libraries. This improvement further boosted the run-time by a significant margin, which is clearly visible with the darker curve that is shown in total run-time tested on 56 Core machine in Fig. 6.24. The total run-time that is required to produce one full-sky photometric catalogue imposed with *Euclid* redshift distribution in 12 photo-z shells is shown in Fig. 6.24 for increasing number of galaxies. The extreme right datapoints represent the case for the expected *Euclid* photo-z sample. Since in the current version of this pipeline, ideal catalogue processed from SI is read by `SMG` module in SII. Keeping the large number of objects in perspective, I parallelised reading and writing with `fitsio` routine in `python` which, as I have mentioned in § 6.3.4, is also used to modify the `PCL` code for on-the-fly

---

[8] https://numba.pydata.org/

reading of catalogue during validation of SI and SII. However in Fig. 6.24, the file `I/O` time is neglected as the parallelisation to I/O option is not yet added to SI core code `FLASK`.

Also, as one can notice from all run-time charts shown in this section, I have used 'Walltime' instead of `CPU` time. The reason for this is, since I run these as 'jobs' on multiple nodes[9] of the `INFN-Roma3` cluster, wall clock-time is easier to keep track of the timeline to produce the set of required mock catalogues. In principle, walltime is a sum of three terms: `CPU` time (normalised with the number of threads used on a multi-core machine), `I/O` time, and the communication delay between different nodes. In contrast to `CPU` time, which measures only the time when processor 'actively' computes the task, walltime measures the total time for the job to complete.

The ultimate goal of this pipeline is to generate a large number of mocks to assess errors and their covariance. Indeed, the accuracy by which a covariance matrix is estimated depends on the number of mock catalogues available. For the cosmological analyses that will be carried out with the *Euclid* data, a reference number for the number of required mocks is of the order $\sim 10^4$. The exact value depends on the type of analysis, the properties of the sample and the possibility to adopt hybrid approach, in which the covariance matrix can be determined by complementing the brute force approach of using mock catalogues with additional theoretical input on the statistical properties of the sample. Using the same resources that I used for run-time tests i.e. considering at least 10 nodes of the cluster are available, the expected total Walltime required to generate a sample of 10^4 mocks is ~ 200 hours (for 100 mocks it takes approximately 2 hours by distributing a stack of 10 mocks on 10 nodes with 56 cores).

In terms of storage occupancy on the disk, each mock catalogue utilises $\sim$ `300MB`, summing to `30TB` of data storage for all the mocks. Using a more compressed file format '`fits.gz`', which is already available option is `FLASK`, this number shrinks further by 50%.

6.5 Summary, Conclusions and Future Work

My goal in this aspect of the thesis was to construct a fully validated, fast and user simplified pipeline to mass produce Lognormal mock galaxy catalogues for next generation photometric surveys that will utilise tomographic clustering approach for extracting cosmological information. The main feature of these mock catalogues, was ability to preserve cross-correlation when a 3D photo-z volume is sliced radially, given the width of the shell is larger than the redshift error.

---

[9] Each node has multiple cores that can be used to distribute threads by using `OpenMP` for `C++`-based code or `Numba` for `Python`-based code.

More importantly, the first dataset which was expected to be tested with this pipeline, was that of the forthcoming *Euclid* photo-z survey, that is ‘being’ designed to produce datasets containing more than 1 billion objects up to redshift $z = 2$. Number of objects of this order would require very accurate modelling of statistical and systematic uncertainties, meaning to construct covariance matrix using at least 10^4 simulated samples that imitate the statistical and observable properties of the survey. In this regard, this order of mock catalogues that are required to fine-tune the error covariance, must be produced in no-time, keeping eye on the survey launch time.

Keeping these goals in mind, I designed the pipeline, which is described in this chapter, by assembling a set of reliable and specific public codes, and adding an originally developed module that specifically contains characteristic features of a photometric wide-survey. The pipeline (Fig. 6.1) consists of two independent stages, both of which I have carefully tested, modified and optimised for efficient production of photo-z mock catalogues. The `Stage-I` produces ideal set of Lognormal mock catalogues featuring `FLASK` and `HEALPix` modules at its core, and for generating the input of angular power spectrum that sets statistical and cosmological properties of the density field, the pipeline communicates with `CLASS` code. `Stage-I` is validated independently, before the development of `Stage-II`. The results of this validation are presented in § 6.2.9.

`Stage-II` of the pipeline was designed to assign the survey properties to archetypal mock catalogues designed in `Stage-I`. In this stage, the core module is `SMG`, which is my original program in the pipeline. `SMG` has 3 different processing elements that post-process the ideal catalogues to realistic ones: *i)* Adding the photo-z information; *ii)* Imposing survey redshift distribution; *iii)* Masking the catalogue objects with survey footprint. All of these elements have been tested thoroughly in § 6.3, along with the final validation of `Stage-II` in § 6.3.4. The χ^2 analysis results from both validations are featured in Table 6.1. The results in the table infer that, the pipeline is authentic to produce photometric mock catalogues specifically optimised to be used for tomography.

In last Section § 6.4, I presented the computational budget of the pipeline, together with discussion on performance improvement and calibration for efficient mock production. Using this pipeline, considering that at least 10 nodes embedded with 56 or more cores are available on the computing cluster/machine, it is possible to produce up to 10^4 mocks well inside 10 days. The pipeline has been used to produce the first batch of Lognormal photo-z catalogue, which is being used in the *Euclid* collaboration for internal validation activities. The full-sky sample consists of 12 shells in photo-z in range $z_p = [0.4, 1.5]$, each containing the expected number of objects by the *Euclid* survey.

As anticipated, the current pipeline is undergoing several improvements. I have already mentioned the possibility of switching from ASCII to binary output and that of using a more compressed data format. Another feature that will be added is that of non-Gaussian redshift dependent photo-z error. Together with this, by using more sophisticated and state-of-the-art codes for generating angular power spectra, the next version of the pipeline should provide the option of specifying the shape of the redshift shell window function to predict the angular power spectrum in the photo-z space.

In terms of scientific goals, next is to explore the cross-correlation between galaxy density field δ and weak-lensing fields such as cosmic shear γ and WL convergence κ, which is an important aspect of the *Euclid* photo-z survey;

Chapter 7

General Conclusions and Future Outlook

Over the past decade, and even more in the next future, increasingly large fraction of the Universe will be observed using large dedicated facilities. The motivation behind this huge effort is the need to understand the 'Nature' of the 'Dark' components that seem to dominate the mass-energy budget today. Universe. As we know already, one of them, the Dark Energy, is possibly driving the accelerated expansion of the Universe, however its own existence, and consequently its physical nature is much debated. Precious information on the nature of this elusive component can be obtained from the study of the distribution of galaxies and their motions. As a result, many large galaxy surveys have been carried out and more, much larger, are expected to be performed in the next decade. In parallel, a significant effort has been made to improve the accuracy of theoretical models as well as the statistical methods to analyse the distribution of the extragalactic sources.

In this Thesis I considered, as a case study, two such large surveys. The first one, in the X-ray band, is being performed by the recently launched *eROSITA* satellite. Its goal is to trace the galaxy clusters over the full sky in next 4 years. The second one, are the surveys that will be carried out by the upcoming *Euclid* satellite. The main focus here in this thesis is on the *Euclid* photometric redshift survey, rather than the spectroscopic one. In addition to these two future datasets, for which quantitative studies can only be performed through simulated data, I also considered an existing dataset, the recently released TGSS survey in the radio band, that allows one to study the angular clustering properties of radio objects over a large fraction of the sky.

The motivation for considering such different types of datasets, whose only similarity is the wide sky coverage, was to show that the use of multiple probes can provide complementary constraints that, although with different precision, can be used to measure fundamental cosmological parameters and identify possible systematic errors by comparing their results with model predictions. Also, besides the cosmological aspect, these combined analyses, were also important to address astrophysical issues (i.e. the nature of radio objects in the TGSS sample) and also to identify potential issues with the dataset itself.

The first analysis presented in Chapter 4, I used galaxy clusters as a cosmological probe and analyse their 3D spatial distribution to measure the linear growth rate of density fluctuations at moderate redshifts ($z \leq 1$) from the redshift space distortions in their 2-point clustering statistics. The goal was to assess if and how well these massive cosmic tracers can be used to probe the linear growth rate and provide an estimate of this key quantity, that is independent and alternative, to that obtained from galaxy samples. This analysis, which was based on simulated data of cluster-sized haloes selected with mass threshold of $M_{\rm halo} > 1 \times 10^{14} h^{-1}\ {\rm M}_\odot$, was designed to match the characteristics of the *eROSITA* survey, i.e. when the spectroscopic follow-up surveys of purely X-ray based *eROSITA* mission would bring us large 3D datasets of clusters to perform clustering analyses as cosmological case studies in near future. And therefore one of the motives of this analysis was to provide a forecast for these datasets with the estimates of the redshift distortion parameter β from 2-point correlation analyses performed in configuration and Fourier space.

The main result of the analysis are:

- In all analyses I performed, linear theory predictions matched the measured 2-point statistics over a large range of scales (corresponding to $k < 0.08h\ {\rm Mpc}^{-1}$ in Fourier space and $> 50\ h^{-1}$ Mpc in configuration space). The adequacy of the linear theory in describing the clustering properties of galaxy clusters is one of the main benefits in considering these as cosmological probes.

- From the measurement of the anisotropic cluster power spectrum at $z = 0$ and $z = 0.5$ one would obtain an unbiased estimate of β using linear perturbation theory. Uncertainties are within 10% of the estimated value. The $z = 0$ sample that I analysed, matched the number density of the expected Deeper-*eROSITA* catalogue of $\sim 90,000$ clusters corresponding to mass scale $M_{\rm halo} > 5 \times 10^{13} h^{-1}\ {\rm M}_\odot$, calculated at mean redshift of $z = 0.35$. The $z = 0.5$ one, however matched the number of objects which is $\sim 40,000$ for the brighter Focused-*eROSITA* sample with more massive, i.e. $M_{\rm halo} > 5 \times 10^{14} h^{-1}\ {\rm M}_\odot$, clusters calculated at mean redshift of $z = 0.46$.

- The same analysis that was performed in configuration space confirmed that the measured β parameter is free of systematic errors. Random errors, however, are larger than in Fourier space by $\sim 50\%$.

- Overall these results showed that if a large sample of clusters selected above a given mass threshold, which is possible in the X-ray band, thanks to the tight correlation between X-ray luminosity and mass, and also thanks to their high correlation signal, these massive objects can be used to estimate the linear growth rate of density fluctuations from the measured β parameter and the halo bias model. This measurement, although less precise, is independent than the one that is obtained from galaxy samples and can be used to validate and corroborate the results of galaxy clustering analyses.

In Chapter 5 I presented another type of analysis that is already a co-authored publication. Triggered by the excess of power on the dipole moments of the NVSS catalogue, a collection of radio objects, were investigated by performing the angular clustering analysis on a newer, wide radio survey: the TGSS, to search for similar excess clustering on very comparative angular scales. Unlike the cluster catalogue that was analysed in Chapter 4, the sample in this case was purely two-dimensional, since radio datasets lack the information on the distance of the objects, and was composed by different type of objects, ranging from relatively faint and nearby star forming galaxies to very bright distant quasars. This study extended previous analyses of the same datasets that have focused in small angular scales (Rana and Singh Bagla, 2018) and on the dipole of angular power spectrum (Bengaly et al., 2018). The main results of the investigations, that have been published in (Dolfi et al., 2019) are:

- From the positive cross-correlation with the SDSS-QSO catalogues it was inferred that the distribution of TGSS objects extends to high redshifts.

- After computing the angular auto-spectrum of the TGSS objects, we detected an auto-correlation signal at all angular scales, like in the NVSS case. However, the power at multipoles below $\ell \sim 40$ is significantly higher in the TGSS sample.

- The results remained robust to tests performed for different flux cuts as well as for checking the geometry mask effects.

- After assuming realistic models for the population of TGSS objects, their bias and redshift distribution we obtained theoretical predictions for the TGSS angular power spectrum, and found that it fails to match the large power detected at low multipoles. The mismatch is highly significant, also when one takes into account theoretical uncertainties in modelling the TGSS angular spectrum.

- We concluded that the excess power at large angular separation is not genuine but probably reflects some undetected systematic uncertainties in the TGSS catalogues

> and its selection criteria. Further studies are required to identify the origin of this mismatch. These results were obtained through a team work in which my main contribution was the modelling of the bias for the various type of radio objects included in the TGSS catalogues. These bias models were used to predict the angular power spectrum of TGSS galaxies and compare it to the estimated one.

The third and final analysis presented in this thesis, which is also different from the previous ones, can be found in 6. The theme of this particular project relies on investigating the clustering properties of objects for which the angular position and photometric redshifts, instead of spectroscopic ones, are used to infer the radial distance of the objects. Comparatively larger errors in the estimate of the redshift using photometric information only, induce large random errors that erase a significant fraction of the clustering information along the radial direction. In this case a full three dimensional analysis of the galaxy distribution would not be justified, on the basis of its high computation time as well as noise level in the dataset. Instead, a strategy to divide sample into radial shells and perform a tomographic analysis would be ideal approach in such case.

The next on-board space-based *Euclid* satellite mission is expected to generate a very large, photo-z galaxy catalogue. On these futuristic samples, a tomographic approach, in which auto- and cross- angular spectra are measured in various redshift shells, will be performed to infer fundamental cosmological parameters.

For such large datasets, the problem arises while estimating random and systematic errors, for the enormous size of the covariance matrix needed to assess the level of accuracy required in this case. The typical approach is that of generating numerous realistic random catalogues, perform the same type of tomographic analyses and assess the amplitude of both systematic and random errors. Ideally one would use synthetic catalogues that are obtained from time consuming N-body simulations. However, a simpler and also much faster approach would be, to exploit the fact that the 1-point probability distribution function of the galaxy density field can be approximated by a Lognormal function. This model assumption, would then be used to generate simplified mock catalogues of objects, that are tested to be accurate enough to characterise errors and their covariance from large samples of data.

My work consisted in setting up an efficient pipeline for the generation of a large number of Lognormal mock catalogues, with objects and their photometric redshift estimates, that are optimised for performing tomographic analyses. Alternatively the objective was to measure their angular spectra to spot the presence of systematic errors and quantify the magnitude of the random errors. To build this pipeline I assembled publicly available codes (to generate correlated Lognormal catalogues of objects and map their ‘true’ distribution) and added my own code to simulate the effects of known selection criteria and observational biases of the photometric survey sample under consideration. It also features built-in validation tests designed to check whether the angular correlation properties of the output mock objects match expectations. As described above, I used *Euclid* photo-z survey as a first test case for this pipeline. During this work, I optimised the performance of this pipeline, which, using the computational resources on a high performance machine, can now produce 10^4 mock catalogues in 10 days time. This work has been performed in the framework of the *Euclid* collaboration and the first batch of 1000 mock catalogues has already been delivered to the community and is currently being used to carry out further tests. Also this is a very much work in progress. An updated version of the pipeline, currently under construction, will feature geometry mask effects, non-Gaussian redshift-dependent photo-z errors and the possibility to generate self consistently objects’ position as well as weak lensing shear map for a more ambitious auto- and cross- correlation analysis.

References

Abate, A. et al. (2012). Large Synoptic Survey Telescope: Dark Energy Science Collaboration.

Abazajian, K. N. et al. (2009). The Seventh Data Release of the Sloan Digital Sky Survey. *Astrophys. J. Suppl.*, 182:543–558.

Abbott, T. et al. (2005). The dark energy survey.

Ade, P. A. R. et al. (2014). Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.*, 571:A16.

Ade, P. A. R. et al. (2016a). Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13.

Ade, P. A. R. et al. (2016b). Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts. *Astron. Astrophys.*, 594:A24.

Aghamousa, A. et al. (2016). The DESI Experiment Part I: Science,Targeting, and Survey Design.

Aghanim, N. et al. (2018). Planck 2018 results. VI. Cosmological parameters.

Agrawal, A., Makiya, R., Chiang, C.-T., Jeong, D., Saito, S., and Komatsu, E. (2017). Generating log-normal mock catalog of galaxies in redshift space. , 10:003.

Alam, S. et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Mon. Not. Roy. Astron. Soc.*, 470(3):2617–2652.

Allison, R., Lindsay, S. N., Sherwin, B. D., de Bernardis, F., Bond, J. R., Calabrese, E., Devlin, M. J., Dunkley, J., Gallardo, P., Henderson, S., Hincks, A. D., Hlozek, R., Jarvis, M., Kosowsky, A., Louis, T., Madhavacheril, M., McMahon, J., Moodley, K., Naess, S., Newburgh, L., Niemack, M. D., Page, L. A., Partridge, B., Sehgal, N., Spergel, D. N., Staggs, S. T., van Engelen, A., and Wollack, E. J. (2015). The Atacama Cosmology Telescope: measuring radio galaxy bias through cross-correlation with lensing. , 451:849–858.

Alonso, D., Salvador, A. I., Sánchez, F. J., Bilicki, M., García-Bellido, J., and Sánchez, E. (2015). Homogeneity and isotropy in the Two Micron All Sky Survey Photometric Redshift catalogue. , 449:670–684.

Ananthakrishnan, S. (1995). The Giant Meterwave Radio Telescope / GMRT. *Journal of Astrophysics and Astronomy Supplement*, 16:427.

Anderson, L. et al. (2013). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample. *Mon. Not. Roy. Astron. Soc.*, 427(4):3435–3467.

Anderson, L. et al. (2014). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples. *Mon. Not. Roy. Astron. Soc.*, 441(1):24–62.

Asorey, J., Crocce, M., and Gaztañaga, E. (2014). Redshift-space distortions from the cross-correlation of photometric populations. , 445:2825–2835.

Asorey, J., Crocce, M., Gaztanaga, E., and Lewis, A. (2012). Recovering 3D clustering information with angular correlations. *Mon. Not. Roy. Astron. Soc.*, 427:1891.

Balaguera-Antolínez, A., Bilicki, M., Branchini, E., and Postiglione, A. (2018). Extracting cosmological information from the angular power spectrum of the 2MASS Photometric Redshift catalogue. *Mon. Not. Roy. Astron. Soc.*, 476(1):1050–1070.

Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. , 340(4-5):291–472.

Batista, R. C. and Marra, V. (2017). Clustering dark energy and halo abundances. *JCAP*, 1711(11):048.

Becker, R. H., White, R. L., and Helfand, D. J. (1995). The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. , 450:559.

Bengaly, C. A. P., Maartens, R., and Santos, M. G. (2018). Probing the Cosmological Principle in the counts of radio galaxies at different frequencies. , 4:031.

Bennett, C. L., Banday, A., Gorski, K. M., Hinshaw, G., Jackson, P., Keegstra, P., Kogut, A., Smoot, G. F., Wilkinson, D. T., and Wright, E. L. (1996). Four year COBE DMR cosmic microwave background observations: Maps and basic results. *Astrophys. J.*, 464:L1–L4.

Berlind, A. A., Narayanan, V. K., and Weinberg, D. H. (2001). Biased Galaxy Formation and Measurements of β. , 549:688–701.

Bersanelli, M., Mandolesi, N., Butler, R. C., Mennella, A., Villa, F., Aja, B., Artal, E., Artina, E., Baccigalupi, C., Balasini, M., and et al. (2010). Planckpre-launch status: Design and description of the low frequency instrument. *Astronomy and Astrophysics*, 520:A4.

Best, P. N., Arts, J. N., Röttgering, H. J. A., Rengelink, R., Brookes, M. H., and Wall, J. (2003). CENSORS: A Combined EIS-NVSS Survey Of Radio Sources - I. Sample definition, radio data and optical identifications. , 346:627–683.

Beutler, F., Blake, C., Colless, M., Jones, D. H., Staveley-Smith, L., Poole, G. B., Campbell, L., Parker, Q., Saunders, W., and Watson, F. (2012). The 6dF Galaxy Survey: $z \approx 0$ measurement of the growth rate and σ_8. *Mon. Not. Roy. Astron. Soc.*, 423:3430–3444.

Bianchi, D., Guzzo, L., Branchini, E., Majerotto, E., de la Torre, S., Marulli, F., Moscardini, L., and Angulo, R. E. (2012). Statistical and systematic errors in redshift-space distortion measurements from large surveys. , 427:2420–2436.

Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., and Steward, L. (2013). 2MASS Photometric Redshift catalog: a comprehensive three-dimensional census of the whole sky. [Astrophys. J. Suppl.210,9(2014)].

Blain, A. W., Smail, I., Ivison, R. J., Kneib, J. P., and Frayer, D. T. (2002). Submillimeter galaxies. *Phys. Rept.*, 369:111–176.

Blake, C. et al. (2011). The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9. *Mon. Not. Roy. Astron. Soc.*, 415:2876.

Blake, C., Ferreira, P. G., and Borrill, J. (2004a). The angular power spectrum of NVSS radio galaxies. , 351:923–934.

Blake, C., Mauch, T., and Sadler, E. M. (2004b). Angular clustering in the Sydney University Molonglo Sky Survey. , 347:787–794.

Blake, C. and Wall, J. (2002a). A velocity dipole in the distribution of radio galaxies. , 416:150–152.

Blake, C. and Wall, J. (2002b). Measurement of the angular correlation function of radio galaxies from the NRAO VLA Sky Survey. , 329:L37–L41.

Blanton, M. R. et al. (2017). Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies and the Distant Universe. *Astron. J.*, 154(1):28.

Blas, D., Lesgourgues, J., and Tram, T. (2011). The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes. *JCAP*, 1107:034.

Bock, D. C.-J., Large, M. I., and Sadler, E. M. (1999). SUMSS: A Wide-Field Radio Imaging Survey of the Southern Sky. I. Science Goals, Survey Design, and Instrumentation. , 117:1578–1593.

Bonaldi, A., Bonato, M., Galluzzi, V., Harrison, I., Massardi, M., Kay, S., De Zotti, G., and Brown, M. L. (2019). The Tiered Radio Extragalactic Continuum Simulation (T-RECS). , 482:2–19.

Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N. (1991). Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. , 379:440.

Braun, R., Bourke, T., Green, J. A., Keane, E., and Wagg, J. (2015). Advancing Astrophysics with the Square Kilometre Array. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 174.

Cabré, A. and Gaztañaga, E. (2009). Clustering of luminous red galaxies - II. Small-scale redshift-space distortions. , 396:1119–1131.

Cai, Y.-C. and Bernstein, G. (2012). Combining weak-lensing tomography and spectroscopic redshift surveys. *Monthly Notices of the Royal Astronomical Society*, 422(2):1045–1056.

Campagne, J. E., Neveu, J., and Plaszczynski, S. (2017). Angpow: a software for the fast computation of accurate tomographic power spectra. *Astron. Astrophys.*, 602:A72.

Carroll, S. M. (2019). *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press.

Chen, S. and Schwarz, D. J. (2016). Angular two-point correlation of NVSS galaxies revisited. , 591:A135.

Chiang, C.-T. et al. (2013). Galaxy redshift surveys with sparse sampling. *JCAP*, 1312:030.

Chuang, C.-H. and Wang, Y. (2012). Measurements of H(z) and D_A(z) from the two-dimensional two-point correlation function of Sloan Digital Sky Survey luminous red galaxies. , 426:226–236.

Cole, S. et al. (2005). The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications. *Mon. Not. Roy. Astron. Soc.*, 362:505–534.

Coles, P. and Jones, B. (1991). A lognormal model for the cosmological mass distribution. , 248:1–13.

Coles, P. and Lucchin, F. (1995). *Cosmology: The Origin and evolution of cosmic structure*.

Colin, J., Mohayaee, R., Rameez, M., and Sarkar, S. (2017). High-redshift radio galaxies and divergence from the CMB dipole. , 471:1045–1055.

Colless, M. et al. (2001). The 2dF Galaxy Redshift Survey: Spectra and redshifts. *Mon. Not. Roy. Astron. Soc.*, 328:1039.

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., and Broderick, J. J. (1998). The NRAO VLA Sky Survey. , 115:1693–1716.

Cress, C. M., Helfand, D. J., Becker, R. H., Gregg, M. D., and White, R. L. (1996). The Angular Two-Point Correlation Function for the FIRST Radio Survey. , 473:7.

Davis, M. and Peebles, P. J. E. (1983). A survey of galaxy redshifts. V - The two-point position and velocity correlations. , 267:465–482.

de Gasperin, F., Intema, H. T., and Frail, D. A. (2018). A radio spectral index map and catalogue at 147-1400 MHz covering 80 per cent of the sky. , 474:5008–5022.

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., and Valentijn, E. A. (2012). The kilo-degree survey. *Experimental Astronomy*, 35(1-2):25–44.

de la Torre, S., Guzzo, L., Peacock, J. A., Branchini, E., Iovino, A., Granett, B. R., Abbas, U., Adami, C., Arnouts, S., Bel, J., Bolzonella, M., Bottini, D., Cappi, A., Coupon, J., Cucciati, O., Davidzon, I., De Lucia, G., Fritz, A., Franzetti, P., Fumana, M., Garilli, B., Ilbert, O., Krywult, J., Le Brun, V., Le Fèvre, O., Maccagni, D., Małek, K., Marulli, F., McCracken, H. J., Moscardini, L., Paioro, L., Percival, W. J., Polletta, M., Pollo, A., Schlagenhaufer, H., Scodeggio, M., Tasca, L. A. M., Tojeiro, R., Vergani, D., Zanichelli, A., Burden, A., Di Porto, C., Marchetti, A., Marinoni, C., Mellier, Y., Monaco, P., Nichol, R. C., Phleps, S., Wolk, M., and Zamorani, G. (2013). The VIMOS Public Extragalactic Redshift Survey (VIPERS) . Galaxy clustering and redshift-space distortions at z 0.8 in the first data release. , 557:A54.

de la Torre, S., Jullo, E., Giocoli, C., Pezzotta, A., Bel, J., Granett, B. R., Guzzo, L., Garilli, B., Scodeggio, M., Bolzonella, M., Abbas, U., Adami, C., Bottini, D., Cappi, A., Cucciati, O., Davidzon, I., Franzetti, P., Fritz, A., Iovino, A., Krywult, J., Le Brun, V., Le Fèvre, O., Maccagni, D., Małek, K., Marulli, F., Polletta, M., Pollo, A., Tasca, L. A. M., Tojeiro, R., Vergani, D., Zanichelli, A., Arnouts, S., Branchini, E., Coupon, J., De Lucia, G., Ilbert, O., Moutard, T., Moscardini, L., Peacock, J. A., Metcalf, R. B., Prada, F., and Yepes, G. (2016). The VIMOS Public Extragalactic Redshift Survey (VIPERS). Gravity test from the combination of redshift-space distortions and galaxy-galaxy lensing at $0.5z1.2$. *ArXiv e-prints*.

Di Dio, E., Montanari, F., Lesgourgues, J., and Durrer, R. (2013). The CLASSgal code for relativistic cosmological large scale structure. , 11:044.

Dodelson, S. (2003). *Modern cosmology*.

Dolfi, A., Branchini, E., Bilicki, M., Balaguera-Antolínez, A., Prandoni, I., and Pand it, R. (2019). Clustering properties of TGSS radio sources. , 623:A148.

Doran, M. (2005). CMBEASY: an object oriented code for the cosmic microwave background. *JCAP*, 0510:011.

Doré, O. et al. (2018). WFIRST Science Investigation Team "Cosmology with the High Latitude Survey" Annual Report 2017.

Einstein, A. (1916). Die grundlage der allgemeinen relativitätstheorie. *Annalen der Physik*, 354(7):769–822.

Eisenstein, D. J. et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *Astrophys. J.*, 633:560–574.

Eisenstein, D. J. et al. (2011). SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems. *Astron. J.*, 142:72.

Eisenstein, D. J. and Zaldarriaga, M. (2001). Correlations in the Spatial Power Spectra Inferred from Angular Clustering: Methods and Application to the Automated Plate Measuring Survey. , 546:2–19.

Ellis, G. F. R. and Baldwin, J. E. (1984). On the expected anisotropy of radio source counts. , 206:377–381.

Feldman, H. A., Kaiser, N., and Peacock, J. A. (1994). Power-spectrum analysis of three-dimensional redshift surveys. , 426:23–37.

Fernández-Cobos, R., Vielva, P., Pietrobon, D., Balbi, A., Martínez-González, E., and Barreiro, R. B. (2014). Searching for a dipole modulation in the large-scale structure of the Universe. , 441:2392–2397.

Ferramacho, L. D., Santos, M. G., Jarvis, M. J., and Camera, S. (2014). Radio galaxy populations and the multitracer technique: pushing the limits on primordial non-Gaussianity. , 442:2511–2518.

Ferraro, S., Schmidt, F., and Hu, W. (2011). Cluster abundance inf(r)gravity models. *Physical Review D*, 83(6).

Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386.

Friedrich, O., Seitz, S., Eifler, T. F., and Gruen, D. (2016). Performance of internal Covariance Estimators for Cosmic Shear Correlation Functions. *Mon. Not. Roy. Astron. Soc.*, 456(3):2662–2680.

Gibelyou, C. and Huterer, D. (2012). Dipoles in the sky. , 427:1994–2021.

Górski, K. M., Hivon, E., Banday, A. J., Wand elt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. , 622(2):759–771.

Gregory, P. C. and Condon, J. J. (1991). The 87GB catalog of radio sources covering delta between O and + 75 deg at 4.85 GHz. , 75:1011–1291.

Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., and Dalla Vecchia, C. (2016). Gaussian covariance matrices for anisotropic galaxy clustering measurements. , 457(2):1577–1592.

Guo, H. et al. (2013). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Luminosity and Color Dependence and Redshift Evolution. *Astrophys. J.*, 767:122.

Guzzo, L. et al. (2008). A test of the nature of cosmic acceleration using galaxy redshift distortions. *Nature*, 451:541–545.

Guzzo, L. et al. (2013). VIPERS: An Unprecedented View of Galaxies and Large-Scale Structure Halfway Back in the Life of the Universe. *The ESO Messenger*, 151:41–46.

Hale, C. L., Jarvis, M. J., Delvecchio, I., Hatfield, P. W., Novak, M., Smolčić, V., and Zamorani, G. (2018). The clustering and bias of radio-selected AGN and star-forming galaxies in the COSMOS field. , 474:4133–4150.

Hamilton, A. J. S. (1992). Measuring Omega and the real correlation function from the redshift correlation function. , 385:L5–L8.

Hamilton, A. J. S. (1997). Linear redshift distortions: A Review. In *Ringberg Workshop on Large Scale Structure Ringberg, Germany, September 23-28, 1996*.

Hatton, S. and Cole, S. (1999). Estimating β from redshift-space distortions in the 2dF galaxy survey. , 310:1137–1146.

Hawken, A. J., Aubert, M., Pisani, A., Cousinou, M.-C., Escoffier, S., Nadathur, S., Rossi, G., and Schneider, D. P. (2019). Constraints on the growth of structure around cosmic voids in eBOSS DR14.

Hawkins, E. et al. (2003). The 2dF Galaxy Redshift Survey: Correlation functions, peculiar velocities and the matter density of the universe. *Mon. Not. Roy. Astron. Soc.*, 346:78.

Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., Nolta, M. R., Halpern, M., Hill, R. S., Odegard, N., and et al. (2013). Nine-yearwilkinson microwave anisotropy probe(wmap) observations: Cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2):19.

Hopkins, A. M., Afonso, J., Chan, B., Cram, L. E., Georgakakis, A., and Mobasher, B. (2003). The Phoenix Deep survey: The 1.4 GHz microJansky catalog. *Astron. J.*, 125:465.

Howlett, C., Lewis, A., Hall, A., and Challinor, A. (2012). CMB power spectrum parameter degeneracies in the era of precision cosmology. *JCAP*, 1204:027.

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.

Hurley-Walker, N. (2017). A Rescaled Subset of the Alternative Data Release 1 of the TIFR GMRT Sky Survey. *ArXiv e-prints*.

Hurley-Walker, N., Callingham, J. R., Hancock, P. J., Franzen, T. M. O., Hindson, L., Kapińska, A. D., Morgan, J., Offringa, A. R., Wayth, R. B., Wu, C., Zheng, Q., Murphy, T., Bell, M. E., Dwarakanath, K. S., For, B., Gaensler, B. M., Johnston-Hollitt, M., Lenc, E., Procopio, P., Staveley-Smith, L., Ekers, R., Bowman, J. D., Briggs, F., Cappallo, R. J., Deshpande, A. A., Greenhill, L., Hazelton, B. J., Kaplan, D. L., Lonsdale, C. J., McWhirter, S. R., Mitchell, D. A., Morales, M. F., Morgan, E., Oberoi, D., Ord, S. M., Prabu, T., Shankar, N. U., Srivani, K. S., Subrahmanyan, R., Tingay, S. J., Webster, R. L., Williams, A., and Williams, C. L. (2017). GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey - I. A low-frequency extragalactic catalogue. , 464:1146–1167.

Intema, H. T., Jagannathan, P., Mooley, K. P., and Frail, D. A. (2017). The GMRT 150 MHz all-sky radio survey. First alternative data release TGSS ADR1. , 598:A78.

Intema, H. T., van der Tol, S., Cotton, W. D., Cohen, A. S., van Bemmel, I. M., and Röttgering, H. J. A. (2009). Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results. , 501:1185–1205.

Jain, B. and Zhang, P. (2008). Observational Tests of Modified Gravity. *Phys. Rev.*, D78:063503.

Jarvis, M., Bernstein, G., and Jain, B. (2004). The skewness of the aperture mass statistic. , 352:338–352.

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., and Yoshida, N. (2001). The mass function of dark matter haloes. , 321:372–384.

Jeong, D., Dai, L., Kamionkowski, M., and Szalay, A. S. (2015). The redshift-space galaxy two-point correlation function and baryon acoustic oscillations. , 449:3312–3322.

Joachimi, B. and Bridle, S. L. (2010). Simultaneous measurement of cosmology and intrinsic alignments using joint cosmic shear and galaxy number density correlations. , 523:A1.

Joachimi, B., Taylor, A. N., and Kiessling, A. (2011). Cosmological information in Gaussianised weak lensing signals. *Mon. Not. Roy. Astron. Soc.2011*, 418:145–169. [Mon. Not. Roy. Astron. Soc.418,145(2011)].

Jones, B. J. T. (2017). *Precision Cosmology: The First Half Million Years*. Cambridge University Press.

Kaiser, N. (1987). Clustering in real space and in redshift space. , 227:1–21.

Kayo, I., Taruya, A., and Suto, Y. (2001). Probability distribution function of cosmological density fluctuations from Gaussian initial condition: comparison of one- and two-point log-normal model predictions with n-body simulations. *Astrophys. J.*, 561:22–34.

Kolb, E. W. and Turner, M. S. (1990). The Early Universe. *Front. Phys.*, 69:1–547.

Kwan, J., Lewis, G. F., and Linder, E. V. (2012). Mapping Growth and Gravity with Robust Redshift Space Distortions. , 748:78.

Lahav, O. and Suto, Y. (2004). Measuring our universe from galaxy redshift surveys. *Living Rev. Rel.*, 7:8.

Landy, S. D. and Szalay, A. S. (1993). Bias and variance of angular correlation functions. , 412:64–71.

Laureijs, R. et al. (2011). Euclid Definition Study Report.

Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47:49–59.

Lesgourgues, J. (2011a). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *arXiv e-prints*, page arXiv:1104.2932.

Lesgourgues, J. (2011b). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *ArXiv e-prints*.

Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476.

Limber, D. N. (1953). The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. , 117:134.

Linder, E. V. (2005). Cosmic growth history and expansion history. , 72(4):043529.

Linder, E. V. (2017). Cosmic growth and expansion conjoined. *Astroparticle Physics*, 86:41–45.

Lindsay, S. N., Jarvis, M. J., Santos, M. G., Brown, M. J. I., Croom, S. M., Driver, S. P., Hopkins, A. M., Liske, J., Loveday, J., Norberg, P., and Robotham, A. S. G. (2014). Galaxy and Mass Assembly: the evolution of bias in the radio source population to z
1.5. , 440:1527–1541.

Loan, A. J., Wall, J. V., and Lahav, O. (1997). The correlation function of radio sources. , 286:994–1002.

Magliocchetti, M., Maddox, S. J., Hawkins, E., Peacock, J. A., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., de Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C. A., Jones, B., Lahav, O., Lewis, I., Lumsden, S., Norberg, P., Peterson, B. A., Sutherland, W., Taylor, K., and 2dFGRS Team (2004). The 2dF galaxy redshift survey: clustering properties of radio galaxies. , 350:1485–1494.

Magliocchetti, M., Popesso, P., Brusa, M., Salvato, M., Laigle, C., McCracken, H. J., and Ilbert, O. (2017). The clustering properties of radio-selected AGN and star-forming galaxies up to redshifts z
3. , 464:3271–3280.

Mandelbaum, R., Li, C., Kauffmann, G., and White, S. D. M. (2009). Halo masses for optically selected and for radio-loud AGN from clustering and galaxy-galaxy lensing. , 393:377–392.

Marulli, F., Veropalumbo, A., Moscardini, L., Cimatti, A., and Dolag, K. (2017). Redshift-space distortions of galaxies, clusters, and AGN. Testing how the accuracy of growth rate measurements depends on scales and sample selections. , 599:A106.

Merloni, A., Predehl, P., Becker, W., Böhringer, H., Boller, T., Brunner, H., Brusa, M., Dennerl, K., Freyberg, M., Friedrich, P., Georgakakis, A., Haberl, F., Hasinger, G., Meidinger, N., Mohr, J., Nandra, K., Rau, A., Reiprich, T. H., Robrade, J., Salvato, M., Santangelo, A., Sasaki, M., Schwope, A., Wilms, J., and German eROSITA Consortium, t. (2012). eROSITA Science Book: Mapping the Structure of the Energetic Universe. *ArXiv e-prints*.

Mo, H., van den Bosch, F. C., and White, S. (2010). *Galaxy Formation and Evolution*.

Mo, H. J. and White, S. D. M. (1996). An analytic model for the spatial clustering of dark matter haloes. , 282(2):347–361.

Mohammad, F. G., de la Torre, S., Bianchi, D., Guzzo, L., and Peacock, J. A. (2016). Group-galaxy correlations in redshift space as a probe of the growth of structure. , 458:1948–1963.

Moster, B. P., Naab, T., and White, S. D. M. (2013). Galactic star formation and accretion histories from matching galaxies to dark matter haloes. , 428:3121–3138.

Myers, S. T. and VLASS Survey Team, S. S. G. . (2018). The VLA Sky Survey (VLASS): Overview and First Results. In *American Astronomical Society Meeting Abstracts #231*, volume 231 of *American Astronomical Society Meeting Abstracts*, page 231.08.

Negrello, M., Magliocchetti, M., and De Zotti, G. (2006). The large-scale clustering of radio sources. , 368:935–942.

Neyrinck, M. C. (2011). Rejuvenating the Matter Power Spectrum III: The Cosmology Sensitivity of Gaussianized Power Spectra. *Astrophys. J.*, 742:91.

Norris, R. P., Hopkins, A. M., Afonso, J., Brown, S., Condon, J. J., Dunne, L., Feain, I., Hollow, R., Jarvis, M., Johnston-Hollitt, M., Lenc, E., Middelberg, E., Padovani, P., Prandoni, I., Rudnick, L., Seymour, N., Umana, G., Andernach, H., Alexander, D. M., Appleton, P. N., Bacon, D., Banfield, J., Becker, W., Brown, M. J. I., Ciliegi, P., Jackson, C., Eales, S., Edge, A. C., Gaensler, B. M., Giovannini, G., Hales, C. A., Hancock, P., Huynh, M. T., Ibar, E., Ivison, R. J., Kennicutt, R., Kimball, A. E., Koekemoer, A. M., Koribalski, B. S., López-Sánchez, Á. R., Mao, M. Y., Murphy, T., Messias, H., Pimbblet, K. A., Raccanelli, A., Randall, K. E., Reiprich, T. H., Roseboom, I. G., Röttgering, H., Saikia, D. J., Sharp, R. G., Slee, O. B., Smail, I., Thompson, M. A., Urquhart, J. S., Wall, J. V., and Zhao, G.-B. (2011). EMU: Evolutionary Map of the Universe. , 28:215–248.

Nusser, A. and Tiwari, P. (2015). The Clustering of Radio Galaxies: Biasing and Evolution versus Stellar Mass. , 812:85.

Nørgaard-Nielsen, H. U. (2018). Confirmation of the detection of B modes in the Planck polarization maps. *Astron. Nachr.*, 339(6):432–439.

Oka, A., Saito, S., Nishimichi, T., Taruya, A., and Yamamoto, K. (2014). Simultaneous constraints on the growth of structure and cosmic expansion from the multipole power spectra of the SDSS DR7 LRG sample. *Mon. Not. Roy. Astron. Soc.*, 439:2515–2530.

Okumura, T. (2008). An Analysis of SDSS Data Using the Galaxy Anisotropic Correlation Function. *Astronomical Herald*, 101:589–597.

Okumura, T. and Jing, Y. P. (2011). Systematic Effects on Determination of the Growth Factor from Redshift-space Distortions. , 726:5.

Overzier, R. A., Röttgering, H. J. A., Rengelink, R. B., and Wilman, R. J. (2003). The spatial clustering of radio sources in NVSS and FIRST; implications for galaxy clustering evolution. , 405:53–72.

Padmanabhan, T. (2002). *Theoretical Astrophysics - Volume 3, Galaxies and Cosmology*, volume 3.

Pâris, I., Petitjean, P., Aubourg, É., Myers, A. D., Streblyanska, A., Lyke, B. W., Anderson, S. F., Armengaud, É., Bautista, J., Blanton, M. R., Blomqvist, M., Brinkmann, J., Brownstein, J. R., Brandt, W. N., Burtin, É., Dawson, K., de la Torre, S., Georgakakis, A., Gil-Marín, H., Green, P. J., Hall, P. B., Kneib, J.-P., LaMassa, S. M., Le Goff, J.-M., MacLeod, C., Mariappan, V., McGreer, I. D., Merloni, A., Noterdaeme, P., Palanque-Delabrouille, N., Percival, W. J., Ross, A. J., Rossi, G., Schneider, D. P., Seo, H.-J., Tojeiro, R., Weaver, B. A., Weijmans, A.-M., Yèche, C., Zarrouk, P., and Zhao, G.-B. (2018). The Sloan Digital Sky Survey Quasar Catalog: Fourteenth data release. , 613:A51.

Peacock, J. A. and Bilicki, M. (2018). Wide-area tomography of CMB lensing and the growth of cosmological density fluctuations. *Mon. Not. Roy. Astron. Soc.*, 481(1):1133–1148.

Peacock, J. A. and Nicholson, D. (1991). The large-scale clustering of radio galaxies. , 253:307–319.

Peebles, P. J. E. (1973). Statistical Analysis of Catalogs of Extragalactic Objects. I. Theory. , 185:413–440.

Peebles, P. J. E. (1980). *The large-scale structure of the universe.*

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419–421.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and Project, T. S. C. (1999). Measurements of and from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565–586.

Pezzotta, A., de la Torre, S., Bel, J., Granett, B. R., Guzzo, L., Peacock, J. A., Garilli, B., Scodeggio, M., Bolzonella, M., Abbas, U., Adami, C., Bottini, D., Cappi, A., Cucciati, O., Davidzon, I., Franetti, P., Fritz, A., Iovino, A., Krywult, J., Le Brun, V., Le Fèvre, O., Maccagni, D., Małek, K., Marulli, F., Polletta, M., Pollo, A., Tasca, L. A. M., Tojeiro, R., Vergani, D., Zanichelli, A., Arnouts, S., Branchini, E., Coupon, J., De Lucia, G., Koda, J., Ilbert, O., Mohammad, F. G., Moutard, T., and Moscardini, L. (2016). The VIMOS Public Extragalactic Redshift Survey (VIPERS): The growth of structures at 0.51.2 from redshift-space distortions in the clustering of the PDR-2 final sample. *ArXiv e-prints.*

Pillepich, A., Porciani, C., and Reiprich, T. H. (2012). The X-ray cluster survey with eROSITA: forecasts for cosmology, cluster physics, and primordial non-Gaussianity. *Mon. Not. Roy. Astron. Soc.*, 422:44–69.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., and et al. (2016). Planck 2015 results. XIII. Cosmological parameters. , 594:A13.

Prandoni, I. and Seymour, N. (2015). Revealing the Physics and Evolution of Galaxies and Galaxy Clusters with SKA Continuum Surveys. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 67.

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. , 187:425–438.

Qin, F., Howlett, C., and Staveley-Smith, L. (2019). The redshift-space momentum power spectrum – II. Measuring the growth rate from the combined 2MTF and 6dFGSv surveys. *Mon. Not. Roy. Astron. Soc.*, 487(4):5235–5247.

Rana, S. and Singh Bagla, J. (2018). Angular clustering of point sources at 150 MHz in the TGSS survey. *ArXiv e-prints.*

Reid, B. A. et al. (2010). Cosmological Constraints from the Clustering of the Sloan Digital Sky Survey DR7 Luminous Red Galaxies. *Mon. Not. Roy. Astron. Soc.*, 404:60–85.

Reid, B. A. et al. (2012). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measurements of the growth of structure and expansion rate at z=0.57 from anisotropic clustering. *Mon. Not. Roy. Astron. Soc.*, 426:2719.

Reiprich, T. (2016). Follow-up of eROSITA and Euclid Clusters. In *XMM-Newton: The Next Decade*, page 7.

Reiprich, T. H. (2017). Follow-up of eROSITA and Euclid Galaxy Clusters with XMM-Newton. *ArXiv e-prints*.

Rengelink, R. B., Tang, Y., de Bruyn, A. G., Miley, G. K., Bremer, M. N., Roettgering, H. J. A., and Bremer, M. A. R. (1997). The Westerbork Northern Sky Survey (WENSS), I. A 570 square degree Mini-Survey around the North Ecliptic Pole. , 124:259–280.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009–1038.

Roche, N. and Eales, S. A. (1999). The angular correlation function and hierarchical moments of ˜70000 faint galaxies to R=23.5. , 307:703–721.

Ross, A. J., Percival, W. J., and Manera, M. (2015). The information content of anisotropic Baryon Acoustic Oscillation scale measurements. , 451:1331–1340.

Rozo, E., Wechsler, R. H., Koester, B. P., McKay, T. A., Evrard, A. E., Johnston, D., Sheldon, E. S., Annis, J., and Frieman, J. A. (2007). Cosmological Constraints from SDSS maxBCG Cluster Abundances. *Submitted to: Astrophys. J.*

Rubart, M. and Schwarz, D. J. (2013). Cosmic radio dipole from NVSS and WENSS. , 555:A117.

Salazar-Albornoz, S., Sánchez, A. G., Padilla, N. D., and Baugh, C. M. (2014). Clustering tomography: measuring cosmological distances through angular clustering in thin redshift shells. , 443:3612–3623.

Salvato, M. and Merloni, A. (2015). The eROSITA All-Sky Survey and its spectroscopic follow-up. In *Demographics and Environment of AGN from Multi-Wavelength Surveys*, page 92.

Samushia, L. et al. (2014). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring growth rate and geometry with anisotropic clustering. *Mon. Not. Roy. Astron. Soc.*, 439(4):3504–3519.

Sanchez, A. G. et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: combining correlated Gaussian posterior distributions. *Mon. Not. Roy. Astron. Soc.*, 464(2):1493–1501.

Sargent, W. L. W. and Turner, E. L. (1977). A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. , 212:L3–L7.

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., Spyromilio, J., Riess, A. G., Filippenko, A. V., Hamuy, M., Smith, R. C., Hogan, C., Stubbs, C., Diercks, A., Reiss, D., Gilliland, R., Tonry, J., Maza, J., Dressler, A., Walsh, J., and Ciardullo, R. (1998). The High-Z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type IA Supernovae. , 507:46–63.

Schneider, P. (2006). *Extragalactic Astronomy and Cosmology: An Introduction.* Springer, Berlin.

Schwarz, D. J., Bacon, D., Chen, S., Clarkson, C., Huterer, D., Kunz, M., Maartens, R., Raccanelli, A., Rubart, M., and Starck, J. L. (2015). Testing foundations of modern cosmology with SKA all-sky surveys. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 32.

Seljak, U. and Warren, M. S. (2004). Large-scale bias and stochasticity of haloes and dark matter. , 355(1):129–136.

Seljak, U. and Zaldarriaga, M. (1996). A Line of sight integration approach to cosmic microwave background anisotropies. *Astrophys. J.*, 469:437–444.

Seo, H.-J., Sato, M., Takada, M., and Dodelson, S. (2012). Dark Energy from the log-transformed convergence field. *Astrophys. J.*, 748:57.

Sheth, R. K., Mo, H. J., and Tormen, G. (2001). Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. , 323:1–12.

Sheth, R. K. and Rossi, G. (2010). Convolution and deconvolution based estimates of galaxy scaling relations from photometric redshift surveys. *Mon. Not. Roy. Astron. Soc.*, 403:2137.

Sheth, R. K. and Tormen, G. (1999). Large-scale bias and the peak background split. , 308(1):119–126.

Shimwell, T. W., Röttgering, H. J. A., Best, P. N., Williams, W. L., Dijkema, T. J., de Gasperin, F., Hardcastle, M. J., Heald, G. H., Hoang, D. N., Horneffer, A., Intema, H., Mahony, E. K., Mandal, S., Mechev, A. P., Morabito, L., Oonk, J. B. R., Rafferty, D., Retana-Montenegro, E., Sabater, J., Tasse, C., van Weeren, R. J., Brüggen, M., Brunetti, G., Chyży, K. T., Conway, J. E., Haverkorn, M., Jackson, N., Jarvis, M. J., McKean, J. P., Miley, G. K., Morganti, R., White, G. J., Wise, M. W., van Bemmel, I. M., Beck, R., Brienza, M., Bonafede, A., Calistro Rivera, G., Cassano, R., Clarke, A. O., Cseh, D., Deller, A., Drabent, A., van Driel, W., Engels, D., Falcke, H., Ferrari, C., Fröhlich, S., Garrett, M. A., Harwood, J. J., Heesen, V., Hoeft, M., Horellou, C., Israel, F. P., Kapińska, A. D., Kunert-Bajraszewska, M., McKay, D. J., Mohan, N. R., Orrú, E., Pizzo, R. F., Prandoni, I., Schwarz, D. J., Shulevski, A., Sipior, M., Smith, D. J. B., Sridhar, S. S., Steinmetz, M., Stroe, A., Varenius, E., van der Werf, P. P., Zensus, J. A., and Zwart, J. T. L. (2017). The LOFAR Two-metre Sky Survey. I. Survey description and preliminary data release. , 598:A104.

Shimwell, T. W., Tasse, C., Hardcastle, M. J., Mechev, A. P., Williams, W. L., Best, P. N., Röttgering, H. J. A., Callingham, J. R., Dijkema, T. J., de Gasperin, F., Hoang, D. N., Hugo, B., Mirmont, M., Oonk, J. B. R., Prandoni, I., Rafferty, D., Sabater, J., Smirnov, O., van Weeren, R. J., White, G. J., Atemkeng, M., Bester, L., Bonnassieux, E., Brüggen, M., Brunetti, G., Chyży, K. T., Cochrane, R., Conway, J. E., Croston, J. H., Danezi, A., Duncan, K., Haverkorn, M., Heald, G. H., Iacobelli, M., Intema, H. T., Jackson, N., Jamrozy, M., Jarvis, M. J., Lakhoo, R., Mevius, M., Miley, G. K., Morabito, L., Morganti, R., Nisbet, D., Orrú, E., Perkins, S., Pizzo, R. F., Schrijvers, C., Smith, D. J. B., Vermeulen, R., Wise, M. W., Alegre, L., Bacon, D. J., van Bemmel, I. M., Beswick, R. J., Bonafede, A., Botteon, A., Bourke, S., Brienza, M., Calistro Rivera, G., Cassano, R., Clarke, A. O., Conselice, C. J., Dettmar, R. J., Drabent, A., Dumba, C., Emig, K. L., Enßlin, T. A., Ferrari, C., Garrett, M. A., Génova-Santos, R. T., Goyal, A., Gürkan, G., Hale, C., Harwood, J. J., Heesen, V., Hoeft, M., Horellou, C., Jackson, C., Kokotanekov, G., Kondapally, R., Kunert- Bajraszewska, M., Mahatma, V., Mahony, E. K., Mandal, S., McKean, J. P., Merloni, A., Mingo, B., Miskolczi, A., Mooney, S., Nikiel- Wroczyński, B., O'Sullivan, S. P., Quinn, J., Reich, W., Roskowiński, C., Rowlinson, A., Savini, F., Saxena, A., Schwarz, D. J., Shulevski, A., Sridhar, S. S., Stacey, H. R., Urquhart, S., van der Wiel, M. H. D., Varenius, E., Webster, B., and Wilber, A. (2018). The LOFAR Two-metre Sky Survey - II. First data release. *arXiv e-prints*.

Shirasaki, M., Takada, M., Miyatake, H., Takahashi, R., Hamana, T., Nishimichi, T., and Murata, R. (2017). Robust covariance estimation of galaxy–galaxy weak lensing: validation and limitation of jackknife covariance. *Mon. Not. Roy. Astron. Soc.*, 470(3):3476–3496.

Singal, A. K. (2011). Large Peculiar Motion of the Solar System from the Dipole Anisotropy in Sky Brightness due to Distant Radio Sources. , 742:L23.

Smith, R. E. (2009). Covariance of cross-correlations: towards efficient measures for large-scale structure. , 400:851–865.

Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., and Couchman, H. M. P. (2003). Stable clustering, the halo model and non-linear cosmological power spectra. , 341:1311–1332.

Smoot, G. F. et al. (1992). Structure in the COBE differential microwave radiometer first year maps. *Astrophys. J.*, 396:L1–L5.

Spergel, D. N. et al. (2003). First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *Astrophys. J. Suppl.*, 148:175–194.

Swarup, G. (1991). Giant metrewave radio telescope (GMRT). In Cornwell, T. J. and Perley, R. A., editors, *IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications*, volume 19 of *Astronomical Society of the Pacific Conference Series*, pages 376–380.

Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., and Oguri, M. (2012). Revising the Halofit Model for the Nonlinear Matter Power Spectrum. , 761:152.

Tegmark, M. et al. (2004). The 3-D power spectrum of galaxies from the SDSS. *Astrophys. J.*, 606:702–740.

Tegmark, M. et al. (2006). Cosmological Constraints from the SDSS Luminous Red Galaxies. *Phys. Rev.*, D74:123507.

Thomas, S. A., Abdalla, F. B., and Lahav, O. (2011). The angular power spectra of photometric Sloan Digital Sky Survey luminous red galaxies. *MNRAS*, 412:1669–1685.

Tingay, S. J., Goeke, R., Bowman, J. D., Emrich, D., Ord, S. M., Mitchell, D. A., Morales, M. F., Booler, T., Crosse, B., Wayth, R. B., Lonsdale, C. J., Tremblay, S., Pallot, D., Colegate, T., Wicenec, A., Kudryavtseva, N., Arcus, W., Barnes, D., Bernardi, G., Briggs, F., Burns, S., Bunton, J. D., Cappallo, R. J., Corey, B. E., Deshpande, A., Desouza, L., Gaensler, B. M., Greenhill, L. J., Hall, P. J., Hazelton, B. J., Herne, D., Hewitt, J. N., Johnston-Hollitt, M., Kaplan, D. L., Kasper, J. C., Kincaid, B. B., Koenig, R., Kratzenberg, E., Lynch, M. J., Mckinley, B., Mcwhirter, S. R., Morgan, E., Oberoi, D., Pathikulangara, J., Prabu, T., Remillard, R. A., Rogers, A. E. E., Roshi, A., Salah, J. E., Sault, R. J., Udaya-Shankar, N., Schlagenhaufer, F., Srivani, K. S., Stevens, J., Subrahmanyan, R., Waterson, M., Webster, R. L., Whitney, A. R., Williams, A., Williams, C. L., and Wyithe, J. S. B. (2013). The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. , 30:e007.

Tiwari, P. (2016). Radio spectral index from NVSS and TGSS. *ArXiv e-prints*.

Tiwari, P. and Aluri, P. K. (2018). Large angular scale multipoles at redshift ˜0.8. *arXiv e-prints*.

Tiwari, P. and Jain, P. (2015). Dipole anisotropy in integrated linearly polarized flux density in NVSS data. , 447:2658–2670.

Tiwari, P., Kothari, R., Naskar, A., Nadkarni-Ghosh, S., and Jain, P. (2015). Dipole anisotropy in sky brightness and source count distribution in radio NVSS data. *Astroparticle Physics*, 61:1–11.

Tiwari, P. and Nusser, A. (2016). Revisiting the NVSS number count dipole. , 3:062.

Tocchini-Valentini, D., Barnard, M., Bennett, C. L., and Szalay, A. S. (2012). A Method to Extract the Redshift Distortion β Parameter in Configuration Space from Minimal Cosmological Assumptions. , 757:131.

van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., de Bruyn, A. G., Nijboer, R., Swinbank, J., Fallows, R., Brentjens, M., Nelles, A., Beck, R., Falcke, H., Fender, R., Hörandel, J., Koopmans, L. V. E., Mann, G., Miley, G., Röttgering, H., Stappers, B. W., Wijers, R. A. M. J., Zaroubi, S., van den Akker, M., Alexov, A., Anderson, J., Anderson, K., van Ardenne, A., Arts, M., Asgekar, A., Avruch, I. M., Batejat, F., Bähren, L., Bell, M. E., Bell, M. R., van Bemmel, I., Bennema, P., Bentum, M. J., Bernardi, G., Best, P., Bîrzan, L., Bonafede, A., Boonstra, A.-J., Braun, R., Bregman, J., Breitling, F., van de Brink, R. H., Broderick, J., Broekema, P. C., Brouw, W. N., Brüggen, M., Butcher, H. R., van Cappellen, W., Ciardi, B., Coenen, T., Conway, J., Coolen, A., Corstanje, A., Damstra, S., Davies, O., Deller, A. T., Dettmar, R.-J., van Diepen, G., Dijkstra, K., Donker, P., Doorduin, A., Dromer, J., Drost, M., van Duin, A., Eislöffel, J., van Enst, J., Ferrari, C., Frieswijk, W., Gankema, H., Garrett, M. A., de Gasperin, F., Gerbers, M., de Geus, E., Grießmeier, J.-M., Grit, T., Gruppen, P., Hamaker, J. P., Hassall, T., Hoeft, M., Holties, H. A., Horneffer, A., van der Horst, A., van Houwelingen, A., Huijgen, A., Iacobelli, M., Intema, H., Jackson, N., Jelic, V., de Jong, A., Juette, E., Kant, D., Karastergiou, A., Koers, A., Kollen, H., Kondratiev, V. I., Kooistra, E., Koopman, Y., Koster, A., Kuniyoshi, M., Kramer, M., Kuper, G., Lambropoulos, P., Law, C., van Leeuwen, J., Lemaitre, J., Loose, M., Maat, P., Macario, G., Markoff, S., Masters, J., McFadden, R. A., McKay-Bukowski, D., Meijering, H., Meulman, H., Mevius, M., Middelberg, E., Millenaar, R., Miller-Jones, J. C. A., Mohan, R. N., Mol, J. D., Morawietz, J., Morganti, R., Mulcahy, D. D., Mulder, E., Munk, H., Nieuwenhuis, L., van Nieuwpoort, R., Noordam, J. E., Norden, M., Noutsos, A., Offringa, A. R., Olofsson, H., Omar, A., Orrú, E., Overeem, R., Paas, H., Pandey-Pommier, M., Pandey, V. N., Pizzo, R., Polatidis, A., Rafferty, D., Rawlings, S., Reich, W., de Reijer, J.-P., Reitsma, J., Renting, G. A., Riemers, P., Rol, E., Romein, J. W., Roosjen, J., Ruiter, M., Scaife, A., van der Schaaf, K., Scheers, B., Schellart, P., Schoenmakers, A., Schoonderbeek, G., Serylak, M., Shulevski, A., Sluman, J., Smirnov, O., Sobey, C., Spreeuw, H., Steinmetz, M., Sterks, C. G. M., Stiepel, H.-J., Stuurwold, K., Tagger, M., Tang, Y., Tasse, C., Thomas, I., Thoudam, S., Toribio, M. C., van der Tol, B., Usov, O., van Veelen, M., van der Veen, A.-J., ter Veen, S., Verbiest, J. P. W., Vermeulen, R., Vermaas, N., Vocks, C., Vogt, C., de Vos, M., van der Wal, E., van Weeren, R., Weggemans, H., Weltevrede, P., White, S., Wijnholds, S. J., Wilhelmsson, T., Wucknitz, O., Yatawatta, S., Zarka, P., Zensus, A., and van Zwieten, J. (2013). LOFAR: The LOw-Frequency ARray. , 556:A2.

Veropalumbo, A., Marulli, F., Moscardini, L., Moresco, M., and Cimatti, A. (2014). An improved measurement of baryon acoustic oscillations from the correlation function of galaxy clusters at z
0.3. , 442:3275–3283.

Waddington, I., Dunlop, J. S., Peacock, J. A., and Windhorst, R. A. (2001). The LBDS Hercules sample of mJy radio sources at 1.4 GHz - II. Redshift distribution, radio luminosity function, and the high-redshift cut-off. , 328:882–896.

White, R. L., Becker, R. H., Helfand, D. J., and Gregg, M. D. (1997). A Catalog of 1.4 GHz Radio Sources from the FIRST Survey. *Astrophys. J.*, 475:479.

White, S. V., Franzen, T. M. O., Wong, O. I., Kapinska, A. D., Riseley, C., Hancock, P., Callingham, J., Hunstead, R., Hurley-Walker, N., Wu, C., Seymour, N., Swan, J., Wayth, R., Morgan, J. S., Chhetri, R., Jackson, C., Weston, S., and Mauch, T. (2018). The MWA GLEAM 4-Jy (G4Jy) Sample. *arXiv e-prints*.

Willott, C. J., Rawlings, S., Blundell, K. M., Lacy, M., and Eales, S. A. (2001). The radio luminosity function from the low-frequency 3CRR, 6CE and 7CRS complete samples. , 322:536–552.

Wilman, R. J., Miller, L., Jarvis, M. J., Mauch, T., Levrier, F., Abdalla, F. B., Rawlings, S., Klöckner, H.-R., Obreschkow, D., Olteanu, D., and Young, S. (2008). A semi-empirical simulation of the extragalactic radio continuum sky for next generation radio telescopes. , 388:1335–1348.

Wright, A. E., Griffith, M. R., Burke, B. F., and Ekers, R. D. (1994). The Parkes-MIT-NRAO (PMN) surveys. 2: Source catalog for the southern survey (delta greater than -87.5 deg and less than -37 deg). , 91:111–308.

Xavier, H. S., Abdalla, F. B., and Joachimi, B. (2016). Improving lognormal models for cosmological fields. , 459:3693–3710.

York, D. G. et al. (2000). The Sloan Digital Sky Survey: Technical Summary. *Astron. J.*, 120:1579–1587.

Zaldarriaga, M. and Seljak, U. (1998). Gravitational lensing effect on cosmic microwave background polarization. *Phys. Rev.*, D58:023003.

Zehavi, I. et al. (2005). The Luminosity and color dependence of the galaxy correlation function. *Astrophys. J.*, 630:1–27.

Zentner, A. R. (2007). The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth. *Int. J. Mod. Phys.*, D16:763–816.