UNIVERSITY OF CALIFORNIA
RIVERSIDE

Measurement of the Galaxy Angular Power Spectrum Using Simulated Data From the Legacy Survey of Space and Time

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Zilong Du

September 2023

Dissertation Committee:

Dr. John Ellison, Chairperson
Dr. Simeon Bird
Dr. Laura Sales

The Dissertation of Zilong Du is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First and foremost, I am grateful to my advisor, John Ellison, without whose support during these six years, I would not have been here. I want to thank him for providing help, guidance and the freedom to explore opportunities and interests. Working with him has allowed me to develop the qualities to become a better researcher and student.

Secondly, I'd like to acknowledge members from the LSSTDESC collaboration. In particular, I would like to thank Javier Sánchez, who helped me tremendously throughout the project and is always patient to answer my questions. I want to thank the LSS group conveners, David Alonso, Andrina Nicola and Boris Leistedt, for organizing monthly teleconferences, providing opportunities to present my work and to receive constructive feedback. Thanks to Elisa Chisari for leading and organizing the development of CCL, which I'm fortunate to be part of. I also want to thank all other members who I had the chance to interact with.

In addition, I'd like to thank the faculty and staff members at the UCR Physics and Astronomy department. In particular, I would like to acknowledge Simeon Bird and Laura Sales for providing advice and feedback as members of my qualifying and final defense committees. I want to thank Derek Beving for helping with all aspects of the graduate program.

Lastly, I would be remiss in not mentioning my family, friends and roommates. They have kept my spirits and motivation up at difficult times. I would also like to thank the cats for emotional support.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Measurement of the Galaxy Angular Power Spectrum Using Simulated Data From the Legacy Survey of Space and Time

by

Zilong Du

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, September 2023
Dr. John Ellison, Chairperson

The Legacy Survey of Space and Time (LSST) carried out by the Vera C. Rubin Observatory is one of the most ambitious optical surveys planned in the near future. It will generate $\sim$ 20 TB of data every night during its 10 year survey duration. Simulated data sets ("Data Challenges") have been generated by the LSST Dark Energy Science Collaboration (DESC) to allow development of efficient analysis algorithms, evaluation of their performance and scaling capabilities, and to study the impact of systematic uncertainties. In this work we focus on the analysis of Large Scale Structure using galaxies selected from the DESC Data Challenge 2 (DC2) corresponding to a survey area of 300 square degrees and five years of observing. We develop and perform an end-to-end analysis of the galaxy Large Scale Structure with the DC2 data, validating the extraction of galaxy power spectra and galaxy bias in the presence of complex systematics. We provide a range of strong consistency checks of both the simulated data and the analysis techniques.

Based on matching selected galaxies with the DC2 truth catalog, we optimize selection to obtain a clean sample of galaxies. After generating and applying survey masks,

we produce galaxy density maps binned in redshift, with six bins from $z = 0.2 - 1.4$. We generate the power spectra measurements from the processed data and fit the data using theoretical predictions to estimate the galaxy bias as a function of redshift. Finally, we study the effect of systematic uncertainties on estimation of the galaxy bias.

We find that the relative uncertainty σ_ℓ/C_ℓ in the measured angular power spectra is smaller than 10% in most cases, and smaller than 4% at multipoles above 1500. Changes in the angular power spectra due to systematic uncertainties are found to be on average of order 0.5%, i.e. $\Delta C_\ell/C_\ell \approx 0.5\%$, with a maximum deviation of 1.5%. Systematic effects are small compared to the statistical uncertainty of the measured angular power spectra: $\Delta C_\ell \sim 0.1\sigma_\ell$ on average with a maximum of $0.25\sigma_\ell$.

Contents

List of Figures

List of Tables

Chapter 1

The standard model of cosmology

Based on the Big bang model [9], the Cosmological Principle and the General Theory of Relativity [10], the ΛCDM model or the Standard Model of Cosmology is one of the most widely accepted cosmological models. The model can be further extended and complemented with more speculative theories like Cosmic Inflation [11, 12] and Quintessence [13]. In this chapter, we briefly describe the ΛCDM model.

1.1 The Cosmological Principle

The standard model of modern cosmology is mainly founded based on the Cosmology Principle: on sufficiently large scales (typically 1Gly or greater), the universe is homogeneous and isotropic. Homogeneity implies that all positions are equivalent, and isotropic means that all directions are equivalent. This simple and elegant presumption allows us to model the evolution of the universe using a few cosmological parameters, which can be measured through observations.

To understand the cosmological principle mathematically, we can discuss its implication on the overdensity field:

$$\delta(\vec{x}) \equiv \frac{\rho(\vec{x}) - \langle\rho\rangle}{\langle\rho\rangle} \tag{1.1}$$

where $\rho(\vec{x})$ is a density field and $\langle\rho\rangle$ is its average. Some of the most common usages of the density field are to describe CMB background temperature, matter density distribution or galaxy number density.

We can also define the density 2-point correlation function:

$$\xi(\vec{x}_1, \vec{x}_2) \equiv \langle\delta(\vec{x}_1)\delta(\vec{x}_2)\rangle \tag{1.2}$$

This function reflects how densities at different locations are correlated to each other on average: if the function value is negative, then one point has overdensity ($\delta > 1$) and the other point has underdensity ($\delta < 1$); if the function value is positive, then both points have overdensities or underdensities.

Consider the 2-point correlation function in the context of the Cosmological principle: homogeneity implies that $\xi(\vec{x}_1, \vec{x}_2)$ can only depend on the separation vector $\vec{r} \equiv \vec{x}_2 - \vec{x}_1$ (exact locations $\vec{x}_1$, $\vec{x}_2$ don't matter). Hence, we can rewrite $\xi(\vec{x}_1, \vec{x}_2)$ as $\xi(\vec{r}) = \xi(\vec{x}_2 - \vec{x}_1) = \langle\delta(\vec{x})\delta(\vec{x} + \vec{r})\rangle$. Due to isotropy, $\xi(\vec{r})$ can only depend on the magnitude of the separation vector (all directions are equivalent). We then have $\xi(\vec{r}) = \xi(r)$. Usually, $\xi(r)$ is a decreasing function of r in cosmology, as the correlation is strong within typical dense regions and becomes weaker at larger scales.

1.2 The FRW Metric

In the 1920s and 1930s, the physicists Alexander Friedmann, Howard Robertson, and Arthur Walker independently developed the so-called FRW metric [14]. Their work was driven by the desire to construct a mathematical model based on General Relativity that could account for the Cosmological Principle (isotropy and homogeneity) and the observed expansion of the universe. The FRW metric can be written in the form [15]:

$$ds^2 = -c^2dt^2 + a(t)^2[dr^2 + S_\kappa(r)^2 d\Omega^2] \tag{1.3}$$

where

$$d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2 \tag{1.4}$$

represents the differential angular separation. Due to the Cosmological Principle, the metric is required to be spatially isotropic and homogeneous. The geometry can be specified by two quantities, κ - the *curvature constant* and R - the *radius of curvature* (if $\kappa \neq 0$). For a given radius of curvature R, we have

$$S_\kappa(r) = \begin{cases} R\sin(r/R) & (\kappa = 1) \\ r & (\kappa = 0) \\ R\sinh(r/R) & (\kappa = -1) \end{cases}$$

where $\kappa = 0, +1, -1$ correspond to flat, positively curved and negatively curved spaces.

Based on this metric, the distance between any two points is directly proportional to $a(t)$ - the scale factor, which describes the expansion of the universe. A normalization factor can be applied to $a(t)$ since different units can be used to describe distances. We

adapt the commonly used normalization convention and set it's value to 1 at the current time, i.e. $a(t_0) = 1$.

1.3 Dynamics and Cosmological Parameters

Once the scale factor $a(t)$ and curvature constant κ are determined, the evolution of the universe is described by the FRW metric. According to general relativity, the distribution of matter and energy in the universe determines the curvature of spacetime, and the dynamics of this curvature govern the behavior of cosmic expansion. The equation that describes the relation between κ, $a(t)$, R and the energy density $\rho(t)$ is the *Friedmann equation*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\rho(t) - \frac{\kappa c^2}{R_0 a(t)^2} + \frac{\Lambda}{3} \tag{1.5}$$

Here Λ is the *cosmological constant*, one possible form of dark energy. The Friedmann equation is a statement of energy conservation. Under the context of Newtonian approximation, it states that the sum of kinetic energy of the expansion and the gravitational potential energy is a constant. We can define the *Hubble parameter* $H(t) \equiv \dot{a}/a$ and its present value $H_0 = H(t_0) = 100h$ km s^{-1}Mpc^{-1} is known as the *Hubble constant* [16], where $h \approx 0.7$ is the dimensionless Hubble constant.

Consider the first law of thermodynamics under the context of general relativity; we can also derive the *fluid equation*:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p) \tag{1.6}$$

Using Eq.(1.5) and Eq.(1.6), we can write:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3p) \tag{1.7}$$

where we have included the dark energy density $\rho_\Lambda \equiv \frac{\Lambda c^2}{8\pi G}$ in the total energy density ρ. This is the so-called *acceleration equation*, which tells us how the expansion of the universe speeds up or slows down over time based on the composition of energy contents. Furthermore, we treat different energy contents as independent non-interacting components. For most cosmological substances, we can model them as ideal fluids and define the equation of state:

$$p \equiv w\rho \tag{1.8}$$

where w is a dimensionless number that depends on the physical properties of the substance. Combined with the fluid equation we get:

$$\rho_i(t) \propto a(t)^{-3(1+w_i)} \tag{1.9}$$

Three types of components are usually considered:

1. Non-relativistic matter: from the ideal gas law, we have $p = \rho kT/(\mu c^2) = v_{rms}^2/(3c^2)\rho$, where μ is the mass of the gas particle and v_{rms} is the root-mean-square speed of the gas particles. Since the gas is non-relativistic, $v_{rms} \ll c$ and $w \approx 0$. Therefore, $\rho \propto a^{-3}$.

2. Radiation or a gas of photons: for ultra-relativistic bosons we have $p = 1/3\rho$ or $w = 1/3$ and $\rho \propto a^{-4}$.

3. Dark energy or the cosmological constant Λ: since the energy density is a constant, we have $\dot{\rho} = 0$ and $p = -\rho$ or $w = -1$. Therefore, $\rho \propto a^0$.

If we define the dark energy density $\rho_\Lambda \equiv \Lambda/(8\pi G)$ along with matter density ρ_m and radiation density ρ_r, Eq.(1.5) can be written as

$$H^2 = \frac{8\pi G}{3c^2}\sum_i \rho_i - \frac{\kappa c^2}{R^2} \tag{1.10}$$

It's useful to also define the critical density ρ_c, whose value is found by setting the curvature constant κ to zero:

$$\rho_c \equiv \frac{3c^2}{8\pi G}H^2 \tag{1.11}$$

Then we can define the *density parameters* for each component as the ratio of its density to the critical density: $\Omega_i \equiv \rho_i/\rho_c$. Moreover, we define the *total density parameter* $\Omega \equiv \sum_i \rho_i/\rho$. If $\Omega = 1$, the universe is spatially flat ($\kappa = 0$). If $\Omega > 1$, the universe is positively curved ($\kappa > 0$). If $\Omega < 1$, the universe is negatively curved ($\kappa < 0$).

Now we can rewrite the Friedmann equation as:

$$H^2 = H_0^2(\Omega_{m,0}a^{-3} + \Omega_{r,0}a^{-4} + \Omega_{\Lambda,0} + \Omega_{\kappa,0}a^{-2}) \tag{1.12}$$

where we have defined $\rho_\kappa \equiv \frac{3c^2}{8\pi G}\frac{-kc}{R^2}$ and $\Omega_\kappa = \frac{\rho_\kappa}{\rho_c}$ accordingly. The subscripts 0 indicate the present values of each cosmological parameter. If we connect the equations of state of each component to Eq.(1.3) and use $H = \frac{\dot{a}}{a}$, we get:

$$q(t) \equiv -\frac{\ddot{a}a}{\dot{a}^2} = \frac{1}{2}\sum_i \Omega_i(t)(1 + 3w_i) \tag{1.13}$$

where Ω_κ is excluded in the sum. This is the so-called *deceleration parameter*. If $q < 0$, then $\ddot{a} > 0$ and the universe will accelerate outward. Among matter, radiation and dark enregy, only dark energy satisfies $w_i < -1/3$ and contributes to a negative q value.

1.4 Distance and Redshift

In cosmology, distance measurements are usually related to redshifts. Consider the light waves emitted by a distant galaxy. During its journey to us its wavelength increases due to the cosmic expansion and scales proportional to $a(t)$:

$$\frac{\lambda_e}{a(t_e)} = \frac{\lambda_0}{a(t_0)} \tag{1.14}$$

Here subscripts e and 0 represent the value of the physical quantity when the light wave is emitted and received, respectively. Due to the expansion of the universe, $a(t_0) > a(t_e)$ and $\lambda_0 > \lambda_e$, i.e. the wavelength becomes longer when received by an observer and the galaxy appears "redder". The *redshift* of the galaxy is defined as:

$$z \equiv \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{\nu_e - \nu_0}{\nu_0} \tag{1.15}$$

combined with Eq.(1.14), we get

$$1 + z = \frac{a(t_0)}{a(t_e)} = \frac{1}{a(t_e)} \tag{1.16}$$

or

$$a(t_e) = a = \frac{1}{1+z} \tag{1.17}$$

The observed redshift only depends on the scale factor when the light was emitted.

Most distance measurements in cosmology are made through indirect inference instead of direct measurements. In a curved expanding universe, the concept of distance becomes complicated. One can get different distance definitions based on the physical processes used to estimate them. Since distance measurements are fundamental to cosmological observations and constraining the cosmological model, it's important to distinguish different

distance measurements and discuss their relation with the cosmological models. Some of the commonly used definitions are [15, 17]:

1. **Comoving distance.** Consider a photon emitted from a distant galaxy. It travels along the null geodesic: $ds^2 = 0$. Since the photon travels in the radial direction, $d\Omega = 0$. Combine these conditions with the FRW metric, we get $0 = -c^2dt^2 + a(t)^2dr^2$. Simplifying and integrating:

$$r = \int_0^r dr' = c\int_{t_e}^{t_0} \frac{dt}{a(t)} \tag{1.18}$$

Using $dt = da/\dot{a}$, $H = \dot{a}/a$ and $dz = -da/a^2$, we can rewrite it in terms of redshift:

$$\chi(z) \equiv r = c\int_0^z \frac{dz'}{H(z')} \tag{1.19}$$

This is the *comoving distance* to the galaxy. It is the distance measured using the comoving coordinates (r) and doesn't change over time with the scale factor and the cosmic expansion.

2. **Proper distance** The *proper distance* between two points is the length of the spacial geodesic when time t is fixed. We set dt to zero and ignore angular dependence in the FRW metric, we get:

$$ds = a(t)dr \tag{1.20}$$

Integrating both sides yields the proper distance:

$$d_p(t) = a(t)r = a(t)\chi \tag{1.21}$$

By fixing time t, it's implied that distance measurements should be made simultaneously along the geodesic. Therefore, it's not practical to set up experiments to

measure the distance based on this definition. Other distances are defined based on different observational methods, and their relations to proper distance can be derived mathematically.

3. **Luminosity distance** In static Euclidean space, light propagates according to the inverse square law, so the flux f measured by an observer is

$$f = \frac{L}{4\pi d^2} \tag{1.22}$$

where L is the luminosity of the light source at distance d away from the observer. The factor $1/4\pi d^2$ means that the energy from the source is distributed isotropically to the spherical surface area centered at the source. Based on this relation, if the luminosity of an object is known, we can compute its *luminosity distance* using the measured flux f:

$$d_L \equiv \sqrt{\frac{L}{4\pi f}} \tag{1.23}$$

However, in a curved expanding universe, several generalizations need to be made to the inverse square law: the proper surface area of a sphere is given by $A_p(t_0) = 4\pi S_\kappa(r)^2$; the energy of each photon is decreased by a factor of $1+z$ due to redshift; the detection rate of photons is also decreased by a factor of $1+z$ due to an increase in their separation from cosmic expansion. The observed flux after considering all effects is:

$$f = \frac{L}{4\pi S_\kappa(r)^2 (1+z)^2} \tag{1.24}$$

Comparing with Eq.(1.23), we have:

$$d_L(z) = r(1+z) = \chi(z)(1+z) \tag{1.25}$$

where we've approximated our universe to be spatially flat so $S_\kappa(r) = r$.

4. **Angular diameter distance** In Euclidean space, an object with physical transverse size (proper distance) l at distance d away from an observer has an angular size of $\delta\theta \approx l/d$ when $d \gg l$. The *angular diameter distance* is defined according to this relation:

$$d_A \equiv \frac{l}{\delta\theta} \tag{1.26}$$

In general, the physical transverse size can be approximated with $l \approx ds = a(t_e)S_\kappa(r)\delta\theta$ by setting dt and dr to zeros in the FRW metric. Therefore,

$$d_A = \frac{r}{1+z} = \frac{\chi(z)}{1+z} \tag{1.27}$$

for a spatially flat universe ($S_\kappa(r) = r$).

To summarize, if the universe is spatially flat, we have

$$(1+z)d_A = \chi(z) = \frac{d_L}{1+z} \tag{1.28}$$

1.5 Observational Evidence

The name "ΛCDM model" is associated with the major energy compositions of the universe: "Λ" represents the cosmological constant related to dark energy; "CDM" stands for cold dark matter; ordinary matter and radiation are also included in the model. It's also known as the standard model of cosmology as it is the simplest model that provides reasonably good explanations to observations such as: the existence and anisotropies of the cosmic microwave background, the abundance of light elements, the accelerated expansion of the universe, and the large-scale structure of the distribution of matter.

1.5.1 Cosmic microwave background

One key aspect to the thermal history of the universe is the decoupling of some species from the other components in the cosmic fluid. We can figure out when the decoupling happens by comparing the scattering rate of the species to the cosmic expansion rate. One of the key moments is the time at which *photon decoupling* occurs: i.e. when the interaction rate of photons with electrons becomes less than the Hubble parameter. It happens soon after the epoch of *recombination*: when the temperature of the universe was low enough (~ 0.3 eV, at 380,000 years) to allow electrons to bind with nuclei to form the first neutral atoms. Shortly after decoupling comes the epoch of *last scattering*, after which photons can travel freely with little scattering, forming the Cosmic Microwave Background (CMB) that we observe today.

CMB was first discovered in 1964 by Penzias and Wilson [18]. Since then, observations have shown that the CMB follows a black body spectrum of $T \approx 2.725$ K, and has an almost uniform distribution with fluctuations to the level of $\sim 10^{-4}$ [19]. The fact that the CMB is almost isotropic is one piece of evidence of the Cosmological Principle.

There are two types of anisotropy of the CMB categorized based on the source: primary anisotropy is caused by effects happen at or before the last scattering such as the pressure conflict in the baryon-photon plasma, or Baryon Acoustic Oscillation (BAO); secondary anisotropy is caused by effects such as interactions of the CMB photons with distributions of high energy electrons (Sunyaev-Zel'dovich effect [20]) or gravitational potentials (Integrated Sachs-Wolfe effect [21]), which occur between the last scattering surface

and the observer. The structure of the CMB anisotropy has been measured in high resolutions and accuracy. See Figure 1.1 and 1.2.

1.5.2 Abundance of light elements

The abundance of light elements is another piece of evidence that strongly supports the standard model. The standard model predicts the abundance of light elements produced during the early phases of the universe. The process is refered to as the *Big Bang nucleosynthesis* (BBN) [22]. About 3 minutes after the Big Bang, the temperature of the universe cooled down to about 10^9 K or an energy scale of 10 MeV. This temperature is on par with the typical binding energy per nucleon and therefore is low enough to allow protons and neutrons colliding to produce the first neucli: deuterium, helium and lithium. However, the universe then cooled rapidly due to expansion, and the production of heavier elements couldn't take place. At around 100 keV, neucleosynthesis stopped. The BBN theory predicts that the mass of the universe consisted of $\sim 25\%$ helium and $\sim 75\%$ hydrogen, which is in line with observations. Elements heavier than helium are thought to have been produced inside stars that formed long after.

1.5.3 Accelerated expansion

Since redshift is related to the lookback time, one of the most robust methods to estimate the expansion rate of the universe is via the measurements of redshifts and distances of astronomical objects. Redshifts can be measured rather straightforwardly through spectroscopy. One way of measuring distances is to use the special properties of objects

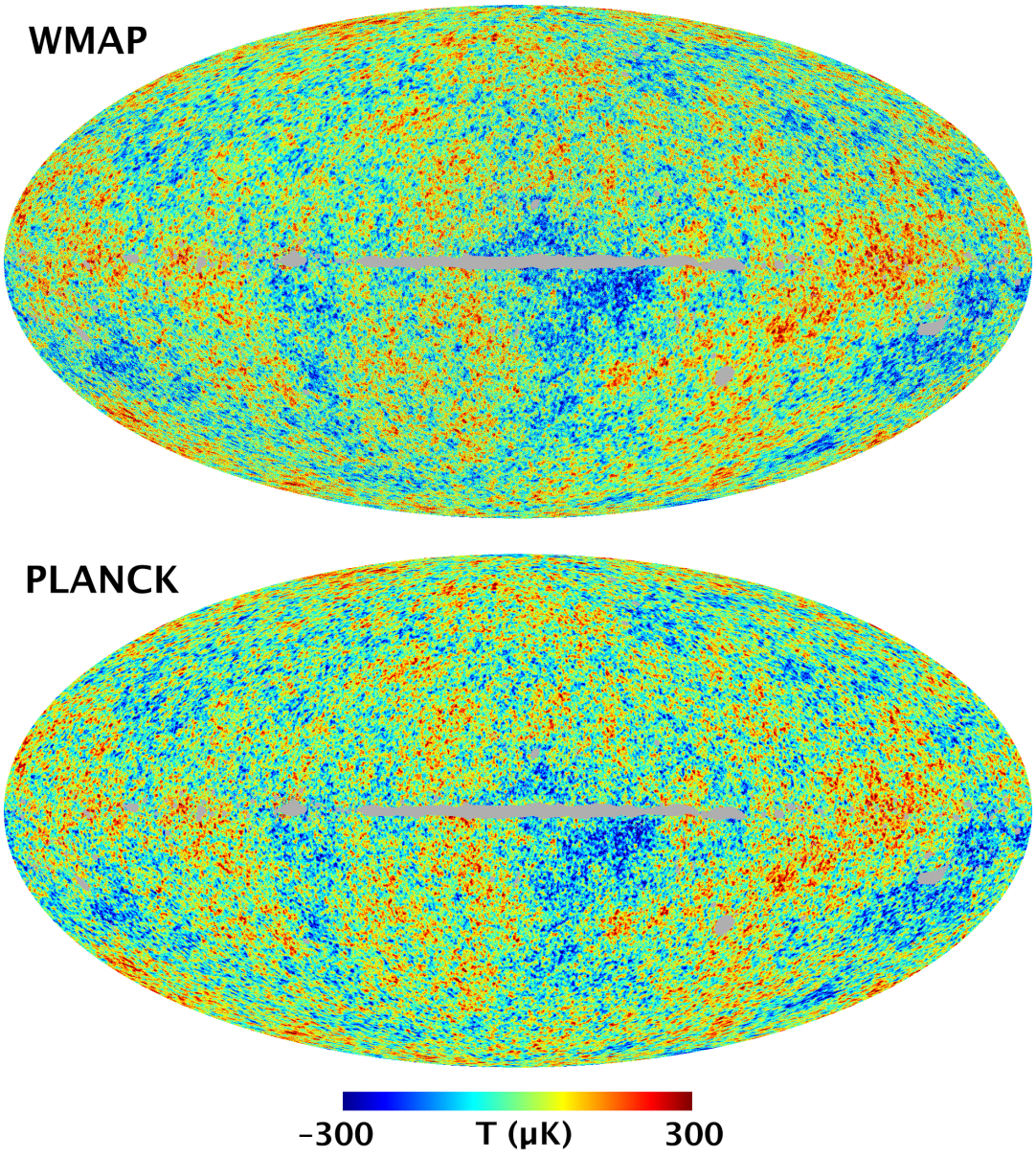


Figure 1.1: CMB anisotropy map in Galactic coordinates, as measured by the WMAP [1] 9 year probe (Top) and Planck with SMICA algorithm [2] (bottom). Both maps are foreground-cleaned, WMAP by subtracting a linear least squares fit to the Planck dust and low-frequency templates. The resolution of Planck result is up to 0.16°. Image Credit: NASA/LAMBDA Science Team.

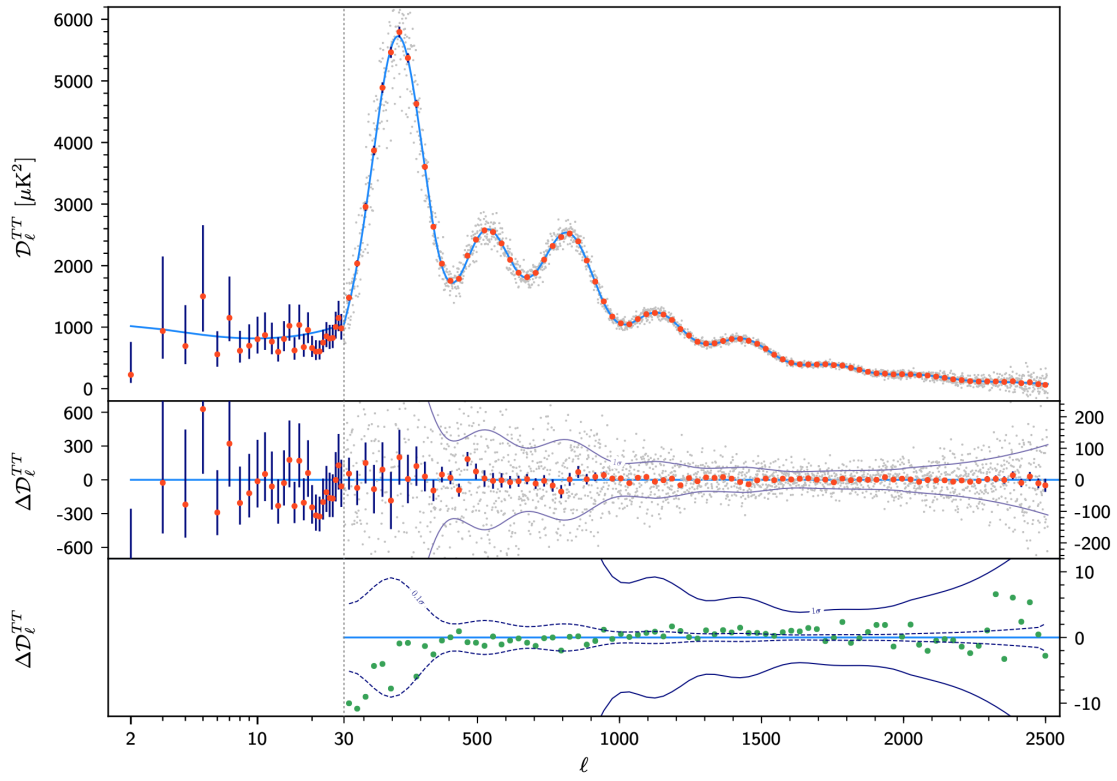


Figure 1.2: Best fit of Planck 2018 temperature power spectrum [3].

with a fixed luminosity, the *standard candle*. Standard candles are useful since they allow us to infer luminosity distances ($d_L = \sqrt{L/(4\pi f)}$) by simply measuring their fluxes f.

For standard candles to be observed over cosmological distances, they also need to be bright enough. The best and mostly used standard candles in cosmology are type Ia supernovae. Supernovae are classified according to their spectra: Type II contain strong hydrogen absorption lines; Type I contain no hydrogen absorption lines while Type Ia an ionized silicon line and Type Ib contain a helium line. Type Ia supernovae are believed to be white dwarf stars that accrete matter from a stellar companion, going over the Chandrasekhar limit, and triggering a runaway nuclear explosion. The luminosity of Type Ia supernovae is well studied. It's known that their peak luminosity is almost constant and are strongly correlated with the shape of their light curves.

The Hubble law [16] states that in the local universe, the recession velocity of an astronomical object due to cosmic expansion is proportional to its proper distance. The observation does not generalize to high-redshifts necessarily. Through the observation of high-redshift type Ia supernovae, it was discovered that the universe currently undergoes an accelerated expansion, which led to the dark energy component being formulated in the ΛCDM model (see Figure 1.3).

1.6 Inflation

We have seen that there is a considerable amount of evidence that supports the standard cosmology model. However, it is not without flaws. There are 3 underlying problems, namely, the flatness problem, the horizon problem and the monopole problem.

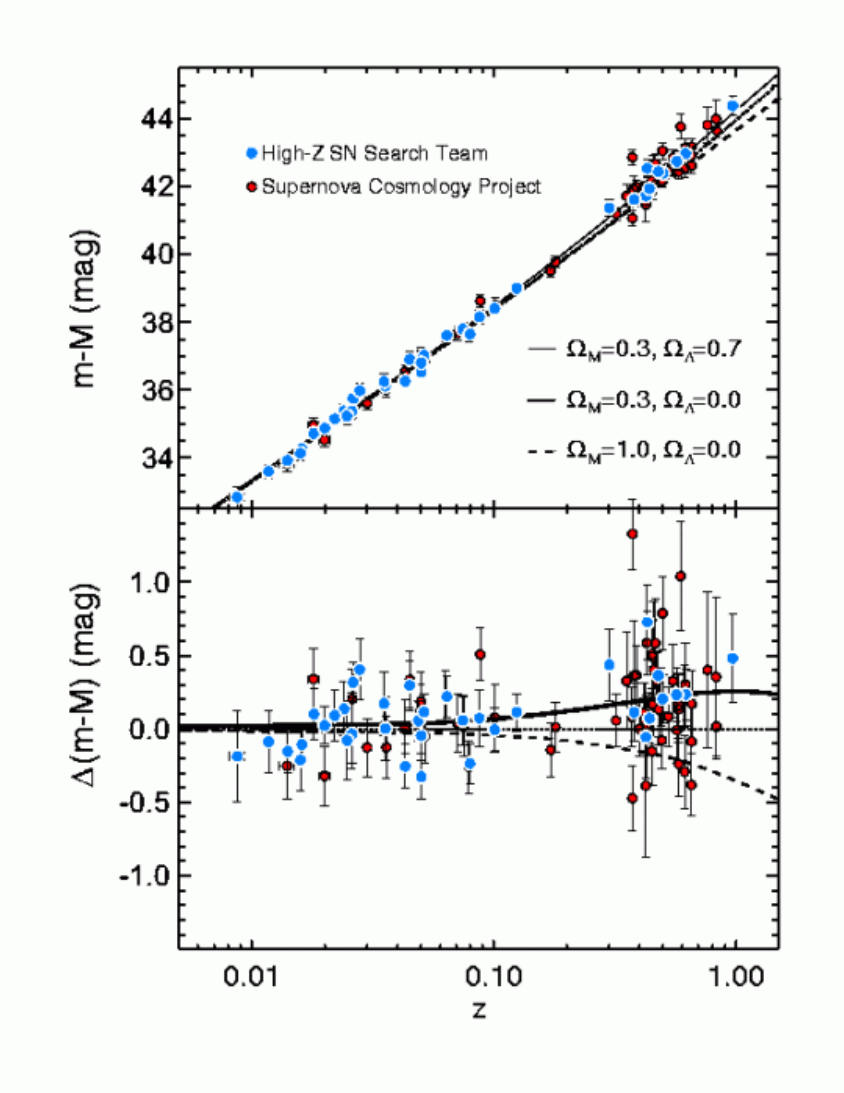


Figure 1.3: Plot of Hubble Diagram from Type Ia supernovae. The graph should be a perfect straight line if the Hubble law can be generalized to high redshifts. However, in both panels, the data points with $z > 0.5$ deviate somewhat from the straight line. Image credit: High-z supernova search team.

1.6.1 The flatness problem

We can derive the relation between the spatial curvature κ and the density parameter $\Omega(t)$ by rearranging the Friedmann equation (Eq.(1.10)):

$$1 - \Omega(t) = -\kappa \left(\frac{c}{H(t)a(t)R_0} \right)^2 \tag{1.29}$$

where $\Omega(t)$ represents the density parameter including all energy contents other than curvature. Present observations including CMB and Type Ia supernovae results show that Ω_0 is very close to 1 [23], specifically $|1 - \Omega_0| \leq 0.005$. One may explain this phenomenon as a coincidence. However, we can consider the same relation in the early universe:

1. Radiation dominated universe: $a \propto t^{1/2} \Rightarrow H(t) \propto t^{-1}$ and $|1 - \Omega(t)| \propto t$.

2. Matter dominated universe: $a \propto t^{2/3} \Rightarrow H(t) \propto t^{-1}$ and $|1 - \Omega(t)| \propto t^{2/3}$.

Using these relations, we can extrapolate the value of Ω_0 back to the matter-radiation equality era, which yields $|1 - \Omega_{mr}| \leq O(10^{-6})$. Similarly, we require $|1 - \Omega(t)| \leq O(10^{-15})$ at the time of the Big Bang neucleosynthesis, and $|1 - \Omega(t)| \leq O(10^{-62})$ at Planck time $t_P \sim 10^{-44}$ s. Attempts to explain these fine-tuned conditions with coincidence would be especially far-fetched. This is the flatness problem, and new physical mechanisms are needed to resolve it.

1.6.2 The horizon problem

Consider the comoving distance from us to the surface of last scattering:

$$d_p(t_0) = c \int_{t_{ls}}^{t_0} \frac{dt}{a(t)} = c \int_0^{z_{ls}} \frac{dz}{H(z)} \tag{1.30}$$

The angular diameter distance to the last scattering surface is:

$$d_A(t_0) = \frac{d_p(t_0)}{1 + z_{ls}} \tag{1.31}$$

Let's define the particle horizon $d_{hor}(t)$:

$$d_{hor}(t) = a(t) \int_{t_*}^{t} \frac{cdt'}{a(t')} \tag{1.32}$$

which is the proper distance from which light could have traveled from the beginning of the universe. It represents the maximum distance to another point that is causally connected to the particle. For a photon at the last scattering surface, we have:

$$d_{hor}(t_{ls}) = a(t_{ls}) \int_{t_*}^{t_{ls}} \frac{cdt}{a(t)} = \frac{1}{1 + z_{ls}} \int_{z_{ls}}^{\infty} \frac{cdz}{H(z)} \tag{1.33}$$

The epoch of last scattering happened in a matter-dominated era: $H(z) \propto (1+z)^{\frac{3}{2}}$. Using $z_{ls} \approx 1100$, we can calculate the angular separation of points on the last scattering surface that are separated by the particle horizon:

$$\theta_{hor} = \frac{d_{hor}(t_{ls})}{d_A(t_0)} \approx \frac{1}{\sqrt{z}_{ls}} \approx 0.03 \text{ rad} \approx 2^\circ \tag{1.34}$$

This result implies that only photons within regions smaller than 2° are causally connected and had enough time to reach thermal equilibrium. However, the CMB is observed to have a uniform temperature across the sky, with relative fluctuations of only $\sim 10^{-5}$. This is the horizon problem, and new physical mechanisms are needed to explain it.

1.6.3 An accelerated phase as a solution

One possible solution to both problems is to assume an accelerated expansion phase during the early universe known as $inflation$, when $a(t) \sim e^{H_{\text{inf}}t}$. Here the Hubble

constant H_{inf} is fixed during the inflation phase. Let t_i and t_f be the starting and ending time of this phase. We also define the number of e-foldings as $N \equiv H_{\text{inf}}(t_f - t_i)$.

How does inflation solve the flatness problem? In Eq.(1.29), substitute $H(t) = H_{\text{inf}}$ and $a(t) = e^{H_{\text{inf}}t}$, and we have

$$|1 - \Omega(t)| \propto e^{-2H_{\text{inf}}t} \Rightarrow |1 - \Omega(t_f)| = e^{-2N}|1 - \Omega(t_i)| \tag{1.35}$$

Therefore the difference between the density parameter $\Omega(t)$ and 1 decreased exponentially during the inflation phase. Any value of $N \geq 70$ corresponds to $|1 - \Omega(t_f)| \sim 10^{-60}$, sufficient to resolve the flatness problem.

How does inflation solve the horizon problem? Recall that the horizon problem arises from the ratio $d_{hor}(t_{ls})/d_A(t_0) \approx 0.03$, resulting in the calculated causal regions of the CMB being much smaller than observed. Inflation resolves this contradiction by increasing $d_{hor}(t_{ls})$ drastically while keeping $d_A(t_0)$ the same. Recall that

$$d_{hor}(t_{ls}) = a(t_{ls}) \int_{t_*}^{t_{ls}} \frac{cdt}{a(t)} \tag{1.36}$$

$$d_A(t_0) = a(t_{ls}) \int_{t_{ls}}^{t_0} \frac{cdt}{a(t)} \tag{1.37}$$

Using the normalization convention $a(t_0) = 1$, the second integral stays roughly the same when the inflation phase is incorporated, since it happens long before the last scattering and does not effect the evolution of $a(t)$ afterwards. However, the first integral increases drastically as the scale factor is now much smaller before inflation due to the exponential growth of $a(t)$ (as demonstrated in Figure 1.4). As a result, we now have $d_{hor}(t_{ls})/d_A(t_0) \gg 1$. This solves the horizon problem.

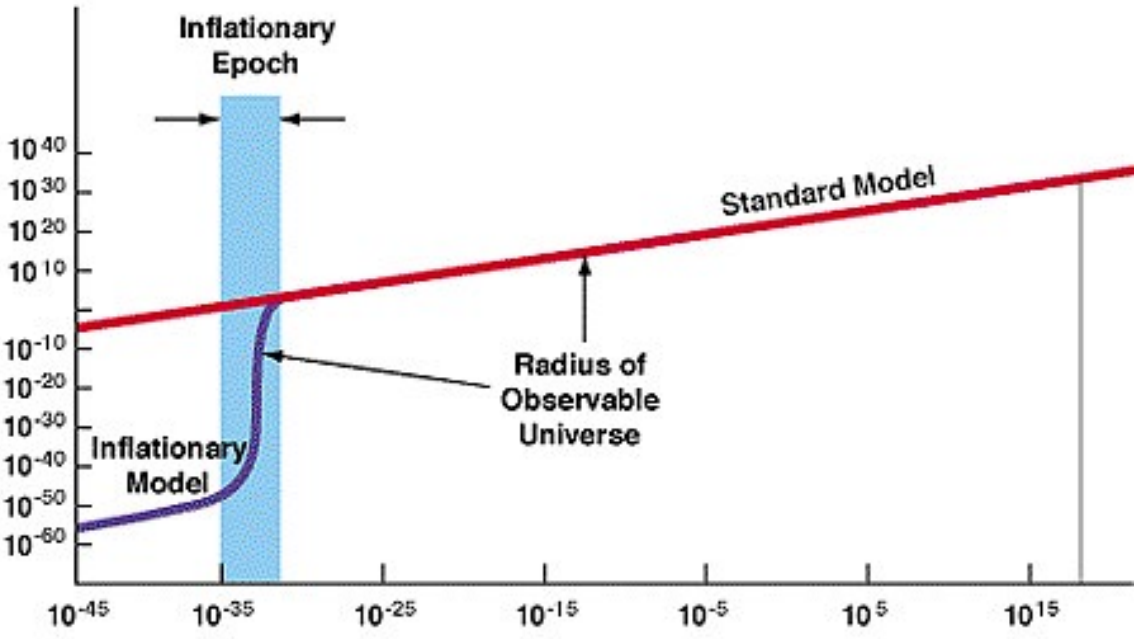


Figure 1.4: Scheme of the evolution of the particle horizon with inflation included vs. excluded in the standard model. We can see the exponential expansion happens in a short period during the very early universe.

1.6.4 The inflation field

The simplest way to implement inflation field is using a scalar field $\phi(\vec{x}, t)$. For simplicity, consider the action of the field under flat space ($\kappa = 0$):

$$S = \int d^3x\, dt\, a^3(t) \left[\frac{1}{2}\dot{\phi}^2 - \frac{c^2}{2a^2(t)} \nabla\phi \cdot \nabla\phi - V(\phi) \right] \tag{1.38}$$

Here $V(\phi)$ is the potential of the inflation. One common choice is $V(\phi) = \frac{1}{2}m^2\phi^2$. Assuming spatial homogeneity, we have $\phi(\vec{x}, t) = \phi(t)$ and $\nabla\phi = 0$. The action becomes:

$$S = \int d^3x\, dt\, a^3(t) \left[\frac{1}{2}\dot{\phi}^2 - V(\phi) \right] \tag{1.39}$$

Varying the action and integrating by parts, we have:

$$\delta S = \int d^3x\, dt\, a^3(t) \left[\dot{\phi}\delta\dot{\phi} - \frac{\partial V(\phi)}{\partial\phi}\delta\phi \right] = \int d^3x\, dt\, \left[-\frac{d}{dt}\left(a^3(t)\dot{\phi} \right) - \frac{\partial V(\phi)}{\partial\phi} \right] \delta\phi \tag{1.40}$$

Setting $\delta S = 0$ for all variations $\delta\phi$, we obtain the equation of motion of the inflation field:

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial V}{\partial\phi} = 0 \tag{1.41}$$

In order to describe the evolution of the scale factor during inflation, we need the energy density of the inflation field:

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{1.42}$$

Using this definition in the fluid equation (1.6) and considering the equation of motion, we get the pressure of the inflation field:

$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{1.43}$$

The pressure and energy density of the inflation field do not follow the usual equation of state: $p = w\rho$. However, we can impose the condition $V(\phi) \gg \dot{\phi}^2$, known as *slow roll*

approximation. Under this approximation, we have:

$$\frac{p}{\rho} = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} \approx -1 \tag{1.44}$$

which implies that the inflation field acts like dark energy ($w \approx -1$) and leads to an accelerated expansion phase. In order to achieve a substantial e-folding value, it is necessary that the inflation field stays in a slow roll phase. This can be achieved through limiting the rate of change of $\dot{\phi}$, namely $\ddot{\phi} \approx 0$ or $3H\dot{\phi} \approx -\partial V/\partial\phi$.

Eventually, by the end of inflation, the inflation field rolls towards the minima of $V(\phi)$ and oscillates around it, ending the slow roll phase. If the inflation field is coupled to radiation and matter, the oscillation will produce ordinary particles, transferring the energy of the inflation field in the process. This process is known as *reheating* [24] (See Figure 1.5). Afterwards, the standard model of cosmology can be used to describe the evolution of the universe.

1.7 Dark Energy

The standard and simplest model of dark energy is the cosmological constant Λ. Since Λ remains constant during the cosmic expansion, it is interpreted as the energy of the vacuum. In quantum field theory, the vacuum energy can be estimated using the Planck energy $E_P = \sqrt{\hbar c/G}$, Planck scale $l_p = \sqrt{\hbar G/c^2}$ and $\rho_{\text{vac}} = E_P/l_P^3 = c^5\hbar/G^2 \sim 10^{96}$ kg m^{-3}. However, the observed value of dark energy density is about $\rho_\Lambda \approx 0.7\rho_c \approx 10^{-26}$ kg m^{-3}, which is a disagreement of ~ 120 orders of magnitude with the theoretical prediction. This is known as the *cosmological constant* problem.

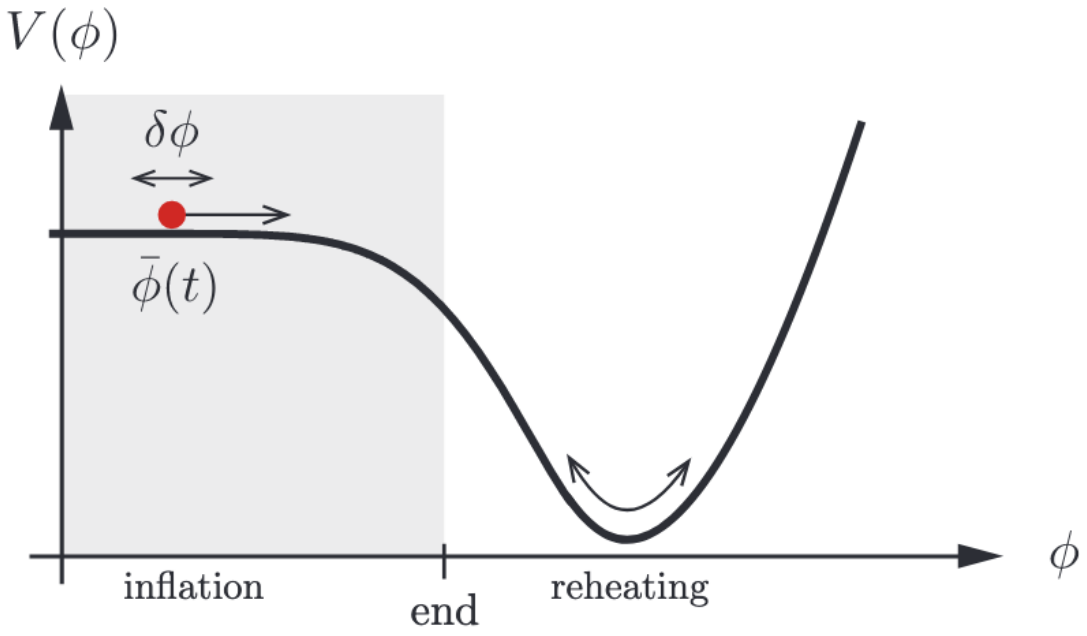


Figure 1.5: The scalar field rolling down its potential. The shaded region represents the slow-roll inflation. The field eventually ends up oscillating at the bottom of the potential, where reheating and the production of ordinary particles happen. Image credit: Daniel Baumann's lecture notes.

Another issue with the cosmological constant model is the fact that the dark energy density is very close to the matter energy density at the present time, a special epoch that we happen to live in. This is known as the *cosmic coincidence* problem.

Due to these issues with the cosmological constant, there are other models proposed for dark energy. One alternative is known as *quintessence*: a scalar field ϕ with potential $V(\phi)$ which takes the same form as the inflation field and differs on time and energy scales. Similar to the inflation field, the equation of state parameter is given by

$$w = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}.$$

In the limit that $\frac{1}{2}\dot{\phi}^2 \ll |V(\phi)|$, $w \approx -1$ and the scalar field acts like a cosmological constant. More generally, the equation of state parameter $w = p_\phi/\rho_\phi$ evolves with time depending on the details of the theory. The time dependence is usually parameterized as $w = w_0 + w_a(1-a)$. Due to this time dependence, its energy density does not have to be small compared to radiation and matter during the early universe, which partly solves the cosmological constant problem.

Phantom energy is another hypothetical form of dark energy satisfying an equation of state parameter $w < -1$. It indicates that the energy density of dark energy is increasing over time instead of being a cosmological constant. Also under such a model, the universe will end in a big rip, meaning that the scale factor will reach infinity in a finite amount of time. Vacuum instabilities is a problem with this model.

Chapter 2

Cosmology and Large Scale Structure

The cosmological principle states that the universe is homogeneous and isotropic on sufficiently large scales ($\sim$ 500 Mpc). At smaller scales, different structures can be observed ranging from solar systems (AU) and galaxies (kpc) to clusters of galaxies (Mpc). It's believed that these structures grew from inhomogeneities in the early universe by gravitational collapse. Inflationary models provide a convincing explanation of the source of the inhomogeneities as density perturbations which are generated by Gaussian quantum fluctuations in the inflation field. The evolution of the inhomogeneities to large scale structures (LSS) that we observe today can be modeled through perturbation theory, which models the energy content of the universe as Newtonian fluids with slight inhomogeneities.

2.1 Perturbation theory and Matter distribution

To model density perturbations, one can consider two extremes of approximation in terms of the particle mean free path:

1. The mean free path of particles is very short and can be modeled as ideal fluids. This approximation works well on large scales.

2. The mean free path of particles is very long. This approximation works well when we study structures on much smaller scales, such as stars and galaxies.

We focus on the first case since we are more interested in large scales.

2.1.1 Power spectrum

Inhomogeneities in the matter distribution, or density perturbations, can be described by the power spectrum $P(\vec{k})$:

$$\langle \delta(\vec{k})\delta(\vec{k'})\rangle \equiv (2\pi)^3 \delta_D(\vec{k}-\vec{k'})P(\vec{k}) \tag{2.1}$$

where δ_D is the 3d Dirac delta function, and $\delta(\vec{k})$ is the overdensity field in the Fourier space:

$$\delta(\vec{k}) = \int d^3x\, e^{i\vec{k}\cdot\vec{x}}\delta(\vec{x}) \tag{2.2}$$

with $\delta(\vec{x})$ the overdensity field in the real space. The power spectrum can also be expressed as the Fourier transform of the correlation function $\xi(\vec{r})$:

$$P(\vec{k}) = \int d^3r\, e^{-i\vec{k}\cdot\vec{r}}\xi(\vec{r}) \tag{2.3}$$

The correlation function is more directly connected to observations, while the power spectrum is more expressive for theoretical predictions. It can be difficult to determine $\xi(\vec{r})$ accurately for large scales, where $\xi(\vec{r})$ has small values. Therefore, $P(\vec{k})$ is often used to discuss inhomogeneities in large scales, or low k values, while small scale density perturbations are more commonly discussed using $\xi(\vec{r})$.

If we assume statistical isotropy, $P(\vec{k})$ is independent of the direction of $\vec{k}$, i.e. $P(\vec{k}) = P(k)$ (similarly, $\xi(\vec{r}) = \xi(r)$). $P(k)$ has the dimension of volume, while $\xi(r)$ is dimensionless. One can define the dimensionless power spectrum:

$$\mathcal{P}(k) \equiv \frac{k^2}{2\pi^2} P(k) \tag{2.4}$$

Inflation predicts a scale-invariant primordial power spectrum, i.e. $\mathcal{P}_\Phi(k) =$ constant, where $\mathcal{P}_\Phi(k)$ is the dimensionless power spectrum of the gravitational potential. In terms of overdensity power spectrum, this is equivalent to $P(k) \propto k\mathcal{P}_\Phi(k) \propto k$. Deviations from the scale-invariant power spectrum can be quantified using the *scalar spectral index*:

$$P(k) \propto k^{n_s} \tag{2.5}$$

where $n_s \sim 1$. An almost scale-invariant power spectrum corresponds to the slow roll condition and slow varying Hubble constant. Since fluctuations in the inflation field set the initial conditions for anisotropies in CMB and LSS, n_s can be constrained by related observations, which are in agreement with the prediction. For example, $n_s = 0.9626 \pm 0.0057$ from Planck 2018 results [25].

2.1.2 Newtonian linear perturbation theory

In Newtonian theory, the equations that describes an ideal fluid are:

$$\frac{\partial \rho}{\partial t}|_{\vec{r}} + \nabla_{\vec{r}} \cdot (\rho \vec{u}) = 0 \tag{2.6}$$

$$\rho \left(\frac{\partial}{\partial t}|_{\vec{r}} + \vec{u} \cdot \nabla_{\vec{r}} \right) \vec{u} = -\nabla_{\vec{r}} p - \rho \nabla_{\vec{r}} \Phi \tag{2.7}$$

$$\nabla^2_{\vec{r}} \Phi = 4\pi G \rho \tag{2.8}$$

where $\rho(\vec{x}, t)$ is the energy density, $p(\vec{x}, t)$ is the pressure, $\vec{u}(\vec{x}, t)$ is the velocity field and $\Phi(\vec{x}, t)$ is the gravitational field experienced by the fluid. The subscript $\vec{r}$ indicates that the equations should be viewed in terms of physical (proper) coordinates. Eq. (2.6) is the continuity equation, which captures conservation of energy. Eq. (2.7) is Euler's equation, which can be viewed as Newton's 2nd law for fluids. Eq. (2.8) is Poisson's equation for the gravitational field.

It is generally easier to perform linear perturbation calculations in comoving coordinates $\vec{x}(t)$, which is related to physical coordinates via $\vec{r}(t) = a(t)\vec{x}(t)$. Their derivatives are related by:

$$\nabla_{\vec{r}} = \frac{1}{a} \nabla_{\vec{x}} \tag{2.9}$$

$$\frac{\partial}{\partial t}|_{\vec{r}} = \frac{\partial}{\partial t}|_{\vec{x}} + \frac{\partial \vec{x}}{\partial t} \cdot \nabla_{\vec{x}} = \frac{\partial}{\partial t}|_{\vec{x}} - H\vec{x} \cdot \nabla_{\vec{x}} \tag{2.10}$$

From now on, we perform our calculations in comoving coordinates and omit the subscripts $\vec{x}$.

To introduce perturbations to the ideal fluid equations, we expand the proper velocity in terms of the peculiar velocity $\vec{v}(\vec{x}, t) = a(t)\dot{\vec{x}}(t)$: $\vec{u}(\vec{x}, t) = H\vec{r} + \vec{v}(\vec{x}, t)$ and decompose the density field in terms of the overdensity field $\rho(\vec{x}, t) = \bar{\rho}(r)(1 + \delta(\vec{x}, t))$,

where $\bar{\rho}$ represents the homogeneous background density. In comoving coordinates, the continuity equation becomes:

$$\frac{\partial \rho}{\partial t} + 3H\rho + \frac{1}{a}\nabla \cdot (\rho \vec{u}) = 0 \tag{2.11}$$

For non-relativistic matter, we have $p \ll \rho$, combining with the fluid equation (1.6), yields that $\dot{\bar{\rho}} + 3H\bar{\rho} = 0$. To derive the continuity equation for perturbations, we apply the linear approximation, keeping first order terms of $\vec{v}$ or δ and dropping 2nd or higher order terms. We get:

$$\dot{\delta} = -\frac{1}{a}\nabla \cdot \vec{v} \tag{2.12}$$

Similarly, perturbing the pressure and gravitational field about the background ($p = \bar{p} + \delta p$, $\Phi = \bar{\Phi} + \delta\Phi$), we get the linear version of Euler's equation and Poisson's equation in comoving coordinates:

$$\bar{\rho}a(\dot{\vec{v}} + H\vec{v}) = -\nabla\delta p - \bar{\rho}\nabla\delta\Phi \tag{2.13}$$

$$\nabla^2\delta\Phi = 4\pi G\bar{\rho}a^2\delta \tag{2.14}$$

Combining the 3 linearized equations (2.12), (2.13) and (2.14), we get an equation that describes the evolution of the density perturbation $\delta(\vec{x}, t)$:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}(\nabla^2 + k_J^2)\delta = 0 \tag{2.15}$$

where c_s is the speed of sound ($\partial p/\partial\rho = c_s^2/c^2$) and $k_J \equiv 4\pi G a^2\bar{\rho}/(c^2 c_s^2)$ is the (co-moving) Jeans wavenumber.

2.1.3 Growth of structure

Using Eq. (2.2) to convert $\delta(\vec{x}, t)$ in Eq. (2.15) to Fourier space we obtain

$$\ddot{\delta}(\vec{k}, t) + 2H\dot{\delta}(\vec{k}, t) + \frac{c_s^2}{a^2}(k^2 - k_J^2)\delta(\vec{k}, t) = 0 \tag{2.16}$$

The solutions behave differently depending on the scale:

1. On small scales, $k \gg k_J$ and $k^2 - k_J^2 > 0$. The equation (2.16) represents a damped harmonic oscillation with an amplitude that decreases exponentially over time. This solution represents a decaying mode, where pressure force overcomes gravity.

2. On large scales, $k \ll k_J$ and $k^2 - k_J^2 \approx k_J^2$. The general solution is:

$$\delta(\vec{k}, t) = c_1(\vec{k})H(t) + c_2(\vec{k})H(t)\int_0^t \frac{dt'}{H^2(t')a^2(t')} \tag{2.17}$$

   Usually, the first half represents a decay mode. We are more interested in the second half of the solution, which represents a growth mode, also known as the *growth factor*:

$$D(a) = \frac{5}{2}\Omega_M H_0^2 H(a)\int_0^a \frac{da'}{[a'H(a')]^3} \tag{2.18}$$

   where normalization $D(a \to 0) \to a$ is used. For example, in a matter dominated universe, we have $a \propto t^{2/3}$ and $H = \dot{a}/a \propto t^{-1}$. The first half yields $\delta \propto t^{-1}$, and the second half yields $\delta \propto t^{2/3}$. If we remove expansion and set $H = 0$ in equation (2.15), the solution yields $\delta \propto \exp(c_s K_J t/a)$, a faster growth mode compared to a solution with a non-zero Hubble constant. This deceleration of structure growth due to expansion is known as *Hubble friction*.

   Another observation we can make about solution (2.17) is that the growth factor is independent of the scale k, and each Fourier mode evolves following the same pattern. Therefore, on large linear scales, the shape of the power spectrum stays constant.

2.1.4 Galaxy bias

In practice, it's much more difficult to obtain data of the matter overdensity δ_m than the galaxy overdensity δ_g, since most surveys measure the galaxy distribution directly. Therefore, a complete description of the relation of the two distributions is desirable. In general, this relation is very complicated, as it encodes the physics of galaxy formation, which is a complex process that happens over long periods of time, during which non-linear interaction of structures take place.

Generally, the mapping from δ_g to δ_m is described by the *bias function*: $\delta_g = F[\delta_m, \nabla_{ij}\Phi_v]$, where $\nabla_{ij}\Phi_v$ are 2nd order derivatives of the velocity potential. It relates δ_g to δ_m beyond the smoothing scale, and is generally a non-local, non-linear function. Remarkably, for sufficiently large scales and Gaussian initial conditions such contribution is negligible, allowing the bias function to be approximated by a Taylor expansion in terms of the matter fluctuations:

$$\delta_g(\vec{k}, t) = b(k, z)\delta_m(\vec{k}, t) \tag{2.19}$$

This is the linear bias approximation, whose validity relies on the fact that structure formation is dominated by gravity on large scales [26]. By equation (2.1), we can express the galaxy power spectrum using the matter power spectrum:

$$P_g(k, z) = b^2(k, z)P_m(k, z) \tag{2.20}$$

The galaxy bias is greater at higher redshifts, or earlier epochs of galaxy formation, since the first galaxies formed will collapse in the most dense regions as the structures grow overtime [27, 28]. Eventually, galaxies will become unbiased tracers of the mass distribution ($b \to 1$ as $t \to \infty$).

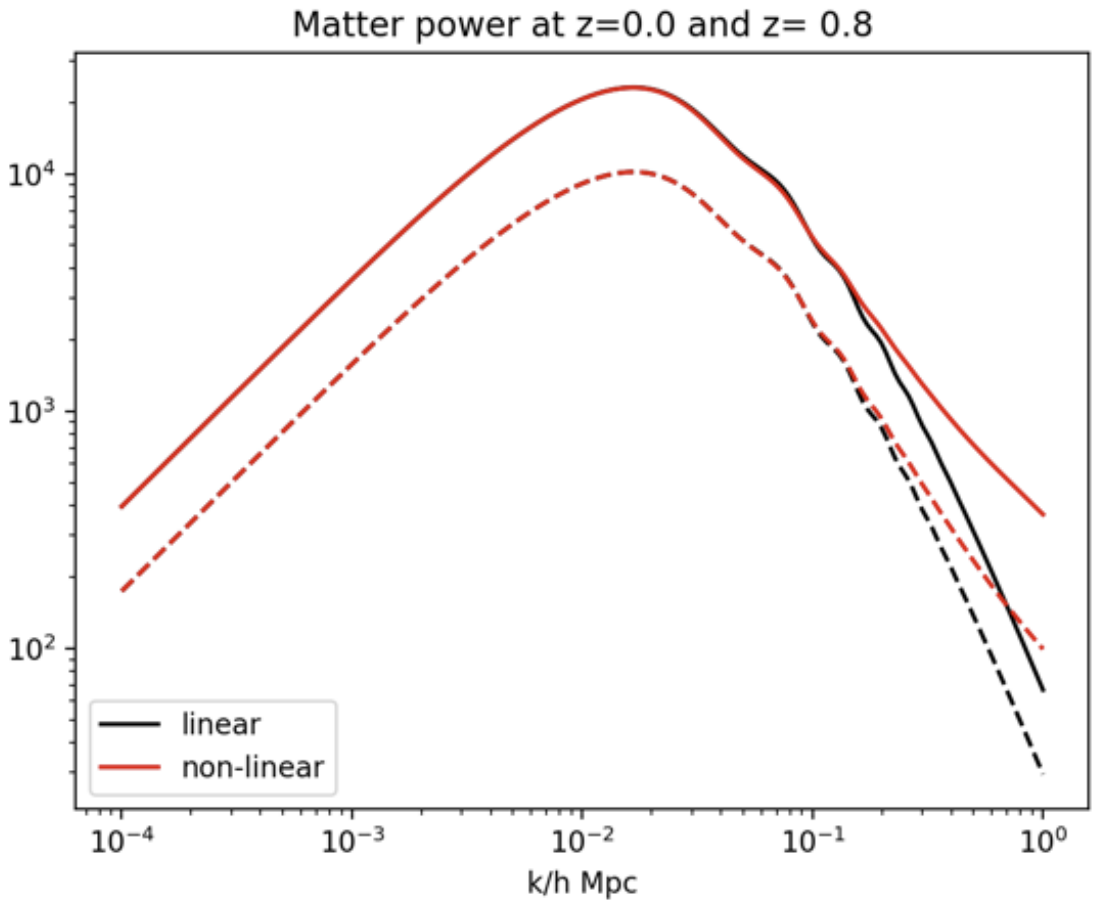


Figure 2.1: Matter power spectra at different redshifts, as calculated by `CAMB` (Code for Anisotropies in the Microwave Background [4]). The cosmological parameters from Planck 2018 are used.

2.1.5 Velocity field

From Eq. (2.17) and (2.18), $\delta(a) \propto D(a)$ under the growth mode. We can define the *growth rate*:

$$f \equiv \frac{d\ln D}{d\ln a} = \frac{\dot{D}}{D}\frac{a}{\dot{a}} = \frac{\dot{\delta}}{H\delta} \tag{2.21}$$

The growth rate can be modelled as $f(z) \approx \Omega_m^{\gamma}(z)$, where γ (~ 0.55) is the *growth index* and stays mostly constant over a long range of time. Combine with Eq. (2.12) yields:

$$\nabla \cdot \vec{v} = -aHf\delta(\vec{x}) \tag{2.22}$$

Going to Fourier space and using the galaxy bias approximation, we have:

$$-ikv_{\vec{k}} = aH\beta\delta_{g\vec{k}} \tag{2.23}$$

where $\beta \equiv f/b$. Since k and $v_{\vec{k}}$ are inversely proportional, the effect of density perturbation on peculiar velocities is more significant on large scales (lower wave number k), making probes based on peculiar velocities robust when studying matter distributions.

2.2 Redshift space distortion

Distance information is rarely measured directly in astronomy. One of the different ways to estimate distances to objects is by using their redshifts. If the cosmology is known, the comoving distances to objects can be calculated from redshifts using Eq. (1.19) or estimated by the Hubble relation: $z = H_0 r$ for $z \ll 1$. The redshift measurement can be done using spectroscopy (measuring shifts in emission spectra) or photometry. However, due to the presence of radial peculiar velocities, the measured redshifts also reflect Doppler

effects on top of the cosmic expansion:

$$z = H_0 r + \frac{v_r}{c} \tag{2.24}$$

where v_r is the radial peculiar velocities. This discrepancy between distance space and redshift space is known as the *redshift space distortion* (RSD). Since the radial peculiar velocities are affected by local gravity, RSD is correlated to the overdensity field and measured matter distribution by redshifts.

2.2.1 Kaiser effect and Finger of God

Consider a spherical shell of galaxies a certain distance away from an observer undergoing gravitational collapse due to overdensity at the center of the shell. The galaxies farther away have negative peculiar velocities (moving towards the observer), while the closer galaxies have positive peculiar velocities (moving away from the observer). On linear scales ($\delta(\vec{x}) \ll 1$), the contribution on measured redshifts from peculiar velocities does not overcome the cosmic expansion ($H_0 r \gg v_r/c$), and the shell is flattened in the radial direction in redshift space. This mechanism is known as the Kaiser effect [29]. However, on small scales, the peculiar velocities can compensate the Hubble Law ($H_0 r \ll v_r/c$). The sphere will appear elongated in redshift space, with the far end of the sphere (in real space) appears closer to us (in redshift space) and vice versa. This is known as the "Finger of God" [29]. These effects are demonstrated in Figure 2.2.

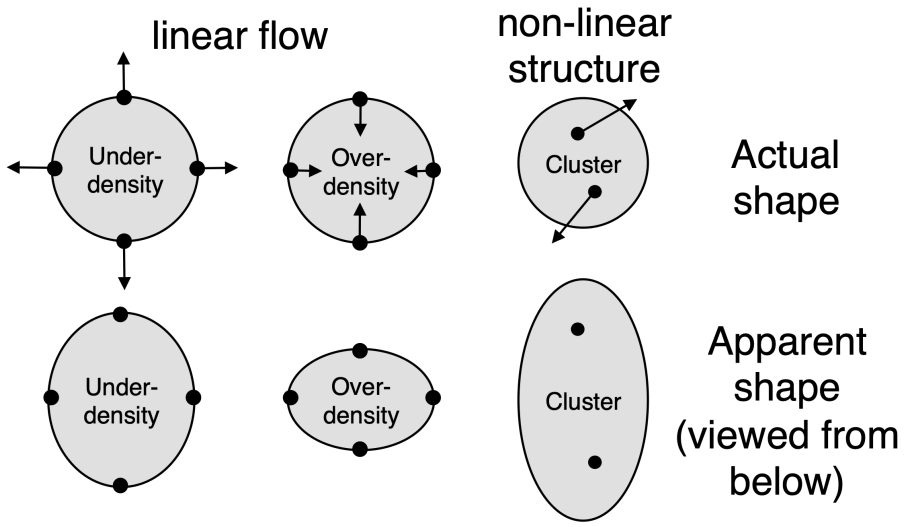


Figure 2.2: Illustrative diagrams representing how spherical matter distributions (top row) appear in redshift space (bottom row) [5]. In the diagram, up-down represents the radial direction, and left-right represents the transverse direction.

2.2.2 Kaiser approximation

On linear scales, the Kaiser effect can be analysed quantitatively by considering the relation between the matter power spectra in real and redshift space [29]:

$$P^{(z)}(\vec{k}) = (1 + \beta \cos^2 \theta_{\vec{k}})^2 P(k) \tag{2.25}$$

where $\theta_{\vec{k}}$ is the angle between wave vector $\vec{k}$ and radial vector $\vec{r}$ of the observed region. As expected, we can see that the redshift space power spectrum $P^{(z)}(\vec{k})$ has an angular dependence, even though the real space power spectrum $P(k)$ is isotropic. The angular dependence can be further explained using a spherical multipole expansion:

$$P^{(z)}(\vec{k}) \equiv P^{(z)}(k, \cos\theta_{\vec{k}}) = \sum P_l^{(z)} L_\ell(\cos\theta_{\vec{k}}) \tag{2.26}$$

where $L_l(\cos\theta_{\vec{k}})$ are Legendre polynomials. The only non-zero multipoles $P_l^{(z)}$ can be calculated as:

$$P_0^{(z)}(k) = (1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2)P(k) \tag{2.27}$$

$$P_2^{(z)}(k) = (\frac{4}{3}\beta + \frac{4}{7}\beta^2)P(k) \tag{2.28}$$

$$P_4^{(z)}(k) = \frac{8}{35}\beta^2 P(k) \tag{2.29}$$

These multiploes can be measured using redshift distributions from galaxy surveys. We can use the equations to solve for β and $P(k)$. The multipoles are illustrated in figure 2.3 in terms of correlation functions.

2.3 Baryon Acoustic Oscillations

During the early universe, before photon decoupling and reionization, baryons were tightly coupled with photons in a hot plasma. Quantum fluctuations during cosmic inflation led to density fluctuations, which created imbalance with in the primordial plasma. The competing forces of pressure gradient and gravitational collapse generated sound waves in the baryon-photon plasma, also known as Baryon Acoustic Oscillations (BAO).

After photon decoupling, photons ceased to interact with baryons and streamed freely, forming the Cosmic Microwave Background that we observe. The universe was cold enough for baryons to combine into neutral atoms (recombination). However, the acoustic oscillation presented earlier imprints on the matter distribution, creating overdensity at a particular distance scale, i.e. the comoving sound horizon at photon decoupling:

$$r_s(t) = \int_0^t \frac{c_s(t)}{a(t)} dt \tag{2.30}$$

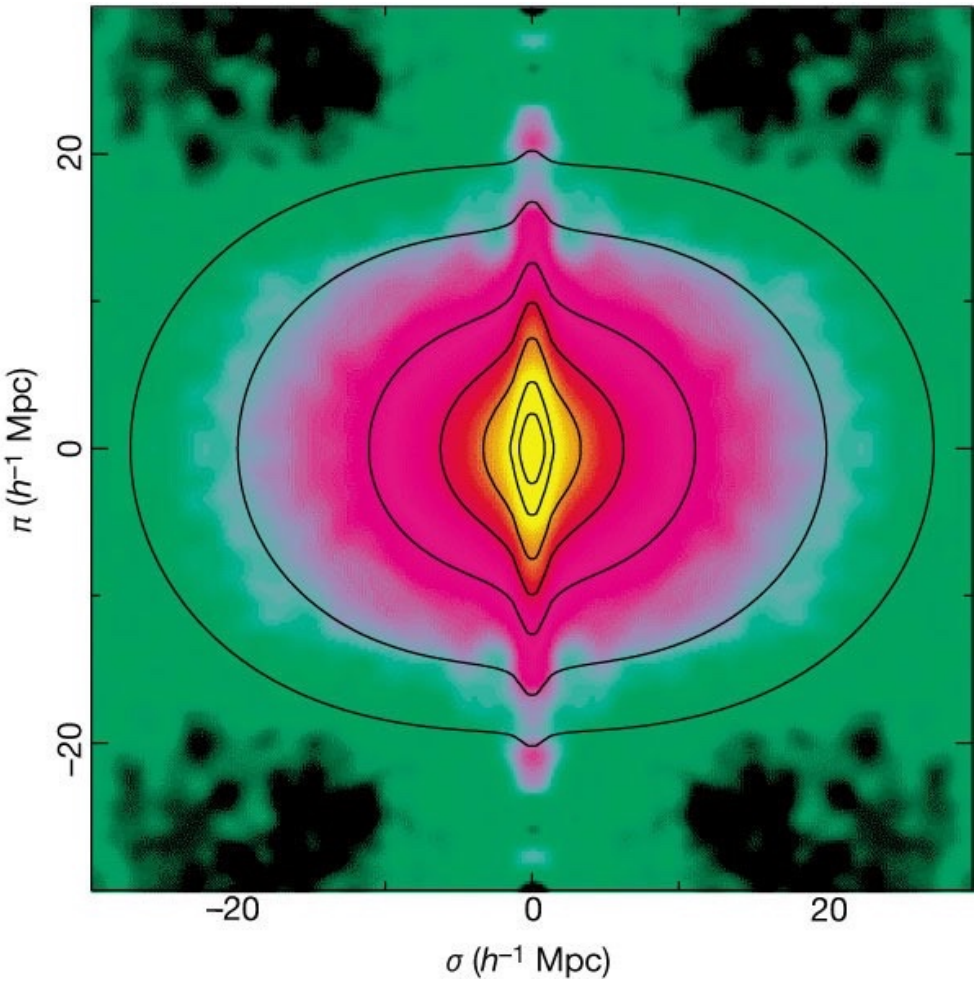


Figure 2.3: The correlation function in redshift space, measured by the 2dFGRS survey, plotted as a function of radial (π) and transverse (σ) directions [6]. The FOG elongations are demonstrated by the dense regions at low σ and high π, while the Kaiser effect is displayed by high σ and low π regions.

The sound horizon represents the largest possible distance that a sound wave could have travelled since the beginning of the universe: $r_d \equiv r_s(t_d) \approx 110$ Mpc h^{-1}. This length scale is reflected as a"bump" in the correlation function at r_d or "wiggles" in the power spectrum (See Figure 2.1 at $k \sim 0.1$h Mpc) for k up to 0.2 h Mpc^{-1}. It also provides a standard ruler, allowing the geometry of the universe to be measured. BAO measurements serve a crucial role in constraining cosmological parameters and understanding the expansion history of the universe.

Chapter 3

The Vera Rubin Observatory

3.1 Introduction

The Legacy Survey of Space and Time (LSST) is a 10-year survey that will be carried out by the Vera Rubin Observatory. Designed to obtain repeated images of the entire southern sky ($\sim 20,000$ deg^2) in great detail, it is one of the most ambitious surveys in the optical spectrum. The telescope is a ground-based system equipped with wide-deep-fields. It has an 8.4 m primary mirror (6.5 m effective), a 9.6 deg^2 field of view (FOV), a 3.2-gigapixel camera, and six photometric filters (ugrizy) covering the wavelength range of 320–1050 nm [7]. Over the course of the survey period, each part of the survey footprint will be visited ~ 150 times in each band [30].

The Observatory's design centers on 4 main science goals: probing dark energy and dark matter, taking an inventory of the solar system, exploring the transient optical sky, and mapping the Milky Way. To distinguish different models of dark energy and constrain cosmological parameters, the most robust method is to use a combination of probes. With

about 90% of the survey time devoted to the deep-wide-fast mode, LSST will be able to measure the redshift dependencies of the most powerful probes, including:

1. Weak gravitational lensing/cosmic shear (WL).
2. Galaxy power spectrum (large-scale structure, LSS) and baryon acoustic oscillations (BAO).
3. The mass function and clustering of clusters of galaxies.
4. Time delays in strong lensed quasar and supernovae (SN).
5. Photometry of Type Ia supernovae.

In order to perform these measurements accurately, requirements are made on data properties, and the systems are designed accordingly to optimize various parameters.

3.2 The telescope and camera

The site that houses the telescope and its camera is constructed atop Cerro Pachón, a mountain located close to the northern Chilean city of Vicuña, sharing the ridge with Gemini South and the Southern Astrophysical Research (SOAR) Telescope. The system first light is expected in July 2024 and full survey operations are aimed to begin in October 2024. The data collected is scheduled to become fully public after two years.

3.2.1 Telescope

The effective etendue, a metric that measures both the width and depth that a telescope can capture, will be around 320 m^2deg^2. It is achieved through a reflective 3-mirror design (See Figure 3.2):

1. The primary mirror (M1) is a ring shaped concave mirror. It has an outer diameter of 8.4 m and an inner diameter of 5.0 m, corresponds to an effective 6.5 m aperture.

2. After the incident light is collected by M1, it is reflected and focused further by the the secondary mirror (M2), a convex mirror with a diameter of 3.4 m, and the tertiary mirror (M3, concave with a diameter of 5.0 m).

3. M1 an M3 are designed as a single surface, with the inner diameter of M1 coincide with M3. This design allows M1 and M3 to be manufactured and polished from a single substrate.

After reflected by all 3 mirrors, the incident light is focused onto the three refractive lenses of the camera, and captured by the CCD sensors at the focal plane.

3.2.2 Camera

The LSST Camera is designed to provide a 9.6 deg^2 field of view, with a 0.2 arc-second pixel resolution for optimized pixel sensitivity. To achieve the design requirements, the focal plane CCD array follows a hierarchical layout: 9 4K $\times$ 4K 10 μm CCD sensors form a 3 $\times$ 3 raft (Figure 3.3 3.4), and the 21 rafts form the 3.2-gigapixel focal plane.

To minimize noise and dark current, the optimal operating temperatures of the CCD sensors lie in the range of -100°C to -80°C. A vacuum cryogenic refrigeration system

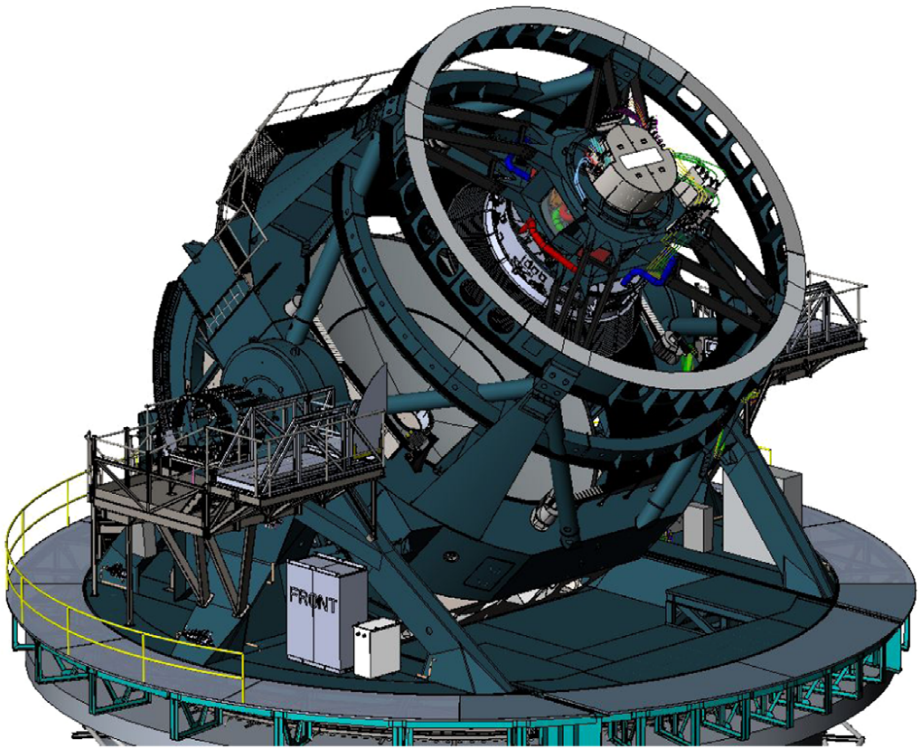


Figure 3.1: Structural design of the LSST telescope [7].

that operates at -130°C is specially designed to maintain the electronics at optimal operating temperatures [1]. It is located behind the L3 lens and is composed of a cluster of 6 parallel cooling circuits, each with a 85W cooling capacity.

To change the camera optical filters in short time spans (90 to 120s), a mechanical shutter and a carousel assembly which holds five large optical filters are mounted at the front of the camera body. Any of the five filters can be swapped with a sixth filter during daylight hours.

To achieve a fast readout time (2s) of the images obtained by the entire focal plane, 3 front-end electronic boards are embedded in each raft for read-in parallelization.

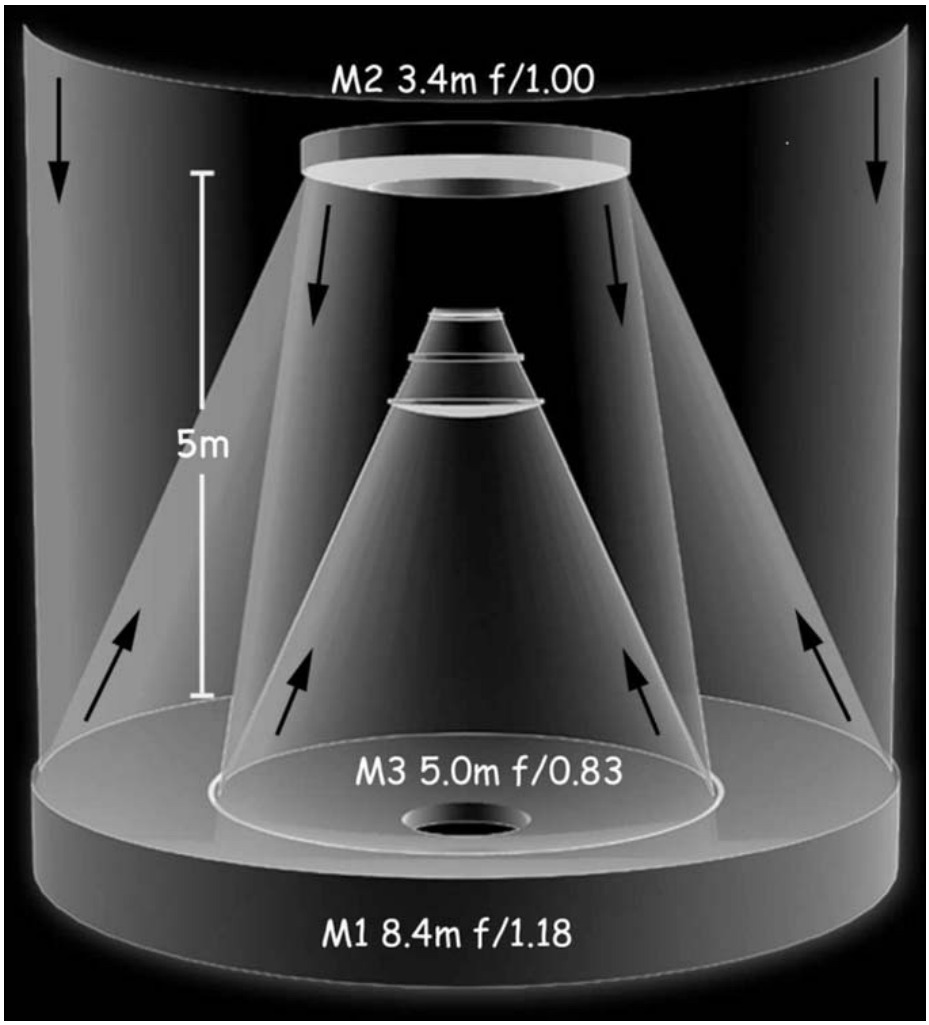


Figure 3.2: The three-mirror design of the telescope's optical system. The M1 and M3 mirrors are polished from a single layer of material and form a smooth surface [7].

3.3 Data requirement and system design

To meet the science goals described previously, various constraints on data properties and survey parameters are required. For example,

1. The single-visit depth needs to reach $r \sim 24.5$, and the single-visit exposure time needs to be less than a minute (30 s is chosen). The depth requirement is partly driven by the co-added depth required by WL and LSS probes. The exposure time requirement allows the removal of systematics in the galaxy shape measurements due to the point-spread functions (PSFs).

2. The total number of visits of any area of the sky should be ~ 1000, and the co-added survey depth should reach $r \sim 27.5$ with high signal-to-noise ratio in all bands.

3. The number of filters should be at least 6, ranging from 320 nm to 1050 nm, with no large gaps in between filters. The number of visits per filter should be roughly evenly distributed across all bands. This enables accurate and precise photometric redshifts, separation of stellar populations. Cosmological parameter estimation from different probes also requires uniform coverage of the sky.

4. Observing conditions (the median seeing from the atmosphere in r-band is $\sim 0.65''$ at the site) should remain the main limitation for image quality. The hardware should not depreciate the image quality significantly. This constraint is driven by WL, LSS, and survey depth requirements of point like objects.

[1] `https://www.lsst.org/about/camera`

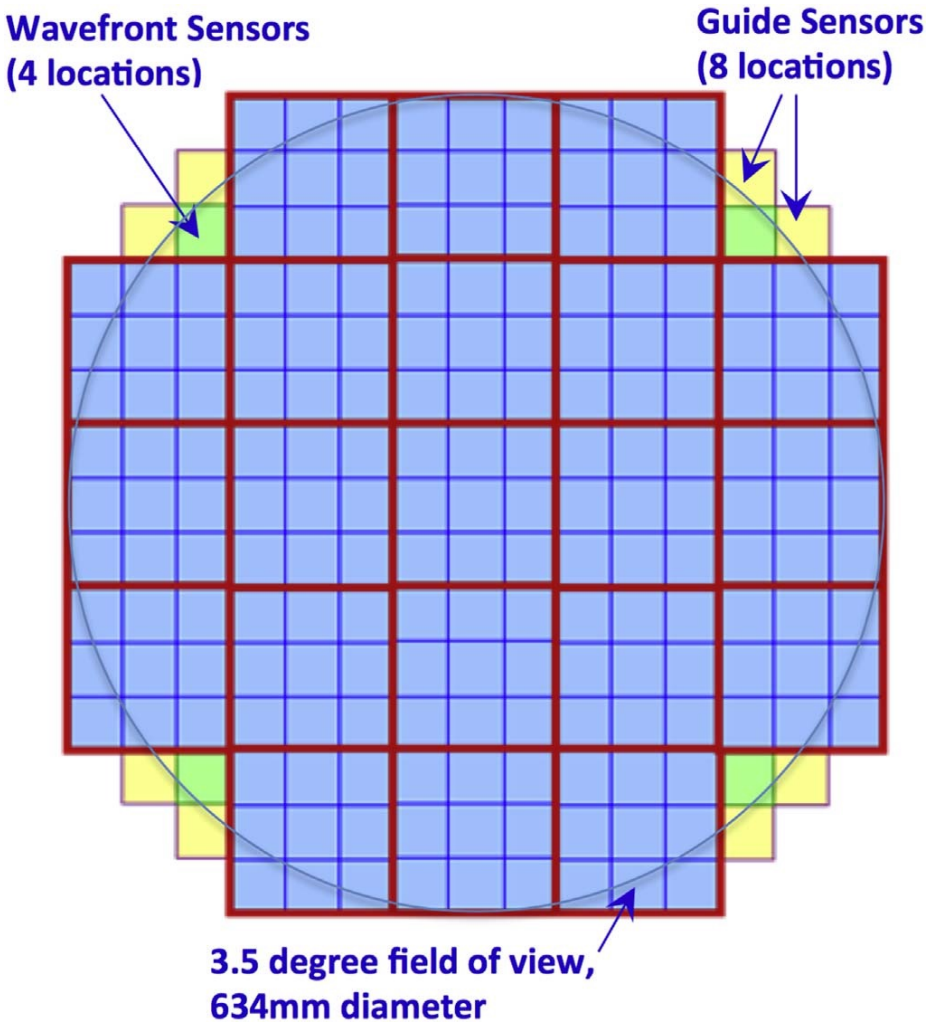


Figure 3.3: Focal plane array of the LSST Camera. Each small square (outlined in blue) represent a 4K×4K pixel CCD sensor. 9 sensors are assembled into a raft (outlined in red). There are 21 rafts in total, resulting in 3.2 gigapixels from the 189 CCD sensors. [7]

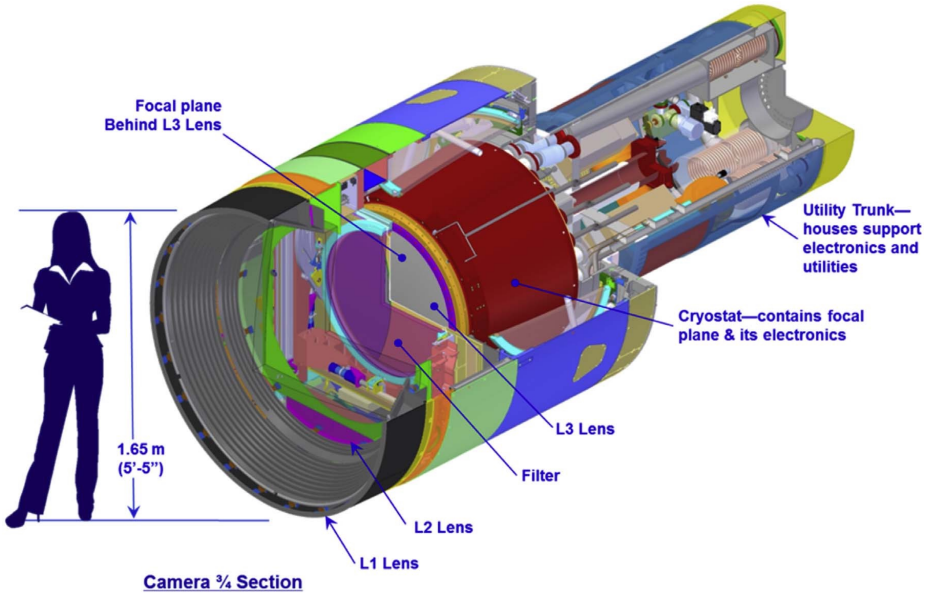


Figure 3.4: Cross-sectional view of the LSST camera [7]

A more detailed description of constraints can be found in the LSST Science Book [30] and LSST System Science Requirements Document [31]. The system design parameters are optimized based on these constraints. For example: the 6.5 m effective diameter of the primary mirror is the minimum required to satisfy both the depth (single visit: $r \sim 24.5$, co-added: $r \sim 27.5$) and cadence (3–4 days revisit time, with 30 s exposure time per visit) requirements; the addition of y-band measurement that can detect wavelengths up to ~ 1000 nm is based on the increased redshift range achievable with deeper survey depth.

3.4 Survey strategy

The baseline survey strategy is designed to maximize the return on the four main science goals, which corresponds to about 90% of the observing time. The current baseline cadence assumes that images are acquired as pairs of back-to-back, 15-second exposures to

enable removal of cosmic rays. The pair of images is a single observation of ~ 10 deg^2 of sky through a single filter, also known as a "visit". The detailed ordering of these visits in time and allocation of them among the six filters are still under discussion within the LSST community.

The Operation Simulator (OpSim) [32] is a software application developed to run survey simulations. Specifically, it models the telescope control system, site conditions, and observation schedules. A ranking algorithm is used to optimize observation schedules. After each visit, the algorithm assigns scores to all possible next visits based on their locations, times, observing conditions and filters according to the scientific requirements. For example, to ensure uniform coverage, regions with fewer previous observations will score higher than those that have been visited more; to benefit WL measurements, preference is given to r and i-band observations in times of good seeing and and low airmass.

Using the baseline cadence, the anticipated total number of visits over a 10 year survey period is about 2.45 million (about 4.9 million exposures over six filters).

Besides the main deep-wide-fast survey via the baseline cadence, about 10% of the survey time will be dedicated to other strategies, such as allocating high number of visits to a set of Deep Drilling Fields. The deep drilling field program will generate data sets that are excellent for a variety of science topics, such as planetary and interstellar studies.

3.5 Data management

The LSST survey program will produce about 15 TB of raw images every night, thanks to its high image resolution and survey cadance. Similar to most modern surveys,

this represents a substantial data generation rate, and data reduction is necessary before analysis can be made systematically. The LSST Data Management system (DM) [33] was designed to reduce the data collected in real time to catalogs and images with scientific significance.

To convert the raw survey data into formats that's suitable for scientific analysis, static and self-consistent Data Releases (DR) will be generated periodically through out the duration of the survey. The DRs will include single-epoch images, deep co-adds of the images, catalogs of sources (the detection, measured properties and characterizations of objects on individual visits). The object detection is performed using the co-adds, and the measured properties are determined through simultaneously fitting models to all single-epoch observations. Suitable software, application programming interfaces (APIs), and computing infrastructure will be provided at the LSST data access centers. By running user generated code and post-process the DR catalogs, customized catalogs can be made to suit different analysis needs.

The LSST Software Stack is developed to process and analyse the data generated in an automated fashion. It's capable of LSST data processing tasks such as co-addition, calibration and single-frame processing [34]. While the software development is mainly built upon the refined frameworks of prior cosmological surveys such as SDSS and DES, a substantial portion of the code base has been newly written, due to the consideration of performance, scalability and long term maintainability.

3.6 Simulating the LSST system

Before the telescope starts operating, in order to demonstrate that the LSST system is capable to achieve the designed performance described in the SRD, simulations have been carried out by the LSST DESC Collaboration, where design decisions and system requirements can be optimized. The simulations and data pipelines are developed as a series of "data challenges" (DCs) with increasing scale and complexity. In this work, we mainly use the data product released by the second Data Challenge (DC2) [8].

The simulation pipeline consists of 4 primary stages:

1. The simulation and optimization of the observation scheduler (OpSim).
2. N-body simulations that generate extragalactic object catalogs (cosmoDC2).
3. The simulated astronomical instance and truth catalogs (CatSim).
4. A system for generating observation and realistic LSST images of a given area of the sky (PhoSim). DM catalogs are generated through image processing.

3.6.1 Extragalactic object catalogs

The first part of the workflow consists of simulations under a cosmological model with parameters: $\omega_{cdm} = \Omega_{cdm}h^2 = 0.1109$, $\omega_b = \Omega_b h^2 = 0.02258$, $n_s = 0.963$, $h = 0.71$, $\sigma_8 = 0.8$ and $w = -1.0$, where σ_8 is the normalization factor of the power spectrum and describes the standard deviation of the matter-density distribution in spheres of radius 8 Mpc/h. The product of this stage is an extensive extragalactic object catalog: cosmoDC2. A more detailed description of its generation and final products can be find in [35].

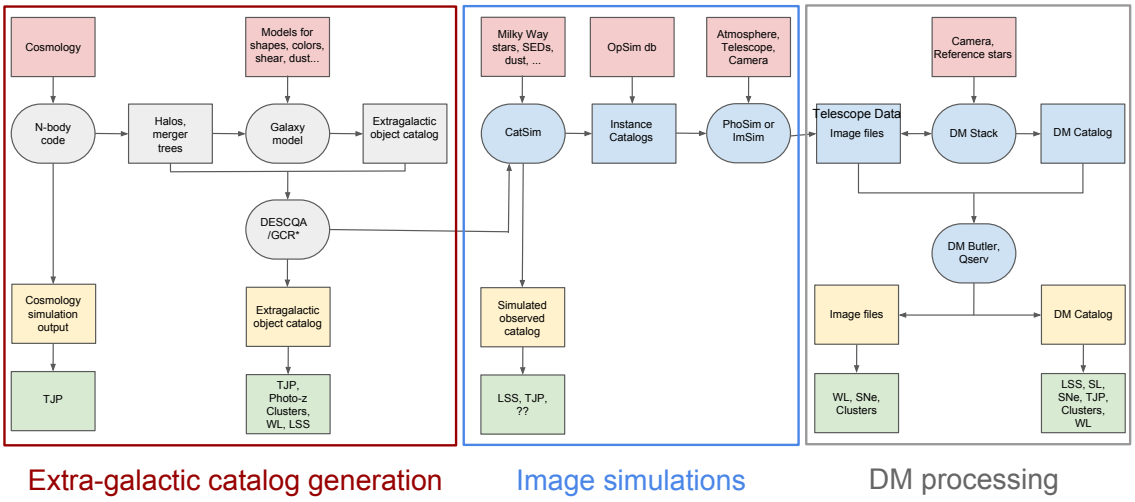


Figure 3.5: Overview of the DC2 Pipeline. The red blocks are external inputs consist of theoretical models, cosmological parameters and external data. The grey and blue blocks represent the main work flow and intermediate products. Final outputs are shown in yellow and green. [8]

Two N-body simulations involving only gravitational interactions were used to create the catalogs in this stage:

1. The "Outer Rim" simulation [36] serves as the basis of the final synthetic galaxy catalog. It covers $(4.225\ \text{Gpc})^3$ with $10{,}240^3$ particles, leading to a particle mass of $m_p = 2.6 \cdot 10^9 M_\odot$. Particle light cones were created so that weak lensing effects can be estimated using a ray-tracing code. The halo catalogs are generated based on the Friends-of-Friends (FOF) algorithm.

2. AlphaQ is a down-scaled simulation ($\sim$ 1600 times smaller) implemented with the Hardware/Hybrid Accelerated Cosmology Code (HACC [37]). It serves as a prototype that permits fast iterations of troubleshooting the codes essential to generate larger catalogs. It's also used to generate and assign complex galaxy properties to the final catalog.

After obtaining halo merger tree data from the simulations, two different methods are used to populate the synthetic galaxies:

1. The Empirical Model approach combines the merger-tree data with Monte Carlo re-sampling of galaxies from the UniverseMachine [38]. The UniverseMachine uses empirical model to predict the formation history of galaxies. Assuming galaxy formation is correlated to the development of the underlying dark matter halo, the empirical model uses a simple scaling relation between them, and fits the model parameters to observational data. The intermediate product is the baseDC2 catalog, which includes all simulated galaxies with limited properties, such as position, stellar mass, LSST colors and magnitudes.

2. The semi-analytic model (SAM) approach inputs merger-trees data from AlphaQ simulation to Galacticus SAM [39]. Base on the assembly history of each halo, the SAM approach adds baryon-specific processes to simulate galaxy formation, which is absent from the N-body gravity only simulations.

Next, the weak lensing maps are generated via a ray-tracing algorithm that simulates the bending of photons as they traverse the density field. The maps are used to assign weak lensing properties to each galaxy in baseDC2 catalog. Finally, the baseDC2 galaxies are matched to those in the Galacticus SAM library to complete the assignment of a realistic and complex set of properties.

3.6.2 Instance and truth catalogs

The next stage creates instance catalogs that serve as the input to image simulation. For each pointing of the telescope, the instance catalog returns the astrophysical sources that lie within its footprint. This concept, inspired by the image simulation tool PhoSim [40], permits the addition of time variability. The end product of this stage includes both instance catalogs and truth catalogs, achieved by the LSST software CatSim [41, 42]. For each object in the catalogs, a range of LSST specific attributes can be accessed, including position on the camera focal plane, apparent position on the celestial sphere, luminosity distance, reddening and extinction from Milky Way dust, LSST magnitudes/fluxes, shape parameters and uncertainty estimates. The instance catalogs consist of objects detected and identified after co-addition. The truth catalogs contain truth tables that describe the true time-averaged properties of objects.

3.6.3 Image simulations and processing

The image simulation tools process the data contained in the instance catalogs, and outputs pixel data expected to be obtained from the LSST camera. The image software imSim [2] is used for both Run 2 and 3 of DC2. The input of isSim includes object attributes from the instance catalogs, along with information about how the light is distorted before being observed, including lensing and extinction information. The systematic effects from the atmosphere on the observed PSF are also modeled. Run 2 covers the entire 300deg^2 footprint, and mainly focuses on static objects (galaxies without AGN, non-varying stars) in the wide-fast-deep (WFD) area that are the primary interests the static dark energy probes. On the other hand, Run 3 is designed for time-domain probes and adds time-varying objects (AGN, SNe) to the DDF. The WFD and DDF regions of each run is shown in Figure 3.6.

For image processing, the LSST Data Release Production (DRP) pipeline is consisted of four steps: single-frame processing, joint calibration, image co-addition and co-add processing. Readers can refer to [34] for a more detailed description of the workflow. For DC2 Run 2, where most visits happen during WFD observations, image processing is limited to the co-added images. DDF region and its vicinity are excluded due to the consideration of the cost of computational resources and rewards. The final results consist of objects catalogs required by the WL, LSS, and clustering probes, which added up to less than 2.5 TB.

[2] `https://github.com/LSSTDESC/imSim`

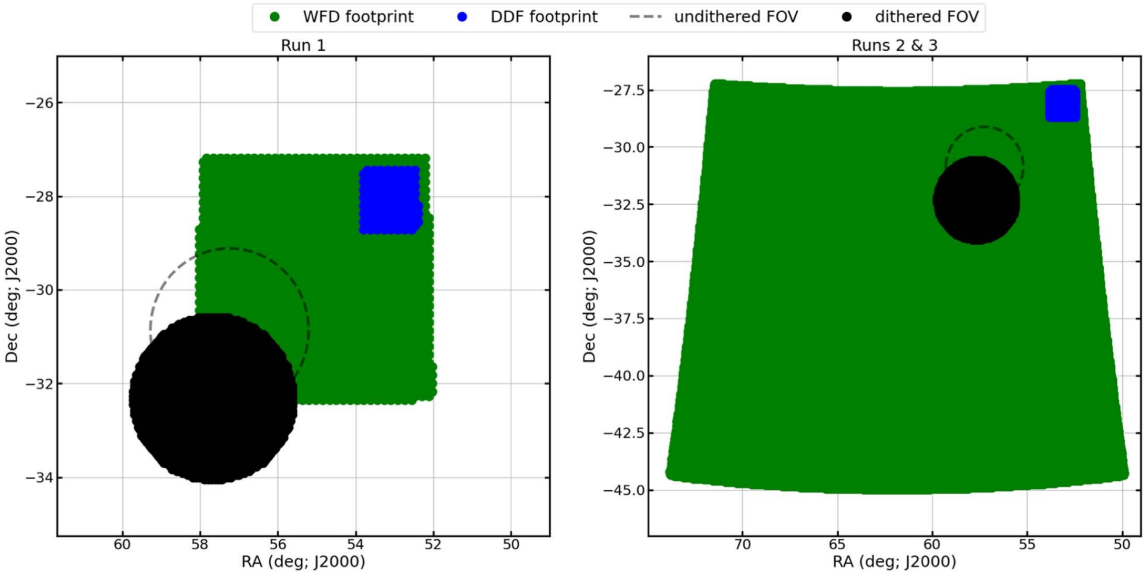


Figure 3.6: The footprint of DC2 simulations. Left: Run 1 footprint. Right: Run 2 and Run 3 footprints. DDF and WFD regions are shown in blue and green [8].

3.6.4 Data Releases

The final step in the workflow generates data product as defined in the LSST System Engineering Data Products Definition Document (DPDD) [3]. After image processing and validation, the catalog files are collected, merged, and translated into database tables for end users. These Object tables are stored in Apache Parquet format to support modern data access, and can be accessed directly or via the Generic Catalog Reader (GCR) [4] interface. As the main data access interface for end users, GCRCatalogs provides convenient methods to customize columns, filter data, and iterate over filtered data without using traditional databases. Each data table is registered in GCRCatalogs through YAML files containing the file paths. This approach allows highly customized data access through interface function

[3] `https://lse-163.lsst.io`
[4] `https://github.com/LSSTDESC/gcr-catalogs`

parameters without the need of custom data ingestion code or input file path that is often required for accessing traditional databases.

Chapter 4

LSS analysis with LSST

The Wide-Fast-Deep (WFD) region in the 2nd DESC data challenge (DC2) spans 300 $\deg^2$ and contains 5 years of survey data (data release 6 or DR6 in DESC) in 6 broadband filters ($ugrizy$). In this chapter, we present the Large Scale Structure (LSS) analysis of this data set. First, we describe the raw sample and the optimized selection criteria used to filter out the less ideal objects. Next, we introduce survey masks to define the geometry and handle associated systematic uncertainties. After that, we discuss the generation of survey property maps and the binning of the sample objects using photometric redshift. Finally, we present the calculation of the galaxy angular power spectra and discuss further corrections due to deprojection of survey property contamination.

Due to the amount of data involved and the availability of interface to access the data, the analysis is carried out on Cori, a KNL architecture-based supercomputer at the National Energy Research Scientific Computing Center (NERSC).

4.1 Selection cuts

In this work, we mainly use the `dc2_object_run2.2i_dr6` catalog (hereafter referred to as the *object catalog*); this is a static object catalog for Run 2.2i DR6 which can be accessed via GCRCatalog, the Python package that serves as a repository of various galaxy catalogs and sky catalogs for the LSST DESC. The first step is making selection cuts to produce the magnitude-limited sample which we describe in this section.

4.1.1 Basic quality cuts

In order to get a clean and well-behaved sample before starting the analysis, we impose several basic quality cuts to remove objects with poorly measured attributes such as fluxes, ellipticities, positions etc. Most of the quality cuts are given as flags (true or false values) of object attributes. In this section we describe the methods, mainly based on DC1 LSS analysis [43], used to determine which flags should be included in our selection.

Taking advantage of the simulated nature of our data and the knowledge about truth information, we first match the object catalog to the cosmoDC2/truth catalog, a large synthetic galaxy catalog designed to support precision dark energy science. We use the spatial and magnitude matching [43], hereafter denoted as SM matching, where for each observed object in the object catalog we find all the objects in the truth catalog within $0.6''$ radius and select the one that has the closest i-band cModel magnitude (`mag_i_cModel`) [1], which is a linear combination of the best fit exponential and de Vaucouleurs profile. If the i-band magnitude difference Δi between the observed and truth objects is less than 1, we consider a match has been found, otherwise the observed object is unmatched. This

[1] `https://classic.sdss.org/dr6/algorithms/photometry.php`

approach can be implemented with a `KDTree` or a Friends-of-Friends (FOF) matching algorithm with a linking strength of $0.6''$. Both algorithms are found to give similar matching results, in terms of the number of groups and the number of objects per group.

After the matching process we calculate the fraction of observed objects that are flagged in the unmatched population: $f_u = N_{flag,u}/N_{tot,u}$ (u stands for unmatched) and compare to the corresponding fraction in the matched population: $f_m = N_{flag,m}/N_{tot,m}$ (m stands for matched). If the ratio $f_u/f_m > 20$ while $f_m < 0.01$, then we include the flag in our quality cuts. Roughly speaking, these criteria mean that a much higher fraction of the unmatched/badly-observed objects are eliminated compared with matched/well-observed objects (20:1), and most of the matched/well-observed objects are retained (99%).

In addition to the selection flags, we also apply cuts on the Signal-to-Noise ratio (SNR). We require the i-band SNR to be greater than 6 (`snr_i_cModel`> 6) since we are also imposing magnitude cuts in i-band, and `snr_{b}_cModel`> 3 in at least 2 of the other 5 photometric bands were `b` $\in \{u, g, r, z, y\}$.

In conclusion, we select a clean sample of all observed objects using the cuts listed in Table 4.1. After all standard quality cuts, about 93 million objects are selected from a total of 147 million objects (a 63% selection rate) in the *object catalog*.

4.1.2 Star/galaxy classification

For the galaxy clustering analyses, we are only interested in the spatial distribution of galaxies in the sample. Therefore, it is important that we can efficiently separate stars from galaxies while attaining maximum survey depth. In this analysis, the classifier

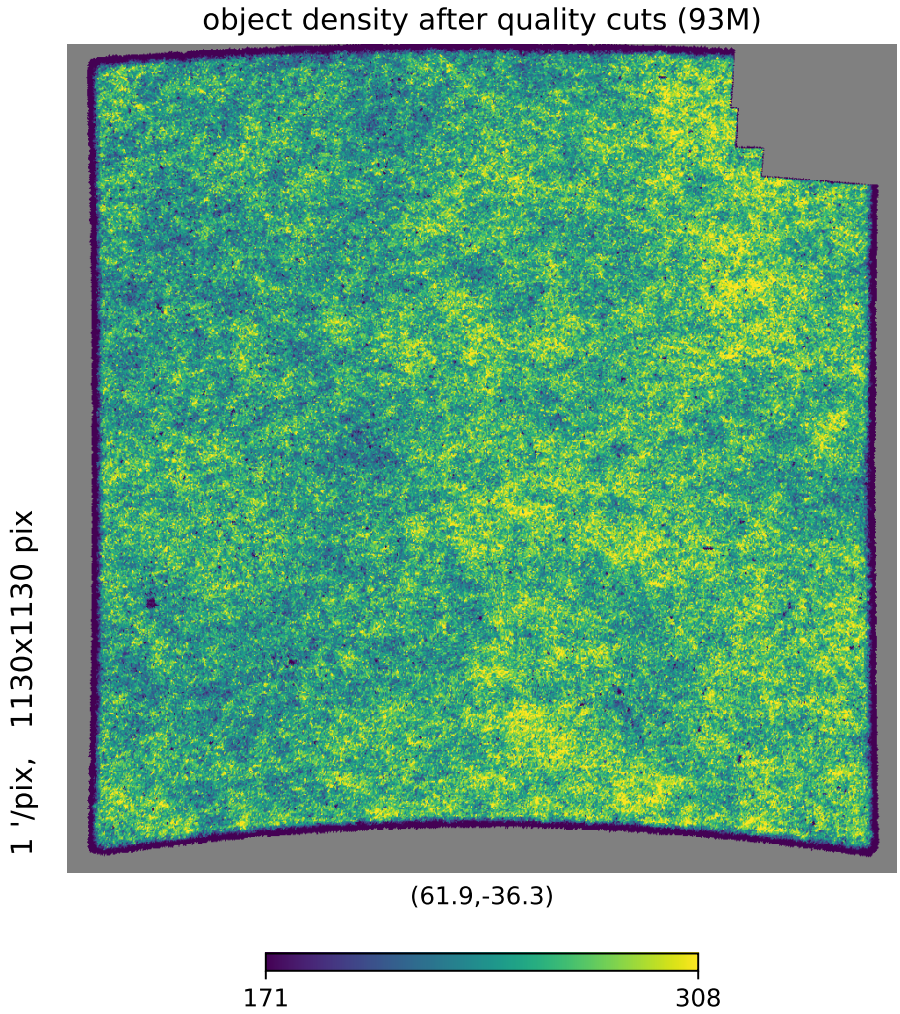


Figure 4.1: Map of number of objects per pixel after quality cuts. The map is in HEALPix format with Nside = 2048.

Quality Cuts	
Cuts	**Description**
detect_isPrimary modelfit_CModel_flag_badCentroid=False base_SdssCentroid_flag=False base_PixelFlags_flag_edge=False base_PixelFlags_flag_interpolatedCenter=False base_PixelFlags_flag_saturatedCenter=False base_PixelFlags_flag_bad=False base_PixelFlags_flag_suspectCenter=False deblend_skipped=False base_PsfFlux_flag=False base_SdssShape_flag_psf=False modelfit_DoubleShapeletPsfApprox_flag=False base_ClassificationExtendedness_flag=False	Quality cuts, eliminating objects with bad center, deblending, PSF shape, PSF flux, extendedness etc.
[ugrizy]_base_PsfFlux_flag=False [gri]_base_ClassificationExtendedness_flag=False [ugrizy]_base_PixelFlags_flag_edge=False [ugrizy]_base_PixelFlags_flag_saturatedCenter=False	Quality cuts per photometric band
base_Blendedness_abs<=0.42169650342	Blendedness cut
snr_i_cModel > 6	SNR cut in i-band
snr_[ugrzy]_cModel > 3	SNR cuts (required in at least 2 bands)

Table 4.1: Basic quality cuts to select a clean sample.

used for star/galaxy separation is called *extendedness*, which is defined based on the difference between the PSF magnitude and the CModel magnitude. The value of *extendedness* is usually zero for stars and notably different from zero for galaxies, since the CModel mag-

nitude is an adequate proxy as a universal magnitude for all types objects and the PSF models point source or isolated stars very well. In the object catalog, this quantity is converted to a flag `base_ClassificationExtendedness` or `extendedness` via the condition: $m_{\mathrm{PSF}} - m_{\mathrm{CModel}} = 0.0164$ [44]. The flag `extendedness` has a value of 1 if an object is classified as a galaxy. Otherwise, if an object has `extendedness` $= 0$, then it is classified as a star. Here we study the performance of this classifier as a function of i-band magnitude and choose our magnitude selection cuts based on the results.

We utilize the truth information and first match the object catalog to the input catalog using the SM matching method. Then we use the following metrics from [45] to determine the quality of the classifier, namely completeness (TPR) and purity (PPV):

$$\text{Completeness } (TPR) = \frac{TP}{TP + FN} \tag{4.1}$$

$$\text{Purity } (PPV) = \frac{TP}{TP + FP} \tag{4.2}$$

Here TP stands for "True Positive", FP stands for "False Positive" and FN stands for "False negative". In our case, when considering galaxies, TP is the number of objects classified and matched to galaxies, FP is the number of objects classified as galaxies but matched to stars and FN is the number of objects classified as star but matched to galaxies. The completeness is the fraction of true galaxies that are selected, while the purity is the measure of the contamination of the sample by misclassified objects. Similar definitions can be made for stars.

The results are shown in Figure 4.2. Using `extendedness` as the star/galaxy classifier, we see that the completeness of galaxies starts to fall to around 80% at `mag_i_cModel` $\sim$ 25. Above `mag_i_cModel` $= 25$, the efficiency for selecting galaxies drops off significantly,

and we start to lose galaxies in the selected sample. The purity of the galaxies remains high across the whole range of magnitude. Therefore, we choose

$$17 < \texttt{mag\_i\_cModel} < 25.3 \tag{4.3}$$

as the magnitude selection cut. Combining with the galaxy selection flag `extendedness` == 1, we obtain the selected galaxy sample of 51 million objects with only 1.1% impurity from stars, which is sufficient for our purpose. A completeness of $\sim 80\%$ around high i-band magnitude and a overall completeness of 95% indicate that the selected sample is a good representation of the true galaxy population. We don't lose too much sample size from the more strict magnitude cut. Hereafter we refer to the sample after the quality cuts and magnitude selection cut as the *LSS sample*.

4.2 Survey masks

LSST, being a large scale photometric survey, will image the entire available southern hemisphere sky, about $18,000$ square degrees, once every few nights in the course of its 10-year survey duration. This results in non-uniform survey conditions across the survey area and a complicated survey geometry. For example, some area of the sky could be affected by bright objects (stars, satellites etc) which might result in poor object detection in the region surrounding such bright objects, and some patches of the sky might see less exposure than others, which results in a shallower depth. We utilize survey masks to eliminate the regions affected by these phenomena and describe the survey geometry quantitatively. Essentially, these masks are sky maps that carry information as to whether or not to include

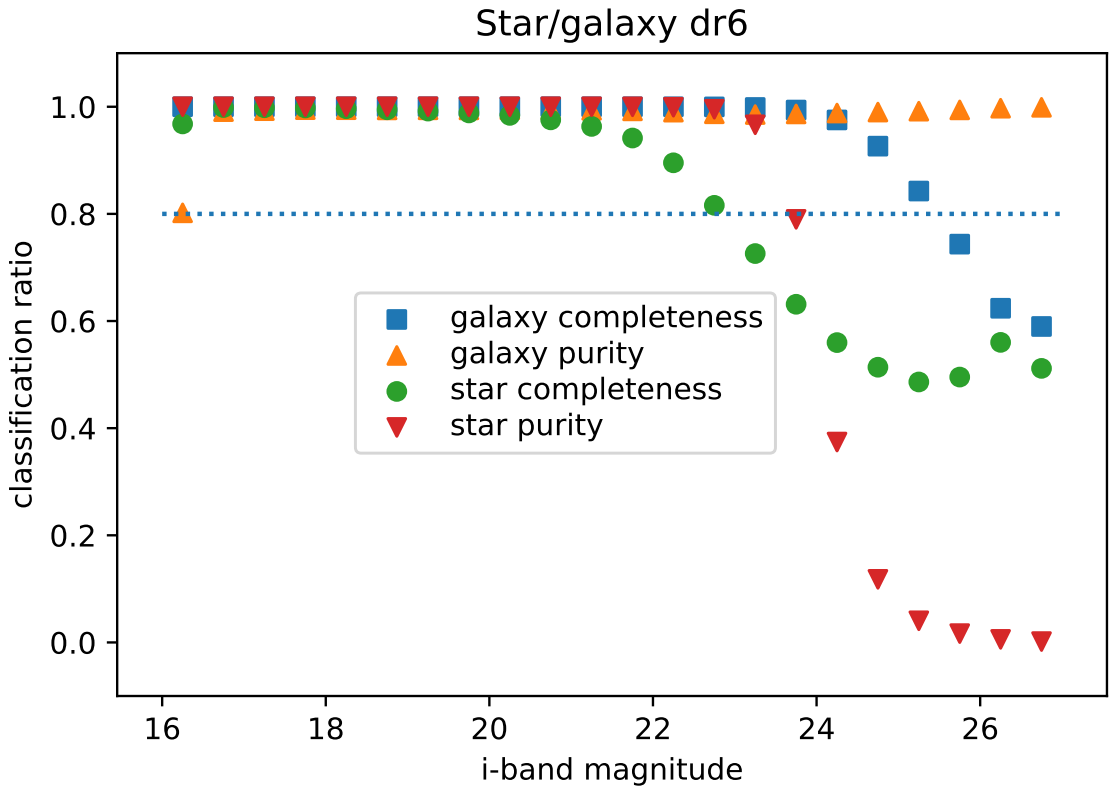


Figure 4.2: Various metrics for testing the performance of `extendedness` plotted as a function of i-band magnitude. The blue squares and orange triangles represent the completeness (TPR) and purity (PPV) of galaxies, respectively. The green circles and red triangles represent the completeness (TPR) and purity (PPV) of stars, respectively.

certain areas in the analysis. Specifically, we make (i) depth masks which eliminate areas that have lower depth than the minimum requirement and (ii) bright-object masks which mask areas around bright objects like stars. In this section, we describe the generation of these masks and their effects on the survey footprint.

The maps and masks generated in the analysis are mostly made in the HEALPix [2] format. HEALPix stands for Hierarchical Equal Area isoLatitude Pixelation of a sphere. It's a pixelation scheme that produces a partition of the spherical surface, where every pixel covers the same surface area. The sphere is hierarchically divided into curvilinear quadrilaterals. The lowest resolution mapping consists of 12 base pixels. This configuration corresponds to `Nside` = 1, a parameter that describes the size of the pixels and can only have values of powers of 2. Each time `Nside` doubles, every pixel is divided into four new ones with equal area, increasing the resolution ($\texttt{Npix} = 12\texttt{Nside}^2$).

4.2.1 Depth mask

Due to the different observing conditions between each exposure, the depth varies across the survey field. To quantify these variations, we first generate depth maps in HEALPix format with `Nside` = 1024, giving a total number of pixels of $12\texttt{Nside}^2 \approx 12.6$ megapixels. We compare three different methods to estimate the 5σ survey depth as a function of angular position:

1. Using magnitude of stars [43]: after quality cuts, for each pixel we calculate the mean i-band magnitude for objects classified as stars that fall in it and with signal-to-noise ratios between 4 and 6. Compared to the following methods, this method leads to

[2] `https://healpix.sourceforge.io`

higher correlations between generated depth maps and the galaxy number density map.

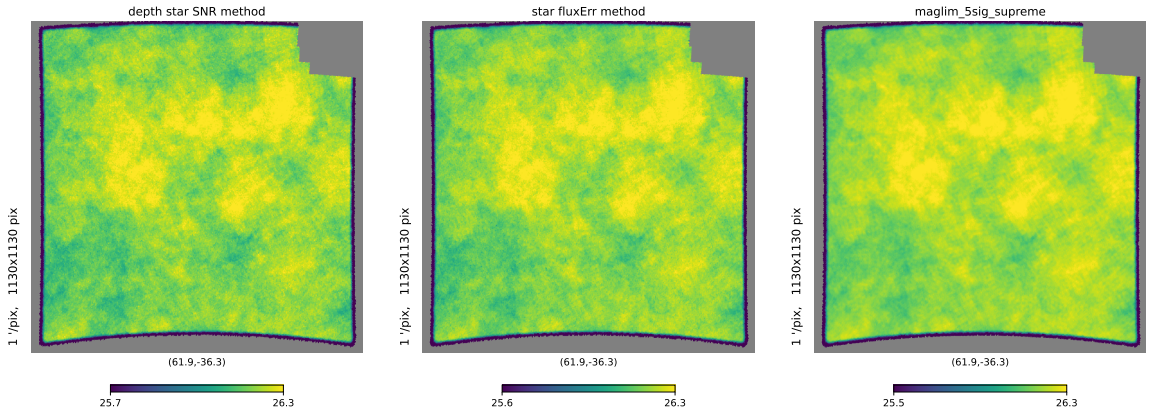
2. Using flux error of stars [46]: we calculate the mean i-band cModel flux error of all objects classified as stars in each pixel, then multiply the value by 5 and transform it into a magnitude.

3. From observating conditions: we use `i_maglim_psf_wmean` maps generated by *supreme* [3], which uses *healsparse*[4] to make survey maps for LSST. Essentially, these maps are created by multiplying noise (dominated by sky noise) and `psf_area` values in each pixel, taking the square root and convert 5 times of its values to magnitudes.

The maps are shown in Figure 4.3. We see that they are visually consistent to one another. To demonstrate this quantitatively, for each pair of maps, we subtract one map from the other pixel-wise, then make a histogram of the pixel differences. For example, we can consider the histogram of the map differences generated by methods 2 and 3: about 86% of the pixels have a difference in depth less than 0.25. The result is shown in Figure 4.4. Since all 3 maps agree with others well, we pick the depth maps generated by method 3 in our analysis.

To remove regions with shallower survey depth, the depth masks are generated by eliminating all the pixels in the depth maps that have a limiting magnitude $i \geq 25.3$. In other words, we only use objects that reside in cells with i-band 5σ-depth less than 25.3. This removes $\sim 3\%$ of the area from the galaxy sample footprint.

[3] `https://github.com/LSSTDESC/supreme`
[4] `https://github.com/LSSTDESC/healsparse`



(a) Method 1 (b) Method 2 (c) Method 3

Figure 4.3: Depth maps generated using 3 different methods.

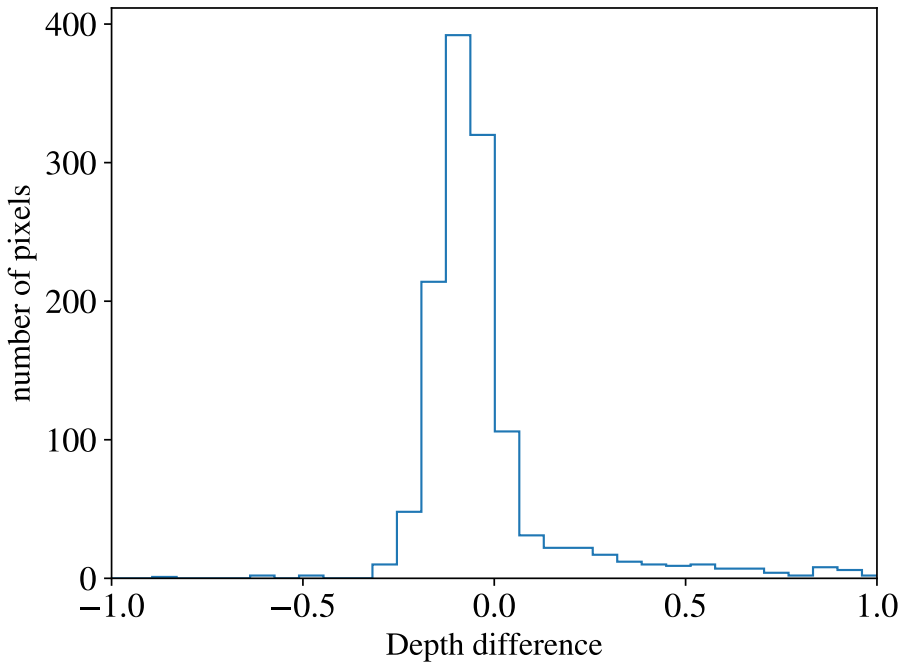


Figure 4.4: Histogram of the per-pixel differences of depth maps generated by methods 2 and 3.

4.2.2 Bright object mask

The measurement and detection of any object can be affected significantly if there are bright objects in its surroundings. One of the effects from the presence of a bright object is obscuring the neighboring sources which results in a reduction of observed objects around it. To eliminate this effect, we can mask the area around bright objects even at the cost of creating a more complex geometry. Again, we use survey masks to describe this geometry. Only stars (i.e. not galaxies) are considered as bright objects to avoid possible biases towards dimmer galaxies.

For each identified bright star (`mag_i_cModel` < 22 and `extendedness` $= 0$), we exclude a circular region ("disk") around it from our analysis. To determine what disk radius should be used, we follow the method detailed in [47]:

1. We divide the bright stars according to their i-band cModel magnitudes into 5 bins: $i < 17$, $17 < i < 18$, $18 < i < 20$, $20 < i < 22$, $22 < i < 24$.

2. For the bright stars in each bin, we count the number of objects in the *LSS sample* that lie within a certain distance θ to any of the bright stars and calculate the object density $N_{\mathrm{disk}}(\theta)$ in that region. We compute this value for a set of radii.

3. We calculate the object density N_{dens} in the *LSS sample* over the full survey footprint and then compute the ratio $N_{\mathrm{disk}}(\theta)/N_{\mathrm{dens}}$ for each bin. For each magnitude bin i, we identify the radius θ_i where the ratio $N_{\mathrm{disk}}(\theta)_i/N_{\mathrm{dens}} = 0.9$ and choose this as the disk radius for the bright objects in bin i.

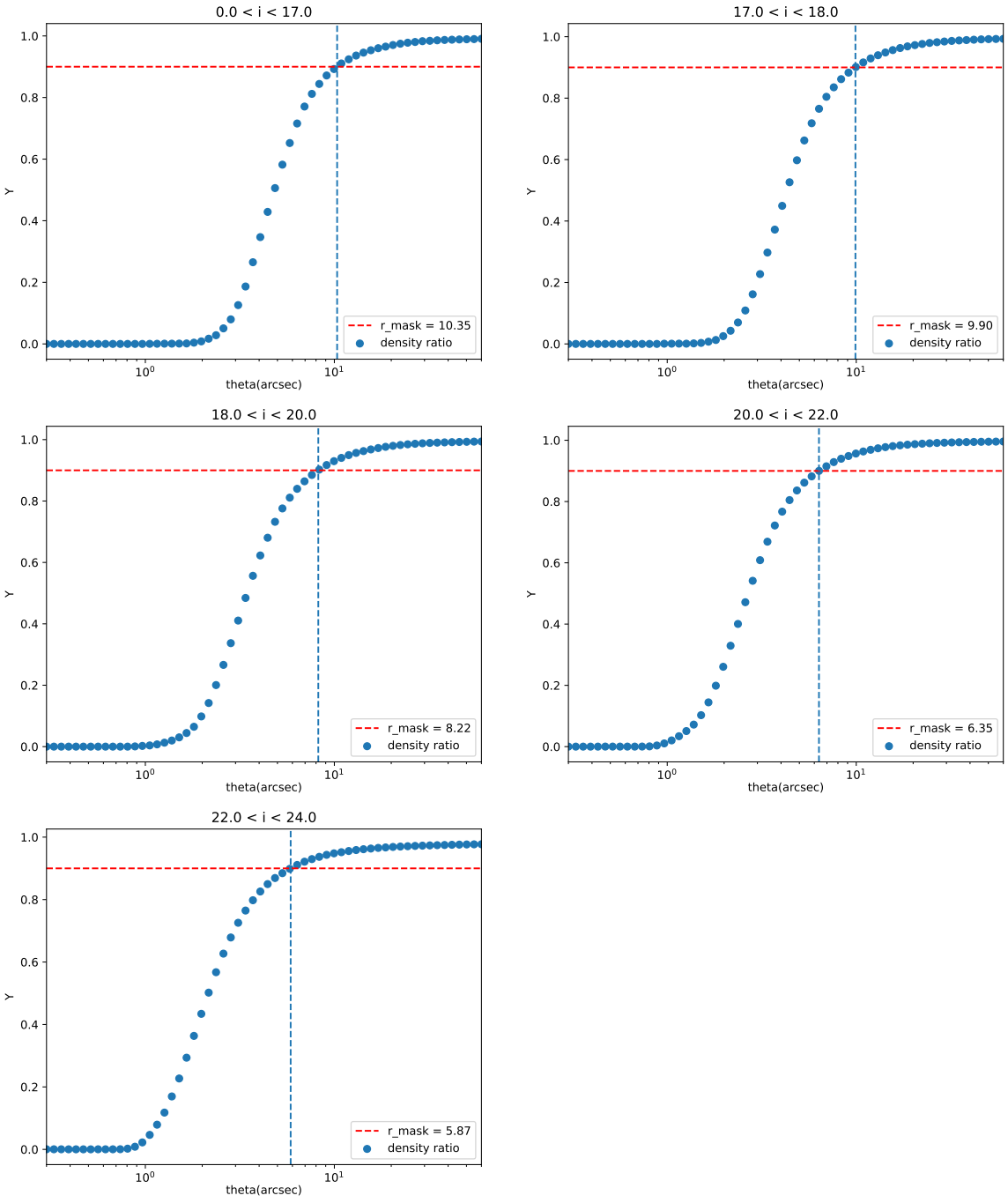


Figure 4.5: Plots of $N_{\rm disk}(\theta)/N_{\rm dens}$ as a function of the disc radius θ (dots) for each i-band magnitude bin, with the radii chosen for the discs based on the condition $N_{\rm disk}(\theta)_i/N_{\rm dens} = 0.9$.

The plots of $N_{\rm disk}(\theta)/N_{\rm dens}$ and the radii chosen for each magnitude bin are shown in Figure 4.5. After choosing the radii for the disks, we can create the bright object mask. First, we make a blank high resolution HEALPix map (`Nside` $= 2^{16} = 65,536$) and mask out all pixels whose centers lie within the disks; we also remove all objects within these pixels from our *LSS sample*. Then we downgrade this map to a lower resolution (`Nside` $= 1024$) by an averaging procedure and eliminate all pixels that have more than 13% of their area masked. This results in a loss of 10% of the total survey area.

4.3 Survey property maps and systematic uncertainties

Apart from the effects from bright objects and variations of survey depth across the field, there are other observational systematic uncertainties that could bias the observed object number density. In order to investigate and mitigate these systematic effects, we first generate HEALPix maps of survey properties or observing conditions using *supreme*. The procedure to generate the maps can be summarized as following: first we accumulate the per-frame metadata for each exposure, then we calculate the weighted mean values of each pixel using the given weights of the CCDs, finally we repeat these steps for each observing condition.

For each photometric band, we map the following observing conditions that could bias our galaxy number density:

1. **PSF size** (`psf_size`) Due to the blurring effect of the atmosphere, a point source can appear as a blob/disk in the image captured by telescopes. Astronomers refer to this distortion as seeing. To quantify the seeing effect, we estimate the Full-Width-

Half-Maximum (FWHM) of the point spread function (psf) of point sources, or "PSF size". A large seeing can affect the star-galaxy classification of objects by increasing their extendedness.

2. **Airmass** (`airmass`) We consider the *relative airmass*, which is the optical distance at the line of sight divided by the optical distance at zenith. The formula in [Hardie (1962)] is used for the calculation of airmass:

$$\text{airmass} = \sec(z) - 0.0018167(\sec(z)-1) - 0.002875(\sec(z)-1)^2 - 0.0008083(\sec(z)-1)^3 \tag{4.4}$$

where z is the zenith angle of the line of sight. A larger airmass corresponds to a thicker layer air that the light ray needs to travel through before being detected by the telescope, causing the extinction of stellar light. Therefore, airmass can affect the observed magnitudes and detection of objects, which in turn biases the object number density.

3. **Sky background** (`bgmean`) Sky background refers to the mean brightness of a patch of sky after the removal of all direct light sources (stars, cosmic rays, etc). Similar to large airmasses, a large sky background can affect the photometry and detection of faint objects.

4. **PSF magnitude limit** (`maglim_psf`) This is the 5-σ magnitude limit. We convert the product of sky-noise and psf area into a flux, multiply this value by 5 then convert it to a magnitude. Specifically, the following formula is used to calculate the 5-σ

magnitude limit:

$$\text{maglim} = zp - 2.5\log_{10}\left(5\sqrt{\frac{A_{\text{psf}}}{W}}\right) \tag{4.5}$$

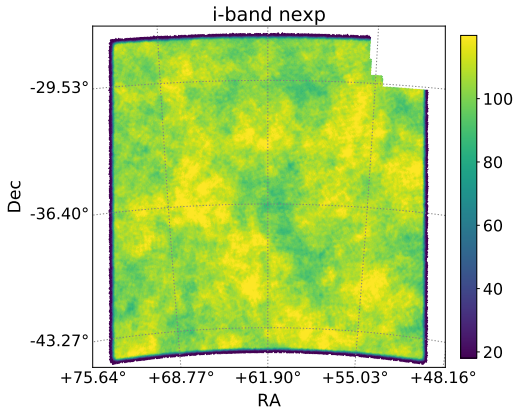
where zp is the zero point of the detector, A_{psf} is the psf area and W is the CCD weight. A low magnitude limit indicates a high total noise in the psf area and can affect the detection of faint objects.

5. **Number of exposures** (`nexp`): This gives the number of times each pixel in the sky map is visited. For each pixel in the HEALPix map, we calculate the number of exposures that overlap with it. A higher number of visits can increase the signal to noise ratio of the detected objects.

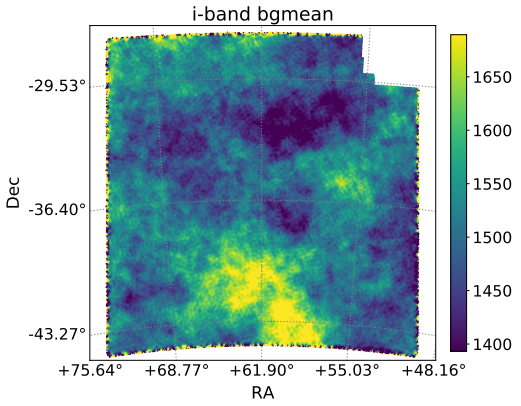
Figure 4.6 shows plots of $i-$band survey property maps made by *supreme*.

4.4 Photometric redshift

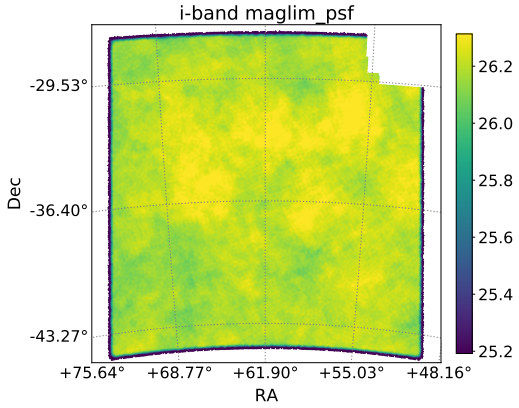
As well as angular coordinates of the observed galaxies, the analysis requires the distance/redshift information of each object. Since LSST is a photometric survey, it stores the spectral information as fluxes in 6 broad optical bands (*ugrizy*), rather than using more precise spectroscopic measurements. The estimation of an object's redshift from these broad band fluxes, known as the "photometric redshift" (photo-z), is extensively used in modern observational cosmology. The redshift in characteristic features (e.g. Lyman and Balmer breaks) of the galaxy spectral energy distributions (SED's) can be reflected as color changes in the broad bands and are often the primary sources of such estimations.



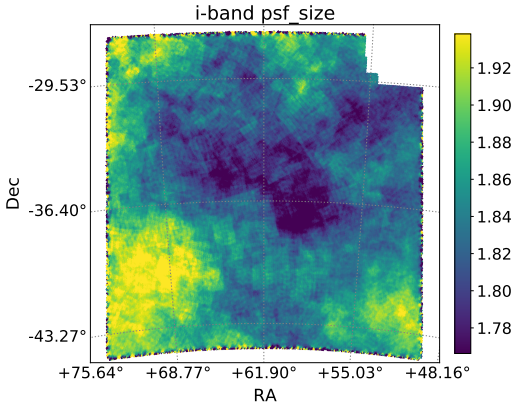
(a) Number of exposures



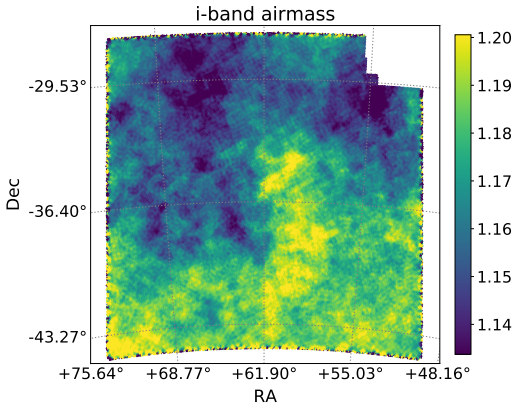
(b) Background mean



(c) PSF magnitude limit



(d) PSF size



(e) Airmass

Figure 4.6: Survey property maps for the *i*-band. The number of exposures map is the sum over all survey visits, while the other maps show the weighted average over all survey visits.

In order to determine the photometric redshift of each galaxy, the LSST colaboration has developed two codes based on photo-z algorithms: a Bayesian Photometric Redshift (BPZ) code called `BPZPipe`, which is an SED template-based method that employs a prior; and a machine learning code called `FlexZPipe` that performs a conditional density estimate; both were trained with a representative data sample complete to $i < 25.0$. More detailed descriptions of these methods can be found in [48, 49]. In this analysis we will use the photometric redshifts generated by the `BPZPipe` code.

4.4.1 Tomographic samples

We divide the *LSS sample* into 6 tomographic samples of equal redshift intervals using the `photoz_mode` redshift estimator as markers. The `photoz_mode` is the mode of the posterior probability distributions for each galaxy computed by the `BPZPipe` photometric redshift code. The bin ranges for the intervals are $z = 0.2-0.4$, $0.4-0.6$, $0.6-0.8$, $0.8-1.0$, $1.0-1.2$ and $1.2-1.4$.

4.4.2 Redshift distributions

The redshift distributions $p^i(z)$ of the tomographic samples play a crucial role in the theory used to compute the angular power spectra and constrain the cosmological parameters, so it's important that we make accurate estimation of them. Here we compare two methods used for the estimation.

We can again benefit from the fact that the data is generated through simulation and get the redshift truth. We match our sample against the truth catalog using the SM matching method, then use the redshifts of the matched objects in the truth catalog as

the true redshifts for each selected galaxy. The $p^i(z)$ of each tomographic bin can then be estimated by making a histogram of these true redshifts.

We also test an alternative to estimate the redshift distributions by "stacking" the posterior probability distributions, a method that is useful for real survey data where direct access to the true redshifts isn't available. In this approach, to produce the redshift distribution estimate, we add the `photoz_pdf` of all objects in each tomographic sample. Here `photoz_pdf` is the posterior probability distribution (pdf) for individual galaxies computed on a redshift grid by the photo-z codes. Even though this approach to estimate the $p^i(z)$ is not mathematically sound, we can compare the distributions with those made from using true redshifts and investigate the uncertainties of $p^i(z)$'s produced with this approach.

As shown in Figure 4.7, these two methods produce different redshift distributions for each tomographic bin. While visually the center locations of the $p^i(z)$s are consistent between the two methods, the redshift distributions produced by "stacking" are narrower, taller, and have less overlaps between adjacent bins, compared to those produced by the true redshifts.

4.4.3 Photometric redshift performance

To evaluate the photometric performance of LSST, we use several figures of merit based on previous simulation studies [30]. They are:

1. The bias (median) in $e_z = (z_{photo} - z_{spec})/(1 + z_{spec})$. Figure 4.8 shows e_z vs. photo-z and the bias in e_z is plotted in figure 4.9. We find the bias in e_z is below 0.05 for

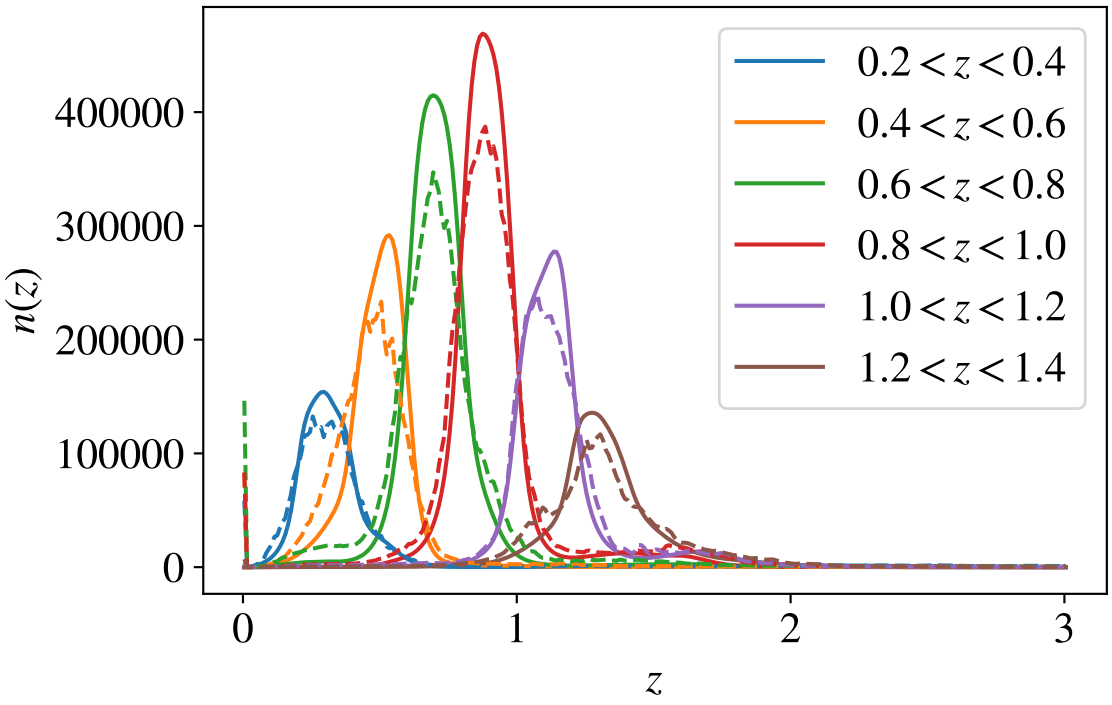


Figure 4.7: Photometric redshift distributions of the *LSS sample* objects in each tomographic bin. The solid curves are estimations of $p(z)$'s made by stacking the photo-z pdf's. The dashed curves are the histograms of the redshifts of the matched objects in the truth catalog.

the range of photo-z used in this analysis. The requirement from the Science Book is 0.003, which is met for part of the range in photo-z.

2. The root-mean-square scatter of e_z in photometric redshifts, $\sigma_F = \sigma_z/(1+z)$. This is shown in figure 4.10 and shows values of less than approximately 0.1 in the photo-z range of interest, somewhat higher than the requirement of 0.05.

3. The fraction of 3σ outliers. This quantity, shown in figure 4.11, is found to be below 10%, which is the requirement for this metric.

For the purpose of the analysis, we only check if these requirements are met for the *LSS sample* in the range of $0 < z < 1.5$, which contains the redshifts of all of the tomographic samples. The photometric redshift performance based on these figures of merit is found to close to the requirements quoted in the LSST Science Book.

4.5 Angular power spectra

4.5.1 Algorithm

In this section we outline the method used to calculate the angular power spectrum. For simplicity, we introduce equations and definitions in the context of auto power spectra in this subsection. In case of cross power spectra, they can be derived similarly by considering a pair of fields.

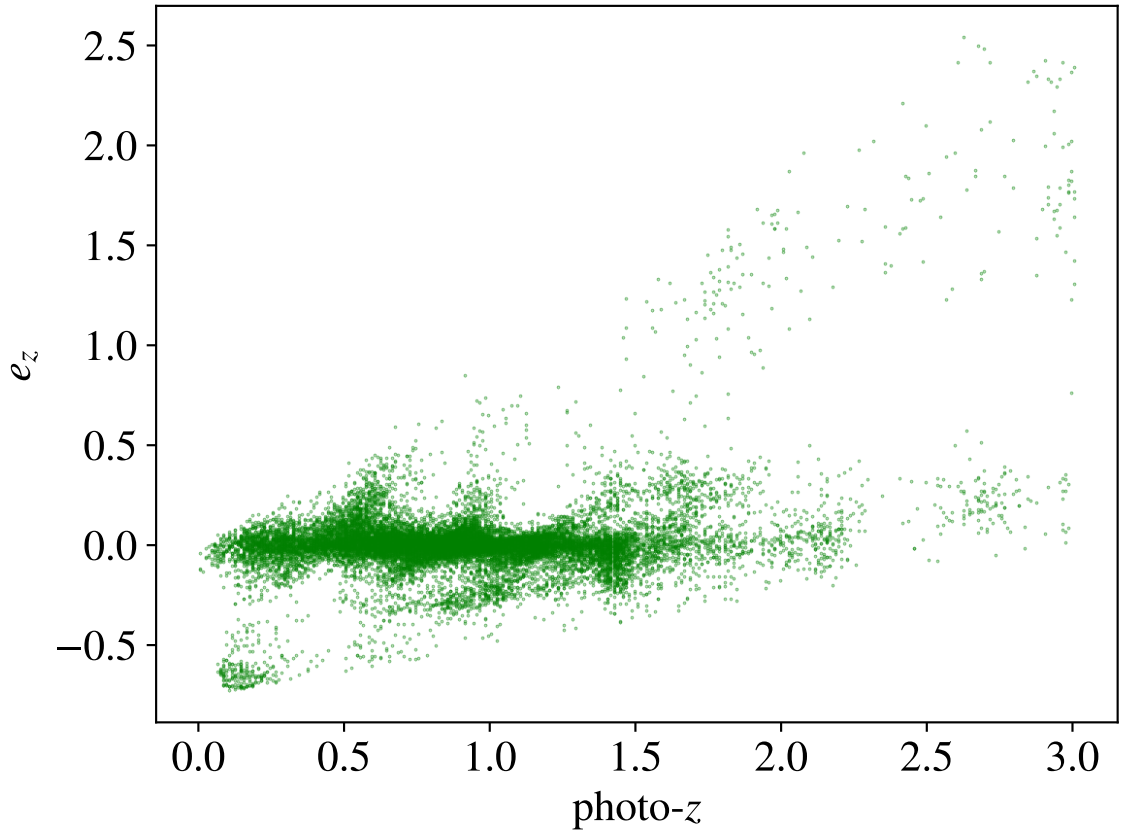


Figure 4.8: Scatter plot of e_z vs. photometric redshift.

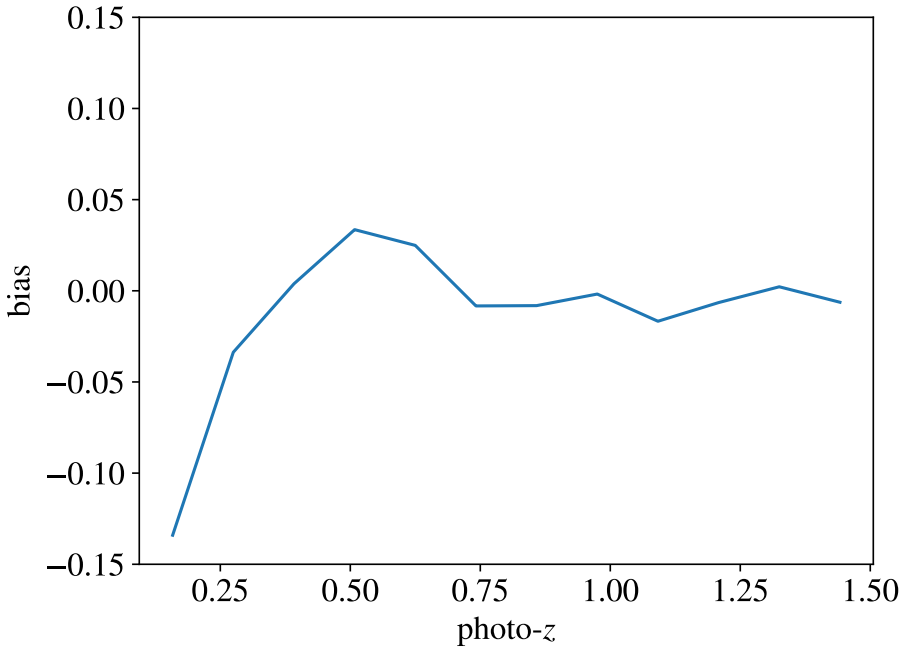


Figure 4.9: The median value of e_z (bias) vs. photometric redshift.

Given the information of the full sky overdensity, we can expand the overdensity in spherical harmonics:

$$\delta(\vec{\theta}) = \delta(\theta, \phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi) \tag{4.6}$$

with harmonic coefficients:

$$a_{\ell m} = \int d\Omega \, Y^*_{\ell m}(\theta, \phi) \delta(\vec{\theta}) \tag{4.7}$$

The angular power spectrum is defined using:

$$\langle a^*_{\ell m} a^*_{\ell' m'} \rangle \equiv \delta_{\ell \ell'} \delta_{m m'} C_\ell \tag{4.8}$$

An unbiased estimatior of the angular power spectrum can be given by:

$$C_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{l} |a_{\ell m}|^2 \tag{4.9}$$

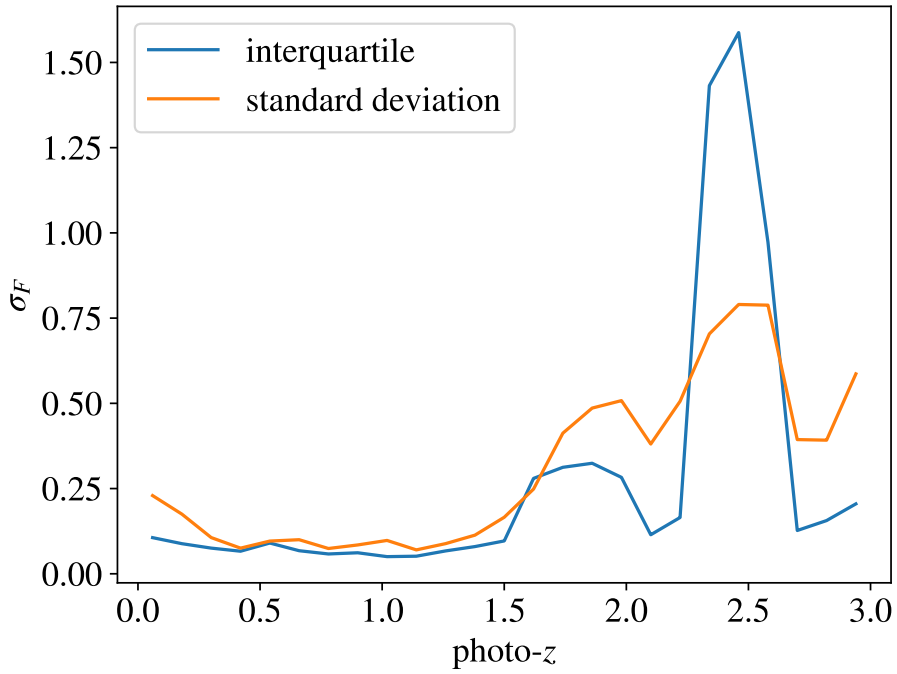


Figure 4.10: The rms scatter (determined from the interquartile range) $\sigma_F = \sigma_z/(1+z)$ of e_z as a function of photometric redshift.

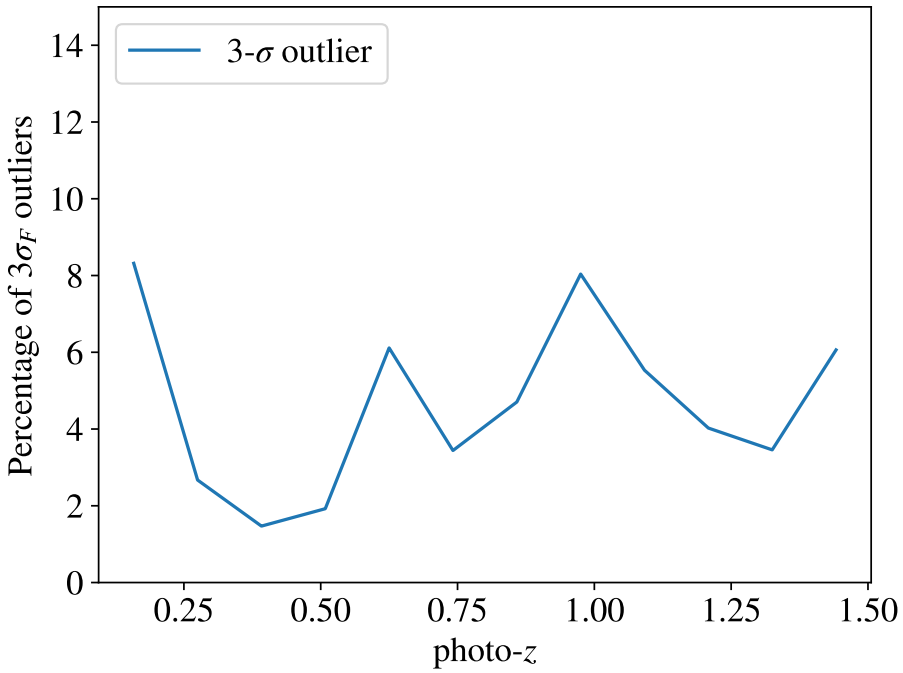


Figure 4.11: The fraction of 3σ outliers as a function of redshift.

In practice, overdensity maps cannot be measured over the full-sky due to the restriction from ground based surveys or galactic contamination for instance. The partial coverage of the sky makes different ℓ modes coupled, and the estimator becomes biased. Furthermore, a position dependent weight $w(\theta)$ (mask) is often applied to the data to correct observational uncertainties, edge effects or simply to remove unobserved pixels. In this work, the power spectra are calculated using `NaMaster` [50], a software package based on the pseduo-C_ℓ algorithm [51, 52], which computes the bias of the estimator and makes corrections accordingly. The algorithm can be outlined as following:

1. The measured data is converted to a overdensity map in HEALPix: using $\delta_i = N_i/(\bar{N}w_i) - 1$, where N_i and w_i are the number of objects and weight of the i-th

pixel, respectively; $\bar{N} = \langle N_i/w_i \rangle$ is the weighted average of the number of objects per pixel.

2. The pseduo-C_ℓ of the partial sky map is then computed using Equation (4.9) with harmonic coefficients:

$$\tilde{a}_{\ell m} = \int d\,\Omega Y^*_{\ell m}(\theta,\phi) w(\vec{\theta}) \delta(\vec{\theta}) \approx \Omega_p \sum_p Y^*_{\ell m}(p) w(p) \delta(p) \tag{4.10}$$

where the integral is approximated with a discrete Fourier transform - a sum over all pixels p with an identical solid angle Ω_p.

3. We can express the ensemble averaged pseudo angular power spectrum measured within the partial sky as a linear combination of the ensemble averaged angular power spectrum over the full sky through the mode-coupling matrix:

$$\left\langle \tilde{C}_\ell \right\rangle = \sum_{\ell'} M_{\ell\ell'} \left\langle C_{\ell'} \right\rangle \tag{4.11}$$

where $M_{\ell\ell'}$ is the matrix that reflects the mode-mode coupling of the partial sky and only depends on the geometry of the partial sky map, and is used to obtain a unbiased estimate of $\langle C_{\ell'} \rangle$ - the measured power spectrum. However, in most cases, it is not possible to invert the mode-coupling matrix directly and uncover the unbiased C_ℓ's. To reduce the dimensions of the coupling-matrix and make it invertible, a set of binning in ℓ modes (bandpowers) is introduced, and the computed power spectrum becomes a step function that is a constant in each binning interval. The binning also reduces the correlation and error in the measured C_ℓ's.

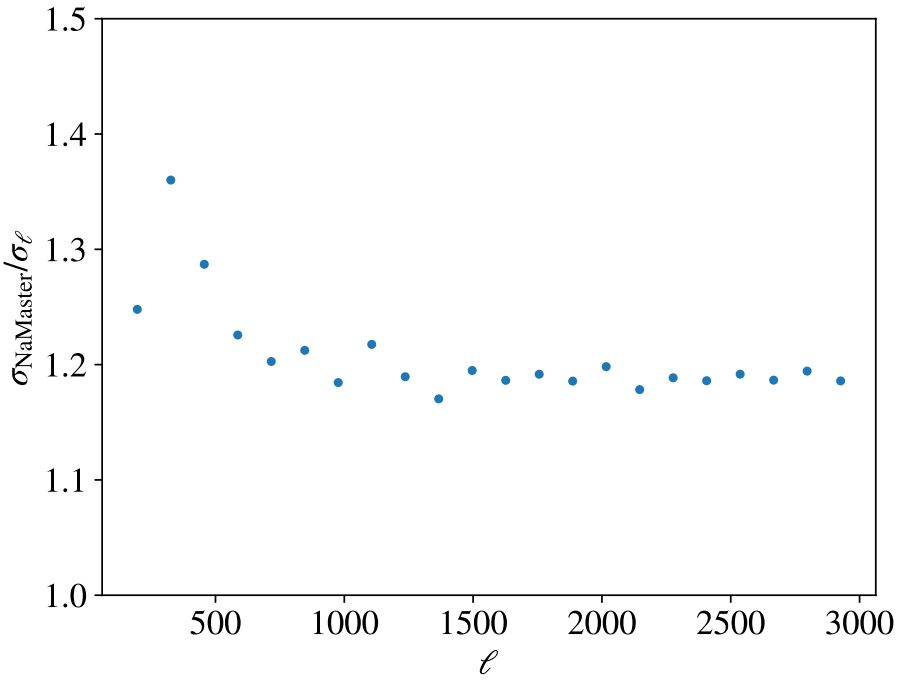


Figure 4.12: The Gaussian covariance error computed analytically by `NaMaster` divided by the approximation $\sigma_\ell = \sqrt{2/((2\ell+1)f_{sky})}C_\ell$ in the third redshift bin.

4. The Gaussian covariance matrices of the measured C_ℓ's are approximated analytically (see [46, 53, 54]), using the covariance of a Gaussian random field with the same power spectrum as the measured data.

A comparison between the angular power spectrum Gaussian covariance error obtained from `NaMaster` and the theoretical approximation $\sigma_\ell = \sqrt{2/((2\ell+1)f_{sky})}C_\ell$ [55, 56] is depicted in Figure 4.12.

4.5.2 Theoretical prediction and fitting

To compute the theoretical angular power spectra, we use the Core Cosmology Library (CCL) [57] developed by DESC.

First, we create a cosmology with the CosmoDC2 parameter set, given by $\omega_{cdm} = 0.1109$, $\omega_b = 0.02258$, $n_s = 0.963$, $h = 0.71$, $\sigma_8 = 0.8$, and $w = -1.0$, with galaxy bias set to a constant that's independent of redshift. The 3D power spectrum is then computed using the non-linear HALOFIT code. Next, the number count tracers of each bin are generated using the distribution of true redshifts of the sample, with redshift space distortion added. Finally, the theoretical angular C_ℓ's are computed with Limber's approximation [58], which can be expressed as the following equation:

$$C_\ell^{th} = \frac{2}{2\ell+1} \int dz \left(\frac{dn}{dz}\right)^2 b^2(z) H^2(z) P(k = \frac{\ell+1}{r}, z) \tag{4.12}$$

where dn/dz is the distribution of true redshifts, $b(z)$ is the galaxy bias and $P(k, z)$ is the theoretical 3D angular power spectrum.

We then fit the measured C_ℓ's against the theoretical power spectra by minimizing the χ^2 function and varying only the galaxy bias parameters in the theoretical power spectra:

$$\chi^2 = \sum_{\ell,\ell'} (C_\ell^{meas} - C_\ell^{th}) \mathrm{Cov}^{-1}(C_\ell, C_{\ell'}) (C_{\ell'}^{meas} - C_{\ell'}^{th}) \tag{4.13}$$

where C_ℓ^{meas} and C_ℓ^{th} are the measured pseudo-C_ℓ and theoretical angular power spectrum, respectively. $\mathrm{Cov}(C_\ell, C'_\ell)$ is the Gaussian covariance of the pseudo-C_ℓ. To account for the inaccuracy of the theoretical power spectra at small scales, we restrict the fitting range mostly to the linear regime, with ℓ values below ℓ_{max} defined by the condition $k_{max} = \ell_{max}\chi(z) < 0.3\,\mathrm{Mpc}^{-1}$, where $\chi(z)$ is the comoving angular distance at the median redshift

of each photometric redshift bin. On the other hand, the limited survey area increases the uncertainty of measured power spectra at large scales and we limit the lower bound of our fit to $\ell_{min} = 1/f_{sky} \approx 138$.

The fit can be either done on individual redshift bins (auto or cross correlation) using Equation (4.13) directly, or as a full fit on all redshift bins combined. Here we use an example consists of 2 photometric redshift bins (0 and 1) to demonstrate the procedure of full fit:

1. We concatenate all measured C_ℓ's to one vector $C_{\ell,all} = (C_\ell^{00}, C_\ell^{01}, C_\ell^{11})$.

2. The full covariance matrix can be written as

$$\mathrm{Cov}_{all} = \begin{pmatrix} \mathrm{Cov}^{00,00} & \mathrm{Cov}^{00,01} & \mathrm{Cov}^{00,11} \\ \mathrm{Cov}^{01,00} & \mathrm{Cov}^{01,01} & \mathrm{Cov}^{01,11} \\ \mathrm{Cov}^{11,00} & \mathrm{Cov}^{11,01} & \mathrm{Cov}^{11,11} \end{pmatrix}$$

3. We plug $C_{\ell,all}$ and Cov_{all} in Equation (4.13) then perform χ^2 fitting. The biases of cross correlations are inferred from the biases of auto correlation using $b_{ij}^2 = b_{ii} b_{jj}$. Therefore, we only consider auto correlation biases as the fitting parameters in the full fit.

The fit with n redshift bins can be generalized accordingly. The results of full fits will be discussed in more detail in the following subsections.

4.6 True redshift vs. Stacking photo-z pdf

One of the most significant systematic uncertainty effects arises from the photometric redshift estimation. We've shown that the photometric redshift model performs well

in terms of serveral figure of merits, such as the root-mean-square scatter $\sigma_z/(1+z)$ and bias in $e_z = (z_{photo} - z_{spec})/(1+z_{spec})$. To further investigate this uncertainty, we perform the whole analysis twice: once using true redshift, and again using the measured photometric redshift, while keeping all other parameters fixed. Here we compare the galaxy bias estimation resulting from both methods.

Method	b_1	b_2	b_3	b_4	b_5	b_6
Full	0.71 ± 0.02	0.81 ± 0.02	0.84 ± 0.01	0.94 ± 0.01	1.12 ± 0.01	1.29 ± 0.01
Individual	0.81 ± 0.03	0.88 ± 0.02	0.89 ± 0.01	0.99 ± 0.01	1.18 ± 0.01	1.34 ± 0.01

Table 4.2: Comparison of galaxy biases obtained through a full simultaneous fit and from fitting individual redshift bins. Here stacking photo-z pdfs is used to calculate the redshift distribution of objects.

We perform both individual fit and full fit of galaxy biases using redshift distributions obtained from stacking photo-z pdfs and true redshifts. The results are shown in Table 4.2 and Table 4.3. Compared to stacking photo-z pdfs, we find much better consistency between the fitted values of galaxy bias when fitting to the auto-power spectra and the cross-power spectra when using the true photo-z distribution. Although stacking photo-z pdfs is a procedure commonly done in analysis when it comes to determining the redshift distributions, it is not mathematically sound [59]. In our case, given the availability of truth data, we opt to use the true redshift in the final fiducial measurements described in this work. Since we are concerned with the statistical uncertainty in the extraction of the power

spectrum and galaxy bias and in systematics from observing conditions, this does not affect the overall conclusions of this thesis.

4.7 Deprojection of contaminants

We have previously covered the methods used to reduce the effects of bright objects and variations in survey depth on the galaxy angular power spectrum in Section 4.2. Variations in survey properties such as seeing, airmass, sky background etc. across the survey area may also lead to biases (or "contamination") of the measured galaxy angular power spectrum. In this section we describe the methods used to correct for these and the uncertainty associated with such systematic shifts.

We use a technique called "template deprojection" that accounts for the effects of sky contaminants to the estimated power spectrum. Deprojection is performed with all survey property maps described in Section 4.3, in order to correct for the contamination of the measured number density of objects (and hence the angular power spectrum) caused by these survey properties. We use `NaMaster` for template deprojection calculations.

4.7.1 Algorithm

If the effect of contaminants on the measured signal is local in space, the contamination can be approximated as linear on large scales [50]:

$$\delta^{\rm obs}(\vec{\theta}) = \delta^{\rm true}(\vec{\theta}) + \sum_i \alpha_i f_i(\vec{\theta}) \tag{4.14}$$

where $\delta^{\rm obs}(\vec{\theta})$ and $\delta^{\rm true}(\vec{\theta})$ are the observed and true number density maps. $f_i(\vec{\theta})$ are templates of known contaminants, and α_i are unknown coefficients. Roughly speaking, con-

taminant deprojection projects the observed map onto a subspace that's orthogonal to all the templates. One can compute the best-fits of the unknown coefficients using:

$$\hat{\alpha}_i = \sum_j F_{ij} \int d\vec{\theta} f_j^*(\vec{\theta}) \delta^{\rm obs}(\vec{\theta}) \tag{4.15}$$

where

$$(\mathbf{F}^{-1})_{ij} = \int d\vec{\theta} f_i^*(\vec{\theta}) f_j(\vec{\theta}) \tag{4.16}$$

is the covariance matrix of templates $f_i(\vec{\theta})$. Consider a special case where all templates are pair-wise orthogonal (with their variance normalized to 1), the solution is reduced to

$$\hat{\alpha}_i = \int d\vec{\theta} f_i^*(\vec{\theta}) \delta^{\rm obs}(\vec{\theta}) \tag{4.17}$$

i.e, the best-fits only keep the component of the observed map that is orthogonal to all templates. Based on Equation (4.14), we can produce the clean version of the number density map. It is defined as

$$\delta^{\rm clean}(\vec{\theta}) \equiv \delta^{\rm obs}(\vec{\theta}) - \sum_i \hat{\alpha}_i f_i(\vec{\theta}) \tag{4.18}$$

However, this definition of the clean map makes it a biased estimator of the true map

$$\delta^{\rm clean}(\vec{\theta}) = \delta^{\rm true}(\vec{\theta}) - \sum_{ij} f_i(\vec{\theta}) F_{ij} \int d\vec{\theta} f_j^*(\vec{\theta}) \delta^{\rm true}(\vec{\theta}) \tag{4.19}$$

This discrepancy of number density maps leads to a loss of modes in the estimated power spectra, the "deprojection bias", which is corrected recursively and analytically in `NaMaster`.

4.7.2 Shot noise

For auto-power spectra, a noise bias term should be subtracted due to shot noise. We compute the theoretical shot noise power spectrum with

$$N_l = \Omega_p \frac{\bar{w}}{\bar{N}}, \tag{4.20}$$

a constant value over all multiples l, where Ω_p is the pixel area, $\bar{w}$ is the mean weight and $\bar{N}$ is the weighted mean of the number of objects per pixel. Then we convolve it with the coupling matrix and bin the multiples to bandpowers. We also estimate the deprojection bias and correct for both contributions.

4.8 Results

Here we present the measured galaxy angular power spectrum C_l and the galaxy bias from our analysis of the *LSS sample*. Figure 4.13 shows the galaxy power spectra in all redshift bins including the deprojection of all 30 survey property maps. Also shown are the results of fitting theoretical predictions from CCL. We perform a full combined fit using all measured auto- and cross-power spectra in all redshift bins. As a cross check, we also fit each individual redshift bin separately. We find good fits in all bins in both cases.

The results of the fitted value of galaxy bias from both methods can be found in Table 4.3. We can see that the two fitting methods are consistent with one another.

The results shown in Figure 4.13 are the final results and are referred to as the *fiducial measurement*. We also include a plot of the galaxy bias as a function of redshift in Figure 4.14.

The fractional uncertainty σ_ℓ/C_ℓ in the measured angular power spectra is shown in Figure 4.15. It is smaller than 10% in most cases, and smaller than 4% at higher multipoles. The full LSST survey will cover a much larger survey area ($f_{sky} \approx 1/2.3$) compared to the

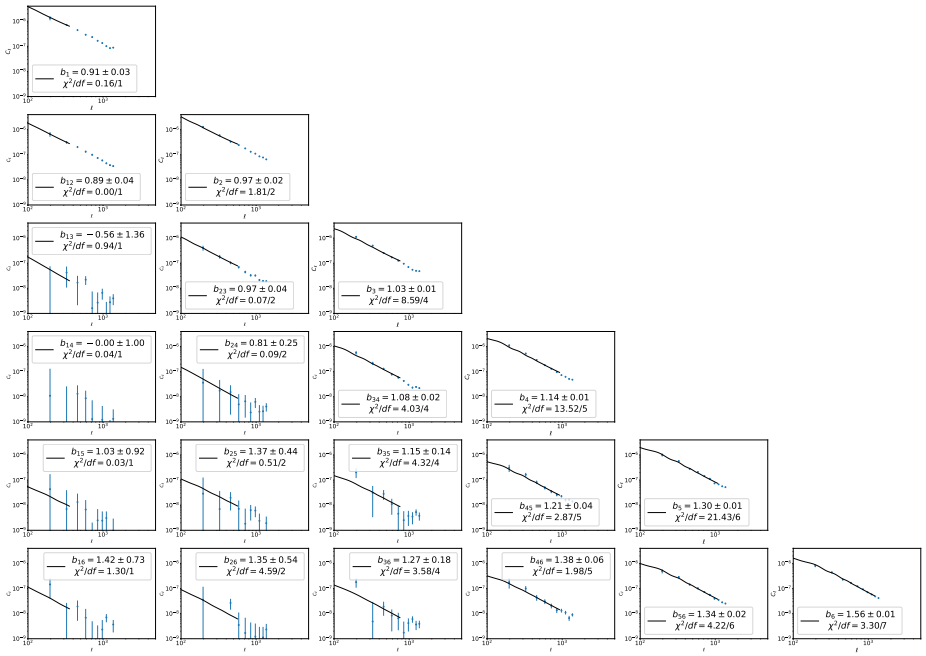


Figure 4.13: Measured galaxy auto- and cross-power spectra for all photometric redshift bins (points with error bars). Also shown are the fits to the data using CCL and the extracted values of the galaxy bias and χ^2 per degree of freedom of fit in each bin (solid line). The galaxy biases are obtained through minimizing χ^2s and the error bars are computed using the differences between the bias values corresponding to $\chi^2 = \chi^2_{min} + 1$ and the best fits. The plot shows results for individual fits in each bin. Similar results are obtained from a full fit of all bins combined: $b_1 = 0.92 \pm 0.03$, $b_2 = 0.97 \pm 0.02$, $b_3 = 1.04 \pm 0.01$, $b_4 = 1.15 \pm 0.01$, $b_5 = 1.32 \pm 0.01$, $b_6 = 1.58 \pm 0.01$.

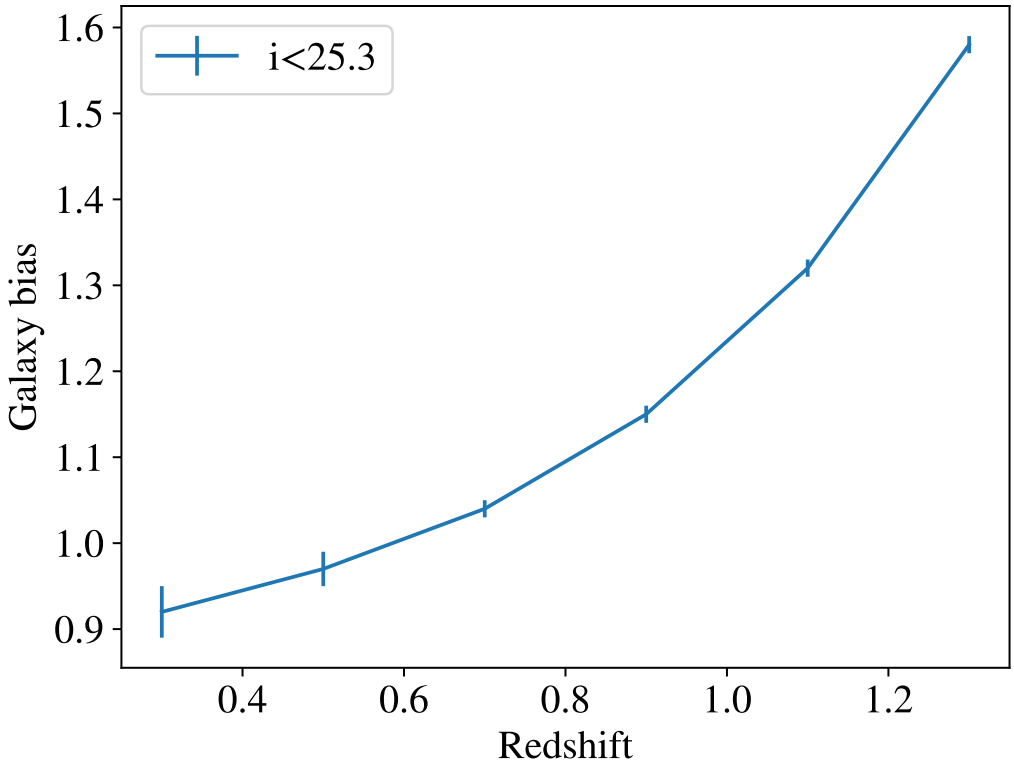


Figure 4.14: Galaxy bias vs. center redshift of each bin from the full fit using true redshifts distributions.

Method	b_1	b_2	b_3	b_4	b_5	b_6
Full	0.92 ± 0.03	0.97 ± 0.02	1.04 ± 0.01	1.15 ± 0.01	1.32 ± 0.01	1.58 ± 0.01
Individual	0.91 ± 0.03	0.97 ± 0.02	1.03 ± 0.01	1.14 ± 0.01	1.30 ± 0.01	1.56 ± 0.01

Table 4.3: Comparison of galaxy biases obtained through a full simultaneous fit and from fitting individual redshift bins. Here true redshifts are used to generate the redshift distribution of objects.

simulated data used here ($f_{sky} \approx 1/138$). We should expect that the relative uncertainty will decrease to 1.3% (0.5% at higher multipoles) with the complete LSST survey.

4.9 Alternative methods for creating deprojection templates

The computation of deprojection bias is the most time-consuming process of the analysis and the time complexity scales as the square of the number of contamination templates. If we include all survey maps described in Section 4.3 for each photometric band, it would result in a total of $5 \times 6 = 30$ templates. The result of deprojection using all 30 maps is shown in figure 4.16. The average fractional change in C_ℓ is found to be around 0.5%, with a maximum change of 1.25%.

The contamination deprojection of one number density map with an area of 300 deg^2 using `Nside` = 1024 and 30 templates takes about 27 CPU hours on a single NERSC Cori node to compute. Since the `NaMaster` code currently utilizes only a single CPU core, it

will be unfeasible to include all contamination templates if we were to scale the analysis to larger survey area and higher map resolution. If we could reduce the number of templates while capturing most of the information stored in them, it can reduce the run time of the deprojection calculation drastically while preventing the resulting power spectra from the risk of overfitting.

Here we explore the effectiveness and performance of two methods to minimize the total computing time by reducing the number of templates used.

4.9.1 Principal Component Analysis (PCA)

In this secttion we describe a solution through Principal Component Analysis (PCA), following the method used in [60]. The procedure to reduce the number of templates can be summarized as follows:

- Normalize all contamination templates to zero mean and unit variance.
- Flatten the templates to vectors and apply PCA.
- Sort the principle component maps according to their eigenvalue or explained variance, then retain the first n maps such that the explained variance percentage from them is $\sim 90\%$. This process reduces the number of maps by $\sim 50\%$, and therefore cuts down the computation time by a factor of ~ 4 .
- When performing deprojection, use these principle component maps instead of the contamination templates.

We compare the power spectra measured using PCA templates with all contamination templates by plotting their differences with the results obtained without deprojection divided

by the 1-σ error, as shown in Figure 4.17. Also presented are the power spectra measured using PCA templates with the results obtained without deprojection (see Figure 4.18).

4.9.2 Contaminant selection based on linearity

As described earlier in section 4.7.1, the deprojection technique is based on the idea of linear expansion. It works well when the contamination effects are adequately small and linear. Here we check if this condition is met for our data. We visualize and study the correlation between the galaxy overdensity maps and contamination templates as follows:

- For each tomographic band, normalize the contamination templates by dividing by its per pixel mean value.
- Divide each normalized template into 9 regions based on the pixel value percentile, where each region contains 10% of the total pixels (we discard the pixels that have values below the 5-th or above the 95-th percentile.)
- Compute the mean galaxy overdensity for each region and estimate the standard error using the leave-one-out Jackknife technique [61].
- For all contamination templates in each tomographic band and redshift bin, plot the mean number overdensity against the mean overdensity of the contamination templates of each region.

For each relation between contamination template and number density map, we perform a linear fit. We then use only contamination templates that have a Pearson corre-

lation coefficient $r > 0.6$. We compare the results with the analysis run where we use the templates selected using these criteria to the fiducial analysis in Figure 4.19. The difference is seen to be relatively small (less than $0.25\sigma_\ell$).

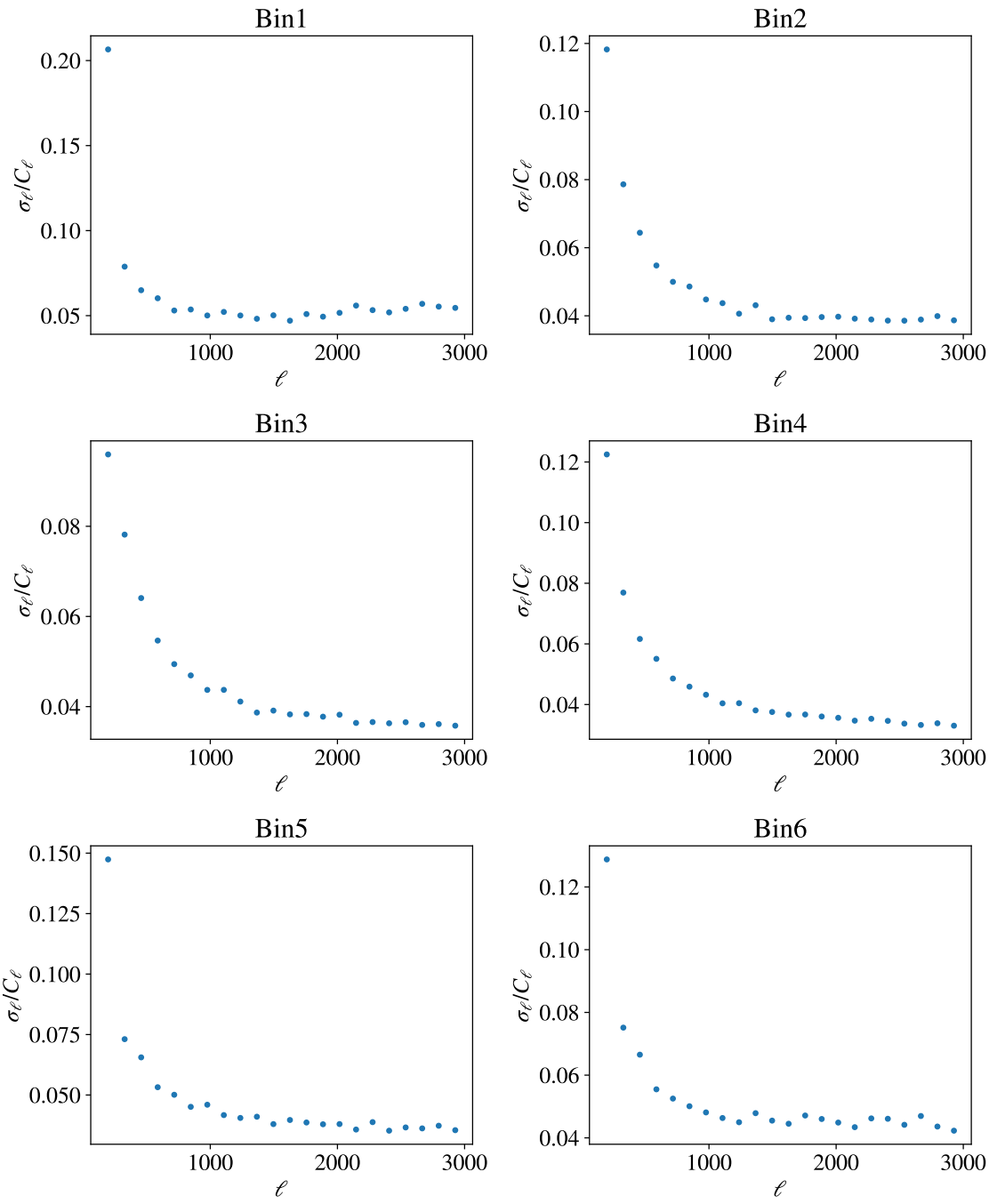


Figure 4.15: Fractional uncertainty in C_ℓ for the all six auto-correlation redshift bins. We can see that, over the whole ℓ range, the signal-to-noise ratio is larger than 10. The signal-to-noise ratio is larger than 20 when $\ell > 700$.

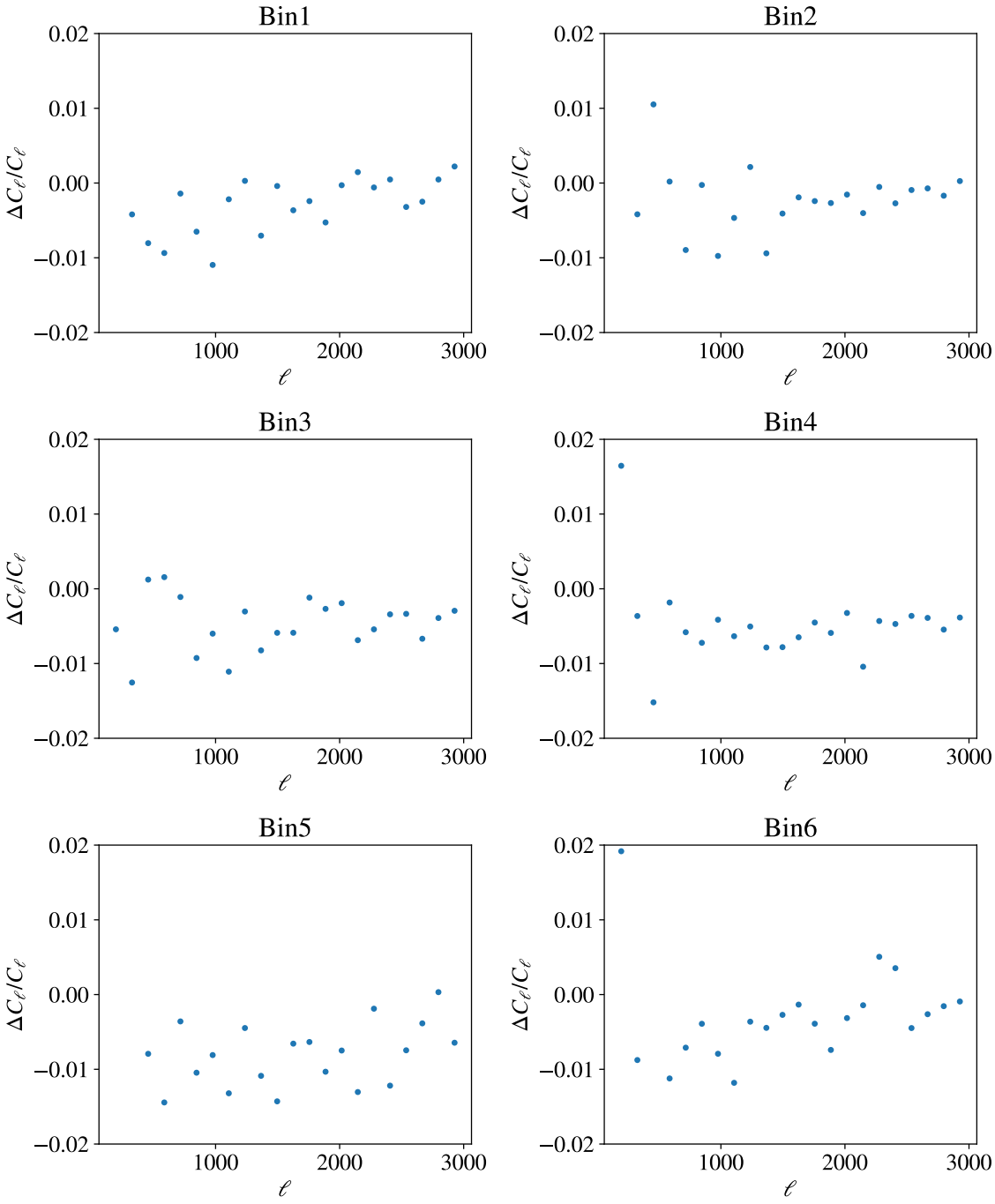


Figure 4.16: Fractional change in the angular power spectrum C_ℓ due to deprojection using all 30 systematics maps for the the auto-correlation redshift bins.

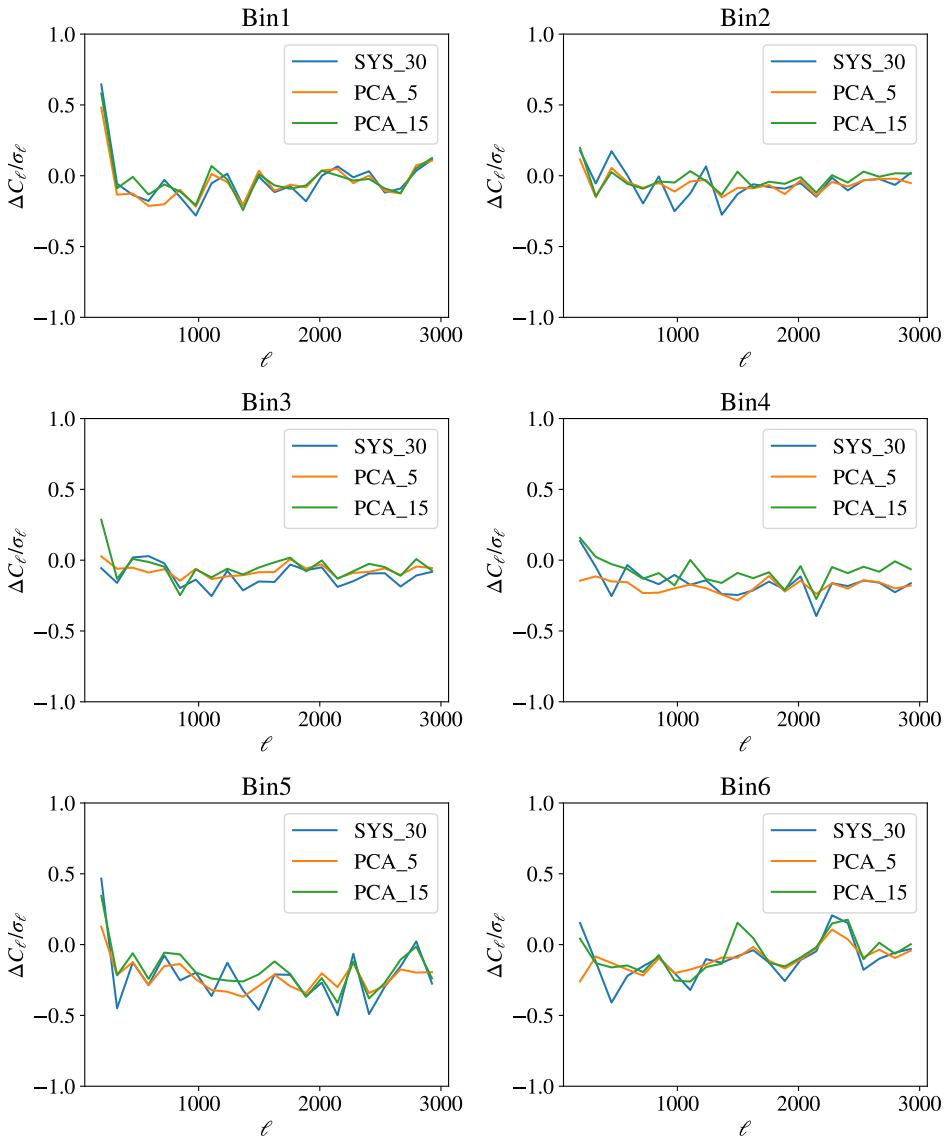


Figure 4.17: Differences between C_l's measured with deprojection templates vs. without deprojection in all auto-correlation redshift bins, normalized by the uncertainty. *Blue:* using all 30 systematic maps; *Orange:* using top 5 maps generated by PCA; *Green:* using top 15 maps generated by PCA.

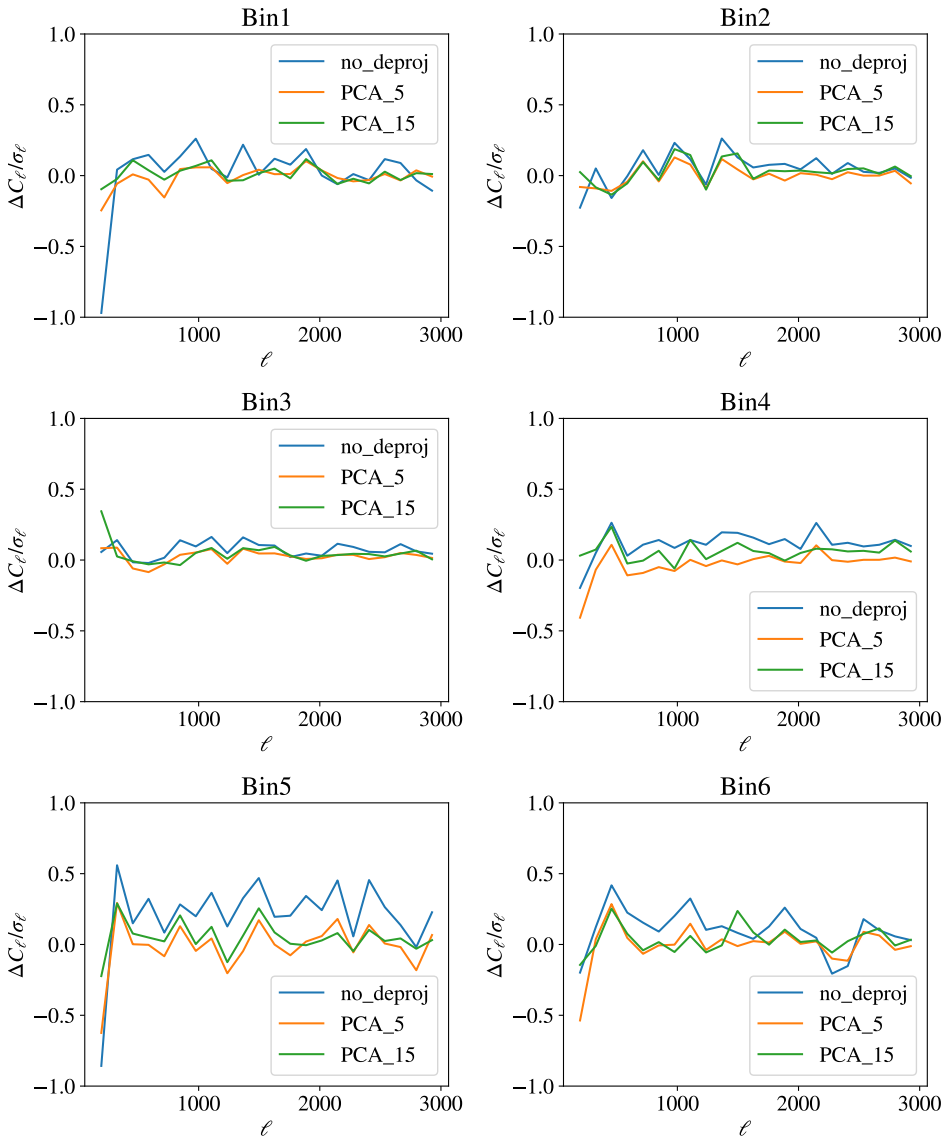


Figure 4.18: Comparison of deprojection methods. Difference between the auto-angular power spectrum C_l measured using PCA deprojection (5 and 15 maps) and C_l measured using deprojection using all 30 maps normalized by the uncertainty are shown in orange and green. The normalized difference between the result without deprojection and the result with deprojection using all 30 maps is shown in blue.

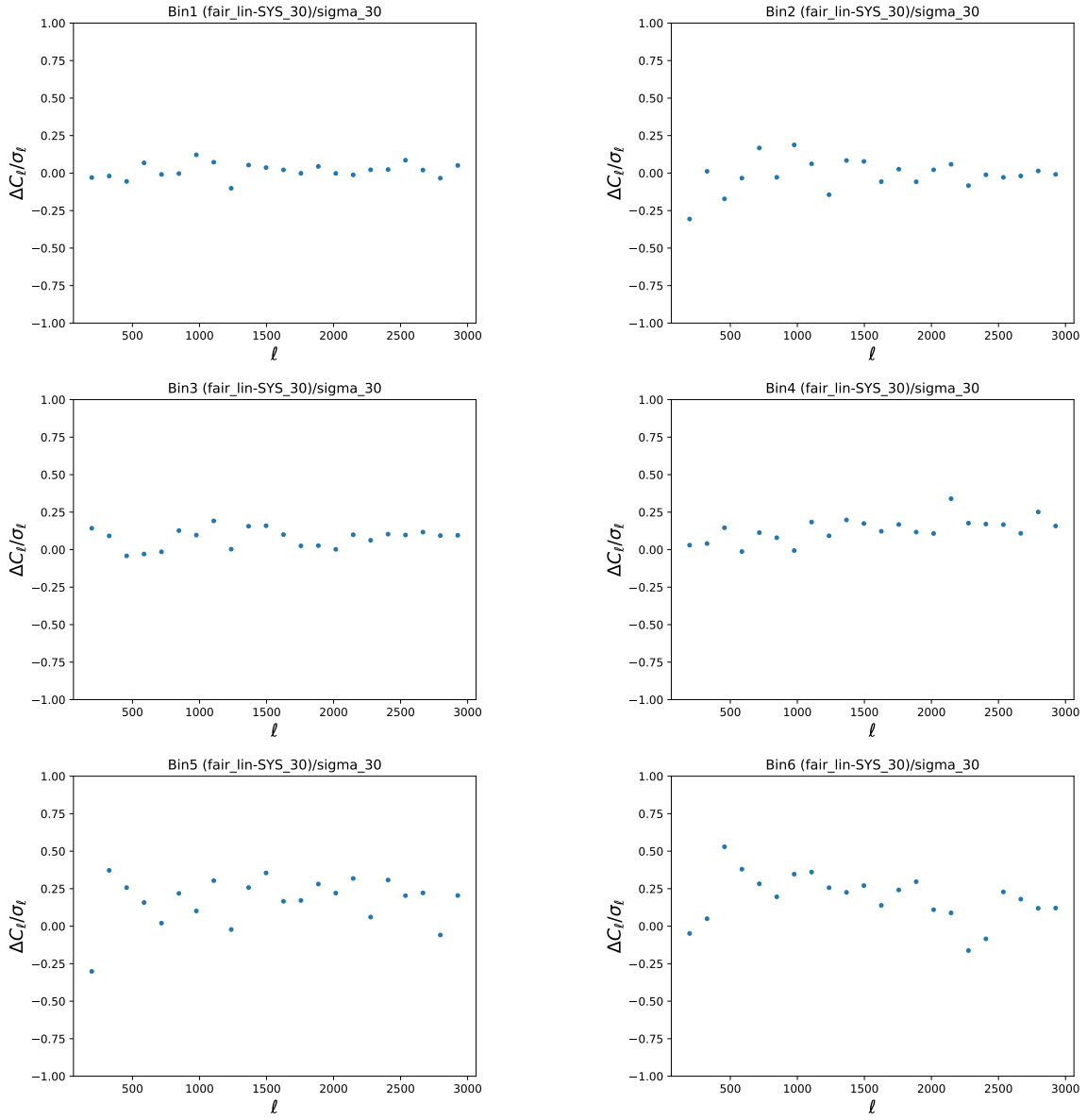


Figure 4.19: Difference between auto-angular power spectrum C_l measured using deprojection of contaminant maps selected using linearity of contaminant with galaxy density and C_l measured using deprojection with all 30 maps normalized by the uncertainty. All 6 photo-z bins are shown.

Chapter 5

Summary and Conclusions

We performed a complete analysis of the galaxy angular power spectrum on simulated data (LSST DC2 catalogs), which correspond to a survey area of 300 deg^2 and an equivalent exposure of 5 years of survey data in 6 broadband filters (*ugrizy*). We find that the photometric redshift performance is close to the requirements quoted in the LSST Science Book up to $z = 1.5$. We use a set of 6 evenly spaced photometric redshift bins from $z = 0.2$ to $z = 1.4$. We apply quality cuts based on studies using matching of selected galaxies to truth information in the catalogs. After magnitude cut, the selected galaxy sample contains 51 million objects with 95% completeness and 99% purity. We then create bright object and depth masks, and generate survey property maps by averaging over all visits. Finally, we compute the angular power spectra for each pair of photometric redshift bins with and without contaminant deprojection and extract the galaxy bias in each bin using a chisquare fit with theoretical precictions from the Core Cosmology Library.

The results show that:

1. The relative uncertainty σ_ℓ/C_ℓ in the measured angular power spectra is smaller than 10% in most cases, and smaller than 4% at higher multipoles. The full LSST survey will cover a much larger survey area ($f_{sky} \approx 1/2.3$) compared to the simulated data used here ($f_{sky} \approx 1/138$). We should expect that the relative uncertainty will decrease to 1.3% (0.5% at higher multipoles) with the complete LSST survey.

2. Changes in the angular power spectrum due to deprojection of contamination from survey properties are found to be on average of order 0.5%, i.e. $\Delta C_\ell/C_\ell \approx 0.5\%$, with a maximum deviation of 1.5%.

3. Deprojection systematic effects are small compared to the statistical uncertainty of the measured angular power spectra ($\Delta C_\ell \sim 0.1\sigma_\ell$ on average with a maximum of $0.25\sigma_\ell$). For the full LSST survey data we expect the effects of the systematics to be significant compared to the statistical uncertainty.

4. We find that the number of survey property maps used to estimate systematic uncertainties can be significantly reduced using principle component analysis or fair linear selection, without affecting the results. This is important to reduce computing resources, especially when scaled to the full LSST survey.

Bibliography

[1] G. Hinshaw, D. Larson, *et al.*, "Nine-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological parameter results," *The Astrophysical Journal Supplement Series* **208** no. 2, (Sep, 2013) 19. https://dx.doi.org/10.1088/0067-0049/208/2/19.

[2] **Planck** Collaboration, Y. Akrami *et al.*, "Planck 2018 results. IV. Diffuse component separation," *Astron. Astrophys.* **641** (2020) A4, arXiv:1807.06208 [astro-ph.CO].

[3] **Planck** Collaboration, N. Aghanim *et al.*, "Planck 2018 results. V. CMB power spectra and likelihoods," *Astron. Astrophys.* **641** (2020) A5, arXiv:1907.12875 [astro-ph.CO].

[4] A. Lewis and S. Bridle, "Cosmological parameters from CMB and other data: A Monte Carlo approach," *Physical Review D* **66** (2002) 103511, arXiv:astro-ph/0205436 [astro-ph].

[5] W. J. Percival, "Large Scale Structure Observations," in *186th Course of International School of Physics 'Enrico Fermi': New Horizons for Observational Cosmology*, vol. 186, pp. 101–135. 2014. arXiv:1312.5490 [astro-ph.CO].

[6] J. A. Peacock *et al.*, "A Measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey," *Nature* **410** (2001) 169–173, arXiv:astro-ph/0103143.

[7] **LSST** Collaboration, v. Ivezić *et al.*, "LSST: from Science Drivers to Reference Design and Anticipated Data Products," *Astrophys. J.* **873** no. 2, (2019) 111, arXiv:0805.2366 [astro-ph].

[8] **LSST Dark Energy Science** Collaboration, B. Abolfathi, D. Alonso, *et al.*, "The lsst desc dc2 simulated sky survey," *The Astrophysical Journal Supplement Series* **253** no. 1, (Mar, 2021) 31. https://dx.doi.org/10.3847/1538-4365/abd62c.

[9] G. Lemaître, "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** (Jan., 1927) 49–59.

[10] A. Einstein, "Die grundlage der allgemeinen relativitätstheorie," *Annalen der Physik* **354** no. 7, (1916) 769–822. https://doi.org/10.1002/andp.19163540702.

[11] A. H. Guth, "Inflationary universe: A possible solution to the horizon and flatness problems," *Phys. Rev. D* **23** (Jan, 1981) 347–356. https://link.aps.org/doi/10.1103/PhysRevD.23.347.

[12] A. Linde, "Inflation and quantum cosmology," *Physica Scripta* **1991** no. T36, (Jan, 1991) 30. https://dx.doi.org/10.1088/0031-8949/1991/T36/004.

[13] R. R. Caldwell, R. Dave, and P. J. Steinhardt, "Cosmological imprint of an energy component with general equation of state," *Phys. Rev. Lett.* **80** (Feb, 1998) 1582–1585. https://link.aps.org/doi/10.1103/PhysRevLett.80.1582.

[14] A. Friedman, "Über die krümmung des raumes," *Zeitschrift für Physik* **10** no. 1, (Dec., 1922) 377–386. https://doi.org/10.1007/bf01332580.

[15] B. Ryden, *Introduction to cosmology*. Addison-Wesley, San Francisco, 2003.

[16] E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," *Proceedings of the National Academy of Sciences* **15** no. 3, (Mar., 1929) 168–173. https://doi.org/10.1073/pnas.15.3.168.

[17] J. Garcia-Bellido, "Cosmology and astrophysics," in *2004 European School of High-Energy Physics*, pp. 267–342. 2, 2005. arXiv:astro-ph/0502139.

[18] A. A. Penzias and R. W. Wilson, "A Measurement of Excess Antenna Temperature at 4080 Mc/s.," *The Astrophysical Journal* **142** (July, 1965) 419–421.

[19] G. F. Smoot, "COBE observations and results," *AIP Conference Proceedings* **476** no. 1, (05, 1999) 1–10, https://pubs.aip.org/aip/acp/article-pdf/476/1/1/11630434/1_1_online.pdf. https://doi.org/10.1063/1.59326.

[20] R. A. Sunyaev and Y. B. Zeldovich, "The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies," *Comments on Astrophysics and Space Physics* **4** (Nov., 1972) 173.

[21] R. K. Sachs and A. M. Wolfe, "Perturbations of a Cosmological Model and Angular Variations of the Microwave Background," *The Astrophysical Journal* **147** (Jan., 1967) 73.

[22] R. A. Alpher, H. Bethe, and G. Gamow, "The origin of chemical elements," *Phys. Rev.* **73** (Apr, 1948) 803–804. https://link.aps.org/doi/10.1103/PhysRev.73.803.

[23] **Planck** Collaboration, N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.* **641** (2020) A6, arXiv:1807.06209 [astro-ph.CO]. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[24] A. D. Linde, "Recent progress in inflationary cosmology," in *1st RESCEU International Symposium on the Cosmological Constant and the Evolution of the Universe*, pp. 133–148. 11, 1995. arXiv:astro-ph/9601004.

[25] **Planck** Collaboration, Y. Akrami *et al.*, "Planck 2018 results. X. Constraints on inflation," *Astron. Astrophys.* **641** (2020) A10, arXiv:1807.06211 [astro-ph.CO].

[26] V. Desjacques, D. Jeong, and F. Schmidt, "Large-scale galaxy bias," *Physics Reports* **733** (Feb, 2018) 1–193. https://doi.org/10.1016%2Fj.physrep.2017.12.002.

[27] J. N. Fry, "The evolution of bias," *The Astrophysical Journal* **461** no. 2, (Apr, 1996) L65. https://dx.doi.org/10.1086/310006.

[28] M. Tegmark and P. J. E. Peebles, "The Time evolution of bias," *Astrophys. J. Lett.* **500** (1998) L79, arXiv:astro-ph/9804067.

[29] N. Kaiser, "Clustering in real space and in redshift space," *Monthly Notices of the Royal Astronomical Society* **227** no. 1, (07, 1987) 1–21, https://academic.oup.com/mnras/article-pdf/227/1/1/18522208/mnras227-0001.pdf. https://doi.org/10.1093/mnras/227.1.1.

[30] L. S. Collaboration, P. A. Abell, *et al.*, "Lsst science book, version 2.0." 2009.

[31] **LSST Dark Energy Science** Collaboration, R. Mandelbaum *et al.*, "The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document," arXiv:1809.01669 [astro-ph.CO].

[32] F. Delgado, A. Saha, S. Chandrasekharan, K. Cook, C. Petry, and S. Ridgway, "The LSST operations simulator," in *SPIE Proceedings*, G. Z. Angeli and P. Dierickx, eds. SPIE, Aug., 2014. https://doi.org/10.1117/12.2056898.

[33] M. Jurić, J. Kantor, *et al.*, "The lsst data management system." 2015.

[34] J. Bosch, Y. AlSayyad, *et al.*, "An overview of the lsst image processing pipelines." 2018.

[35] **LSST Dark Energy Science** Collaboration, D. Korytov *et al.*, "CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST," *Astrophys. J. Suppl.* **245** no. 2, (2019) 26, arXiv:1907.06530 [astro-ph.CO].

[36] K. Heitmann *et al.*, "The Outer Rim Simulation: A Path to Many-Core Supercomputers," *Astrophys. J. Suppl.* **245** no. 1, (2019) 16, arXiv:1904.11970 [astro-ph.CO].

[37] S. Habib, A. Pope, *et al.*, "Hacc: Simulating sky surveys on state-of-the-art supercomputing architectures," *New Astronomy* **42** (2016) 49–65. https://www.sciencedirect.com/science/article/pii/S138410761500069X.

[38] P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy, "Universemachine: The correlation between galaxy growth and dark matter halo assembly from z = 0-10," *Monthly Notices of the Royal Astronomical Society* (2018) . https://api.semanticscholar.org/CorpusID:119385275.

[39] A. J. Benson, "Galacticus: A semi-analytic model of galaxy formation," *New Astronomy* **17** no. 2, (2012) 175–197. https://www.sciencedirect.com/science/article/pii/S1384107611000807.

[40] J. R. Peterson, J. G. Jernigan, *et al.*, "Simulation of astronomical images from optical survey telescopes using a comprehensive photon monte carlo approach," *The Astrophysical Journal Supplement Series* **218** no. 1, (May, 2015) 14. https://dx.doi.org/10.1088/0067-0049/218/1/14.

[41] A. J. Connolly, J. Peterson, *et al.*, "Simulating the LSST system," in *Modeling, Systems Engineering, and Project Management for Astronomy IV*, G. Z. Angeli and P. Dierickx, eds., vol. 7738, p. 77381O, International Society for Optics and Photonics. SPIE, 2010. https://doi.org/10.1117/12.857819.

[42] A. J. Connolly, G. Z. Angeli, *et al.*, "An end-to-end simulation framework for the Large Synoptic Survey Telescope," in *Modeling, Systems Engineering, and Project Management for Astronomy VI*, G. Z. Angeli and P. Dierickx, eds., vol. 9150, p. 915014, International Society for Optics and Photonics. SPIE, 2014. https://doi.org/10.1117/12.2054953.

[43] J. Sánchez, C. W. Walter, *et al.*, "The LSST DESC data challenge 1: generation and analysis of synthetic images for next-generation surveys," *Monthly Notices of the Royal Astronomical Society* **497** no. 1, (07, 2020) 210–228, https://academic.oup.com/mnras/article-pdf/497/1/210/33522604/staa1957.pdf. https://doi.org/10.1093/mnras/staa1957.

[44] J. Bosch, R. Armstrong, *et al.*, "The Hyper Suprime-Cam software pipeline," *Publications of the Astronomical Society of Japan* **70** no. SP1, (10, 2017) S5, https://academic.oup.com/pasj/article-pdf/70/SP1/S5/23692350/psx080.pdf. https://doi.org/10.1093/pasj/psx080.

[45] I. Sevilla-Noarbe, B. Hoyle, *et al.*, "Star–galaxy classification in the Dark Energy Survey Y1 data set," *Monthly Notices of the Royal Astronomical Society* **481** no. 4, (09, 2018) 5451–5469, https://academic.oup.com/mnras/article-pdf/481/4/5451/26075119/sty2579.pdf. https://doi.org/10.1093/mnras/sty2579.

[46] **LSST** Collaboration, A. Nicola *et al.*, "Tomographic galaxy clustering with the Subaru Hyper Suprime-Cam first year public data release," *JCAP* **03** (2020) 044, arXiv:1912.08209 [astro-ph.CO].

[47] J. Coupon, N. Czakon, J. Bosch, Y. Komiyama, E. Medezinski, S. Miyazaki, and M. Oguri, "The bright-star masks for the HSC-SSP survey," *Publications of the*

Astronomical Society of Japan **70** no. SP1, (07, 2017) S7, https://academic.oup.com/pasj/article-pdf/70/SP1/S7/23692103/psx047.pdf. https://doi.org/10.1093/pasj/psx047.

[48] N. Benítez, "Bayesian photometric redshift estimation," *The Astrophysical Journal* **536** no. 2, (Jun, 2000) 571. https://dx.doi.org/10.1086/308947.

[49] D. Coe, N. Benítez, S. F. Sánchez, M. Jee, R. Bouwens, and H. Ford, "Galaxies in the hubble ultra deep field. i. detection, multiband photometry, photometric redshifts, and morphology," *The Astronomical Journal* **132** no. 2, (Jul, 2006) 926. https://dx.doi.org/10.1086/505530.

[50] **LSST Dark Energy Science** Collaboration, D. Alonso, J. Sanchez, and A. Slosar, "A unified pseudo-C_ℓ framework," *Mon. Not. Roy. Astron. Soc.* **484** no. 3, (2019) 4127–4151, arXiv:1809.09603 [astro-ph.CO].

[51] M. Reinecke and D. S. Seljebotn, "Libsharp – spherical harmonic transforms revisited," *Astronomy; Astrophysics* **554** (Jun, 2013) A112. https://doi.org/10.1051%2F0004-6361%2F201321494.

[52] F. Elsner, B. Leistedt, and H. V. Peiris, "Unbiased pseudo-C_ℓ power spectrum estimation with mode projection," *Monthly Notices of the Royal Astronomical Society* **465** no. 2, (10, 2016) 1847–1855, https://academic.oup.com/mnras/article-pdf/465/2/1847/8364799/stw2752.pdf. https://doi.org/10.1093/mnras/stw2752.

[53] G. Efstathiou, "Myths and truths concerning estimation of power spectra," *Mon. Not. Roy. Astron. Soc.* **349** (2004) 603, arXiv:astro-ph/0307515.

[54] C. García-García, D. Alonso, and E. Bellini, "Disconnected pseudo-C_ℓ covariances for projected large-scale structure data," *JCAP* **11** (2019) 043, arXiv:1906.11765 [astro-ph.CO].

[55] F. Lacasa, "Covariance of the galaxy angular power spectrum with the halo model," *Astron. Astrophys.* **615** (2018) A1, arXiv:1711.07372 [astro-ph.CO].

[56] A. Loureiro, B. Moraes, *et al.*, "Cosmological measurements from angular power spectra analysis of BOSS DR12 tomography," *Monthly Notices of the Royal Astronomical Society* **485** no. 1, (01, 2019) 326–355, https://academic.oup.com/mnras/article-pdf/485/1/326/27932728/stz191.pdf. https://doi.org/10.1093/mnras/stz191.

[57] **LSST Dark Energy Science** Collaboration, N. E. Chisari *et al.*, "Core Cosmology Library: Precision Cosmological Predictions for LSST," *Astrophys. J. Suppl.* **242** no. 1, (2019) 2, arXiv:1812.05995 [astro-ph.CO].

[58] M. LoVerde and N. Afshordi, "Extended Limber Approximation," *Phys. Rev. D* **78** (2008) 123506, arXiv:0809.5112 [astro-ph].

[59] A. I. Malz, "How not to obtain the redshift distribution from probabilistic redshift estimates: Under what conditions is it not inappropriate to estimate the redshift distribution N(z) by stacking photo-z PDFs?," *Phys. Rev. D* **103** no. 8, (2021) 083502, arXiv:2101.04675 [astro-ph.CO].

[60] **DES** Collaboration, M. Rodríguez-Monroy *et al.*, "Dark Energy Survey Year 3 results: galaxy clustering and systematics treatment for lens galaxy samples," *Mon. Not. Roy. Astron. Soc.* **511** no. 2, (2022) 2665–2687, arXiv:2105.13540 [astro-ph.CO].

[61] R. G. Miller, "The jackknife–a review," *Biometrika* **61** no. 1, (1974) 1–15. http://www.jstor.org/stable/2334280.