The Distribution of Galaxies as a Test of Primordial Non-Gaussianity

by

Benedict Konrad Wilhelm Kalus

This thesis is submitted in partial fulfilment of
the requirements for the award of the degree of
Doctor of Philosophy of the University of Portsmouth.

17th May 2018

Abstract

Cosmic inflation is a paradigm that successfully seeds large scale structure in a big bang cosmology whilst causing the Universe to be statistically isotropic. Inflation occurred at energy scales that are too high to be accessible with accelerator experiments, thus we have to rely entirely on cosmological observations to rule out classes of inflationary models and get insight into the physics of the early Universe. One way to distinguish these models is by measuring how close their predicted primordial fluctuations are to being Gaussian distributed, described at first order by the parameter $f_{\rm NL}$. Local primordial non-Gaussianity alters the biasing law between dark- matter halos and the underlying mass-density field at the largest scales [1, 2]. Currently, the tightest constraints on the local $f_{\rm NL} = 0.8 \pm 5.0$ come from the cosmic microwave background experiment Planck [3]. Next generation ground-based experiments will be limited by cosmic variance, and we need a different approach to independently confirm these results, and to further narrow down our constraints on the inflationary epoch. Galaxy clustering studies so far could not compete with the precision of the $f_{\rm NL}$ results from the microwave background, but upcoming galaxy surveys will come close to independently confirm the Planck results. Combining future galaxy clustering and Cosmic Microwave Background data will improve $f_{\rm NL}$ constraints such that they will provide physically interesting results.

This will only be possible if some challenges can be addressed properly, of which two are addressed in this thesis. The first problem is that $f_{\rm NL}$ measurements take place at the very largest scales, which are close to maximum scale fitting into the survey volume. This means that we have to rely on a low number of modes and we therefore cannot assume the likelihood of our power spectrum measurement to be Gaussian. The Inverse Cubic Normal distribution is a very good approximation to what we expect for the true likelihood

of the Power Spectrum assuming an almost Gaussian galaxy density field. On top of that, it has the advantage of absorbing the model dependence of the Power Spectrum covariance matrix into the functional form of the posterior distribution function. Thus, one does not have to run simulations to estimate the covariance matrix for each model to be tested [4].

The other problem addressed in this thesis is that galaxy surveys are plagued by systematic contaminants, especially the effect of foreground stars, at the scales interesting for $f_{\rm NL}$ measurements. I discuss two contaminant mitigation techniques, mode deprojection and mode subtraction. Mode deprojection needs a covariance based estimator for the Power Spectrum, such as the Quadratic Maximum Likelihood estimator. This however, is computationally infeasible for a 3-dimensional clustering analysis. Applying mode subtraction naïvely using the simpler Feldman-Kaiser-Peacock Power Spectrum estimator leads to a biased measurement. I introduce an additional factor that unbiases this estimate and show that mode deprojection and subtraction are then in fact identical. This allows a fast error mitigation and Power Spectrum estimation trading against a small amount of information loss [5].

This technique is tested using the twelfth data release of the Baryon Acoustic Oscillation Survey. Previously, even after accounting for the stellar contamination, the power spectrum of an $f_{\rm NL}$ analysis using data of the ninth data release of the Baryon Oscillation Spectroscopic Survey [6] did not agree well with the model for any value $f_{\rm NL}$. Interestingly, even using the new methods, and performing a more careful analysis, the resulting Power Spectrum agrees with the Power Spectrum obtained by applying the methods of [6]. This means that there is still some unexplained discrepancy between our measurement and our model. I therefore discuss other sources of systematic data contamination.

Table of Contents

List of Tables

List of Figures

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

The work within this thesis has been published in two papers. The likelihood study presented in Chapter 3 is published in [4] with Benedict Kalus primarily responsible for statistical tests of different likelihood shapes and the publication itself.

Chapter 4 is based on [5] with Benedict Kalus responsible for adopting the QML estimator for 3D clustering, analytic derivations of contaminant mitigation terms, numerical tests of the different estimators and mitigation techniques, and writing of the article.

Chapter 5 is unpublished and is work in progress.

Word count: 53,165 words.

Acknowledgements

The first person to whom I want to express my gratitude is my first supervisor Will Percival for his support, opportunities, guidance, advice and constructive criticism throughout the last four years.

I also want to thank the other members of my supervisory team, Lado Samushia and David Bacon. In my first year, Lado was a big help in getting me started to research in large scale structure and answered all of my silly and naïve questions with great patience. David was a great source of moral support with his unlimited optimism.

I want to extend my thanks to David Bacon, David Wands and Ben Bose for their valuable discussions and comments on parts of this thesis.

I also want to thank Ben Bose, Hedda Gressel, Rossana Ruggeri, Dario Scovacricchi, Michael Jimmy Tarr and Matt Withers for the good working atmosphere in the office.

The work presented in this thesis would not have been possible without the scientific impact from people outside my supervisory team, most notably Ashley Ross, Eva-Maria Müller, Héctor Gil-Marín and Matteo Tellarini.

Thanks are also due to Katarina Oku Igondova for finding us somewhere to stay for my first three years and to Anne Florath, Monika Vlčnovska, Orla Quinn and Matt Withers for being great people to share a house with.

My parents, Bärbel and Christoph Kalus, deserve acknowledgement for encouraging me to pursue what I want and giving me every opportunity to do so.

Finally, I want to thank Sarah Bahr for all the love and adventures during the last two years and for the time to come.

List of Acronyms

APO	Apache Point Observatory
APOGEE	Apache Point Observatory Galactic Evolution Experiment
BAO	Baryon Acoustic Oscillation
BOSS	Baryon Oscillations Spectroscopic Survey
CAMB	Code for Anisotropies in the Microwave Background
CCD	Charge-Coupled Device
CDF	Cumulative Distribution Function
CFHTLenS	Canada France Hawaii Telescope Lensing Survey
CL	Confidence Level
CLASS	Cosmic Linear Anisotropy Solving System
CMASS	Constant Mass
CMB	Cosmic Microwave Background
COLA	Comoving Lagrangian Acceleration
CosmoMC	Cosmological Monte Carlo
cp	close pair
DES	Dark Energy Survey
DESI	Dark Energy Spectroscopic Instrument
DR	Data Release
eBOSS	extended Baryons Oscillations Spectroscopic Survey
ELG	Emission Line Galaxy
ESA	European Space Agency
EZmocks	Effective Zel'dovich Approximation Mocks
FFT	Fast Fourier Transform
FKP	Feldman-Kaiser-Peacock (estimator)
FLRW	Friedmann-Lemaître-Robertson-Walker
GR	General Relativity

GUT	Grand Unified Theory
HEALPix	Hierarchical Equal Area isolatitude Pixelisation of a sphere
HETDEX	Hobby-Eberly Telescope Dark Energy Experiment
ICN	Inverse Cubic Normal (distribution)
iid	independent and identically distributed
JLA	Joint Light-curve Analysis
KiDS	Kilo Degree Survey
KL	Kullback-Leibler
LRG	Luminous Red Galaxy
LOWZ	Low Redshift
LS	Landy-Szalay (estimator)
Lyα	Lyman α
MARVELS	Multi-object Apache Point Observatory Radial Velocity Exoplanet Large-area Survey
MCMC	Markov Chain Monte Carlo
MeerKAT	Meer Karoo Array Telescope
MeerKLASS	Meer Karoo Array Telescope Large Area Synoptic Survey
MGS	Main Galaxy Sample
NGC	North Galactic Cap
noz	no redshift
NRAO	National Radio Astronomy Observatory
NVSS	National Radio Astronomy Observatory Very Large Array Sky Survey
OLN	Offset Log-Normal (distribution)
PATCHY	Perturbation Theory Catalog Generator of Halo and Galaxy Distributions
PDF	Probability Distribution Function
PH	Peebles-Hauser (estimator)
PINOCCHIO	Pinpointing Orbit-Crossing Collapsed Hierarchical Objects
PSF	Point Spread Function
QML	Quadratic Maximum Likelihood (estimator)
Quasar	Quasi Stellar Object
QSO	Quasi Stellar Object
RSD	Redshift Space Distortions

SDSS	Sloan Digital Sky Survey
SFD	Schlegel-Finbeiner-Davis
SKA	Square Kilometre Array
SN(e)	Supernova(e)
SNLS	Supernova Legacy Survey
UHF	Ultra High Frequencies
VLA	Very Large Array
wCDM	Model of a Universe filled with a general Dark Energy with an equation of state w and Cold Dark Matter
Y1	Year 1
ΛCDM	Model of a Universe filled with a cosmological constant Λ and Cold Dark Matter
2LPT	second order Lagrangian Perturbation Theory

Chapter 1

Introduction

The discovery of an accelerated Hubble expansion, of dark matter, of the Cosmic Microwave Background and constraints on Big Bang Nucleosynthesis put tight limits on our present standard model of cosmology, the Concordance Model. Our Universe most likely originates from a Hot Big Bang, has an energy density described by the Lambda Cold Dark Matter (ΛCDM) model and its contents interact mostly through gravity as described by the Theory of General Relativity. These concepts are introduced in Sec. 1.1 to describe the background cosmology. Full sky observations of the Cosmic Microwave Background tell us that our observable Universe is highly isotropic on large scales. Combined with the Cosmological Principle, that states that our position in the Universe is not special, I derive within the General Relativity framework, introduced in Sec. 1.1.1, the Friedmann-Lemaître-Robertson-Walker metric (cf. Sec. 1.1.2) that describes the background expansion of the Universe. This will be used to describe the background dynamics of the Universe throughout its history in Sec. 1.1.3. I shall show how the dynamics are linked to the energy content of the Universe. An important concept and tool to do so is the cosmic redshift. As the Universe's expansion causes objects to appear moving away from us, the spectra of the radiation they emit is shifted towards the red. A useful relationship between the redshift and the scale factor is therefore presented in Sec. 1.1.4. The cosmic redshift can be used as a distance indicator. However, there are different ways to define distances in an expanding universe. These are reviewed in Sec. 1.1.5. Distance and redshift measurements are the basis for some of the key observations that provide evidence for the ΛCDM

concordance model reviewed in Sec. 1.1.6. In Sec. 1.2, it will be shown how perturbations of this background can lead to the structures that we see today. Much of the remaining observational evidence for the concordance model, of which I give an overview in Sec. 1.1.6, is based on studies of perturbations around the background cosmology described by the concordance model. As these tests are based on statistics of the over-density field, I define the basic diagnostics, such as the correlation function and the power spectrum, in Sec. 1.2.4, after discussing how the underlying perturbations evolve in Sec. 1.2.2 and 1.2.3.

Our current knowledge of cosmology is not yet complete. For instance, we do not know yet, what the major ingredient of the ΛCDM model, dark energy, actually is in terms of fundamental physics, and the Concordance Model does not explain where structure initially comes from. This, and more problems of standard Hot Big Bang cosmology are discussed in Sec. 1.3, where also a popular solution, Cosmic Inflation, and an alternative, Ekpyrosis, are introduced. Cosmic Inflation is a period of exponential growth of the Universe. This paradigm can be realised by different particle physics mechanisms. Some classes of Inflation predict a non Gaussian initial distribution of perturbations. To be able to compare primordial perturbations with today's galaxy distribution, I review how perturbations evolve after inflation in Sec. 1.4. Observing the galaxy distribution does not provide the matter distribution directly. However, galaxies are a biased tracer of the matter field and the concept of galaxy bias is introduced in Sec. 1.5. In particular, measuring the galaxy bias provides a test for local primordial non-Gaussianity if parameterised by a local parameter $f_{\rm NL}$ which will be discussed in Sec. 1.5.2. The current constraints on local $f_{\rm NL}$ are reviewed in Sec. 1.6, before the remainder of this thesis will be outlined.

1.1 The Concordance Background Model of Cosmology

1.1.1 General Relativity

During almost all of the Universe's history, interactions on cosmic scales are dominated by gravity. The best theory of gravity, to this day, is the theory of General Relativity (GR, [33]), based on the Equivalence Principle

> "At every space-time point in an arbitrary gravitational field it is possible to choose a "locally inertial coordinate system" such that, within a sufficiently small region of the point in question, the laws of nature take the same form as in unaccelerated Cartesian coordinate systems in the absence of gravitation." [34]

In a general space-time, the infinitesimal line-element connecting two vectors can be described using a metric tensor (recurring indices have to be interpreted as summation indices):

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{1.1}$$

In order for this principle to hold in any arbitrary coordinate system, the metric tensor $g_{\mu\nu}$ has to be a dynamical field with a non-vanishing Christoffel symbol

$$\Gamma^\sigma{}_{\lambda\mu} = g^{\nu\sigma}\frac{g_{\nu(\mu,\lambda)} - g_{\mu\lambda,\nu}}{2} \quad [34]. \tag{1.2}$$

As it is common in GR literature, the comma denotes a partial derivative with respect to the coordinate written behind it and the round brackets denote symmetrisation with respect to the two components written in between. The Christoffel symbol is needed to define the covariant derivative that is, unlike the ordinary partial derivative, invariant under coordinate transformations. For a covariant vector a_μ, it reads

$$a_{\mu;\nu} \equiv a_{\mu,\nu} - \Gamma^\rho{}_{\nu\mu}a_\rho, \tag{1.3}$$

and for a type (0,2) tensor $A_{\mu\nu}$ it is

$$A_{\mu\nu;\rho} \equiv A_{\mu\nu,\rho} - \Gamma^\sigma{}_{\rho(\mu}A_{\sigma\nu)}. \tag{1.4}$$

The dynamical behaviour of gravitational fields are described by the Einstein Field Equations

$$G_{\mu\nu} = -\kappa T_{\mu\nu} \quad [34]. \tag{1.5}$$

The right-hand side of Eq. (1.5) describes the density and flux of energy and momentum in space-time using the energy-momentum tensor $T_{\mu\nu}$. For a perfect fluid, this is

$$T_{\mu\nu} = p(t)g_{\mu\nu} + (p(t) + \rho(t))u_\mu u_\nu \quad [34,35], \tag{1.6}$$

As it is common in theoretical discussions of GR, throughout this chapter, I have chosen units in which the speed of light is dimensionless and equal to one. Eq. (1.5) can only be fulfilled if the Einstein tensor $G_{\mu\nu}$ on the left-hand side is symmetric and conserved, because the energy-momentum tensor $T_{\mu\nu}$ is symmetric and conserved, too. If we further restrict the Einstein tensor $G_{\mu\nu}$ to contain maximally second derivatives of the metric $g_{\mu\nu}$, and that in the limit of a weak stationary field produced by non-relativistic matter Einstein's Equation (1.5) reduces to Poisson's Equation, we find that $G_{\mu\nu}$ can only be composed of the Ricci tensor

$$R_{\mu\nu} \equiv \Gamma^\rho{}_{[\mu\rho,\nu]} + \Gamma^\sigma{}_{\mu[\rho}\Gamma^\rho{}_{\sigma\nu]} \quad [35], \tag{1.7}$$

the Ricci curvature scalar

$$R \equiv R^\mu{}_\mu, \tag{1.8}$$

and a cosmological constant Λ multiplied with the metric $g_{\mu\nu}$ itself. We can also relate $\kappa = 8\pi G$ to Newton's gravitational constant G by postulating the equivalence between Einstein's and Poisson's equations in the non-relativistic limit. The Einstein field equations then read:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = -8\pi G T_{\mu\nu} \quad [34]. \tag{1.9}$$

We can now see that Eq. (1.9) relates the geometry of a system on the left-hand side with the system's energy content on the right. Hence, the geometry of the Universe provides information about its content, and vice versa; and studying both provides a test of General Relativity. In order to do so, I continue by deriving the metric of a homogeneous and isotropic universe, which, as we shall see in Sec. 1.1.6, describes our Universe well. The Universe's geometry can then be related to its energy content by deriving the equations of motion that governs the dynamics of the Universe in Sec. 1.1.3.

1.1.2 The Cosmological Principle and the Friedmann-Lemaître-Robertson-Walker Metric

We have good evidence, especially from cosmic microwave background (CMB) observations, that the Universe is isotropic on large scales (e.g. [35]). Under the assumption that our position in the Universe is not special, we expect that the Universe looks isotropic from any position within it, which in turn implies that the Universe is also homogeneous (e.g. [34]). Introducing a fundamental observer, i.e. an observer in the rest frame of the overall cosmological fluid, we can also define the hypersurface of simultaneity such that the 4-velocity of any fundamental observer is orthogonal to it:

$$u^\mu \equiv \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = (1,0,0,0) \quad [35]. \tag{1.10}$$

We call the coordinates $x^0 = t$ cosmic time and the constant x^i along the world-line of the fundamental observer comoving coordinates. This hypersurface of simultaneity can then be used to split the line-element for a fundamental observer in a universe obeying the Cosmological Principle into a temporal and spatial part:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \aleph^2(t)\mathrm{d}\sigma^2, \tag{1.11}$$

where $\aleph(t)$ denotes a cosmic time-dependent scale factor. Due to isotropy, the spatial sub-metric $\mathrm{d}\sigma^2$ is constrained to be of the rotationally invariant form

$$\mathrm{d}\sigma^2 = B(r)\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2(\theta)\mathrm{d}\phi^2. \tag{1.12}$$

Using (1.7), the non-vanishing elements of the Ricci tensor read (no summation)

$$R_{rr} = -\frac{1}{rB}\frac{\mathrm{d}B}{\mathrm{d}r},\ R_{\theta\theta} = \frac{1}{B} - 1 - \frac{r}{2B^2}\frac{\mathrm{d}B}{\mathrm{d}r},\ R_{\phi\phi} = R_{\theta\theta}\sin^2(\theta) \quad [35]. \tag{1.13}$$

A non-Euclidean 3-dimensional space embedded into a 4-dimensional flat space has the metric

$$g_{ij} = C_{ij} + \frac{K}{1 - KC_{kl}x^k x^l}C_{im}x^m C_{jn}x^n \ [34] \tag{1.14}$$

with a constant matrix C_{ij} and a constant curvature K. The Ricci tensor for this metric is

$$R_{jk} = -2Kg_{jk} \quad [35]. \tag{1.15}$$

By equating (1.13) and (1.15), we obtain the equations

$$\frac{1}{rB}\frac{\mathrm{d}B}{\mathrm{d}r} = 2KB$$
$$1 + \frac{r}{2B^2}\frac{\mathrm{d}B}{\mathrm{d}r} - \frac{1}{B} = 2Kr^2, \tag{1.16}$$

whose solution is $B(r) = \frac{1}{1-Kr^2}$ [35]. Substituting $\tilde{r} = |K|^{1/2}r$, $k \equiv \frac{K}{|K|}$ and $a(t) \equiv \frac{\aleph(t)}{|K|^{1/2}}$, we obtain the Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}\tilde{r}^2}{1-k\tilde{r}^2} + \tilde{r}^2\left(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2\right)\right] \quad [35,36,37,38]. \tag{1.17}$$

The curvature parameter k can take three different values depending whether the Universe is flat ($k = 0$), spherical ($k = 1$) or hyperbolic ($k = -1$). To remove the singularity for a spherical universe at $\tilde{r} = 1$, the FLRW metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\mathrm{d}\chi^2 + S^2(\chi)\left(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2\right)\right] \quad [35], \tag{1.18}$$

is commonly expressed in terms of a new coordinate

$$\chi \equiv \int_0^{\tilde{r}} \frac{\mathrm{d}\tilde{r}}{1-k\tilde{r}^2} = \begin{cases} \arcsin(\tilde{r}) & \text{if } k = 1 \\ \tilde{r} & \text{if } k = 0 \\ \operatorname{arsinh}(\tilde{r}) & \text{if } k = -1 \end{cases}, \tag{1.19}$$

with

$$S(\chi) = \begin{cases} \sin(\chi) & \text{if } k = 1 \\ \chi & \text{if } k = 0 \\ \sinh(\chi) & \text{if } k = -1 \end{cases}. \tag{1.20}$$

The dynamics of the Universe are hence entirely described by the scale factor $a(t)$. In the following section, we shall see how the cosmic fluid influences the dynamics of the Universe. Assuming ideal fluids, I shall sketch the evolution of a Big Bang Universe.

1.1.3 Dynamical History of a Homogeneous and Isotropic Universe

Assuming that the cosmic fluid is a perfect fluid, its energy-momentum tensor is given by Eq. (1.6). In its inertial rest-frame, where the 4-velocity $u_\mu =$

$(1,0,0,0)_\mu$ is that of a fundamental observer (cf. Eq. (1.10)), this is just a diagonal matrix with the energy density ρ as the T_{00}-component and the pressure p elsewhere on the diagonal, i.e.

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}_{\mu\nu} \quad [34,35]. \tag{1.21}$$

If we insert $T_{\mu\nu}$ into Einstein's Equation (1.9) and solve for the Ricci tensor, we get

$$\begin{aligned} R_{00} &= -4\pi G(\rho + p) - \Lambda \\ R_{ij} &= \left[\Lambda - 4\pi G(\rho - p)\right] g_{ij}. \end{aligned} \tag{1.22}$$

We can also insert the FLRW metric into Eq. (1.7) and equate the resulting Ricci tensor

$$R_{00} = 3\frac{\ddot{a}}{a},\ R_{ij} = -\frac{a\ddot{a} + 2\dot{a}^2 + 2k}{a^2} g_{ij} \quad [34] \tag{1.23}$$

with Eq. (1.22) to obtain the Friedmann equations

$$\begin{aligned} 3\ddot{a} &= -4\pi G(\rho + 3p)a + \Lambda a \\ 3\dot{a}^2 &= 8\pi G\rho a^2 + \Lambda a^2 - 3k \quad [35]. \end{aligned} \tag{1.24}$$

Defining the Hubble expansion rate

$$H \equiv \frac{\dot{a}}{a}, \tag{1.25}$$

taking the derivative of the second Friedmann Equation and rearranging Eq. (1.24) provides us with the following set of equations that describe the dynamics of the Universe:

$$\begin{aligned} H^2 &= \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \\ \dot{\rho} &= -3H\left(\rho + p\right). \end{aligned} \tag{1.26}$$

We have two equations two solve for the three variables a, p and ρ. Thus, if we want to reconstruct how the Universe evolved in time, we have to introduce a model that gives us an expression for the pressure p. A common assumption is that the cosmic fluid is composed of independent ideal fluids i with a

constant equation of state parameter w_i which relates the pressure of the fluid component p_i with its energy density ρ_i as

$$p_i(t) = w_i \rho_i(t) \quad [35]. \tag{1.27}$$

Recalling $H = \frac{\dot{a}}{a}$ and substituting Eq. (1.27) into Eq. (1.26) gives us the differential equation

$$\frac{\dot{\rho}_i}{\rho_i} = -3\frac{\dot{a}}{a}(1 + w_i) \tag{1.28}$$

whose solution

$$\rho_i \propto a^{-3(1+w_i)} \ [35] \tag{1.29}$$

describes the time evolution of each component of the cosmic fluid. Non-relativistic matter can usually be modelled as pressureless dust, i.e. $w_m = 0$, thus its density just decreases as the volume gets larger ($\rho_m \propto a^{-3}$), as one would expect. For relativistic particles ("radiation"), such as photons, the expansion of the Universe also stretches their wavelengths, leading to an energy that is reduced by an additional factor of a. Its equation of state parameter is therefore $w_r = \frac{1}{3}$.

It is often convenient to express the energy density as a new dimensionless parameter

$$\Omega_i \equiv \frac{\rho_i}{\rho_c}, \tag{1.30}$$

where

$$\rho_c \equiv \frac{3H^2}{8\pi G} \tag{1.31}$$

is the critical density for which, in the absence of a cosmological constant, the universe is flat (cf. Eq. 1.26). The discovery of the present-day accelerated expansion [39,40] suggests that the cosmological constant has an important rôle in defining low-redshift evolution and therefore an equivalent energy density for the cosmological constant

$$\Omega_\Lambda \equiv \frac{\Lambda}{3H^2} \tag{1.32}$$

has been defined, such that the sum of $\Omega_\mathrm{m}, \Omega_\mathrm{r}$ and Ω_Λ indicates whether the Universe is flat or not. In this notation, the cosmological constant Λ can be interpreted as another part of the cosmic fluid, called dark energy, with an equation of state parameter $w_\Lambda = -1$, i.e. the density of dark energy is constant. In a simple extension of the ΛCDM concordance model, this equation

of state can have a value that is slightly different from $w = -1$. In analogy to ΛCDM, such models are often referred to as wCDM. In the presence of curvature, it is also convenient to define an analogous parameter

$$\Omega_k \equiv -\frac{k}{a^2 H^2}. \tag{1.33}$$

One advantage of this parametrisation is that the Hubble expansion rate (cf. Eq. (1.26) can be expressed by today's Hubble rate and today's energy densities

$$H^2 = H_0^2 \sum_i \Omega_{i,0} a^{-3(1+w_i)} \text{ [35]}, \tag{1.34}$$

where today's values of Ω_i and H are written as $\Omega_{i,0}$ and H_0. The scale factor a, implicitly, has been rescaled such that its present value is equal to one. Thus, knowing the present day values of H_0 and of all $\Omega_{i,0}$ allows us to reconstruct the history of the Universe. Table 1.1 lists these values as reported by the Planck Collaboration [7]. Note that Planck does not directly include a value for the radiation density $\Omega_{\mathrm{r},0}$. They do however report the redshift at the time when the radiation density equalled the matter density z_{eq}. We shall see in the next Section 1.1.4 that the redshift is related to the scale factor as $a = \frac{1}{1+z}$ (cf. Eq. (1.43)), so we can use Eq. (1.29) to get the radiation density:

$$\Omega_{\mathrm{r},0}(1+z_{\mathrm{eq}})^4 = \Omega_{\mathrm{m},0}(1+z_{\mathrm{eq}})^3 \Rightarrow \Omega_{\mathrm{r},0} = (9.161 \pm 0.246) \times 10^{-5}. \tag{1.35}$$

Now we have all ingredients at hand to plot Fig. 1.1, where we can see that we are close to the end of a transitioning phase from matter domination to being dominated by dark energy. In the lower panel of Fig. 1.1, one can see that dark energy causes an accelerated expansion in our present epoch, which is the reason why it has been reintroduced into our standard picture of the Universe after the discovery of accelerated expansion in supernova data in 1998 [39, 40]. During the earlier matter dominated era, the Universe's expansion slowed down. Before $\frac{1}{1+z_{\mathrm{eq}}}$, the Universe was dominated by radiation, where the expansion rate was slowing down even more. In this epoch and in the early stages of matter domination, our simple model does not hold yet. The Universe was too dense and hot to have neutral atoms, thus photons Thomson scattered with the free charged particles that later formed the atoms, and therefore radiation and "baryonic" matter were not independent ideal fluids. In cosmological jargon, any form of matter that interacts electromagnetically

Table 1.1: Parameter 68% confidence limits for the base ΛCDM model and for $\Omega_{k,0}$ from Planck CMB power spectra, in combination with lensing reconstruction and external BAO, supernova and Cepheid data [7]. Note that in Planck's base ΛCDM model the Universe is assumed to be flat, i.e. $\Omega_{k,0} = 0$.

H_0	67.74 ± 0.46
$\Omega_{\Lambda,0}$	0.6911 ± 0.0062
$\Omega_{m,0}$	0.3089 ± 0.0062
z_{eq}	3371 ± 23
$\Omega_{k,0}$	$0.0008^{+0.0040}_{-0.0039}$

counts as baryonic, even leptons such as electrons, since most of the mass of "baryonic" matter comes from actual baryons. The epoch of combination, often called recombination for historical reasons, where the Universe cooled down enough to allow neutral hydrogen to form, happened well after radiation-matter-equality, and shortly after making the Universe transparent for the first time, and the radiation released in the process can be observed today as the Cosmic Microwave Background (CMB) radiation. As baryonic matter and dark matter, i.e. matter that does not interact electromagnetically, behaved differently until combination, baryon acoustic oscillations (BAO) were formed after combination. The epoch before (re)combination and BAOs will be discussed in more detail in Sec. 1.4 and 2.1. The very early Universe is the topic of Sec. 1.3. I will first show how the scale factor a can be related to the cosmic redshift z in the next section, which is important in most observations testing the cosmic evolution in this section.

1.1.4 Cosmic Redshift in an Expanding Universe

As the Universe expands, objects move away from us, and the wavelengths of any radiation they emit are elongated. This is a crucial concept to understand most cosmological observations, because, as I shall show in this section, redshift is linked to the scale factor at the time of emission. Moreover, it is also related to the time of emission itself and, due to the finiteness of the speed of light, also to the distance to the object. Therefore, as shown in the previous section, it

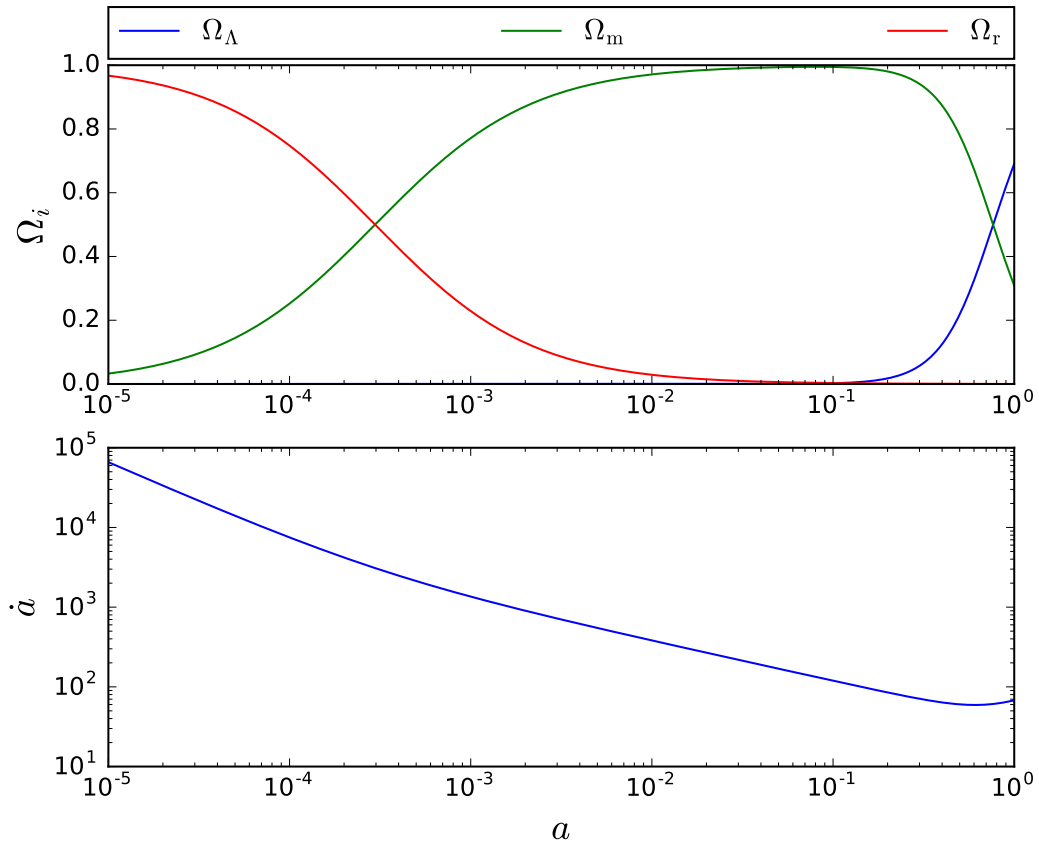

Figure 1.1: The evolution of the density parameters Ω_Λ, Ω_m and Ω_r, as well as the expansion rate $\dot{a}$ with respect to scale factor a.

also depends on the energy content and expansion rate of the Universe, which we commonly want to infer to test certain cosmological models.

Photons are massless and therefore, light travels along null geodesics, which means

$$\mathrm{d}s = 0. \tag{1.36}$$

This means for a radial light-ray in an FLRW universe that

$$\mathrm{d}t = a(t)\mathrm{d}\chi. \tag{1.37}$$

Suppose that an emitter at rest radiates light at a time t_E, then the comoving distance to an observer at $\chi = 0$ and time t_0 can therefore be related to cosmic time and the scale factor by integrating over Eq. (1.37):

$$\chi_\mathrm{E} = \int_0^{\chi_\mathrm{E}} \mathrm{d}\chi = \int_{t_\mathrm{E}}^{t_0} \frac{\mathrm{d}t}{a(t)}. \tag{1.38}$$

Let us further suppose that the emitted light has a frequency ν_E, such that a second light crest is emitted at a time $\frac{1}{\nu_\mathrm{E}}$ later than the first one. Let us furthermore write the frequency at which the radiation is received as ν_0. The comoving distance between the resting emitter and observer is therefore also equal to

$$\chi_\mathrm{E} = \int_{t_\mathrm{E}+\frac{1}{\nu_\mathrm{E}}}^{t_0+\frac{1}{\nu_0}} \frac{\mathrm{d}t}{a(t)}. \tag{1.39}$$

By equating Eq. (1.38) and (1.39) we find

$$\int_{t_0}^{t_0+\frac{1}{\nu_0}} \frac{\mathrm{d}t}{a(t)} = \int_{t_\mathrm{E}}^{t_E+\frac{1}{\nu_\mathrm{E}}} \frac{\mathrm{d}t}{a(t)}. \tag{1.40}$$

Assuming that the frequencies are not extremely low, such that the time between each light crest is small both compared to the time passed between emission and observation, and also compared to the time-scale of the Universe's expansion, we can treat a as constant in these integrals over short time periods. If we again normalise the scale factor such that its present day value is $a(t_0) = 1$, the integrals of Eq. (1.40) can be evaluated to obtain

$$\frac{1}{\nu_0} = \frac{1}{\nu_\mathrm{E}} \frac{1}{a(t_\mathrm{E})}. \tag{1.41}$$

We define the redshift

$$z \equiv \frac{\nu_\mathrm{E} - \nu_0}{\nu_0}. \tag{1.42}$$

as the change in frequency between emission and observation. Using Eq. (1.41), we can find the following useful relationship between the redshift and the scale factor:

$$a(t_{\rm E}) = \frac{1}{1+z}. \tag{1.43}$$

1.1.5 Notions of Distance in an Expanding Universe

What we are still missing to learn about the cosmic history is the time of emission or, equivalently, the distance to the emitter. Distances can be defined differently in an expanding space-time. The distance used in the FLRW metric (cf. Eq. (1.18)) is the **transverse comoving distance**, which is related to the **comoving distance** through Eq. (1.19). There are simple relationships between other distance measures and the transverse comoving distance, such that the latter is related to redshift first and then related to other common distance measures at the end of this subsection.

Going back to Eq. (1.38), we can write the comoving distance to the emitter in terms of the Hubble expansion rate $H = \frac{\frac{{\rm d}a}{{\rm d}t}}{a}$, using the substitution of ${\rm d}t = \frac{{\rm d}a}{aH}$:

$$\chi_{\rm E} = \int_{t_{\rm E}}^{t_0} \frac{{\rm d}t}{a(t)} = \int_{a(t_{\rm E})}^{1} \frac{{\rm d}a}{a^2 H(a)}. \tag{1.44}$$

Taking the derivative of Eq. (1.43)

$$\frac{{\rm d}a}{{\rm d}z} = -\frac{1}{(1+z)^2} = -a^2, \tag{1.45}$$

one can use this result to transform Eq. (1.44) into an integral over redshift:

$$\chi_{\rm E} = \int_0^z \frac{{\rm d}z'}{H(z')}. \tag{1.46}$$

Replacing the scale factor in Eq. (1.34) by the redshift, and inserting it into the previous equation, one obtains an equations that relates the distance, the content of the universe and its expansion by

$$\chi_{\rm E} = \frac{c}{H_0} \int_0^z \frac{{\rm d}z'}{\sqrt{\Omega_{\rm m,0}(1+z')^3 + \Omega_{\rm r,0}(1+z')^4 + \Omega_\Lambda + \Omega_{\rm k}(1+z')^2}}, \tag{1.47}$$

where I explicitly wrote the speed of light c, such that the equation is also valid in the units commonly used in observations. Thus, when we measure the distance and the redshift of many light emitting sources in the Universe,

Eq. (1.43) and (1.47) allow us to test the evolution of energy content and expansion of the Universe.

For some types of observations, other distance measures are more suitable. The comoving and transverse comoving distances have the unintuitive feature of remaining constant even if objects appear to move away from us or from each other. There are two closely related concepts for these two motivations: The **proper distance**

$$d_{\mathrm{p}} = a(t)\tilde{r} \tag{1.48}$$

increases as objects that are at rest with respect to the cosmic fluid recede from us. Similarly, the **angular diameter distance**

$$d_{\mathrm{A}} = a(t)\chi = \frac{\chi}{1+z} \tag{1.49}$$

factors in the universal expansion into the transverse diameter distance. This is especially useful in "standard ruler" experiments, such as measuring the scale of the baryon acoustic oscillations (cf. Sec. 2.1), where the angle θ subtended by a feature, whose physical scale x is known, is measured. The angular diameter distance can then be measured as

$$d_{\mathrm{A}} = \frac{x}{\theta}. \tag{1.50}$$

Another distance measure can be motivated by luminosity measurements. Assuming that an object emits radiation isotropically in Euclidean space, the flux F at a distance d_{L} is related to the intrinsic luminosity L of the object as

$$F = \frac{L}{4\pi d_{\mathrm{L}}^2}. \tag{1.51}$$

In analogy, one can define the **luminosity distance**

$$d_{\mathrm{L}} = \sqrt{\frac{L}{4\pi F}} \tag{1.52}$$

for which Eq. (1.51) holds in a general space-time. As the expansion of the Universe causes a further reduction of the flux due to stretching of the photons' wavelength, the luminosity distance is related to the transverse comoving distance as

$$d_{\mathrm{L}}(z) = (1+z)\chi(z) \text{ [35]}. \tag{1.53}$$

These distance measures are the basis of some of the key observations supporting the ΛCDM concordance model. An overview of these observations follows in the next subsection.

1.1.6 Observational Evidence for the Concordance Model

After having introduced the concordance background model of cosmology, I give a brief overview of some key observations that test it.

The concordance model has been established after the discovery of the accelerated expansion of the Universe through **Supernova Type Ia** (SN Ia) observations [39, 40]. SNe Ia are thermonuclear explosions of stars, which can occur when a white dwarf has burnt all of its nuclear fuel and it has a companion star nearby from which it can accrete enough mass to overcome the Chandrasekhar Mass [41]. As SNe Ia are thought to originate from systems with very similar masses, they are considered standardisable candles, i.e. their intrinsic luminosity L is known after fitting the light curve of a SN. This knowledge can be used to infer the luminosity distance to the supernova from the measured flux F (cf. Eq. (1.53)). Thus, one can fit e.g. Eq. (1.47) to constrain the energy content or, related to that, the equation of state of the Universe. The SN Legacy Survey (SNLS, [42]) gives us > 99.9 per cent confidence that the expansion of the Universe is accelerated, and it puts constraints on the dark energy equation of state of a wCDM universe at $w = -0.91^{+0.16}_{-0.20}(\text{stat})^{+0.07}_{-0.14}(\text{syst})$. This is consistent with a cosmological constant. The SNLS sample has been combined with the SNe observed by the Sloan Digital Sky Survey (SDSS, cf. Sec. 2.2) and reanalysed to form the Joint Light-curve Analysis (JLA, [43]) sample. This provides an equation of state of $w = -1.018 \pm 0.057$, even more consistent with a cosmological constant. JLA [43] also measured a matter energy density of $\Omega_m = 0.295 \pm 0.034$.

The discovery of the existence of the **Cosmic Microwave Background** (CMB) was landmark evidence for the Hot Big Bang model of cosmology [44, 45]. When the Universe became cold enough to form hydrogen atoms, it became transparent for radiation. The Universe had a temperature of about 3000 K at $z \approx 1100$ and the radiation released at that epoch cooled down due to the expansion of Universe and can be observed today as an almost perfectly isotropic background radiation with temperature 2.7260 ± 0.0013 K [46] in the microwave region of the electromagnetic spectrum. Even though it is almost perfectly isotropic, the distribution of deviations of this isotropy are highly informative for cosmological research. The third generation space mission dedicated to measurements of anisotropies in the CMB, called *Planck*,

has put tight constraints on the concordance model [7]. The Planck experiment found that the Universe is remarkably consistent with a spatially flat ΛCDM, and any of the extensions to a simple flat ΛCDM that are considered in [7] are disfavoured by Planck data. Prior to the release of CMB radiation at the time of last scattering, baryons were bound in a plasma, thus forming a strongly coupled fluid with the photons. On account of that they behaved differently from the cold dark matter (cf. Sec. 1.4) and caused a characteristic angular size of fluctuations, called the acoustic scale, in the CMB. These allow us to put constraints on the baryon and cold dark matter content of the Universe, which, in analyses of the CMB alone, are degenerate with the Hubble expansion parameter that is commonly reported as the dimensionless quantity $h \equiv \frac{H_0}{100h\ \mathrm{km/s/Mpc}}$, thus Planck puts constraints on the parameters $\Omega_\mathrm{b}h^2$ and $\Omega_\mathrm{cdm}h^2$. CMB experiments can also measure the Hubble expansion rate H_0, but only in an indirect and highly model dependent way. It is therefore better to combine CMB data with other cosmological data such as lensing, external Baryon Acoustic Oscillation (BAO), SN and Cepheid data. The most important parameters of such a combination also published by the Planck collaboration in [7] has been reproduced in Tab. 1.1.

More accurate measurements of the Hubble expansion rate can be obtained by analysing the imprint of **Baryon Acoustic Oscillations** (BAO, for the physics behind it, see Sec. 1.4) in the distribution of galaxies. For the measuring process I refer to Sec. 2.1, where galaxy redshift surveys will be discussed. BAO measurements are geometric and largely unaffected by non-linear evolution of the matter density field. They are the primary data used to break the degeneracies of the CMB measurements. A summary plot of recent BAO measurements [11, 12, 13, 14, 15, 16] is given in Fig. 1.2. BAO measurements are in good agreement with Planck.

Less consistent with Planck data are measurements of cosmic shear due to **Weak Gravitational Lensing** caused by the large scale structure along the line of sight to observed background galaxies. Without any gravitating foreground objects, averaging over the apparent shape of galaxies should be consistent with a circular shape, even though individual galaxies may have intrinsic ellipticities. Due to the matter field along the line of sight, all background galaxies are affected by lensing, resulting in correlated elliptical shapes.

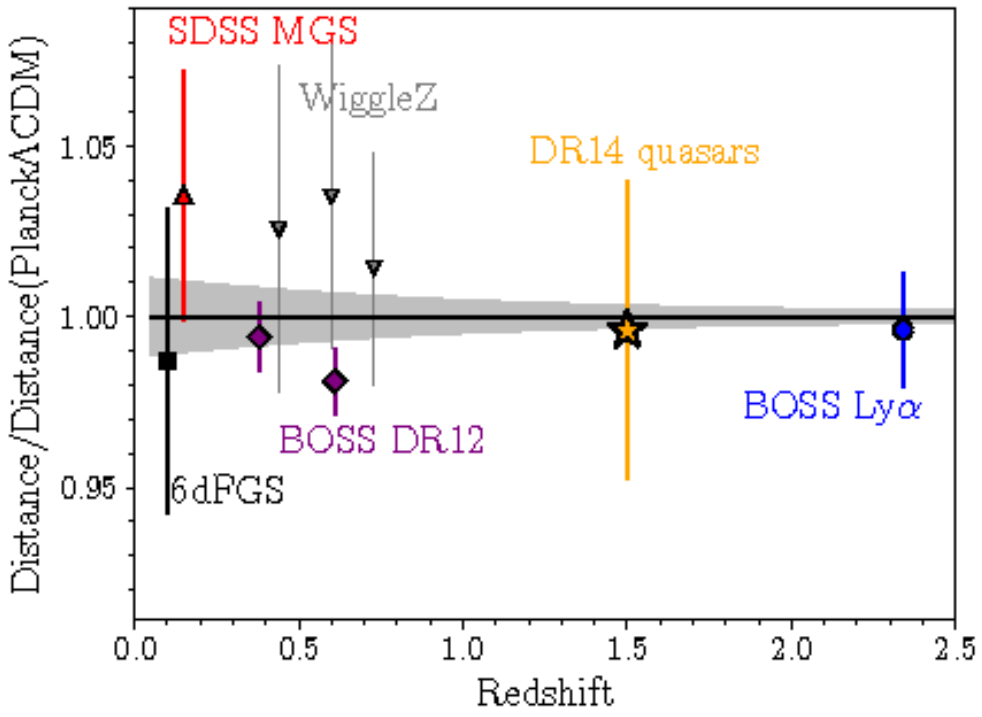

Figure 1.2: Spherically averaged BAO distance measurements (D_{V}) compared to the Planck ΛCDM prediction and extrapolated 68 per cent confidence level (grey region). The eBOSS DR14 quasar sample measurement [11] is shown using a gold star. The other points with 1σ errors are as follows: 6 degree Field Galaxy Survey (6dFGS , [12]); Sloan Digital Sky Survey Main Galaxy Sample (SDSS MGS, [13]); Baryon Oscillation Spectroscopic Survey Data Release 12 (BOSS DR12, [14]); WiggleZ [15]; and BOSS Lyα [16]. Figure taken from [11].

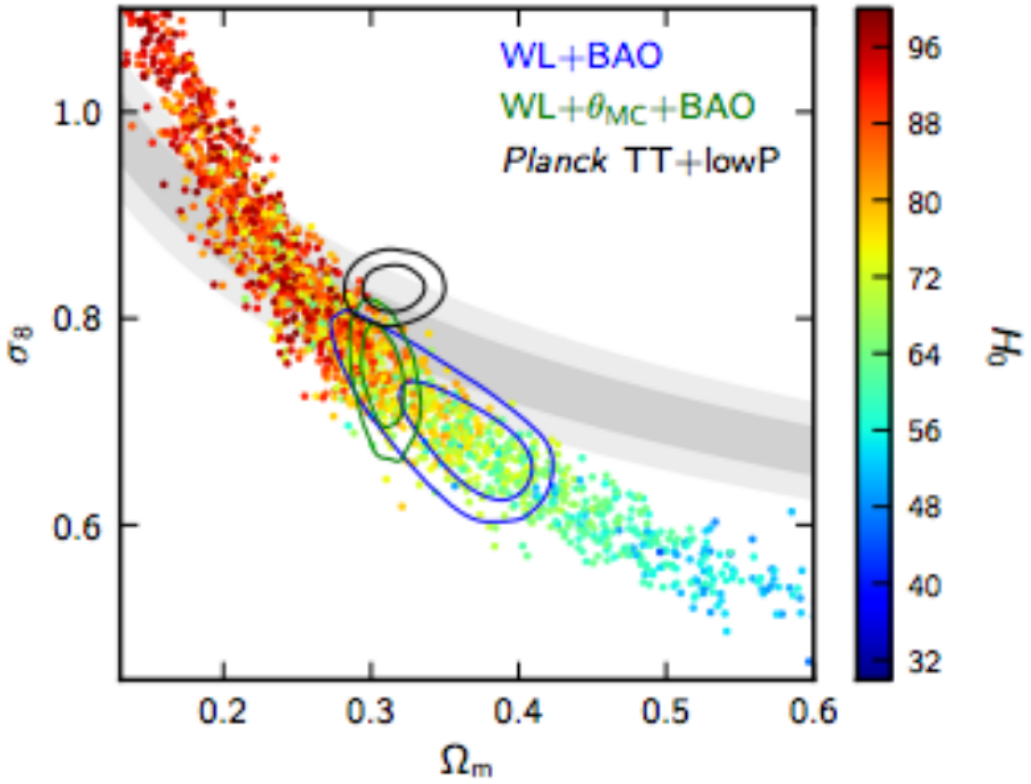

Figure 1.3: Samples in the $\sigma_8 - \Omega_{m,0}$ plane from CFHTLenS data [17], coloured by the value of the Hubble parameter, compared to the joint constraints when the lensing data are combined with BAO (blue), and BAO with the CMB acoustic scale parameter fixed to $\theta_{MC} = 1.0408$ (green). For comparison, the Planck TT+lowP constraint contours are shown in black. The grey bands show the constraint from Planck CMB lensing. Figure taken from [7].

If the sample of background galaxies is split according to their photometric redshifts, one can perform cosmic tomography and obtain information about the 3-dimensional distribution of matter. This can be used to compute the correlation function or power spectrum, as defined in Sec. 1.2.4 and further discussed throughout this thesis for other galaxy clustering data. Similarly as for galaxy clustering and CMB statistics, these can be fitted to theoretical models to obtain cosmological parameters. Cosmic shear data is particularly sensitive to a particular combination of parameters which is $\sigma_8(\Omega_{\mathrm{m},0}/0.27)^{0.46}$, where σ_8 is the variance of the density field smoothed over a scale of 8 Mpc/h, which will be defined in Eq. (1.170). Analysing data from the Canada France Hawaii Telescope Lensing Survey (CFHTLenS, [17, 47]). Heymans *et al.* [17] report a value of

$$\sigma_8(\Omega_{\mathrm{m},0}/0.27)^{0.46\pm0.02} = 0.774 \pm 0.04, \tag{1.54}$$

which is at 2σ discrepant with

$$\sigma_8(\Omega_{\mathrm{m},0}/0.27)^{0.46} = 0.89 \pm 0.03 \tag{1.55}$$

from Planck [7]. The 1σ and 2σ contours of the two measurements is given in Fig. 1.3. The Planck lensing and CFHTLenS contours overlap, but do so at inconsistent H_0 values. However, most of the information in CFHTLenS is in wave numbers where the matter power spectrum is highly non-linear, which makes a direct comparison to CMB and galaxy clustering measurements challenging. A reanalysis [48] of CFHTLenS data that only includes wave numbers $k \leq 1.5h/\mathrm{Mpc}$ is consistent with both CMB and galaxy clustering observations. However, the idea that insufficient modelling of non-linear is the cause for the discrepancy between CMB and lensing data is refuted by Köhlinger *et al.* [18], who found a similar tension using the first 450 square degrees observed by the Kilo Degree Survey (KiDS-450) using only few of the non-linear scales. Their constraints are compared to CFHTLenS, other cosmic shear and CMB measurements in Fig. 1.4.

Distance ladder measurements of local astronomical "standard candles" also favour a larger value of Hubble parameter H_0. Unlike CMB observations, the value they provide is model-independent. However, to obtain the distance to a specific standard candle, one has to anchor its distance scale by comparing

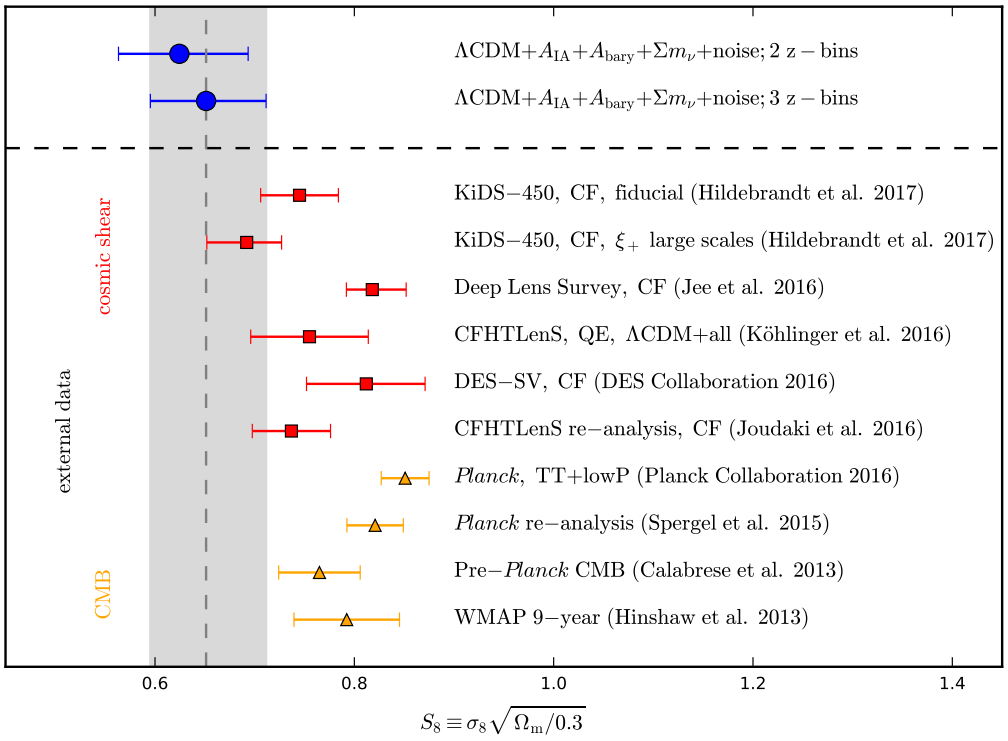

Figure 1.4: The 1σ-constraints of the parameter combination $S_8 \equiv \sigma_8\sqrt{\Omega_{\mathrm{m},0}/0.3}$ from the Kilo Degree Survey (KiDS) for the fiducial model of [18] using a quadratic estimator (QE) of the angular cosmic shear power spectrum in 2 and 3 redshift bins (blue). The next two red points were obtained by a correlation function (CF) analysis of the same survey in 4 redshift bins with and without small scales. The other red points represent 1σ-constraints from other cosmic shear measurements, where the fourth has been obtained by the same authors with similar methods as the blue points. These are compared to constraints from CMB measurements from Planck, WMAP and a combined probe of WMAP, ACT and SPT (Pre-Planck). Figure taken from [18].

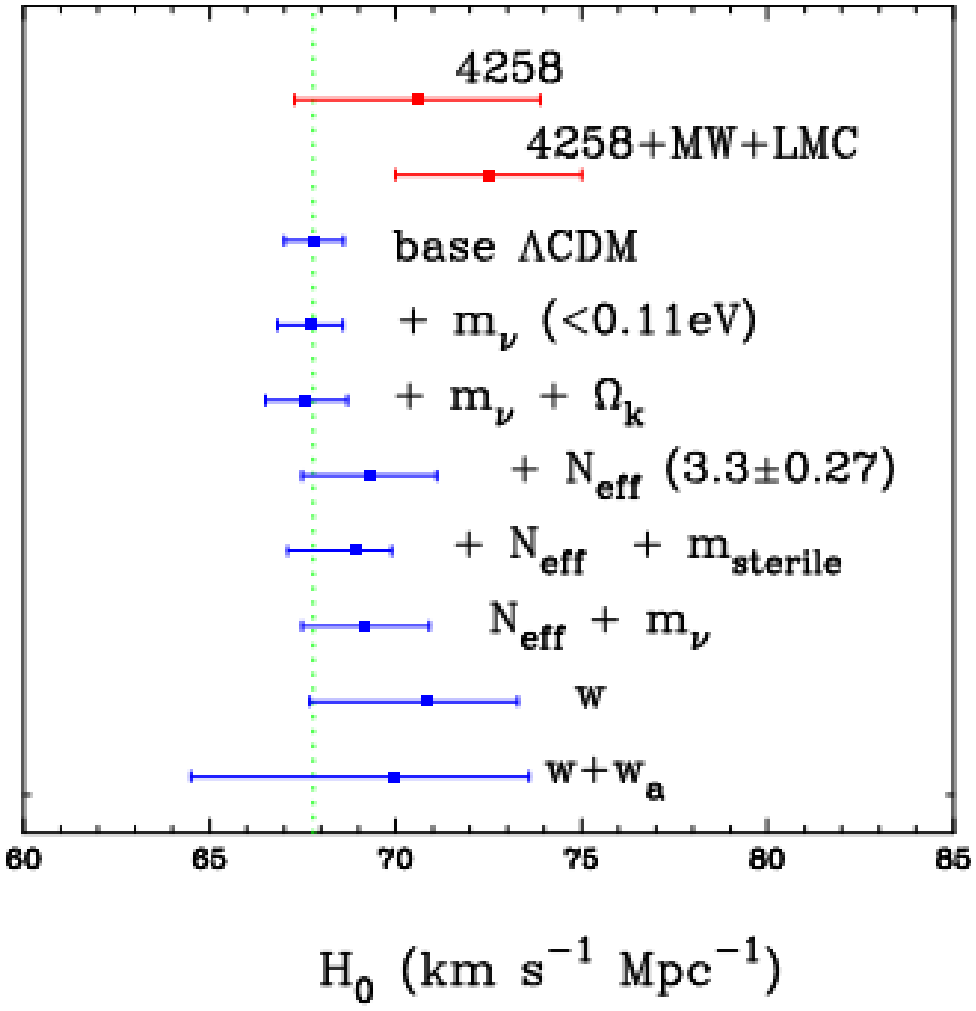

Figure 1.5: The direct estimates (red) of H_0 (together with 1σ error bars) for the NGC 4258 distance anchor and for all three distance anchors. The remaining (blue) points show the constraints from [19] for the base ΛCDM cosmology and some extended models combining CMB data with data from baryon acoustic oscillation surveys. The extensions are as follows: m_ν, the mass of a single neutrino species; $m_\nu + \Omega_k$, allowing a massive neutrino species and spatial curvature; $N_{\rm eff}$, allowing additional relativistic neutrino-like particles; $N_{\rm eff} + m_{\rm sterile}$, adding a massive sterile neutrino and additional relativistic particles; $N_{\rm eff} + m_\nu$, allowing a massive neutrino and additional relativistic particles; w, dark energy with a constant equation of state; $w + w_{\rm a}$, dark energy with a time varying equation of state. Figure taken from [20].

the luminosity of that particular type of standard candle to the luminosity of another type of standard candle within the same galaxy. Such anchoring has to be done several times until one can use objects close enough for parallax to anchor them. Any systematic in one of the lower "rungs" of this distance ladder affects any distance measurement to farther objects. This is why the same data can yield two different values of $H_0 = 73.24 \pm 1.74$ km s^{-1} Mpc^{-1} [49] and $H_0 = 72.5 \pm 2.5$ km s^{-1} Mpc^{-1} [20]. For a particular anchor, NGC 4258, the local H_0 measurement is consistent with its Planck value (cf. Fig. 1.5, [20]). Other direct local H_0 measurements hint towards extending the basic ΛCDM concordance model with, e.g. more relativistic particles or dark energy that is not a cosmological constant (cf. Fig. 1.5). Still, the tension between the two different values of the Hubble parameter does not necessarily mean that ΛCDM is wrong, because as there are density fluctuations in a ΛCDM cosmology, there are also fluctuations in the Hubble parameter. The local Hubble parameter might be different from the global one by the same order of magnitude as the observed discrepancy between the local and CMB measurements [50, 51].

Overall, we have seen that the concordance model is supported by several key observations, including SNe, BAO and CMB experiments. Lensing and local H_0 measurements do not agree with the other measurements, which might hint to a deviation from the concordance model. However, when only taking information from more linear scales, lensing is consistent with BAO and CMB observations, such that one can speculate that the concordance model is a sufficient description of our Universe and the discrepancy between lensing and other pillar observations is due to not well understood non-linear physics. Similarly, the tension between local and geometric measurements of the Hubble parameter H_0 could also point towards new physics, but might also be due to not yet understood systematics or due to fluctuations of the Hubble parameter. The next section will discuss perturbations around the concordance background model, which will also help understanding some of the measurements presented here and will also be needed for $f_{\rm NL}$ measurements.

1.2 From Perturbations to Structure

1.2.1 Perturbing the Continuity, Poisson and Euler Equations

In the previous Section, I have described the background evolution of the Universe, which is homogeneous and isotropic on large scales. On smaller scales however, we see structures such as stars, galaxies or galaxy clusters, thus, there are perturbations around the homogeneous and isotropic background. The proper way to take these perturbations into account is to split the metric tensor, the Einstein tensor, the energy-momentum tensor and the 4-velocity into homogeneous and isotropic background parts and space-time dependent perturbations:

$$\begin{aligned} g_{\mu\nu}(\mathbf{x},t) &= \bar{g}_{\mu\nu}(t) + \Delta g_{\mu\nu}(\mathbf{x},t), \\ u^{\mu}(\mathbf{x},t) &= \bar{u}^{\mu}(t) + \Delta u^{\mu}(\mathbf{x},t), \\ T_{\mu\nu}(\mathbf{x},t) &= \bar{T}_{\mu\nu}(t) + \Delta T_{\mu\nu}(\mathbf{x},t), \\ G_{\mu\nu}(\mathbf{x},t) &= \bar{G}_{\mu\nu}(t) + \Delta G_{\mu\nu}(\mathbf{x},t). \end{aligned} \tag{1.56}$$

Equations to derive the evolution of these perturbations can be derived from the perturbed Einstein equation (cf. Eq. 1.5)

$$\Delta T_{\mu\nu} = -8\pi G \Delta G_{\mu\nu} \tag{1.57}$$

and energy-momentum conservation

$$\Delta T^{\mu\nu}{}_{;\nu} = 0. \tag{1.58}$$

The main problem is that due to the perturbations, one cannot define a unique rest frame for the cosmic fluid and therefore one has to make a "gauge" choice. Otherwise, the derivations are tedious and not very illustrative for the scope of this thesis. Therefore, a Newtonian discussion of perturbation theory is presented here instead and GR corrections are mentioned later without derivation when they are necessary. When talking about perturbations, it is assumed that the perturbation is much smaller than the background quantity.

To facilitate this discussion, let me introduce the conformal time coordinate η with $a\mathrm{d}\eta = \mathrm{d}t$, such that the spatially flat background is just a simple

Minkowski space with the scale factor taking the expansion of the Universe into account:

$$\mathrm{d}\bar{s}^2 = a^2(\eta)(-\mathrm{d}\eta^2 + \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j). \tag{1.59}$$

In the weak field approximation, we can therefore apply classical fluid dynamics in non-expanding space and then embed these by replacing time derivatives with conformal time derivatives and spatial derivatives $\nabla \to \frac{1}{a}\nabla$. Conformal time derivatives are denoted by a prime. To be consistent, one has to define a conformal Hubble parameter $\mathcal{H} \equiv \frac{a'}{a}$ in analogy to the standard Hubble parameter, but with a conformal time derivative. The Newtonian equivalent of perturbing the energy-momentum tensor is perturbing the energy density and the pressure:

$$\begin{aligned} \rho(\mathbf{x}, \eta) &= \bar{\rho}(\eta) + \Delta\rho(\mathbf{x}, \eta), \\ p(\mathbf{x}, \eta) &= \bar{p}(\eta) + \Delta p(\mathbf{x}, \eta). \end{aligned} \tag{1.60}$$

The density perturbations are more commonly expressed by the over-density field or density contrast

$$\delta(\mathbf{x}, \eta) \equiv \frac{\Delta\rho(\mathbf{x}, \eta)}{\bar{\rho}(\eta)} = \frac{\rho(\mathbf{x}, \eta) - \bar{\rho}(\eta)}{\bar{\rho}(\eta)}. \tag{1.61}$$

As the Newtonian equations are only valid for non-relativistic matter for which the background pressure has been assumed negligible, we only have pressure perturbations that, assuming they are only adiabatic, are related to density perturbation by the speed of sound

$$c_\mathrm{s}^2 \equiv \frac{\Delta p}{\Delta\rho}. \tag{1.62}$$

In GR, the metric mediates the gravitational force, thus the Newtonian analogy to scalar metric perturbations is perturbing the gravitational potential

$$\Phi(\mathbf{x}, \eta) = \bar{\Phi}(\mathbf{x}, \eta) + \Delta\Phi(\mathbf{x}, \eta). \tag{1.63}$$

We shall soon see that the background potential indeed can have different values in different spatial positions. The last remaining ingredient are velocity perturbations. As the background is defined as the rest frame of the cosmic fluid, the background velocity is just given by the Hubble flow:

$$\mathbf{u}(\mathbf{x}, \eta) = \mathcal{H}\mathbf{x} + \Delta\mathbf{u}(\mathbf{x}, \eta) \tag{1.64}$$

Now one can apply classical fluid dynamics to find equations that govern the evolution of these perturbations:

Continuity Equation In fluid dynamics, conservation of energy, or in fact any conserved quantity, is realised by the Continuity Equation

$$\rho' + \nabla \mathbf{u}(\rho + p) = 0. \tag{1.65}$$

At background level, $\nabla \bar{\rho} = 0$ and $\nabla \mathbf{u} = 3\mathcal{H}$, such that we find Eq. (1.26). To insert the perturbations, I rearrange Eq. (1.61) to obtain

$$\rho = \bar{\rho}(1 + \delta). \tag{1.66}$$

Substituting Eq. (1.64) into Eq. (1.65) yields

$$\bar{\rho}'(1 + \delta) + \bar{\rho}\delta' + \bar{\rho}(1 + \delta)(3\mathcal{H} + \nabla\Delta\mathbf{u}) + (\mathcal{H}\mathbf{x} + \Delta\mathbf{u})\bar{\rho}\nabla\delta = 0. \tag{1.67}$$

After subtracting off and inserting the background solution (cf. Eq. 1.26), we obtain

$$\delta' + \theta(1 + \delta) + (\mathcal{H}\mathbf{x} + \Delta\mathbf{u})\nabla\delta = 0, \tag{1.68}$$

where the divergence of the velocity perturbation has been defined as a new parameter

$$\theta \equiv \nabla\Delta\mathbf{u}. \tag{1.69}$$

Poisson Equation In Newtonian Gravity, interactions are mediated by a gravitational potential Φ that is connected with the density by Poisson's equation

$$\frac{1}{a^2}\nabla^2\Phi = 4\pi G\rho. \tag{1.70}$$

This is a modified version of Poisson's equation with a factor of $1/a^2$ on the left hand side to account for the expansion of the Universe by a coordinate transformation from a non-expanding space to an expanding one, as motivated below Eq. (1.59). The background equation

$$\nabla^2\bar{\Phi} = 4\pi G a^2\bar{\rho} \tag{1.71}$$

can easily be solved by integration:

$$\bar{\Phi} = \frac{2\pi G}{3}a^2\bar{\rho}\mathbf{x}^2 + C(\eta), \tag{1.72}$$

where $C(t)$ is an arbitrary integration constant that can vary in time though. With the background equation subtracted off, the Poisson equation reads

$$\nabla^2\Delta\Phi = 4\pi G a^2\bar{\rho}\delta. \tag{1.73}$$

Euler Equation The velocity field for an ideal fluid is described by Euler's equation

$$\rho\left(\mathbf{u}' + \frac{1}{2}\nabla\mathbf{v}^2 - \mathbf{v}\times\boldsymbol{\omega}\right) = -\nabla p - \rho\nabla\Phi. \tag{1.74}$$

This is a fluid dynamical formulation of Newton's second law, where the gravitational force is expressed as the divergence of the potential and an additional force term due to pressure acting against the gravitational force. The left hand side is the "material derivative" of the velocity field and includes the rotation of the velocity field called the vorticity $\boldsymbol{\omega} \equiv \nabla \times \boldsymbol{v}$. For the background, we have

$$\bar{\rho}(\mathcal{H}'\mathbf{x} + \mathcal{H}^2\mathbf{x}) = -\bar{\rho}\nabla\bar{\Phi} = -\frac{4\pi G}{3}a^2\bar{\rho}^2\mathbf{x}, \tag{1.75}$$

where the last equality comes from inserting Eq. (1.72). For the background, this yields the Friedmann Equation (1.24)

$$\mathcal{H}' + \mathcal{H}^2 = -\frac{4\pi G}{3}a^2\bar{\rho}. \tag{1.76}$$

Assuming the vorticity can be neglected, the evolution of perturbations follows

$$\Delta\mathbf{u}' + \mathcal{H}\Delta\mathbf{u} + \frac{1}{2}\nabla\Delta\mathbf{u}^2 = -c_\mathrm{s}^2\frac{\nabla\delta}{1+\delta} - \nabla\Delta\Phi \tag{1.77}$$

Knowing the background expansion rate $\mathcal{H}$ and the potential Φ, one can solve Eq. (1.73) to (1.77) to see how structure in the Universe, modelled as perturbations in the energy density ρ. This is done first for linear perturbations only in the next section, before going to higher orders in Sec. 1.2.3.

1.2.2 Linear Perturbations

Having assumed that the perturbations are small compared to the background, it is sufficient, for the scope of this thesis, to linearise Eq. (1.68), (1.73) and (1.77), i.e. omit all products of perturbation terms. The linearised equations we want to solve read

$$\begin{aligned} \delta' + \theta &= 0, \\ \nabla^2\Delta\Phi - 4\pi G a^2\bar{\rho}\delta &= 0, \\ \Delta\mathbf{u}' + \mathcal{H}\Delta\mathbf{u} + c_\mathrm{s}^2\nabla\delta + \nabla\Delta\Phi &= 0. \end{aligned} \tag{1.78}$$

Into the divergence of the linear Euler equation

$$\theta' + \mathcal{H}\theta + c_s^2\nabla^2\delta + \nabla^2\Delta\Phi = 0 \tag{1.79}$$

one can insert the linear Poisson and continuity equations to obtain a closed form expression for δ:

$$\delta'' + \mathcal{H}\delta' - c_s^2\nabla^2\delta + 4\pi G\bar{\rho}a^2\delta = 0. \tag{1.80}$$

This equation provides some intuitive insight into the process of structure formation. The last term describes the growth of structure due to gravitational attraction, which is opposed by the pressure term in front of it. The second term has the form of a friction term and shows that the expansion of the Universe also opposes structure growth. In Fourier-Space, this equation reads

$$\delta_{\mathbf{k}}'' + \mathcal{H}\delta_{\mathbf{k}}' - \left(c_s^2k^2 - 4\pi G\bar{\rho}a^2\right)\delta_{\mathbf{k}} = 0. \tag{1.81}$$

The solution of this equation depends on the value of k. At scales $k > \frac{2a}{c_s}\sqrt{\pi G\bar{\rho}}$, it is given by a pressure-supported sound-wave, at lower k-values, a density perturbation collapses. Thus, this provides an important criterion for whether structures can grow through gravitational collapse and the wave number

$$k_{\mathrm{J}} = \frac{2a}{c_s}\sqrt{\pi G\bar{\rho}} \tag{1.82}$$

is called the Jeans wave number and its corresponding scale

$$\lambda_{\mathrm{J}} = \frac{2\pi}{k_{\mathrm{J}}} = \frac{c_{\mathrm{s}}}{a}\sqrt{\frac{\pi}{G\bar{\rho}}} \tag{1.83}$$

is called the Jeans scale.

After photon decoupling, non-gravitational interaction can be neglected in the cosmic fluid and thence $c_{\mathrm{s}} = 0$. Therefore, the scale-dependent term in Eq. (1.81) vanishes and one can decompose the matter over-density field after photon decoupling as the product of the primordial over-density $\delta_{\mathbf{k}}(0)$, the time dependent linear growth factor $D_1(\eta)$ and a scale dependent part, called the transfer function $T(\mathbf{k})$:

$$\delta_{\mathbf{k}}(\eta) = T(\mathbf{k})D_1(\eta)\delta_{\mathbf{k}}(0). \tag{1.84}$$

The matter transfer function $T(\mathbf{k})$ encapsulates the scale dependent changes to the matter density field before (re)combination that will be discussed in Sec. 1.4. Due to Eq. (1.84), Eq. (1.81) reduces to the differential equation of $D_1(\eta)$

$$D_1''(\eta) + \mathcal{H}D_1'(\eta) - 4\pi G\bar{\rho}a^2D_1(\eta) = 0. \tag{1.85}$$

Using the definition of the matter density parameter Ω_m Eq. (1.30), one can rewrite the zeroth order term such that

$$D_1''(\eta) + \mathcal{H} D_1'(\eta) - \frac{3}{2}\Omega_\mathrm{m}(\eta)\mathcal{H}^2 D_1(\eta) = 0. \tag{1.86}$$

The solution of this equation is given by

$$D_1(\eta) = \frac{5}{2} H(\eta)\Omega_\mathrm{m}(\eta) \int_0^{a(\eta)} \frac{\mathrm{d}a}{a^3 H^3(a)} \text{ [52]} \tag{1.87}$$

When inserting Eq. (1.84) into the first of Eq. (1.78)

$$\theta = -\frac{\mathrm{d}\delta}{\mathrm{d}\eta} = -\frac{\mathrm{d}\delta}{\mathrm{d}D_1}\frac{\mathrm{d}D_1}{\mathrm{d}\eta} = -\frac{\delta}{D_1}\frac{\mathrm{d}D_1}{\mathrm{d}a}\frac{\mathrm{d}a}{\mathrm{d}\eta} \equiv -f\mathcal{H}\delta, \tag{1.88}$$

it is common to introduce the parameter

$$f \equiv \frac{\mathrm{d}\ln(D_1)}{\mathrm{d}\ln(a)}. \tag{1.89}$$

It can be approximated by

$$f(z) \approx \Omega_\mathrm{m}^{\frac{4}{7}}(z) + \frac{\Omega_\Lambda(z)}{70}\left(1 + \frac{\Omega_\mathrm{m}(z)}{2}\right) \text{ [53]}. \tag{1.90}$$

When we compare a cosmological model with data, the model usually does not make predictions for the density field itself, but rather for summary statistics such as the power spectrum. These are going to be discussed in Sec. 1.2.4. But first, I touch on higher perturbative orders in the next subsection.

1.2.3 Lagrangian Perturbation Theory

So far, I have considered linear perturbations in the Eulerian frame. Classical field theory can also be described in the Lagrangian frame, which is a useful concept especially when studying higher order perturbations. The difference between the two frames is that in an Eulerian viewpoint, the focus is on field values at given coordinates in time and space, whereas a Lagrangian observer follows the path line of a fluid parcel. The coordinates of the two reference frames are linked by the displacement field $\mathbf{\Psi}(\mathbf{q}, z)$, which depends on the Lagrangian spatial coordinate $\mathbf{q}$ and on cosmic redshift z. The Eulerian coordinate $\mathbf{x}$ can be obtained by

$$\mathbf{x}(\mathbf{q}, z) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, z). \tag{1.91}$$

The Jacobian of the transformation given in Eq. (1.91) is

$$\mathcal{J} = \det\left[\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\mathbf{q}}\right] = \det\left[\delta_{ij} + \frac{\partial \mathbf{\Psi}_i}{\partial \mathbf{q}_j}\right]. \tag{1.92}$$

Assuming that at an initial redshift z_{in}, when the Eulerian and Lagrangian reference frames were equal, the density was uniform, the Jacobian is directly related to the Lagrangian density contrast $\delta_{\mathrm{L}}(\mathbf{q}, z)$. This is due to mass conservation, i.e. the mass at an arbitrary time in Eulerian coordinates equals the initial uniformly distributed mass in Lagrangian coordinates

$$M = a^3(z)\rho(\mathbf{x}, z)\mathrm{d}\mathbf{x} = a^3(z_{\mathrm{in}})\bar{\rho}(z_{\mathrm{in}})\mathrm{d}\mathbf{q}. \tag{1.93}$$

Transforming the first part into Lagrangian coordinates, one obtains

$$a^3(z)\rho(\mathbf{q}, z)\mathcal{J}\mathrm{d}\mathbf{q} = a^3(z_{\mathrm{in}})\bar{\rho}(z_{\mathrm{in}})\mathrm{d}\mathbf{q}, \tag{1.94}$$

which can be rearranged to find the aforementioned link between the Jacobian and Lagrangian density contrast

$$\delta_{\mathrm{L}}(\mathbf{q}, z) \equiv \frac{\rho(\mathbf{q}, z) - \bar{\rho}(z)}{\bar{\rho}(\mathbf{q}, z)} = \frac{1}{\mathcal{J}} - 1. \tag{1.95}$$

Both the displacement field and the density field can be expanded into perturbative series:

$$\begin{aligned} \mathbf{\Psi}(\mathbf{q}, z) &= \sum_{n=1}^{\infty} \mathbf{\Psi}^{(n)}(\mathbf{q}, z), \\ \delta_{\mathrm{L}}(\mathbf{q}, z) &= \sum_{n=1}^{\infty} \delta_{\mathrm{L}}^{(n)}(\mathbf{q}, z). \end{aligned} \tag{1.96}$$

By equating the two expressions for the Jacobian of Eq. (1.92) and (1.95) and inserting the series up to first order, one finds the first order equation, which is commonly called the **Zel'dovich approximation**

$$\delta_{\mathrm{L}}^{(1)}(\mathbf{q}) = -\nabla_{\mathbf{q}} \mathbf{\Psi}^{(1)} \text{ [54]}. \tag{1.97}$$

Going one order further, the density field in second order Lagrangian Perturbation Theory (**2LPT**) reads

$$\delta_{\mathrm{L}}^{(2)}(\mathbf{q}) = -\nabla_{\mathbf{q}} \mathbf{\Psi}^{(2)} + \frac{1}{2}\left[\delta_{\mathrm{L}}^{(1)}(\mathbf{q})^2 + \sum ij \frac{\partial \mathbf{\Psi}_i^{(1)}}{\partial \mathbf{q}_j} \frac{\partial \mathbf{\Psi}_j^{(1)}}{\partial \mathbf{q}_i}\right] \tag{1.98}$$

The only dynamical variable in the equivalent of Eq. (1.80) in the Lagrangian framework is the time η, thus one can rewrite the displacement field $\mathbf{\Psi}^{(1)}(\mathbf{q},\eta) = D_1(\eta)\mathbf{\Psi}^{(1)}(\mathbf{q},0)$ using the linear growth function $D_1(\eta)$ as in the Eulerian framework. The first order differential equation of motion is then

$$D_1'' + \mathcal{H}D_1' - \frac{3}{2}\Omega_{\rm m}\mathcal{H}^2 D_1 = 0 \tag{1.99}$$

and its solution is the same as in the Eulerian framework (cf. Eq. 1.87). The equation for the second order growth rate D_2 is given by

$$D_2'' + \mathcal{H}D_2' - \frac{3}{2}\Omega_{\rm m}\mathcal{H}^2 D_2 = -\frac{3}{2}\Omega_{\rm m}\mathcal{H}^2 D_1^2, \tag{1.100}$$

to get the quadratic growth function in a flat ΛCDM universe

$$D_2 \approx -\frac{3}{7}D_1^2\Omega_{\rm m}^{-1/143}\ [55]. \tag{1.101}$$

To solve the full development of the density distribution, one has to resort to N-body simulations [56]. Lagrangian Perturbation is, however, a very useful tool to both set up the initial displacement of particles in such simulations and to accelerate the creation of mock catalogues, which I shall go into in Sec. 3.2.

1.2.4 Statistics of Random Fields

I have presented equations describing the time evolution of the density field δ. However, we are not interested in its actual form, but rather in its distribution. Theoretically, the density field is commonly modelled as a random realisation drawn from a suitable probability distribution. To learn about its probability distribution, we have to average over many different realisations of them. A fundamental problem of cosmology is, however, that we can only observe one universe, and that also only from our very own space-time position. One can show that, for modes whose sizes are way shorter than the distance between opposite sides of the survey volume, the ensemble and spatial averages are equivalent. For modes that have sizes closer to the survey scale, the precision of measuring a survey statistic is more and more limited by what we call cosmic variance.

Considering a random field $\varphi(\mathbf{x})$ whose value at each spatial position has a probability given by a homogeneous and isotropic probability distribution function, one can introduce some commonly used survey statistics:

- The **2-point correlation function**

$$\xi(r) \equiv \langle \varphi(\mathbf{x})\varphi(\mathbf{x}+\mathbf{r})\rangle \tag{1.102}$$

describes the probability of finding two fluctuations that are a distance r away from each other. Due to statistical homogeneity and isotropy, ξ only depends on the absolute value $r = |\mathbf{r}|$.

- The **power spectrum**

$$P_{\varphi\varphi}(\mathbf{k}) \equiv (2\pi)^{-3/2} \int \mathrm{d}^3\mathbf{r}\xi(r)e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{1.103}$$

can be defined as the Fourier transform of the 2-point function and is related to the 2-point correlator of the field in Fourier space:

$$\begin{aligned}\langle \varphi(\mathbf{k}_1)\varphi^*(\mathbf{k}_2)\rangle &= \frac{1}{8\pi^3}\int \mathrm{d}^3\mathbf{x}\int \mathrm{d}^3\mathbf{r}\,\langle \varphi(\mathbf{x}+\mathbf{r})\varphi(\mathbf{x})\rangle\, e^{-i\mathbf{k}_1\cdot(\mathbf{x}+\mathbf{r})}e^{i\mathbf{k}_2\cdot\mathbf{x}} \\ &= \frac{\delta_D(\mathbf{k}_1-\mathbf{k}_2)}{\sqrt{8\pi^3}}\int \mathrm{d}^3\mathbf{r}\xi(r)e^{-i\mathbf{k}_1\cdot\mathbf{r}} \\ &= \delta_D(\mathbf{k}_1-\mathbf{k}_2)P_{\varphi\varphi}(\mathbf{k}_1).\end{aligned} \tag{1.104}$$

Due to isotropy, the power spectrum depends only on the absolute value of $\mathbf{k}$.

- In a similar way, higher order correlators are related to the **bispectrum** $B_{\varphi\varphi\varphi}(\mathbf{k}_1,\mathbf{k}_2)$, **trispectrum** $T_{\varphi\varphi\varphi\varphi}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$, etc., through

$$\begin{aligned}\langle \varphi(\mathbf{k}_1)\varphi(\mathbf{k}_2)\varphi(\mathbf{k}_3)\rangle &= \delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B_{\varphi\varphi\varphi}(\mathbf{k}_1,\mathbf{k}_2), \\ \langle \varphi(\mathbf{k}_1)\varphi(\mathbf{k}_2)\varphi(\mathbf{k}_3)\varphi(\mathbf{k}_4)\rangle &= \delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4)T_{\varphi\varphi\varphi\varphi}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3), \text{ etc.}\end{aligned} \tag{1.105}$$

Again due to isotropy, the bispectrum is a function of the shape of the triangle spanned by $\mathbf{k}_1$ and $\mathbf{k}_2$ only, the trispectrum of its respective tetrahedron, and any other polyspectrum depends on the shape of its corresponding polyhedron. Instead of the two vectors $\mathbf{k}_1$ and $\mathbf{k}_2$, one can define the triangle also by the lengths of the three vectors $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3 \equiv \mathbf{k}_1 - \mathbf{k}_2$, thus

$$B_{\varphi\varphi\varphi}(\mathbf{k}_1,\mathbf{k}_2) = B_{\varphi\varphi\varphi}(k_1,k_2,k_3) \tag{1.106}$$

Considering a zero mean Gaussian field φ_G, it has been shown by Isserlis [57] that for odd values of n, the n-point correlators of φ_G are zero and, for even numbers of n, they can be expressed by powers of the power spectrum [57] (This has been applied to products of Gaussian creation and annihilation operators in quantum field theory by Gian-Carlo Wick, and is therefore known as Wick's theorem to the physics community). The bispectrum of a Gaussian field therefore vanishes, and one can therefore use the bispectrum to define a test parameter describing how non-Gaussian the distribution of the field is:

$$f_{\mathrm{NL}}(k_1, k_2, k_3) \equiv \frac{B_{\varphi\varphi\varphi}(k_1, k_2, k_3)}{2\left[P_{\varphi\varphi}(k_1)P_{\varphi\varphi}(k_2) + P_{\varphi\varphi}(k_2)P_{\varphi\varphi}(k_3) + P_{\varphi\varphi}(k_3)P_{\varphi\varphi}(k_1)\right]}. \tag{1.107}$$

This is usually studied in several limits corresponding to different shapes of the k_1-k_2-k_3-triangle (without loss of generality, $k_1 \leq k_2 \leq k_3$ is assumed):

- Squeezed limit ($k_1 \ll k_2 \approx k_3$): In this limit, f_{NL} corresponds to the second-order expression in a perturbative expansion of the field φ as a local function of a Gaussian field φ_G:

  $$\varphi(\mathbf{x}) = \varphi_G(\mathbf{x}) + f_{\mathrm{NL}}\left(\varphi_G^2(\mathbf{x}) - \langle\varphi_G^2\rangle\right) + g_{\mathrm{NL}}\left(\varphi_G^3(\mathbf{x}) - 3\langle\varphi_G^2\rangle\varphi_G(\mathbf{x})\right) + \mathcal{O}(\varphi_G^4) \text{ [58]}. \tag{1.108}$$

  This type of non-Gaussianity is therefore called **local**. It occurs when primordial non-Gaussianity is generated on super-horizon scales (e.g. [19]). In the remainder of this thesis, all mentions of f_{NL} are understood as being of the local type.

- **Equilateral** ($k_1 = k_2 = k_3$) non-Gaussianity is produced at horizon crossing during inflation [59].

- Some string-theory models of inflation predict **folded** configurations ($k_1 + k_2 = k_3$) [60].

- Another configuration is called **orthogonal** because it is a linear combination of the equilateral and folded configurations to which it is orthogonal [61]. This configuration does not correspond to one particular shape and can therefore not be described in terms of k values.

On large scales, local type non-Gaussianity alters the galaxy bias, which shall be introduced in the next subsection.

1.3 The Origin of Cosmic Perturbations

Having described how perturbations evolve in time in the matter dominated era, this section is concerned with the question where the perturbations come from. I begin with introducing three problems that arise in a hot big bang cosmology, before introducing models that create perturbations while solving these problems. As this thesis is about testing primordial non-Gaussianity, I present the prediction of $f_{\rm NL}$ for each of these models.

1.3.1 Problems of Hot Big Bang Cosmology

Flatness Problem

Current observations are compatible with a flat universe. Planck constrains $|\Omega_{k,0}| < 0.0048$ at 68% confidence level (cf. Tab. 1.1, [7]). As the curvature varies as k/a^2 (cf. Eq. (1.33)), its initial value must have been fine-tuned to a value that is several orders of magnitude smaller. A small change would let us observe a completely different, non-flat, Universe today.

Horizon Problem

The isotropy of the CMB could have only been produced if all of it had been in causal contact. However, the horizon at the time of last scattering was $d_{\rm hor} \approx 0.4$ Mpc, which corresponds to an angle of 2° on the CMB map [62]. That means that, in a pure hot big bang cosmology, most of the CMB map has not been in causal contact prior or during the epoch the CMB was produced.

Monopole problem

This problem is of a more hypothetical nature than the previous ones and is given here for historical reasons. The problem arises only if there is a Grand Unified Theory (GUT), which is not established yet. Assuming that there is a valid GUT, going back in time, the Universe was so dense that all fundamental forces would be unified. During GUT phase transitions, massive magnetic monopoles would have been produced with huge number densities [62], which we still should observe nowadays. However, no monopoles have been observed, which could either provide another reason to introduce inflation, or

it just means that there is no GUT. However, the flatness and horizon problems sufficiently motivate inflation, which also acts as the source of initial perturbations.

1.3.2 Inflation

An elegant way to solve the previously mentioned problems, while creating initial perturbations, is inflation, which is defined as a period of accelerated expansion of the scale factor, similar to our present epoch likely dominated by a cosmological constant. According to Eq. (1.34), the Hubble parameter is constant for anything cosmological constant-like with $w = -1$. The scale factor then evolves like

$$a(t) = a_{\mathrm{in}} e^{H(t-t_{\mathrm{in}})}, \tag{1.109}$$

where a_{in} is the scale factor at the time t_{in} when inflation begins. If inflation ends at time t_{end}, the Universe has grown exponentially by

$$\frac{a(t_{\mathrm{end}})}{a(t_{\mathrm{in}})} = e^{H(t_{\mathrm{end}}-t_{\mathrm{in}})} \equiv e^{N}, \tag{1.110}$$

where the *number of e-folds*

$$N \equiv \int_{t_{\mathrm{in}}}^{t_{\mathrm{end}}} \mathrm{d}t H(t) \tag{1.111}$$

has been defined.

This is already enough to solve the flatness problem, as inflation flattens any curvature due to

$$\frac{|\Omega_k(t_{\mathrm{end}})|}{|\Omega_k(t_{\mathrm{in}})|} = e^{-2N}. \tag{1.112}$$

Thus, even if there was a "natural" value of $\Omega_k(t_{\mathrm{in}}) \sim 1$ before inflation, it is suppressed by several orders of magnitudes and, therefore, explains why the Universe is (at least almost) flat today. To get sufficient suppression, the number of e-folds should be about $N \lesssim 60$ [62].

The horizon problem is also solved by inflation, as the horizon

$$d_{\mathrm{hor}}(t) = a(t) \int_0^t \frac{\mathrm{d}t'}{a(t')} \tag{1.113}$$

grows at least exponentially:

$$\begin{aligned} d_{\rm hor}(t_{\rm end}) =& a(t_{\rm end}) \left[\int_0^{t_{\rm in}} \frac{{\rm d}t'}{a(t')} + \int_{t_{\rm in}}^{t_{\rm end}} \frac{{\rm d}t'}{a(t')} \right] \\ =& a(t_{\rm end}) \left[\frac{d_{\rm hor}(t_{\rm in})}{a(t_{\rm in})} + \frac{\int_{t_{\rm in}}^{t_{\rm end}} {\rm d}t' e^{-H(t'-t_{\rm in})}}{a(t_{\rm in})} \right] \\ =& e^N \left[d_{\rm hor}(t_{\rm in}) - \frac{e^{-N}-1}{H} \right], \end{aligned} \tag{1.114}$$

where, in the third equality, I used Eq. (1.110). The e^{-N} is small compared to 1, so it can be omitted:

$$d_{\rm hor}(t_{\rm end}) \approx e^N \left[d_{\rm hor}(t_{\rm in}) + \frac{1}{H} \right]. \tag{1.115}$$

As the Hubble parameter is strictly positive, one obtains the inequality

$$d_{\rm hor}(t_{\rm end}) \gtrsim e^N d_{\rm hor}(t_{\rm in}). \tag{1.116}$$

If the horizon at last scattering without inflation was $d_{\rm hor}(t_{\rm ls}) \approx 0.4$ Mpc, this number increases by at least a factor of e^N. With $N = 60$, which is needed to solve the flatness problem, the horizon at last scattering is then $d_{\rm hor}(t_{\rm ls}) \geq e^{60} \approx 0.4$ Mpc $\approx 10^{26}$ Mpc, i.e. all CMB photons were in causal contact before last scattering [62].

If we believe in a GUT and assume that the GUT phase transition occurred before or at the beginning of inflation, then the monopole number density decreases as the volume expands as

$$n_{\rm MM} \propto a^{-3}, \tag{1.117}$$

because there are no new monopoles produced during inflation. This means that after inflation, again assuming $N = 60$, the density of magnetic monopoles is suppressed by a factor of about 10^{78} explaining why the density is so low that it is hard to detect magnetic monopoles at our present epoch.

1.3.3 Single-Field Slow-Roll Inflation

Having shown how an early epoch of exponential expansion of the universe solves the flatness, horizon and monopole problems, I still have to explain how this can be realised and how this seeds the structure we observe today.

The easiest way to realise inflation is to introduce a homogeneous real scalar field ϕ called the **inflaton** with a matter Lagrangian

$$\mathcal{L}_{\mathrm{M}} = \frac{1}{2}\phi_{,\mu}\phi^{,\mu} - V(\phi). \tag{1.118}$$

Note that in this section, natural units are assumed, such that, as before, the speed of light, and, on top of that, also the reduced Planck constant $\hbar$ are set to unity. The energy-momentum tensor for the Lagrangian of Eq. (1.118) is given by

$$T^{\mu\nu} = \phi^{,\mu}\phi^{,\nu} - g^{\mu\nu}\left(\frac{1}{2}\phi_{,\rho}\phi^{,\rho} - V(\phi)\right) \text{ [63]}. \tag{1.119}$$

As homogeneity has been assumed, the inflaton is a perfect fluid for which, using Eq. (1.21), one can identify its density

$$\rho_\phi = T^{00} = \frac{\dot{\phi}^2}{2} + V(\phi) \tag{1.120}$$

and pressure

$$p_\phi = T^{ii} = \frac{\dot{\phi}^2}{2} - V(\phi). \tag{1.121}$$

Assuming the first **slow-roll** condition

$$\dot{\phi}^2 \ll V(\phi), \tag{1.122}$$

the equation of state is

$$w = \frac{p_\phi}{\rho_\phi} = -1 + \frac{\dot{\phi}^2}{V(\phi)} + \mathcal{O}\left(\frac{\dot{\phi}^4}{V^2(\phi)}\right), \tag{1.123}$$

thus, it imitates a cosmological constant for $\dot{\phi}^2 \ll V$. From Friedmann's Equation (1.25) one can read off the Hubble parameter

$$H^2 = \frac{8\pi G}{3}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) \approx \frac{8\pi G}{3}V(\phi), \tag{1.124}$$

assuming no curvature and Eq. (1.122). The equation of motion of a scalar field is given by the Klein-Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\mathrm{d}V}{\mathrm{d}\phi} = 0. \tag{1.125}$$

Imposing a second slow-roll condition

$$\ddot{\phi} \ll 3H\dot{\phi}, \tag{1.126}$$

one gets

$$\frac{\mathrm{d}V}{\mathrm{d}\phi} = -3H\dot{\phi}. \tag{1.127}$$

The time-derivative of the inflaton field is then

$$\dot{\phi} = -\frac{1}{3H}\frac{\mathrm{d}V}{\mathrm{d}\phi} \approx -\frac{1}{\sqrt{24\pi GV}}\frac{\mathrm{d}V}{\mathrm{d}\phi}, \tag{1.128}$$

where the last equality comes from Eq. (1.124). Taking the time-derivative of Eq. (1.124), one obtains

$$\dot{H} = \frac{1}{2}\sqrt{\frac{8\pi G}{3}}\frac{\dot{V}}{\sqrt{V}} =\approx \frac{1}{2}\sqrt{\frac{8\pi G}{3}}\frac{\dot{\phi}}{\sqrt{V}}\frac{\mathrm{d}V}{\mathrm{d}\phi}. \tag{1.129}$$

Inserting Eq. (1.128) and dividing by Eq. (1.124), one gets the following expression, which is defined as one of the **slow-roll parameters**:

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} \approx \frac{1}{16\pi G}\frac{\left(\frac{\mathrm{d}V}{\mathrm{d}\phi}\right)^2}{V^2}. \tag{1.130}$$

Reinserting Eq. (1.128), one can see that $\varepsilon \propto \frac{\dot{\phi}^2}{V} \ll 1$, and therefore the Hubble parameter is almost constant if Eq. (1.122) holds. The other slow-roll condition (Eq. (1.126)) is fulfilled if the other slow-roll parameter $\eta \equiv \varepsilon - \frac{\ddot{\phi}}{H\dot{\phi}}$ satisfies $|\eta| \ll 1$. Taking the time derivative of the first equation in Eq. (1.128) yields

$$\ddot{\phi} = \frac{\dot{H}}{3H^2}\frac{\mathrm{d}V}{\mathrm{d}\phi} - \frac{1}{3H}\frac{\mathrm{d}\dot{V}}{\mathrm{d}\phi} = \frac{\varepsilon}{3}\frac{\mathrm{d}V}{\mathrm{d}\phi} - \frac{1}{3H}\frac{\mathrm{d}^2V}{\mathrm{d}\phi^2}\dot{\phi}, \tag{1.131}$$

which, after dividing by $H\dot{\phi} = -\frac{1}{3}\frac{\mathrm{d}V}{\mathrm{d}\phi}$, reads

$$\frac{\ddot{\phi}}{H\dot{\phi}} = \varepsilon - \frac{1}{3H^2}\frac{\mathrm{d}^2V}{\mathrm{d}\phi^2} = \varepsilon - \frac{1}{8\pi GV}\frac{\mathrm{d}^2V}{\mathrm{d}\phi^2}, \tag{1.132}$$

where H^2 has been substituted with Eq. (1.124) in the second equality. The second slow-roll parameter is hence related to the potential by

$$\eta = \frac{1}{8\pi GV}\frac{\mathrm{d}^2V}{\mathrm{d}\phi^2}. \tag{1.133}$$

A natural candidate for the inflaton is the Higgs boson, the only scalar field in the standard model of particle physics. There are many problems with simply including the Higgs boson as the inflaton, but its potential

$$V(\phi) = \lambda(\phi^2 - M^2)^2, \tag{1.134}$$

plotted in Fig. 1.6, with some coupling λ and a mass M, serves as a good example of how slow-roll inflation works. Its first slow-roll parameter (while $\phi^2 \ll M^2$ and $M \gg \sqrt{G}$) is given by

$$\varepsilon = \frac{1}{4\pi G}\left(\frac{\phi}{\phi^2 - M^2}\right)^2 \approx \frac{\phi^2}{4\pi G M^2}. \tag{1.135}$$

If inflation starts at $\phi = 0$ and the inflaton mass M is heavy compared to the Planck mass $1/\sqrt{G}$, the slow-roll parameter ε is close to zero and the inflaton slowly rolls towards the global minima at $\phi = \pm M$. The mass of the standard model Higgs boson is actually too light such that ε is large even for small values of ϕ, but, assuming a heavier Higgs boson mass, the shape of its potential well serves as a good example for slow-roll inflation. As ϕ approaches the minima, the slow-roll parameter ε becomes large, thus, inflation ends as the equation of state of Eq. (1.122) starts to differ significantly from $w = -1$. The inflaton starts to oscillate around the minimum. The Hubble friction term in the Klein-Gordon Equation (1.125) dampens the oscillation, the energy decays into relativistic particles which reheat the Universe. This reheating process avoids the dilution of the particles that inflation caused for the magnetic monopoles. This reheating process happens not only for the Higgs potential, but also for any other slow-rolling potentials at the end of inflation.

Even though it was assumed that the inflaton field is homogeneous in the beginning of this section, the Universe was extremely dense and the inflaton has to be described as a quantum field. During the inflationary expansion of space, the quantum fluctuations intrinsic to the quantum field are stretched to macroscopic scales. As soon as they exceed the horizon, they are frozen in, but, eventually, after inflation, they reenter the horizon, seeding the CMB anisotropies and the structures we can observe today. This process is usually calculated in terms of the curvature perturbation

$$\zeta(t, \mathbf{x}) \equiv N(t, \mathbf{x}) - N(t), \tag{1.136}$$

which is the difference between the local and global numbers of e-folds (cf. Eq. 1.111), thus between the local and global integrated expansion rates during inflation. As the number of e-folds in single-field inflation can depend only on ϕ, one can write the curvature perturbation as

$$\zeta = \frac{\mathrm{d}N}{\mathrm{d}\phi}\delta\phi + \frac{1}{2}\frac{\mathrm{d}^2 N}{\mathrm{d}\phi^2}\delta\phi^2 + \mathcal{O}(\delta\phi^3). \tag{1.137}$$

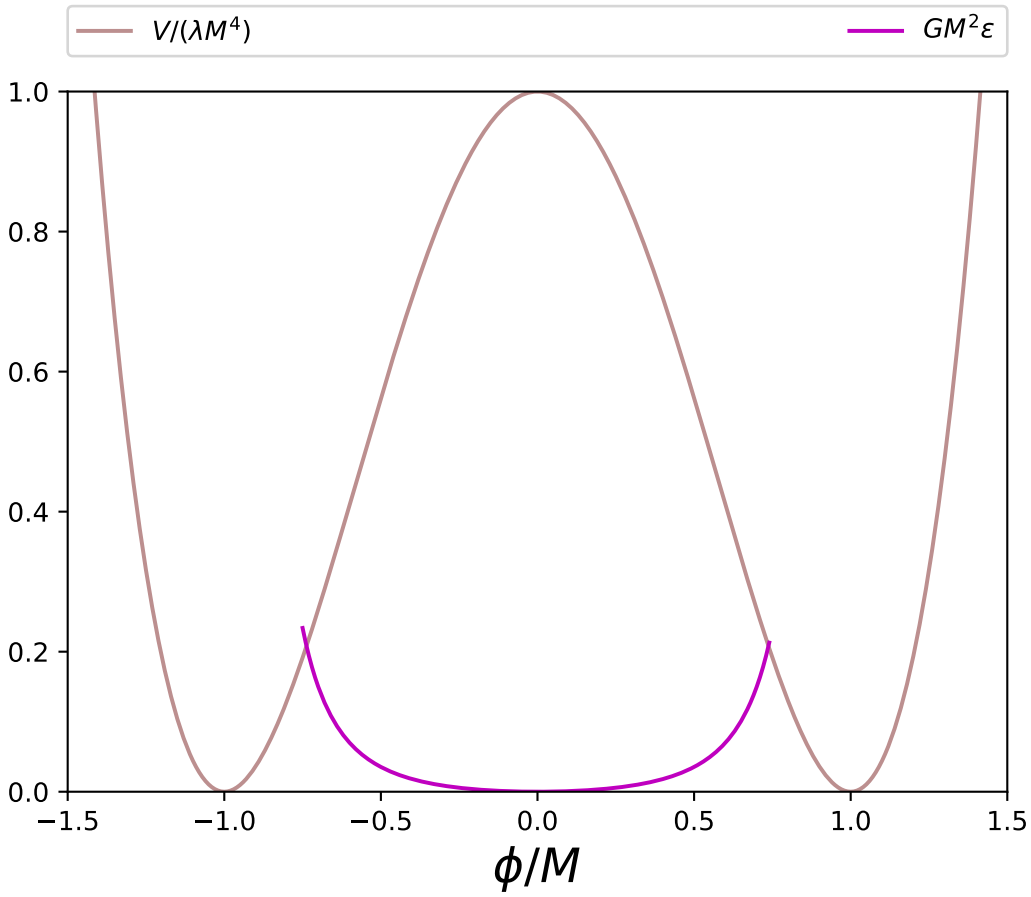

Figure 1.6: The potential of the Higgs field and its respective slow-roll parameter ε where Eq. (1.135) is valid.

The power spectrum of these curvature perturbations

$$P_{\zeta\zeta}(k) \propto k^{2+n_\zeta} \tag{1.138}$$

is predicted to be almost scale invariant, i.e. $n_s \approx 1$, and indeed Planck finds $n_\zeta = 0.9667 \pm 0.0040$ [7]. After inflation, the Universe expands slower than the Hubble radius, thus, the curvature perturbation modes that were produced by inflation and then were frozen in outside of the horizon, reenter the horizon and translate into scalar perturbations of the metric as

$$\Delta\Phi \approx -\frac{3+3w}{5+3w}\zeta \tag{1.139}$$

that can be interpreted as the gravitational potential in our Newtonian treatment. As $\Delta\Phi$ is related to the matter over-density field through the Poisson equation (1.73), we shall see later that on large scales, the matter power spectrum follows Eq. (1.138).

Using the definitions of the power and bispectrum of the curvature perturbations as in Eq. (1.104) and (1.105), one can identify

$$f_{\rm NL} = \frac{5}{6}\frac{\frac{\mathrm{d}^2 N}{\mathrm{d}\phi^2}}{\left(\frac{\mathrm{d}N}{\mathrm{d}\phi}\right)^2}, \tag{1.140}$$

by inserting them into the general definition of $f_{\rm NL}$ in Eq. (1.107) and taking the squeezed limit. This is of the order of the slow roll parameters at the time when the respective modes exited the horizon

$$f_{\rm NL} = \mathcal{O}(\varepsilon, \eta)\ [8, 64]. \tag{1.141}$$

As long modes exit the horizon well before the end of inflation, $f_{\rm NL}$ is very close to zero for single field slow-roll inflation models, and measuring a significant deviation from zero is a strong test to rule out this simple standard model.

1.3.4 Inflation with a Spectator Field: The Curvaton Model

In an alternative scenario of inflation [65, 66, 67, 68, 69], there is an additional weakly coupled scalar field χ, that, during inflation, is light compared to the

Hubble expansion, i.e. $m_\chi \ll H$, and has negligible energy density. Furthermore, χ is a so-called spectator field, i.e. it is not coupled to either the inflaton during inflation nor its decay products at the end of it. Some time after inflation, the Hubble rate drops below m_χ, causing the field to oscillate about its potential minimum and to generate the primordial curvature perturbations. The field is therefore called the curvaton. The curvaton decays into radiation with efficiency R_χ, transferring the curvaton fluctuations into fluctuations in the radiation density. In the simplest curvaton models, the primordial non-Gaussianity parameters then become a function of the decay efficiency:

$$f_{\rm NL} = \frac{5}{4R_\chi}\left(1 - \frac{4R_\chi}{3} - \frac{2R_\chi^2}{3}\right), \text{ and}$$
$$g_{\rm NL} = -\frac{25}{6R_\chi}\left(1 - \frac{R_\chi}{18} - \frac{10R_\chi^2}{9} - \frac{R_\chi^3}{3}\right) \text{ [8, 70, 71, 72].} \tag{1.142}$$

This dependence is plotted in Fig. 1.7. There one can see, that observing values of $f_{\rm NL}$ that are significantly less than $f_{\rm NL} = -\frac{5}{4}$ would rule out this simple curvaton scenario. If a large positive value of $f_{\rm NL}$ would have been observed, one would have to measure the parameter $g_{\rm NL}$, which would be the third order term of Eq. (1.108), to be able to exclude the curvaton model, or to constrain the decay efficiency parameter R_χ.

1.3.5 Modulated Decay

Many string theories include scalar fields with continuous families of global minima of their potentials. Such fields are called moduli fields. One can construct models in which the moduli fields are pressureless and have no density perturbations during inflation. However, the decay rate of the inflaton $\Gamma(\chi)$ could be a function of such moduli fields. Hence, the efficiency of reheating becomes spatially dependent [73]. As this happens after Hubble exit, modulated reheating produces local type non-Gaussianity [8]:

$$f_{\rm NL} = 5\left(1 - \frac{\frac{{\rm d}^2\Gamma}{{\rm d}\chi^2}\Gamma}{\left(\frac{{\rm d}\Gamma}{{\rm d}\chi}\right)^2}\right) \text{ [8, 74].} \tag{1.143}$$

An approximately linear modulation of the decay rate therefore yields $f_{\rm NL} \approx 5$, whereas for $\Gamma \propto \chi^2$ one obtains $f_{\rm NL} = \frac{5}{2}$. Thus, "natural" realisations of this

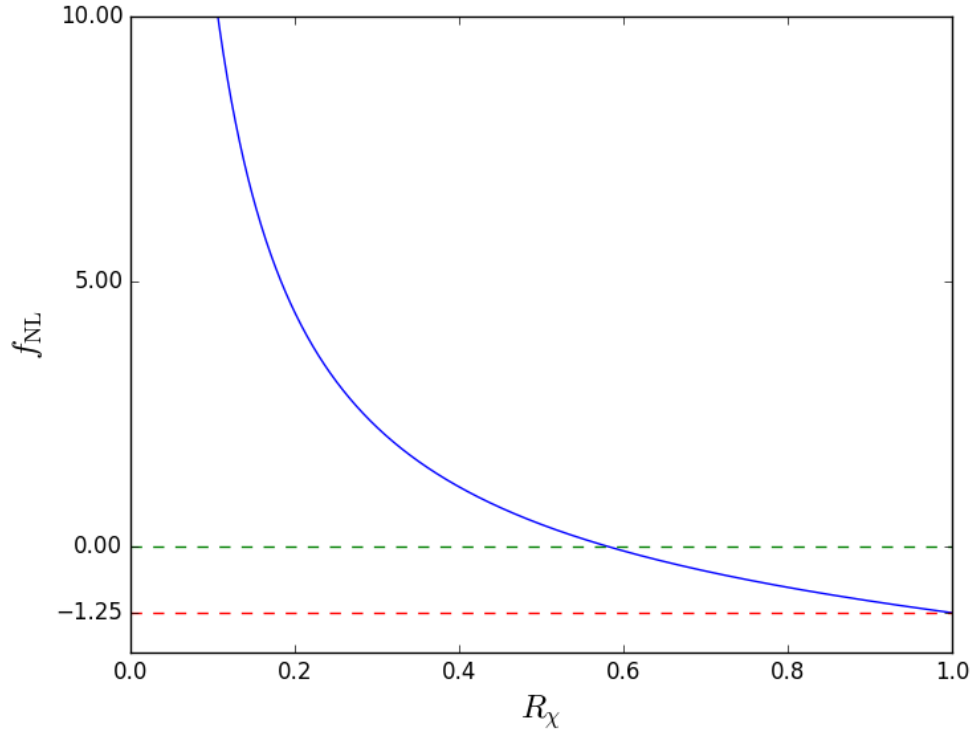

Figure 1.7: Primordial non-Gaussianity parameter $f_{\rm NL}$ generated in the curvaton scenario as a function of the decay efficiency R_χ.

scenario become testable in the foreseeable future (cf. Chapter 2).

1.3.6 Kinetic Inflation

If one considers (one of) the moduli fields as candidates for the inflaton, one finds that their potentials are non-perturbative, i.e. they do not contain any regions where the slow-roll conditions are fulfilled. This can be fixed by introducing non-standard, that is non-quadratic, kinetic terms into their Lagrangians [75]. This also produces significant non-Gaussianity of the field perturbations, but usually in folded configurations [76] rather than in squeezed triangles (cf. 1.2.4). Thus, this serves as an example where we might measure vanishing local $f_{\rm NL}$, even though there exists a primordial bispectrum.

1.3.7 An Alternative to Inflation: Ekpyrotic Cosmology

While inflation successfully solves the flatness, horizon and monopole problems, and explains the presence of structure, it is not the only model doing so. An alternative to inflation is the ekpyrotic model. In an ekpyrotic cosmology, the big bang was actually a big bounce that ended a previous epoch of contraction of the Universe. During the collapse, the Hubble scale decreases and quantum fluctuations evolve into super-Hubble scales similar to as in the inflationary paradigm. The ekpyrotic universe is driven by several scalar fields ϕ_I with steep exponential potentials

$$V(\phi_I) \propto \exp(-c_I \phi_I) \tag{1.144}$$

that lead to an ultra-stiff equation-of-state $w = p/\rho \gg 1$. The potential also exhibits strong self-interaction terms, thus it causes large non-linearities [77]. In the case where the ekpyrotic phase is driven by two fields with potential

$$V = -V_1 \exp(-c_1 \phi_1) - V_2 \exp(-c_2 \phi_2), \tag{1.145}$$

a simple ekpyrotic model [78] gives

$$f_{\rm NL} \approx -\frac{5}{12} c_1^2, \tag{1.146}$$

assuming that ϕ_1 is the field that becomes sub-dominant at late times, without loss of generality [78]. As the parameter c_1 also determines the scale depend-

ence of the primordial power spectrum (cf. Eq. 1.138)

$$n_\zeta - 1 \equiv \frac{1}{2\pi^2}\frac{\mathrm{d}\ln\left(k^3 P_{\zeta\zeta}(k)\right)}{\mathrm{d}\ln(k)} = 4\left(c_1^{-2} + c_2^{-2}\right), \tag{1.147}$$

the Planck constraint of $n_\zeta = 0.9667 \pm 0.0040$ [7] disagrees with the Planck constraint of $f_{\mathrm{NL}} > -3.2$ [3], which, in this simple model, would require $n_\zeta > 1.52$. This illustrates how measuring f_{NL} also provides a test of whether inflation is the right paradigm.

Another test for whether inflation happened comes from the fact that inflation also produces primordial gravitational waves. In standard single-field slow-roll inflation, their spectrum amplitude are directly connected to the energy scale of inflation [79], thus providing complementary information to non-Gaussianity measurements on the physical process causing inflation. The first direct observations of gravitational waves have been made recently by the laser interferometers of the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo [80, 81, 82, 83], however, the amplitude of primordial gravitational waves is expected to be below the sensitivity of current experiments using interferometry to detect gravitational waves induced space-time strain [79]. Nevertheless, primordial gravitational waves imprint a curl-like pattern in the polarisation of the CMB, which in analogy to electromagnetism are called B-modes. These are dominated by polarised thermal emission from diffuse Galactic dust and, therefore, there has not been a detection of primordial gravitational waves yet [84], even though it was claimed by the second generation of Background Imaging of Cosmic Extragalactic Polarization [85, BICEP2].

We have seen in this section that, even though there are non-Gaussian models such as kinetic inflation that do not exhibit their non-Gaussianity in local f_{NL}, measuring local f_{NL} is a powerful test of many other scenarios of inflation, such as the vanilla model of inflation, that is single field slow-roll inflation, curvaton inflation and modulated decay. A summary of the predicted value of local f_{NL} for these models can be found in Tab. 1.2. Models such as multi-field inflation or the ekpyrotic alternative to inflation can be tested using local f_{NL} constraints when combined with other test parameters. I proceed by discussing how the perturbations seeded by inflation evolve until (re)combination and the explain how local f_{NL} can be measured using galaxy

Table 1.2: Values of local-type $f_{\rm NL}$ according to [8] for some popular models of inflation.

Model	$f_{\rm NL}$ prediction
Single-field slow-roll inflation	$f_{\rm NL} = \mathcal{O}(\varepsilon, \eta) \approx 0$
Curvaton inflation	$f_{\rm NL} > -\frac{5}{4}$
Modulated decay with decay rate $\Gamma \propto \chi$	$f_{\rm NL} = 5$
Modulated decay with decay rate $\Gamma \propto \chi^2$	$f_{\rm NL} = \frac{5}{2}$
Kinetic inflation	$f_{\rm NL} = 0$ (but significant folded configurations)

or other large scale clustering data in Sec. 1.5.2.

1.4 From Inflation to Recombination

To learn about inflation from galaxy clustering data, one has to understand how the density field evolved after inflation to what we observe today. As derived in Sec. 1.1.3, the Universe entered a phase where its energy content was dominated by radiation. At a redshift of $z_{\rm eq} \approx 3400$ (cf. Tab. 1.1), the energy density in radiation and non-relativistic matter is equal and matter starts to dominate afterwards. At this stage, the Universe was still so dense that baryonic matter (in the cosmological sense) and photons were coupled due to Compton scattering and formed a plasma that requires to be treated as a single fluid [86]. As the photons and baryons are tightly coupled, they have to be regarded as one fluid, the photon-baryon plasma. As in Sec. 1.2.2, I only consider adiabatic fluctuations in the photon-baryon fluid. Such perturbations have vanishing entropy perturbations $\Delta S = \int \Delta\rho/T = 0$. As the background radiation energy density $\bar{\rho}_{\rm r} \propto T_{\rm r}^4$ and the baryon energy density is proportional to the baryon number density, the entropy is $S \propto \frac{\bar{\rho}_{\rm r}^{3/4}}{\bar{\rho}_{\rm b}}$. Perturbing the entropy, one obtains

$$\frac{\Delta S}{S} = \frac{3}{4}\frac{\Delta\rho_{\rm r}}{\bar{\rho}_{\rm r}} - \frac{\Delta\rho_{\rm b}}{\bar{\rho}_{\rm b}} \text{ [63].} \tag{1.148}$$

For adiabatic perturbations, one can therefore express the baryon perturbations in terms of the radiation perturbations and the baryon to radiation

fraction $R \equiv \frac{3}{4}\frac{\bar{\rho}_\mathrm{b}}{\bar{\rho}_\mathrm{r}}$:

$$\Delta\rho_\mathrm{b} = R\Delta\rho_\mathrm{r}. \tag{1.149}$$

Due to energy-momentum conservation, the total density perturbation in the photon-baryon fluid is

$$\Delta\rho = \Delta\rho_\mathrm{r} + \Delta\rho_\mathrm{b} = (1+R)\Delta\rho_\mathrm{r}. \tag{1.150}$$

As the baryons are pressureless, the total pressure of the fluid is due to the photons only, and with the radiation equation of state we have

$$\Delta p = \Delta p_\mathrm{r} = \frac{1}{3}\Delta\rho_\mathrm{r}. \tag{1.151}$$

Inserting the last two equations into the definition of the speed of sound Eq. (1.62), we find

$$c_\mathrm{s}^2 = \frac{1}{3(1+R)}, \tag{1.152}$$

thus, as the presence of baryons makes the photon-baryon fluid heavier than just radiation, sounds waves propagate slower. With the background equation (1.26), which for a flat universe with negligible cosmological constant relates the background density $\bar{\rho} = \frac{3\mathcal{H}^2}{8\pi G}$ with the Hubble rate, the Jeans scale (cf. Eq. (1.82)) is of the same order of magnitude as the Hubble rate:

$$k_\mathrm{J} = 3(1+R)\sqrt{\frac{3}{2}}a\mathcal{H} = 3(1+R)\sqrt{\frac{3}{2}}H. \tag{1.153}$$

This means that before (re)combination, the radiation pressure of the photons prevents the photon-baryon fluid from clustering on scales within the Hubble radius $\sim 1/H$.

Relativistic correction are important before (re)combination. Choosing conformal Newtonian gauge, there are two scalar fluctuations $\Delta\Phi$ and $\Delta\Psi$ to the metric:

$$\mathrm{d}s^2 = a^2\left[-(1+2\Delta\Phi)\mathrm{d}\eta^2 + (1-2\Delta\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{1.154}$$

These scalar perturbations act similar to the Newtonian potential. For a diagonal energy-momentum tensor one can show that $\Delta\Phi = \Delta\Psi$. The relativistic linear continuity, Poisson and Euler equations in Newtonian gauge read:

$$\begin{aligned} \delta' + (1+w)(\theta + 3\Delta\Phi') &= 0 \\ \nabla^2\Delta\Phi = 4\pi G a^2\bar{\rho}\delta + 3\mathcal{H}(\Delta\Phi' + \mathcal{H}\Delta\Phi) & \\ \Delta\boldsymbol{u}' + \mathcal{H}(1-3c_\mathrm{s}^2)\Delta\boldsymbol{u} + \frac{c_\mathrm{s}^2}{1+w}\nabla\delta + \nabla\Delta\Phi &= 0. \end{aligned} \tag{1.155}$$

Analogously to Sec. 1.2.2, one can derive a relativistic equivalent of Eq. (1.81) by taking the divergence of the relativistic Euler equation:

$$\theta' + \mathcal{H}(1 - 3c_\mathrm{s}^2)\theta + \frac{c_\mathrm{s}^2}{1+w}\nabla^2\delta + \nabla^2\Delta\Phi = 0 \tag{1.156}$$

For the photons, one can insert $w = 1/3$ and $c_\mathrm{s}^2 = \frac{1}{3(1+R)}$:

$$\theta_\mathrm{r}' + \mathcal{H}\frac{R}{1+R}\theta_\mathrm{r} + \frac{1}{4(1+R)}\nabla^2\delta_\mathrm{r} + \nabla^2\Delta\Phi = 0 \tag{1.157}$$

Again analogously to Sec. 1.2.2, the θ terms can be substituted using the continuity equation $\theta_\mathrm{r} = -\frac{3}{4}\delta_\mathrm{r}' - 4\Delta\Phi'$:

$$-\frac{3}{4}\delta_\mathrm{r}'' - 4\Delta\Phi'' - \frac{3}{4}\mathcal{H}\frac{R}{1+R}\delta_\mathrm{r}' - 4\mathcal{H}\frac{R}{1+R}\Delta\Phi' + \frac{1}{4(1+R)}\nabla^2\delta_\mathrm{r} + \nabla^2\Delta\Phi = 0. \tag{1.158}$$

After multiplying with $-4/3$, splitting δ and Φ terms, and Fourier transforming, one obtains an equation for a forced, damped harmonic oscillator

$$-\frac{3}{4}\delta_\mathrm{r}'' - \frac{3}{4}\mathcal{H}\frac{R}{1+R}\delta_\mathrm{r}' + \frac{1}{4(1+R)}k^2\delta_\mathrm{r} = 4\Delta\Phi'' + 4\mathcal{H}\frac{R}{1+R}\Delta\Phi' - k^2\Delta\Phi. \tag{1.159}$$

For the baryon perturbations, we get the same functional form but with different prefactors since for the baryons the equation of state is $w = 0$, thus

$$\theta_\mathrm{b}' + \mathcal{H}\frac{R}{1+R}\theta_\mathrm{b} + \frac{1}{3(1+R)}k^2\delta_\mathrm{b} + k^2\Delta\Phi = 0,$$

$$\theta_\mathrm{b} = -\delta_\mathrm{b}' - 3\Delta\Phi', \text{ and}$$

$$-\delta_\mathrm{b}'' - \mathcal{H}\frac{R}{1+R}\delta_\mathrm{b}' + \frac{1}{3(1+R)}k^2\delta_\mathrm{b} = 3\Delta\Phi'' + 3\mathcal{H}\frac{R}{1+R}\Delta\Phi' - k^2\Delta\Phi. \tag{1.160}$$

These are known as acoustic oscillations and their imprint on the distribution of galaxies (cf. Fig. 1.8) is an important probe of cosmology (cf. Sec. 1.1.6 and Sec. 2.1).

Due to the photon-baryon fluid's oscillations, its perturbations are smoothed out and do not cause any structures before (re)combination. The only component at that epoch that forms structures is cold dark matter, which can be modelled as a pressureless fluid whose linear perturbations according to Eq. (1.81) evolve as

$$\delta_\mathrm{cdm}''(\mathbf{k}) + \mathcal{H}\delta_\mathrm{cdm}'(\mathbf{k}) - 4\pi G\bar{\rho}_\mathrm{cdm}(\mathbf{k})a^2\delta_\mathrm{cdm}(\mathbf{k}) = 0. \tag{1.161}$$

With Eq. (1.30), one can substitute the term $4\pi G\bar{\rho}_{\rm cdm}(\mathbf{k})a^2 = \frac{3}{2}\Omega_{\rm cdm}(\eta)\mathcal{H}^2$. During radiation domination, $\Omega_{\rm cdm}(\eta)$ is small and can be neglected, such that

$$\delta''_{\rm cdm}(\mathbf{k}) + \mathcal{H}\delta'_{\rm cdm}(\mathbf{k}) = 0. \tag{1.162}$$

Thus, during radiation domination, cold dark matter perturbations grow logarithmically as the Jeans scale is $\lambda_{\rm J} = 0$ (cf. Eq. 1.83). To solve Eq. (1.161) also during matter-radiation equality and matter domination, it is convenient to define the variable

$$y \equiv \frac{a}{a_{\rm eq}} \approx \frac{\bar{\rho}_{\rm cdm}}{\bar{\rho}_{\rm r}} \tag{1.163}$$

to rewrite it as the Mészáros equation

$$\frac{{\rm d}^2\delta_{\rm cdm}}{{\rm d}y^2} + \frac{2+3y}{2y(y+1)}\frac{{\rm d}\delta_{\rm cdm}}{{\rm d}y} - \frac{3}{2y(y+1)}\delta_{\rm cdm} = 0 \text{ [87]}, \tag{1.164}$$

whose solution is

$$\begin{aligned}\delta_{\rm cdm}(\mathbf{k}) = A(\mathbf{k})&\left(1+\frac{3}{2}y\right)\\ &+ B(\mathbf{k})\left[3\sqrt{1+y} - \left(1+\frac{3}{2}y\right)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right)\right].\end{aligned} \tag{1.165}$$

In the limit $y \ll 1$, one rediscovers the logarithmic growth on all scales during radiation domination. In the opposite limit, i.e. during matter domination, the A-term provides a linear growing mode, whereas the B-term decays [56]. This means that before matter-radiation-equality, the growth of structure was suppressed as the scale factor was controlled by radiation. This suppression is visible in the matter power spectrum as a turn-over (cf. Fig. 1.8). The structure that we see in the CMB suggests that matter started to dominate much before (re)combination.

After the expansion has cooled the Universe enough to let the baryons form neutral atoms, they decouple from the photons and gravitationally attract the dark matter from the centre, and vice versa, leaving a slightly higher density in a spherical shell whose radius corresponds to the distance

$$r_{\rm d} = \int_0^{\eta_{\rm d}} c_{\rm s}(\eta){\rm d}\eta = \int_0^{t_{\rm d}} \frac{c_{\rm s}(t)}{a(t)}{\rm d}t \tag{1.166}$$

a sound wave can travel until the time $\eta_{\rm d}$ or $t_{\rm d}$ the baryons were decoupled from the photons. Because of that, it is more likely to find objects on structure that evolved from the shells, thus, the two point correlation function has

a peak at a distance corresponding to the shell radius. In Fourier space, superimposing many of these shells introduces a series of oscillations into the power spectrum [88] (cf. Fig. 1.8). As mentioned in Sec. 1.2.2, the transfer function incorporates this scale dependent effect along with other scale dependent effects such as Compton drag, Silk damping and growth suppression due to cold dark matter. Fitting formulae, e.g. by Eisenstein and Hu [88], can provide some insight into how these effects influence the transfer function. In practice, it is more accurate to solve the multi-species Boltzmann equations numerically, using codes such as the Code for Anisotropies in the Microwave Background (CAMB, [22, 89]) or the Cosmic Linear Anisotropy Solving System (CLASS, [90, 91, 92, 93]). A CAMB example power spectrum is plotted in Fig. 1.8. The observational aspect of this will be readdressed in Sec. 2.1. But first, let us go back to observational features of local type primordial non-Gaussianity.

1.5 Galaxy Bias

Up to now, we have seen that perturbations in the matter density field provides a wealth of information about the physics of the Universe. However, the most easily observed abundant extragalactic objects are galaxies and quasi stellar objects (quasars). These can only be observed above a certain brightness threshold. If one assumes that the luminosity and the mass of the objects are related, then the galaxies (or quasars) are biased tracers of the underlying matter density field.

1.5.1 Mass Smoothing and Linear Bias

When observing the large scale structure of the Universe, we do not consider the fluctuations of single particles in the matter density field, but rather in mass concentrations of a particular size. For that reason, one smooths the matter density on a scale

$$R = \sqrt[3]{\frac{3M}{4\pi\bar{\rho}}} \tag{1.167}$$

related to the halo mass scale M assumed to be equivalent to the mass of a sphere with radius R and with the same mass density as the background. The

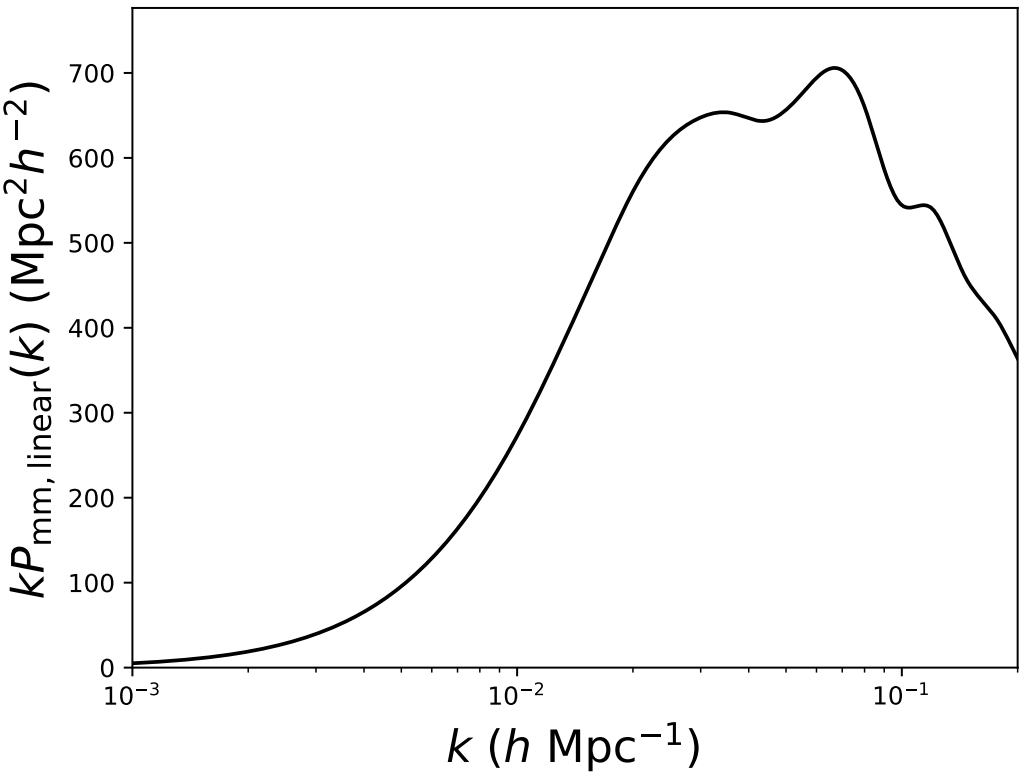

Figure 1.8: Example of a linear model power spectrum obtained using the CAMB code. The power spectrum is multiplied with k to make the oscillations more prominent.

smoothing is performed using a filter $W_R(\mathbf{k})$ such that the smoothed density field reads

$$\delta_R(\mathbf{k}) = W_R(\mathbf{k})\delta(\mathbf{k}). \tag{1.168}$$

The filter

$$W_R(\mathbf{k}) = 3\frac{\sin(kR) - kR\cos(kR)}{(kR)^3} \tag{1.169}$$

is commonly chosen to be the Fourier transform of a top-hat filter in real space, such that $\delta_R(\mathbf{k})$ can be interpreted as the density field of objects in a given mass range. The variance of the real space density smoothed at radius R or equivalent mass scale M is given by

$$\sigma^2(R) = \langle \delta_R^2(\mathbf{k}) \rangle = \int \mathrm{d}k \frac{k^2}{2\pi^2} W_R^2(k) P_{\delta\delta}(k). \tag{1.170}$$

This is an essential building block of Press-Schechter Theory [94], a simplistic, but for this thesis sufficient, model of structure growth at a certain mass scale. Note that the original derivation of the scale dependent bias by Dalal *et al.* [1] contains a discussion about more accurate formalisms such as the Extended Press-Schechter formalism [95], the Matarrese-Verde-Jimenez model [96] and their own simulation based fitting function approach. Dalal *et al.* state that even though the Press-Schechter Theory does not reproduce the correct clustering, one can use it to correctly compute the departure from Gaussianity. For illustration purposes, I continue using Press-Schechter Theory, according to which, virialised objects of mass M can be formed where the Lagrangian density field smoothed at the respective mass scale exceeds a threshold δ_c. The number density of virialised objects

$$\frac{\mathrm{d}n}{\mathrm{d}\ln(M)} \propto \nu \exp\left(-\frac{\nu^2}{2}\right) \tag{1.171}$$

depends on the height parameter $\nu \equiv \frac{\delta_\mathrm{c}}{\sigma(M)}$. To understand the bias between the virialised objects and the underlying density field, it is useful to split the density perturbations into modes with short wavelengths δ_s that cause the collapse of the objects we observe and into modes with long wavelengths δ_ℓ that can be understood as a background field in a perturbation theory sense. This procedure is known as the Peak-Background Split formalism (e.g. [97, 98]). The large scale component δ_ℓ perturbs the threshold δ_c, because the small

scale component $\delta_{\rm s}$ only needs to reach $\delta_{\rm s} = \delta_{\rm c} - \delta_\ell$ to virialise. Hence, long wavelength modes perturb the background number density of haloes

$$n = \bar{n} + \frac{\mathrm{d}\bar{n}}{\mathrm{d}\delta_\ell}\delta_\ell \tag{1.172}$$

and the height parameter

$$\nu = \frac{\delta_{\rm c}}{\sigma(M)} \rightarrow \frac{\delta_{\rm c} - \delta_\ell}{\sigma(M)} \tag{1.173}$$

are also perturbed. The derivative in Eq. (1.172) can be rewritten as

$$\frac{\mathrm{d}\bar{n}}{\mathrm{d}\delta_\ell} = \frac{\mathrm{d}\bar{n}}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\delta_\ell}. \tag{1.174}$$

Taking the derivative of Eq. (1.171) with respect to ν, one obtains

$$\frac{\mathrm{d}\bar{n}}{\mathrm{d}\nu} = \bar{n}\frac{1-\nu^2}{\nu}. \tag{1.175}$$

The derivative of Eq. (1.173) is given by

$$\frac{\mathrm{d}\nu}{\mathrm{d}\delta_\ell} = -\frac{1}{\sigma(M)}. \tag{1.176}$$

Defining the density field of galaxies (or other objects of interest) as

$$\delta_{\rm g} \equiv \frac{n - \bar{n}}{\bar{n}} \tag{1.177}$$

and inserting Eq. (1.172)-(1.176) shows that

$$\delta_{\rm g} = \frac{\nu^2 - 1}{\delta_{\rm c}}\delta_\ell \tag{1.178}$$

traces the background density field δ_ℓ with bias

$$b \equiv \frac{\delta_{\rm g}}{\delta_\ell} = \frac{\nu^2 - 1}{\delta_{\rm c}}. \tag{1.179}$$

On large scales, the linear matter density δ is entirely described by the large scale component δ_ℓ, thus from Eq. (1.179) we know that the linear matter density and the galaxy density are related by the constant linear bias term

$$\delta_{\rm g} = b\delta. \tag{1.180}$$

This is expected because the clustering at these scales is unaffected by local physical processes, thus the clustering of tracers is governed by (linear) gravity.

At epochs where galaxies already exist, baryonic matter and the more massive dark matter comove, thus having the same distribution, and at its peaks, galaxies and other tracers can form. Apart from gravity and local processes, there is another source that influences the bias, namely imprints in the initial conditions. The most prominent example is discussed in Sec. 1.5.2, where we shall see that the presence of primordial non-Gaussianity gives rise to a scale dependent bias and therefore provides a test of primordial non-Gaussianity and of different inflationary scenarios. Discussions of other imprints on the bias, such as from BAO's, as well as a thorough review of the galaxy/halo bias can be found, e.g., in [99].

1.5.2 Observing Primordial Non-Gaussianity in Large Scale Structure

In Sec. 1.5, we have seen that the galaxy density is a biased tracer of the matter density with a linear bias b. In this section, I will show that this relationship only holds in the absence of local primordial non-Gaussianity. The curvature perturbations produced by inflation are proportional to perturbations of the Newtonian potential $\Delta\Phi$. With Poisson's equation (1.73)

$$\nabla^2\Phi_{\rm in} = \frac{3\Omega_m H_0^2}{2c^2}\delta_{\rm in}, \tag{1.181}$$

the primordial Newtonian metric potential $\Phi_{\rm in}$ then determines the initial linear matter density. If one Fourier transforms this equation, the differentiation operator becomes a k-factor and one finds for the linear matter density field

$$\delta_{\rm in}(\mathbf{k}) = \frac{2k^2c^2}{3\Omega_{\rm m}H_0^2}\Phi_{\rm in}(\mathbf{k}). \tag{1.182}$$

As seen in the previous section, there are models of inflation that predict local-type non-Gaussianity. Using the definition of Eq. (1.108), one can then express the primordial Newtonian metric potential with the help of a Gaussian auxiliary potential $\phi_{\rm G}$:

$$\Phi_{\rm in} = \phi_{\rm G} + f_{\rm NL}\left(\phi_{\rm G}^2 - \langle\phi_{\rm G}^2\rangle\right) + \mathcal{O}(\phi_{\rm G}^3). \tag{1.183}$$

Due to the Poisson relation between the density field and the potential, one also has to perform the Peak-Background Split to the potential. Hence, I write

the Gaussian auxiliary potential

$$\phi_{\rm G} = \phi_{{\rm G},\ell} + \phi_{\rm G,s} \tag{1.184}$$

as the sum of a long mode part $\phi_{{\rm G},\ell}$ and a short mode part $\phi_{\rm G,s}$. The overall non-Gaussian potential then reads

$$\Phi_{\rm in} = \phi_{{\rm G},\ell} + f_{\rm NL}\left(\phi^2_{{\rm G},\ell} - \left\langle\phi^2_{{\rm G},\ell}\right\rangle\right) + (1 + 2f_{\rm NL}\phi_{{\rm G},\ell})\phi_{\rm G,s} + f_{\rm NL}(\phi^2_{\rm G,s} - \langle\phi^2_{\rm G,s}\rangle_V), \tag{1.185}$$

where $\langle\phi^2_{\rm G,s}\rangle_V$ is a local expectation evaluated over the volume defined by the scale at which the Peak-Background split occurs. From the Fourier transform of Eq. (1.185), one obtains

$$\begin{aligned}\Phi_{\rm in}(\mathbf{k}) =& \phi_{{\rm G},\ell} + f_{\rm NL}\left(\phi_{{\rm G},\ell} * \phi_{{\rm G},\ell} - \left\langle\phi^2_{{\rm G},\ell}\right\rangle\right) \\ &+ (1 + 2f_{\rm NL}\phi_{{\rm G},\ell})\phi_{\rm G,s} + f_{\rm NL}(\phi_{\rm G,s} * \phi_{\rm G,s} - \langle\phi^2_{\rm G,s}\rangle_V),\end{aligned} \tag{1.186}$$

where the asterisk stands for a convolution, one can read off the background density perturbation

$$\delta_\ell(\mathbf{k}) = \delta_{{\rm G},\ell} + f_{\rm NL}\frac{3\Omega_m H_0^2}{2c^2}k^2\left(\phi_{{\rm G},\ell} * \phi_{{\rm G},\ell} - \left\langle\phi^2_{{\rm G},\ell}\right\rangle\right) \tag{1.187}$$

as well as the peak density

$$\delta_{\rm s}(\mathbf{k}) = (1 + 2f_{\rm NL}\phi_{{\rm G},\ell})\delta_{\rm G,s} + f_{\rm NL}\frac{3\Omega_m H_0^2}{2c^2}k^2(\phi_{\rm G,s} * \phi_{\rm G,s} - \langle\phi^2_{\rm G,s}\rangle_V), \tag{1.188}$$

after defining $\delta_{{\rm G},\ell} = \frac{3\Omega_m H_0^2}{2c^2}k^2\phi_{{\rm G},\ell}$ and $\delta_{\rm G,s} = \frac{3\Omega_m H_0^2}{2c^2}k^2\phi_{\rm G,s}$ analogously to Eq. (1.182). The interesting result here is that, in the presence of non-Gaussianity, the leading term of $\delta_{\rm s}(\mathbf{k})$ in Eq. (1.188) exhibits a modulation of the short mode density field by the long mode auxiliary potential. This causes the bias to become scale dependent and of the form

$$b(k, f_{\rm NL}) = b_0 + \delta b(f_{\rm NL}) + \Delta b(k, f_{\rm NL}), \tag{1.189}$$

where b_0 is the bias in a Universe without primordial non-Gaussianity and

$$\Delta b(k, f_{\rm NL}) \approx (b_0 - 1)f_{\rm NL}A(k)\ [1, 2, 100] \tag{1.190}$$

is the local scale-dependent correction due to the easier halo formation with additional long-wavelength fluctuations, which depends on the critical density

$\delta_c(z)$ in the Peak- Background Split model. As the metric potential and linear over-density field underwent a scale-dependent evolution between inflation and photon decoupling (cf. Sec. 1.4), the bias also contains the matter transfer function $T(k)$ through the parameter

$$A(k,z) = \frac{3\Omega_m \delta_c(z)}{k^2 T(k)} \left(\frac{H_0}{c}\right)^2 . \tag{1.191}$$

Another correction coming from non-Gaussian models of the mass function $\delta b(f_{\rm NL})$ also enters Eq. (1.189), but as $\delta b(f_{\rm NL}) \ll \Delta b(k, f_{\rm NL})$ at the scales we consider [101,102], we neglect $\delta b(f_{\rm NL})$. Note that Eq. (1.190) has been derived assuming that the halo occupation distribution only depends on the halo mass. In some cases, it can also depend on the merging history of the haloes. For instance, there is evidence that quasar activity is triggered by recent mergers. Slosar *et al.* [101] have shown that the scale dependence of the bias can be described more generally by

$$\Delta b(k, f_{\rm NL}) \approx (b_0 - p) f_{\rm NL} A(k), \tag{1.192}$$

where $1 < p < 1 + \delta_{\rm c}$ is the so called halo merger bias. However, the objects that I am going to use have $p = 1$. Allowing different values of p will be important in some of the forecasts for future surveys that are presented in Chapter 2, because the scale dependent bias for such tracers is less sensitive to $f_{\rm NL}$. It will also be important for future work (Chapter 6).

Fig. 1.9 shows the effect of $f_{\rm NL}$ on the galaxy power spectrum at large scales. We plot the galaxy power spectrum P_g divided by the galaxy power spectrum at $f_{\rm NL} = 0$, hence what we plot is proportional to the square of Eq. (1.189). At very low values of k, negative $f_{\rm NL}$ enhances the power spectrum due to the fact that the term proportional to $f_{\rm NL}^2$ dominates the total bias. At slightly higher k, but still at large scales, the term linear in $f_{\rm NL}$ dominates, and the power is enhanced or decreased depending on the sign of $f_{\rm NL}$. At small, yet still linear, scales, $A(k,z)$ becomes small, thus initial local non-Gaussianities do not have an effect on the galaxy power spectrum at these scales.

Due to this scale-dependence of the bias, the galaxy power spectrum provides a good test of local type primordial non-Gaussianity. Compared to $f_{\rm NL}$ constraints from the bispectrum, it has the advantage of being unaffected by non-linear effects on small scales due to gravity. The disadvantage is that the

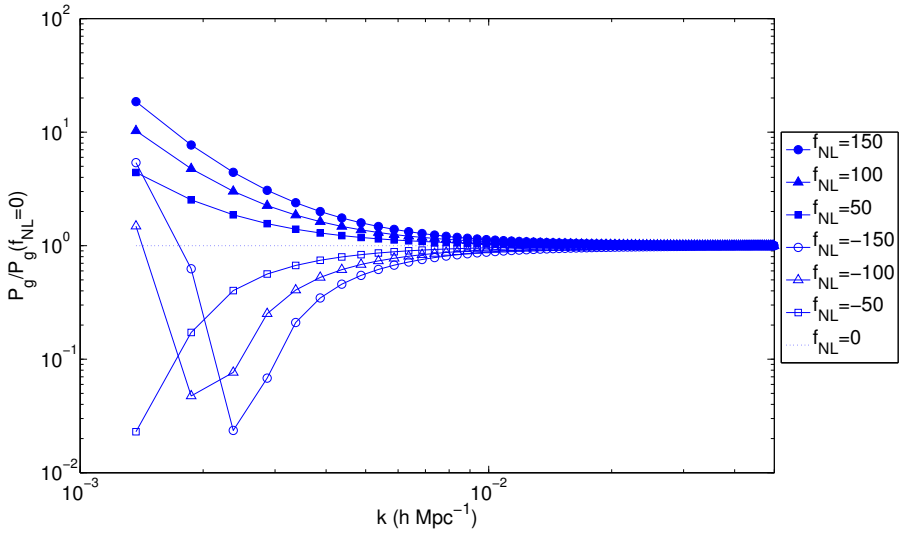

Figure 1.9: Galaxy power spectra P_g calculated for different values of $f_{\rm NL}$ divided by the galaxy power spectrum $P_g(f_{NL}=0)$ of a universe with a Gaussian primordial density field.

power spectrum at the scales where we get the strongest $f_{\rm NL}$ signal is strongly affected by both cosmic variance and large scale systematic effects. The latter will be discussed in Chapters 4 and 5, whereas the former problem can be solved in principle by observing more than one tracer. In such cases, one can consider the ratio of the density fields of two tracers $\delta_1(\mathbf{k})$ and $\delta_2(\mathbf{k})$. As both fields are tracing the same underlying matter density field, whose measurement is limited by cosmic variance, we have

$$\frac{\delta_1(\mathbf{k})}{\delta_2(\mathbf{k})} = \frac{b_1(k, f_{\rm NL})}{b_2(k, f_{\rm NL})}, \tag{1.193}$$

where the matter density field and hence the cosmic variance cancel out. The same argument applies to the ratio of the power spectra, as long as they are measured from the same survey volume and they have the same window function [103]. If the linear bias parameters of both tracers are very similar, the ratio in Eq. (1.193) is $\frac{b_1(k,f_{\rm NL})}{b_2(k,f_{\rm NL})} \approx 1\ \forall k$, thus, $f_{\rm NL}$ cannot be constrained using the multi-tracer technique, as it does not change for different values of $f_{\rm NL}$. Of course, it can still be constrained from each tracer separately. On the other hand, if the linear biases are very different, this provides a powerful tool to obtain tight constraints on $f_{\rm NL}$, if the issue of systematics is addressed thoroughly. Another issue with measuring $f_{\rm NL}$ from the scale dependent bias

is that if one continues Eq. (1.183) to third order and then performs the Peak Background Split, one finds

$$\begin{aligned}\Phi_{\rm in} =&\phi_{{\rm G},\ell} + f_{\rm NL}\left(\phi^2_{{\rm G},\ell} - \langle\phi_{{\rm G},\ell}\rangle\right) + g_{\rm NL}\left(\phi^3_{{\rm G},\ell} - 3\left\langle\phi^2_{{\rm G},\ell}\right\rangle\phi_{{\rm G},\ell}\right)\\ &+\left[1 + 2f_{\rm NL}\phi_{{\rm G},\ell} + 3g_{\rm NL}\left(\phi^2_{{\rm G},\ell} - \langle\phi^2_{{\rm G},\ell}\rangle\right)\right]\phi_{\rm G,s}\\ &+(f_{\rm NL} + 3g_{\rm NL}\phi_{{\rm G},\ell})(\phi^2_{\rm G,s} - \langle\phi^2_{\rm G,s}\rangle_V) + g_{\rm NL}\left(\phi^3_{\rm G,s} - 3\langle\phi^2_{\rm G,s}\rangle_V\phi_{\rm G,s}\right).\end{aligned} \tag{1.194}$$

Thus, from the second line in Eq. (1.194), we see that $g_{\rm NL}$ causes long modes to modulate short modes in a very similar way as $f_{\rm NL}$. $f_{\rm NL}$ and $g_{\rm NL}$ measurements are therefore degenerate. This degeneracy can only be broken with bispectrum measurements or by including a $g_{\rm NL}$-dependent term in Eq. (1.189), which is only valid for highly biased tracers ($b_0 \gtrsim 2.5$), such as quasars, because the $g_{\rm NL}$-dependent bias term strongly depends on the halo occupation distribution for less biased tracers [104]. Given the limited constraining power of current galaxy surveys on that matter, fitting $f_{\rm NL}$ separately is still a useful test, but this is an important caveat for future measurements.

Another important caveat comes from the briefly mentioned gauge dependence of general relativistic perturbation theory. This is due to a certain degree of freedom in how to split the general metric into perturbations and a background. As δ is defined with respect to a background, the bias is a gauge dependent quantity, too. This means that a constant local bias in one particular gauge is generally scale-dependent in another gauge. Therefore, our error budget includes a theoretical error of the order of $\sigma_{f_{\rm NL},{\rm GR}} \sim 1$ [105]. However, we shall see in the next subsection that this is still below of what we can measure today.

1.6 Current Constraints on Inflation and Primordial Non-Gaussianity

The tightest constraints on $f_{\rm NL}$ to date come from the angular bispectrum of fluctuations in the temperature and E-mode polarisation of the cosmic microwave background (CMB) measured by the Planck collaboration [3]. Both the temperature map $\frac{\Delta T}{T}(\mathbf{n})$ and the E-mode polarisation map $E(\mathbf{n})$ are decomposed into spherical harmonics $a^{\rm (T)}_{\ell m}$ and $a^{\rm (E)}_{\ell m}$, respectively. The bispectrum

is then defined analogously to Eq. (1.105), that is

$$B^{m_1 m_2 m_3 X_1 X_2 X_3}_{\ell_1 \ell_2 \ell_3} = \left\langle a^{(X_1)}_{\ell_1 m_1} a^{(X_2)}_{\ell_2 m_2} a^{(X_3)}_{\ell_3 m_3} \right\rangle , \tag{1.195}$$

where $X_i \in \{E, T\}$. f_{NL} can then be measured as defined in Eq. (1.107). This measurement is contaminated by non-Gaussianities due to the Integrated Sachs Wolf effect (ISW), gravitational lensing, extragalactic point sources and cosmic rays hitting the cryogenic detectors of the Planck satellite, hence, the systematic effects are completely different than those affecting galaxy redshift surveys. Galaxy clustering observations are therefore an independent way to confirm the CMB observations. Using temperature data alone, the Planck collaboration found a local f_{NL}-value of $f_{\mathrm{NL}} = 2.5 \pm 5.7$, including also polarisation data the constraints slightly improve to $f_{\mathrm{NL}} = 0.8 \pm 5.0$. This means, that none of the inflation models presented in Sec. 1.3 have been ruled out yet. The next generation ground-based CMB experiment CMB-S4 will tighten the constraint to $\sigma(f_{\mathrm{NL}}) = 2.5$, or $\sigma(f_{\mathrm{NL}}) = 1.8$ if combined with Planck [106]. These are the strongest constraints achievable with ground-based CMB measurements due to cosmic variance. The Cosmic Origins Explorer (CORE), which has been proposed to the European Space Agency (ESA) in October 2016, will be able to slightly improve this value down to $\sigma(f\mathrm{NL}) = 2.1$ [107], but it is not sure yet whether it will be realised. As future CMB experiments will not provide the accuracy we need, a different approach is needed to independently confirm these results, if not improve these constraints to further narrow down our understanding of the physics that governed the inflationary epoch.

Studies of the large scale structure of the universe, using galaxies or quasars, so far could not compete with the precision of the CMB f_{NL} results. Slošar *et al.* [101] found $-31 < f_{\mathrm{NL}} < 70$ at 95 per cent confidence using a combination of spectroscopic and photometric luminous red galaxies as well as photometric quasars from SDSS (cf. Sec. 2.2). Ross *et al.* [6] obtained $-82 < f_{\mathrm{NL}} < 178$ from the ninth data release of the Baryon Oscillation Spectroscopic Survey (BOSS, cf. Sec. 2.2). Leistedt, Peiris and Roth [108] applied a novel error mitigation technique, which is based on the mode deprojection technique described in Sec. 4.3, to the SDSS photometric quasar sample and found $-49 < f_{\mathrm{NL}} < 31$ when constraining f_{NL} individually. They also took the degeneracy of f_{NL} and g_{NL} into account, which weakened their constraints to $-105 < f_{\mathrm{NL}} < 72$.

Upcoming galaxy surveys will come close to independently confirm the Planck results (cf. Sec. 2), and, especially if combined with CMB data, will improve $f_{\rm NL}$ constraints such that they will provide physically interesting results and might rule out some of the models discussed in Sec. 1.3. I shall introduce some promising future surveys in the next section, where I shall also present forecasts of how well they will constrain $f_{\rm NL}$.

1.7 Summary and Thesis Outline

This introductory chapter has outlined the currently favoured model of the Universe, the concordance model, based on General Relativity, the Cosmological Principle, a hot big bang and the assumption of a universe filled with independent perfect fluids with different equations of state corresponding to relativistic radiation, non-relativistic matter and dark energy, that in the simple ΛCDM model is equal to a cosmological constant. I have summarised the key observations that provide evidence for the concordance model. One way of testing the concordance model is to study the distribution of galaxies in the universe. The structures they form arose from fluctuations around the homogeneous and isotropic background. I have shown how these fluctuations evolve and how they can be statistically described by summary statistics such as the 2-point correlation function, the power spectrum or the bispectrum. If the fluctuations are Gaussian distributed, they can be fully described by the power spectrum and the bispectrum vanishes. In the presence of a bispectrum (or other connected polyspectra), one can introduce a phenomenological parameter $f_{\rm NL}$ (and other parameters $g_{\rm NL}, \tau_{\rm NL}$, etc.) to amount the departure from Gaussianity. I have introduced the concept of bias and showed that in the presence of local non-Gaussianity it becomes a scale dependent quantity. The concordance model is not complete yet, as it does not provide a mechanism to seed the initial perturbations. One possible extension is the hypothetical Inflation paradigm, where non-Gaussianity is phenomenologically important and thus can be tested by using large scale structure data. I have introduced some inflationary scenarios focussing on how they can be distinguished by measuring non-Gaussianity. I finished the introduction by giving current constraints on $f_{\rm NL}$.

The following chapter presents ongoing and future large-scale structure surveys that are suitable to improve the constraints on $f_{\rm NL}$ that we currently have. As there are no surveys designed especially to measure non-Gaussianity, I start the chapter by touching on two of the most common science goals, namely baryon acoustic oscillations and redshift space distortions. I shall also briefly discuss other measurements such as the matter, cold dark matter and neutrino densities. The aim of my work is to use the Baryon Oscillation Spectroscopic Survey (BOSS) to constrain $f_{\rm NL}$. Therefore, I do not only present the basic specifications of BOSS, but also describe how galaxies are selected and how systematic errors are treated by the BOSS collaboration. It follows an overview of the extended BOSS (eBOSS), the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX), the Dark Energy Spectroscopic Instrument (DESI), Euclid, the Dark Energy Survey (DES) and some radio surveys. In the remainder of Chapter 2, I describe some common galaxy clustering analysis techniques to generate random galaxy catalogues (Sec. 2.4.1) needed to measure the power spectrum (Sec. 2.4.2), the correlation function (Sec. 2.4.3) and the window function (Sec. 2.4.4).

One major issue that we face when constraining $f_{\rm NL}$ from the power spectrum is that primordial non-Gaussianity affects the power spectrum at large scales only, thus, we have to rely on a low number of large modes, i.e. we cannot make use of the Central Limit Theorem and assume a Gaussian likelihood in the inference process, even if we vary the covariance matrix according to the model to be tested. This problem is addressed in Chapter 3 after a general introduction into the statistics of inference (cf. Sec. 3.1) and the covariance matrix of the power spectrum (cf. Sec. 3.2). In Sec. 3.3, alternative posterior distributions are compared with what we expect as the true posterior, both mode-by-mode and in terms of expected measurements of primordial non-Gaussianity parameterised by $f_{\rm NL}$. Marginalising over a Gaussian posterior distribution $\mathcal{P}_f$ with fixed covariance matrix yields a posterior mean value of $f_{\rm NL}$ which, for a data set with the characteristics of Euclid, will be underestimated by $\triangle f_{\rm NL} = 0.4$, while for the data release 9 (DR9) of the Sloan Digital Sky Survey (SDSS)-III Baryon Oscillation Spectroscopic Survey (BOSS) it will be underestimated by $\triangle f_{\rm NL} = 19.1$. Adopting a different form of the posterior function (cf. Sec. 3.4) means that we do not necessarily require a different

covariance matrix for each model to be tested: this dependence is absorbed into the functional form of the posterior. Thus, the computational burden of analysis is significantly reduced. In Sec. 3.5, a comparison of Taylor coefficients of the true posterior distribution and one of the approximate posterior distributions tested in Sec. 3.4, shows a remarkable agreement between the two. The impact of using these different posterior shapes in measurements of $f_{\rm NL}$ from BOSS DR9 and Euclid is studied in Sec. 3.6.

Another challenge we face at these scales are foreground effects contaminating the data, especially due to stars. In Chapter 4, I suggest a way to mitigate these contaminants. This process is split into three separate stages: (i) removing the contaminant signal, (ii) estimating the uncontaminated cosmological power spectrum, (iii) debiasing the resulting estimates. For (i), we show that removing the best-fit contaminant (*mode subtraction*, cf. Sec. 4.4), and setting the contaminated components of the covariance to be infinite (*mode deprojection*, cf. Sec. 4.3) are mathematically equivalent. For (ii), performing a Quadratic Maximum Likelihood (QML, Sec. 4.2) estimate after *mode deprojection* gives an optimal unbiased solution, although it requires the manipulation of large $N_{\rm mode}^2$ matrices ($N_{\rm mode}$ being the total number of modes), which is unfeasible for recent 3D galaxy surveys. Measuring a binned average of the modes for (ii) as proposed by Feldman, Kaiser & Peacock (FKP, [109]) is faster and simpler, but is sub-optimal and gives rise to a biased solution. In Sec. 4.5, we present a method to debias the resulting FKP measurements that does not require any large matrix calculations. We argue that the sub-optimality of the FKP estimator compared with the QML estimator, caused by contaminants is less severe than that commonly ignored due to the survey window. In Sec. 4.6, I extend the technique to multiple contaminants. This mitigation technique is tested using toy examples in Sec. 4.7, before its application to BOSS CMASS data is discussed in Chapter 5. A general discussion of how to generate systematic templates can be found in Sec. 5.1. The large scale data of BOSS CMASS is strongly affected by the stellar foreground density. Using a single template to mitigate the effect of the stars (cf. Sec. 5.2) results in a power spectrum measurement that does not agree well with our model power spectrum even for non-zero values of $f_{\rm NL}$, but it is consistent with the power spectrum using the standard BOSS weighting technique and with earlier

$f_{\rm NL}$ measurements from BOSS DR9 CMASS data [6]. The analysis is extended in Sec. 5.3 using more than one template for the effect of stars, taking into account the uncertainty of the template building. As this does not remove the tension, we include more templates to account for the effects of seeing, the airmass, galactic extinction, and survey's scanning strategy in Sec. 5.4- 5.7. The inclusion of these additional templates also does not alleviate the discrepancy between our measurement and our model. Therefore, we consider alternative stellar templates in Sec. 5.8 and 5.9. Preliminary work about identifying the source of the remaining excess signal is presented in Sec. 5.10 and leads into Chapter 6, where I present future work and conclude this thesis.

Chapter 2

Galaxy Surveys

This chapter introduces current and upcoming galaxy surveys, their key observations and the basic techniques needed to make these. Optical galaxy surveys fall into two categories: photometric and spectroscopic. For **photometric** surveys, a telescope takes images with different colour filters, whereas for a **spectroscopic** survey, the light of single objects passes through dispersive elements which allows to measure the full spectrum of the source. The former allows to observe many objects simultaneously, but accurate redshift information and identification of the object requires a spectroscopic measurement. The Sloan Digital Sky Survey (SDSS) provides both photometric and spectroscopic observations, first many objects are observed photometrically, of which some are chosen under certain criteria for spectroscopic follow-up observations.

Although galaxy surveys provide a good way to test primordial non-Gaussianity, their primary science goals are usually measurements of the baryon acoustic oscillations (BAO) and redshift-space distortions (RSD). These two effects are first described in the next section along with other power spectrum based parameter measurements, before introducing several optical galaxy surveys in Sections 2.2 and 2.3. The discussion of the Baryon Oscillation Spectroscopic Survey (BOSS) in Sec. 2.2 includes its selection criteria and a description of the standard systematic mitigation technique applied in the BOSS analyses for later reference. Sec. 2.3.6 provides a discussion of radio surveys, as they also allow f_{NL} measurements and, in time to come, will provide a promising complement to optical f_{NL} measurements.

Sec. 2.4 introduces basic concepts and techniques needed in galaxy clustering analyses, including random catalogues, power spectrum and correlation function measuring techniques, as well as a description of the window function.

2.1 Primary Science Goals of Galaxy Surveys

2.1.1 Baryon Acoustic Oscillations (BAO)

As discussed in Sec. 1.4, acoustic oscillations in the photon-baryon plasma before recombination cause a baryonic over-density at the scale of the sound horizon of the baryon-drag epoch $r_{\rm d}$. Even though this is less dominant than the over-density in cold dark matter, which attracts baryons into the centre of potential wells, we can observe this BAO feature imprinted in the distribution of galaxies today as a peak in the two point correlation function, or a series of peaks and troughs in the galaxy power spectrum. An example of a BAO measurement from the Baryon Oscillation Spectroscopic Survey is reproduced in Fig. 2.1. To make the BAO feature more prominent, a smooth model, i.e. one that does not contain the BAO feature, has been subtracted from the correlation function plots and divided out of the power spectrum plots. In the top right panel, the BAO peak can clearly be seen at around 100 h^{-1} Mpc. In the Fourier domain, this corresponds to the ripples that are visible in the top left. As changes to the BAO scale after recombination are almost entirely due to the Universe's expansion, it serves as a standard ruler that provides a means to determine the underlying cosmological model and its expansion rate. By decomposing the two point correlation function $\xi(s_\perp, s_\parallel)$ or the power spectrum $P(k_\perp, k_\parallel)$, into components of separation transverse to and along the line of sight, as shown in the bottom panels of Fig. 2.1, one can relate the BAO feature to basic cosmological parameters. One can define the angular diameter distance $D_{\rm A} \equiv \frac{x}{\theta}$ as the ratio of an object's physical and angular size. Thus, by measuring the angle θ subtended by the BAO feature at a redshift z, one can infer the angular diameter distance

$$D_{\rm A}(z) = \frac{r_{\rm d}}{\theta(1+z)}. \tag{2.1}$$

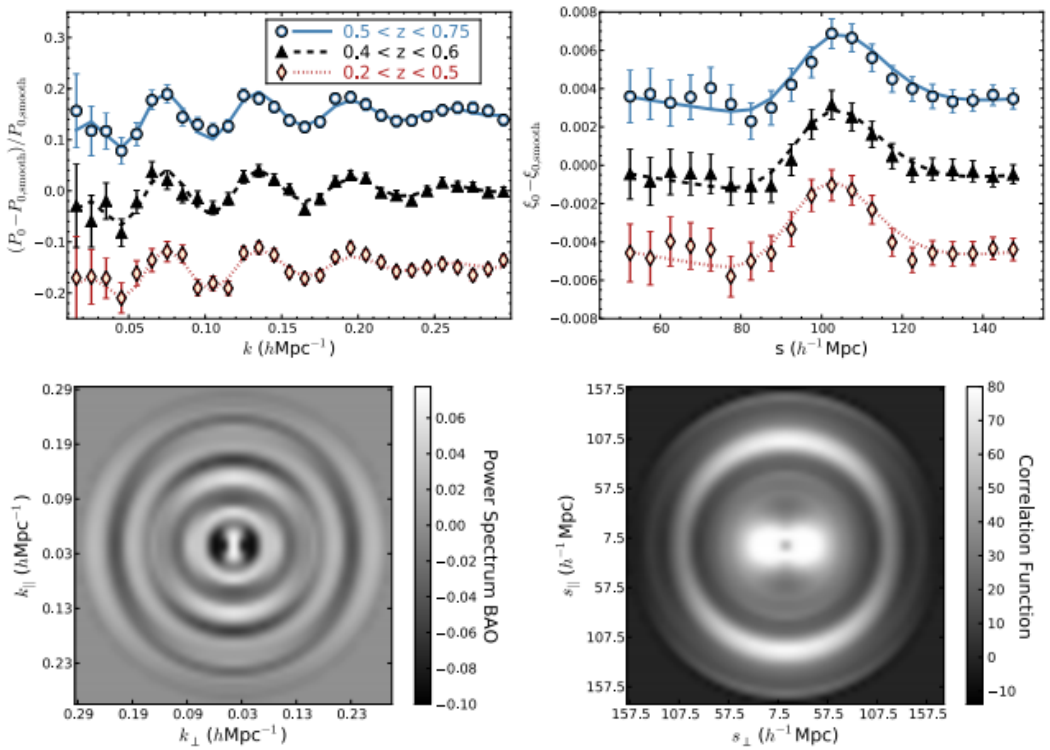

Figure 2.1: BAO signals in the measured post-reconstruction power spectrum (left panels) and correlation function (right panels) and predictions of the best-fit BAO models (curves). To isolate the BAO in the monopole (top panels), predictions of a smooth model with the best-fit cosmological parameters but no BAO feature have been subtracted, and the same smooth model has been divided out in the power spectrum panel. For clarity, vertical offsets of ± 0.15 (power spectrum) and ± 0.004 (correlation function) have been added to the points and curves for the high- and low-redshift bins, while the intermediate redshift bin is unshifted. The bottom panels show the measurements for the $0.4 < z < 0.6$ redshift bin decomposed into the component of the separations transverse to and along the line of sight. Figure taken from [14].

In addition, the Hubble expansion parameter is related to the change in redshift of the BAO feature along the line of sight Δz through

$$H(z) = \frac{c\Delta z}{r_\mathrm{d}}. \tag{2.2}$$

In practice, a fiducial cosmological model has to be assumed to transfer redshifts into distances, and the observed BAO scale is defined relative to this fiducial model. The two **Alcock-Paczyński** [110] parameters

$$\alpha_\perp = \frac{D_\mathrm{A}(z) r_d^{(\mathrm{fid})}}{D_\mathrm{A}^{(\mathrm{fid})}(z) r_d} \tag{2.3}$$

and

$$\alpha_\parallel = \frac{H^{(\mathrm{fid})}(z) r_d^{(\mathrm{fid})}}{H(z) r_d} \tag{2.4}$$

can be understood as dilations of the axes of the plot in the bottom right panel of Fig. 2.1. They can be measured by fitting multipoles of the anisotropic power spectrum, such as the monopole

$$P_0(k) = \frac{1}{2}\left(\frac{r_\mathrm{d}^{(\mathrm{fid})}}{\alpha_\perp^2 \alpha_\parallel r_\mathrm{d}}\right)^3 \int_{-1}^{1} P(k', \mu')\mathrm{d}\mu \tag{2.5}$$

and the quadrupole

$$P_2(k) = \frac{5}{2}\left(\frac{r_\mathrm{d}^{(\mathrm{fid})}}{\alpha_\perp^2 \alpha_\parallel r_\mathrm{d}}\right)^3 \int_{-1}^{1} P(k', \mu')\mathcal{L}_2(\mu)\mathrm{d}\mu, \tag{2.6}$$

where $r_\mathrm{d}^{(\mathrm{fid})}$ is the scale of the sound horizon of the baryon-drag epoch in the fiducial model, $k' \equiv \sqrt{(k'_\perp)^2 + (k'_\parallel)^2}$ is the absolute value of the true wave number that is related to the observed wave number by $k'_\perp = k_\perp/\alpha_\perp$ and $k'_\parallel = k_\parallel/\alpha_\parallel$, μ and μ' are the observed and true cosines of angles to the line of sight, and $\mathcal{L}_2(\mu)$ is the second order Legendre polynomial [111]. Analogously, one can also fit multipoles of the two-point correlation function [112] to obtain the parameter perpendicular to the line of sight $\alpha_\perp$ that probes the comoving distance χ and the parameter parallel to the line of sight $\alpha_\parallel$ that can be used to constrain the Hubble parameter (e.g. [14, 21]). These two parameters are still degenerate with r_d, but r_d has been measured with 0.2 per cent accuracy with Planck from CMB data. As this is 10-20 times more accurate than the final BOSS constraints on $\alpha_\perp$ and $\alpha_\parallel$ [14], they can be regarded as constraints on $H(z)$ and $D_\mathrm{A}(z)$, which is related to the comoving distance by $D_\mathrm{A}(z) = \frac{\chi(z)}{1+z}$.

In the case where only the monopole power spectrum is used to make BAO measurements, one can define an averaged dilation scale parameter

$$\alpha \equiv \sqrt[3]{\alpha_\perp^2 \alpha_\parallel} = \frac{D_V(z) r_d^{(\mathrm{fid})}}{D_V^{(\mathrm{fid})}(z) r_d} \tag{2.7}$$

that relates the ratio of the measured BAO scale and that predicted by the fiducial model $r_d^{(\mathrm{fig})}$ with the ratio of the measurement and fiducial model prediction of a distance function

$$D_V(z) \equiv \sqrt[3]{(1+z)^2 D_\mathrm{A}^2(z) \frac{cz}{H(z)}}, \tag{2.8}$$

Hence one can still obtain a degenerate measurement combining the angular diameter distance to the effective redshift of the survey $D_\mathrm{A}(z)$ and the Hubble expansion rate $H(z)$ from angularly averaged BAO data. A summary of recent spherically averaged BAO distance measurements is given in Fig. 1.2.

We have seen that BAOs provide a standard ruler of the Universe and one can learn about the expansion of the Universe by comparing the BAO distance $D_\mathrm{V}(z)$; or $D_\mathrm{A}(z)$ and $H(z)$ at different redshifts. This is a very accurate cosmological measurement and as they are also very robust to systematic effects [113], they are regarded as one of the pillars of modern cosmology (cf. Sec. 1.1.6).

2.1.2 Redshift-Space Distortions (RSD)

If we have a closer look on the bottom panels of Fig. 2.1, one can see that the (almost) circular symmetry is more squeezed as you approach the centres of the plots. This is due to Redshift Space Distortions (RSD) that leads to an isotropic distribution of galaxies appearing anisotropic due to the fact that not only the expansion of the Universe but also the radial component of peculiar velocities $v_\parallel$ contribute to the redshift. As galaxies are attracted by over-dense regions, they tend to have peculiar velocities towards the centre of mass of the over-dense region. If they are between observer and centre of mass, the peculiar velocity causes an additional redshift and the galaxy appears further away than it actually is. For the same reason, galaxies behind the centre of mass are blue-shifted and hence appear closer. The redshifts of objects, whose peculiar motion is perpendicular to the line of sight, are only due to Hubble expansion

and therefore provide the correct distance information. Consequently, one can relate the redshift-space position (see sketch in Fig. 2.2)

$$\mathbf{s}(\mathbf{r}) = \mathbf{r} + \frac{v_\parallel(\mathbf{r})}{\mathcal{H}} \frac{\mathbf{r}}{|\mathbf{r}|} \tag{2.9}$$

and the real-space position $\mathbf{r}$ using $v_\parallel$, which is the peculiar velocity component along the line of sight. Regardless of the coordinate system used to describe the galaxy density field (written as $\rho_\mathrm{g}(\mathbf{r})$ or $\rho_\mathrm{g}^{(\mathrm{s})}(\mathbf{s})$ in configuration or redshift space, respectively) the total number of galaxies in a given volume is conserved in the mapping

$$\rho_\mathrm{g}(\mathbf{r})\mathrm{d}^3\mathbf{r} = \rho_\mathrm{g}^{(\mathrm{s})}(\mathbf{s})\mathrm{d}^3\mathbf{s}. \tag{2.10}$$

Therefore, one obtains the redshift space density field by multiplying the configuration space density field with the Jacobian

$$\mathcal{J} = \left|\frac{\mathrm{d}^3\mathbf{r}}{\mathrm{d}^3\mathbf{s}}\right| = \left(1 + \frac{\partial}{\partial\mathbf{r}} \frac{v_\parallel(\mathbf{r})}{\mathcal{H}} \frac{\mathbf{r}}{|\mathbf{r}|}\right)^{-1} \left(1 + \frac{v_\parallel(\mathbf{r})}{\mathcal{H}|\mathbf{r}|}\right)^{-2}. \tag{2.11}$$

Making the "distant observer" approximation, i.e. $|\mathbf{r}| \gg v_\parallel(\mathbf{r})/\mathcal{H}$, one can drop the second bracket term in the Jacobian. Furthermore, one can Taylor expand the Jacobian around $v_\parallel(\mathbf{r}) = 0$, such that

$$\mathcal{J} \approx \left(1 - \frac{\partial}{\partial\mathbf{r}} \frac{v_\parallel(\mathbf{r})}{\mathcal{H}} \frac{\mathbf{r}}{|\mathbf{r}|}\right) \tag{2.12}$$

Inserting the divergence of the peculiar velocity from Eq. (1.88) and expressing the cosine of the angle between the velocity vector and the line of sight μ as in the previous subsection, one can find

$$\mathcal{J} \approx \left(1 + \frac{f}{b}\mu^2\right), \tag{2.13}$$

where the factor of $1/b$ comes from the fact that the peculiar velocity is caused by the underlying matter density field, not the biased galaxy density. Inserting $1+\delta_\mathrm{g}^{(\mathrm{s})} = \left(1 + \frac{f}{b}\mu^2\right)(1 + \delta_\mathrm{g})$ into the definition of the power spectrum, it follows that the linear redshift-space galaxy power-spectrum

$$P_\mathrm{gg}^{(\mathrm{s})}(k, \mu) = (b + f\mu^2)^2 P_\mathrm{m}(k) \text{ [114]} \tag{2.14}$$

depends on the angle $\arccos(\mu)$. As f is scale-independent on large scales in General Relativity [115], measuring $f \equiv \Omega_\mathrm{m}^{\gamma(z)}(z)$ is a strong test of General Relativity [116], where $\gamma \approx \frac{4}{7}$ at all redshifts (cf. Eq. (1.90), [53]).

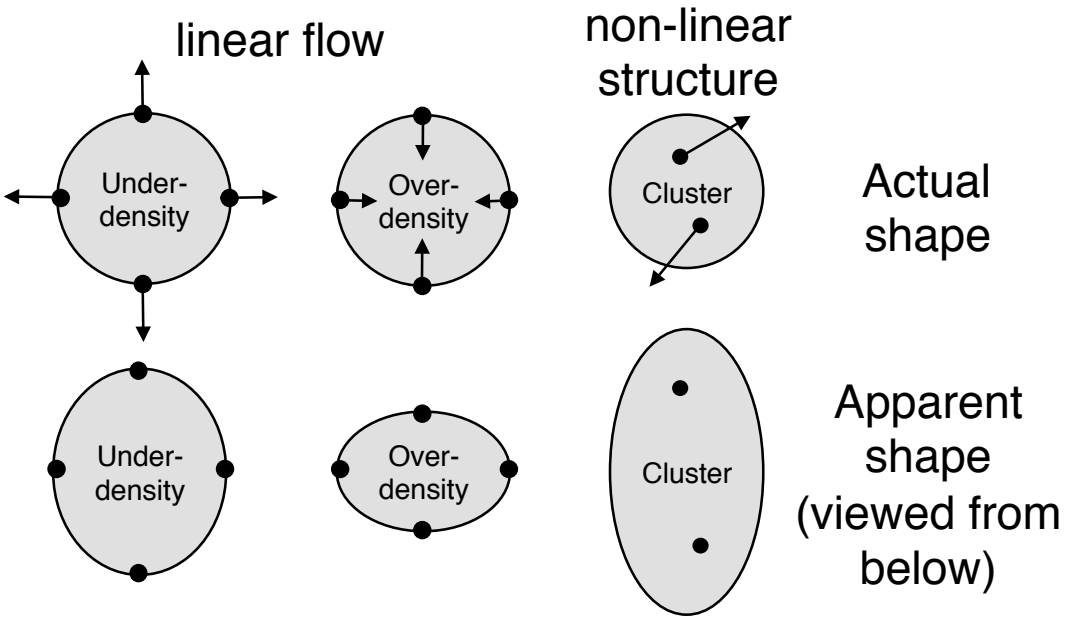

Figure 2.2: Diagram explaining how real-space structures (top) look in redshift-space (bottom) (taken from [21]).

As we are going to measure $f_{\rm NL}$ from the power spectrum at large scales only, we can average over all angles μ in Eq. (2.14) and are not affected by RSD other than causing an increase in the amplitude of the clustering, and a corresponding improvement in the signal to noise. However, Tellarini *et al.* [10] pointed out that primordial non-Gaussianity causes signatures in the galaxy bispectrum that resemble RSD effects.

2.1.3 Other Measurements

Besides BAO and RSD measurements, galaxy surveys also allow full shape measurements of the power spectrum, which in turn provide a probe of all cosmological parameters entering a model power spectrum. For instance, as discussed in Sec. 1.4, density perturbations in a universe dominated by radiation behave differently than in a matter-dominated universe. Therefore, the scale of the particle horizon at matter-radiation equality

$$k_{\rm eq} = \sqrt{2\Omega_{\rm m,0} H_0^2 z_{\rm eq}} \tag{2.15}$$

is imprinted on the matter transfer function [88]. Perturbations on small scales $k > k_{\rm eq}$ are suppressed, as can be seen in the solid lines in Fig. 2.3. The figure

shows linear model power spectra computed with CAMB [22], where the blue and red solid lines are for 10 per cent less and more matter, respectively. The suppression effect is even more clear in the dotted lines that represent changes in the dark matter density only, as changing the baryon density changes the speed of sound of the baryon-photon plasma prior to (re)combination and consequently the sound horizon at the drag epoch $r_{\rm d}$. Hence, the acoustic oscillation peaks in the power spectrum get shifted, as can be seen in the dashed lines in Fig. 2.3. Thus, one can measure $\Omega_{\rm m,0}h^2$ and $\Omega_{\rm b,0}h^2$ from a multi-parameter fit to the full power spectrum $P(\mathbf{k})$ for sufficiently good data and it is possible to obtain more information on cosmology.

A multi-parameter fit to the power spectrum also has the potential to provide new insights into fundamental physics, such as measuring neutrino masses. Neutrinos are so light that for a long time they were assumed to be massless. However, neutrinos can change their flavours and experiments have shown that they must have a finite mass [117]. As they are very light, they remained relativistic for a long time even after recombination: thus they add to the radiation density at the drag epoch. At least two neutrino flavours are massive enough to have become non-relativistic before our current epoch, such that they now contribute to the dark matter [23]. This changing nature of the neutrinos affects the power spectrum non- trivially. As an example, when neutrinos turn non-relativistic, they add to the suppression of the power spectrum due to matter, but this occurs at scales $k > k_{\rm nr}$ that correspond to when they become non-relativistic rather than the matter-radiation equality scale $k_{\rm eq}$ (cf. Fig. 2.4 [23]), leaving a distinctive imprint on the power spectrum. As $k_{\rm nr}$ depends on the mass of the neutrino, one can also infer $\Omega_{\nu,0}h^2$ from the power spectrum. Furthermore, finding $k_{\rm nr}$ for each neutrino flavour can potentially yield absolute values of each neutrino flavour mass [118]. A problem in neutrino mass measurements from the power spectrum is that the scales $k_{\rm nr}$ are expected to be in the non-linear regime of the power spectrum.

The aim of this thesis is to use a bias measurement to constrain the non-Gaussianity parameter $f_{\rm NL}$. Given a theoretical model of the linear matter power spectrum $P_{\rm m}(\mathbf{k})$, with Eq. (1.180), the local linear bias parameter can be defined as

$$P_{\rm gg}(\mathbf{k}) = b^2 P_{\rm m}(\mathbf{k}). \tag{2.16}$$

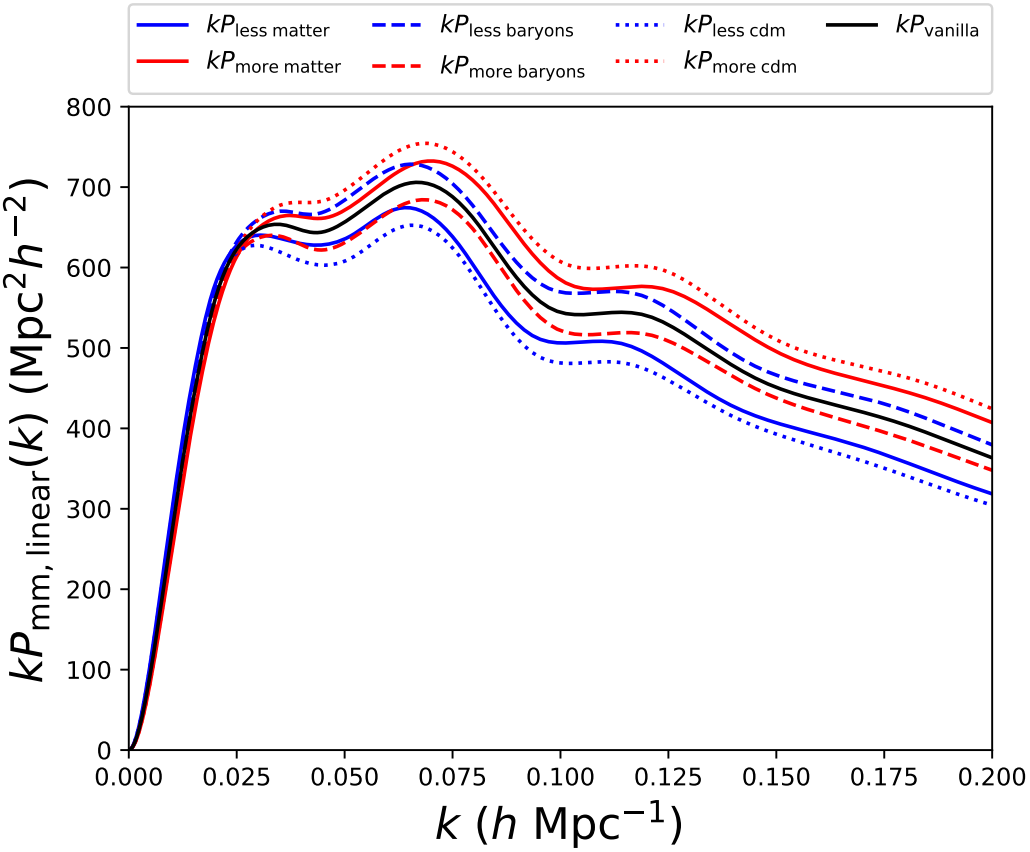

Figure 2.3: Linear matter power spectra generated using CAMB [22]. The black line corresponds to a power spectrum P_{vanilla} with input parameters $H_0 = 70$ km/s/Mpc, $\Omega_{\text{b},0}h^2 = 0.0226$, $\Omega_{\text{cdm},0}h^2 = 0.112$, $\Omega_{\nu,0}h^2 = 0.00064$ and $\Omega_{\text{k}} = 0$. For the blue and red line, the matter density, thus both the $\Omega_{\text{b},0}h^2$ and $\Omega_{\text{cdm},0}h^2$ inputs, has been decreased or increased, respectively, by 10 per cent.

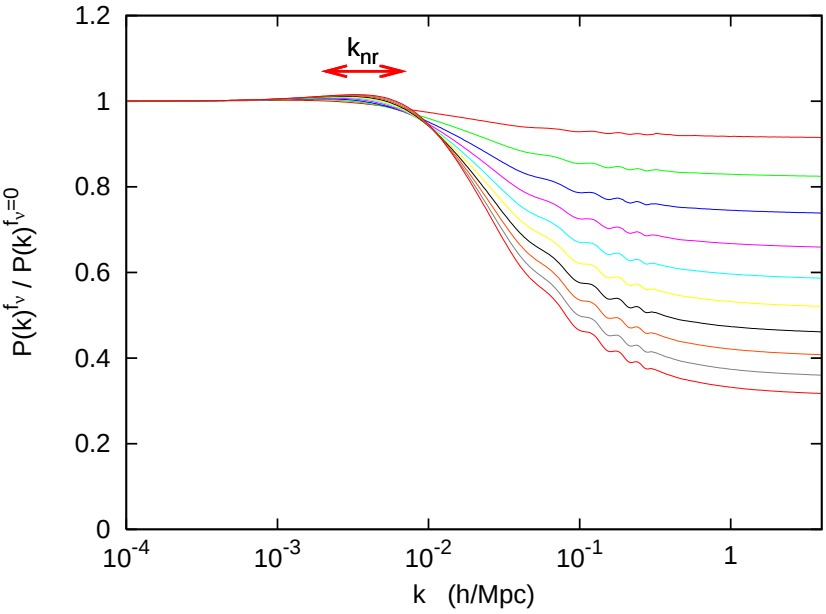

Figure 2.4: Ratio of the matter power spectrum including three degenerate massive neutrinos with density fraction $f_\nu \equiv \Omega_\nu/\Omega_\mathrm{m}$ to that with three massless neutrinos. The parameters $\Omega_{\mathrm{m},0}h^2 = 0.147$ and $\Omega_\Lambda = 0.7$ are kept fixed, and from top to bottom the curves correspond to $f_\nu = 0.01, 0.02, 0.03, ..., 0.1$. Figure taken from [23].

The amplitude of the over-density field, usually parametrised using $\sigma_8 \equiv \sigma(8h^{-1}\text{Mpc})$ (cf. Eq. 1.170) is degenerate with bias measurements. A similar degeneracy is seen for RSD measurements where the relationship between the theory and observed power spectra are modelled by Eq. (2.14) leading to f being degenerate with σ_8. Therefore, some authors consider the combinations $b\sigma_8$ and $f\sigma_8$ rather than b, f and σ_8. The degeneracy can be broken by the multi-tracer technique (cf. Sec. 1.5.2) because σ_8 depends on the cosmology whereas the bias b is specific to the class of tracers. In the presence of local non-Gaussianity, the degeneracy is also broken for a single tracer, because the bias is scale dependent, but σ_8 is not. We therefore should marginalise over b and σ_8 separately when measuring f_{NL}. Before delving further into problems related to f_{NL} measurements in Chapters 3 to 5, I first introduce the surveys and techniques applied when using galaxy clustering to make cosmological inference.

2.2 The Baryon Oscillation Spectroscopic Survey

The Baryon Oscillation Spectroscopic Survey (BOSS) is part of the third stage of the Sloan Digital Sky Survey (SDSS-III, [119, 120]) that uses the Sloan Foundation 2.5-meter Telescope based at Apache Point Observatory (APO) in New Mexico. The other two SDSS-III experiments are the Apache Point Observatory Galactic Evolution Experiment (APOGEE) and the Multi-object Apache Point Observatory Radial Velocity Exoplanet Large-area Survey (MARVELS). The data was taken from 2008 to 2014 and was published in five data releases numbered DR8-DR12, continuing the consecutive numbering of SDSS-I and SDSS-II. All data releases, bar DR11, are publicly available from http://www.sdss.org/data/.

The **SDSS-I** and **SDSS-II** [1] experiments used the drift-scanning technique to perform an imaging survey over 14,055 square degrees. In this method of observing, the telescope remains stationary while a mosaic of 30 charge-coupled device (CCD) cameras [121] mounted on the telescopes observes stripes of

[1] http://www.sdss2.org/

the sky in 5 photometric band passes u, g, r, i, z as the Earth rotates. The stars drifting through the image provided a constant baseline for astrometric calibration. 500 million objects in an area covering about a third of the sky were thus imaged.

In addition to making photometric images within the SDSS-I and SDSS-II experiments, targets were selected from a uniform, photometrically well calibrated 8,400 square degree area for subsequent spectroscopic observation. The sky was split into tiles with a diameter of 3 degrees. For each tile, an aluminium plate was produced with holes to plug 640 optical fibres connected to a spectrograph. This provided the magnitude-limited Main Galaxy Sample (MGS), the near-volume-limited sample of galaxies called Luminous Red Galaxies (LRG), and a magnitude-limited sample of quasars. In DR3, the LRG sample allowed the first detection of the BAO peak in the galaxy correlation function [122].

After SDSS-II finished in 2008, the spectrograph was upgraded for the **SDSS-III** experiment that includes BOSS and is discussed in the remainder of this section. The upgraded SDSS-III spectrograph allowed 1000 spectra to be taken simultaneously. As part of DR8 [123], the first SDSS-III release, the photometric data of SDSS-I and SDSS-II was reprocessed [124] and galaxies were selected for follow-up spectroscopic observation for BOSS. BOSS observed two spectroscopic galaxy samples, the Low Redshift (LOWZ) sample consisting of 361,762 LRGs at $0.16 < z < 0.36$, and the Constant Mass (CMASS) sample, that includes both LRGs and fainter blue galaxies at $0.43 < z < 0.7$. By combining the two red and blue populations into one single sample the shot-noise in the measured density field is reduced. The total number of galaxies in CMASS amounts to 777,202, of which 568,776 are in the Galactic North and the remaining 208,426 galaxies are in the Galactic South [125]. There are also 13,290 "known" galaxy spectra from SDSS-II that fulfil the selection criteria of CMASS and are therefore also included. The number of "known" spectra for the LOWZ sample is much larger, with 153,517 "known" galaxies, mainly SDSS-II LRGs. The number densities of CMASS and LOWZ, including its subsets LOWZE2 and LOWZE3, are plotted in Fig. 2.5. The final footprint of BOSS covers 9329 square degrees and can be seen in Fig. 2.12 and in several figures in Chapter 5. As the CMASS sample probes a larger volume than the LOWZ sample, Ross *et al.* [6] and I have chosen it to constrain $f_{\rm NL}$. Therefore,

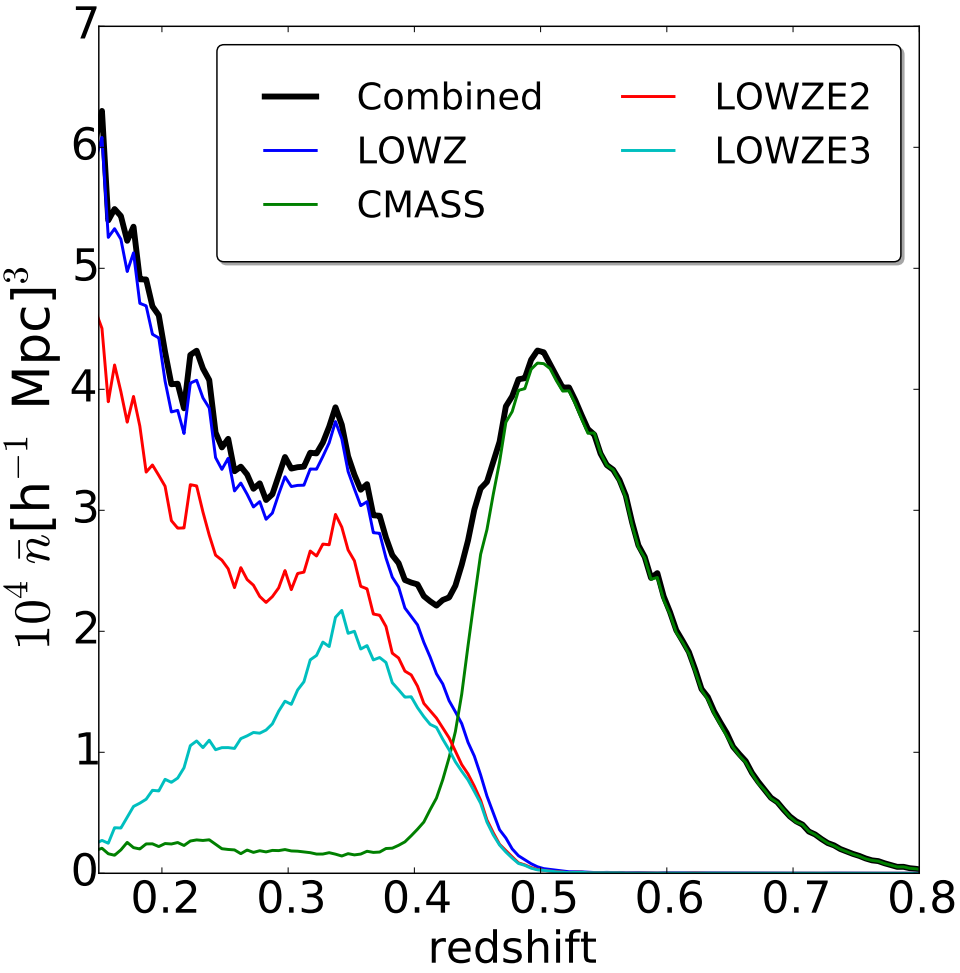

Figure 2.5: Number density of the target classes observed by BOSS as a function of redshift. Figure taken from [14].

I present more details about how galaxies have been selected to be part of the CMASS sample.

The colour criterion used in the selection process for the CMASS sample is dominated by limits on the parameter

$$d_\perp \equiv r_{\rm mod} - i_{\rm mod} - \frac{g_{\rm mod} - r_{\rm mod}}{8}, \tag{2.17}$$

where $g_{\rm mod}$, $r_{\rm mod}$ and $i_{\rm mod}$ are the model g, r and i-band magnitudes adopting either a de Vaucouleurs or an exponential luminosity profile, depending on which of the two fits better in the r-band [126]. Other important quantities in the selection process are

- $i_{\rm cmod}$, which is the model i-band magnitude calculated from the best-fitting linear combination of the de Vaucouleurs and exponential luminosity profiles [127], and
- the i-band magnitude within a $2''$ aperture radius $i_{\rm fib2}$.

The requirements on CMASS galaxies are then given by

$$\begin{aligned} 17.5 < i_{\rm cmod} &< 19.9 \\ r_{\rm mod} - i_{\rm mod} &< 2 \\ d_\perp &> 0.55 \\ i_{\rm fib2} &< 21.5 \\ i_{\rm cmod} &< 19.86 + 1.6(d_\perp - 0.8). \end{aligned} \tag{2.18}$$

Isolated stars can be distinguished from galaxies as they have profiles closer to that of the point spread function (PSF). After fitting the magnitudes $i_{\rm psf}$ and $z_{\rm psf}$ to the point spread function, one can define further criteria to avoid targeting stars:

$$\begin{aligned} i_{\rm psf} &> 4.2 + 0.98 i_{\rm mod} \\ z_{\rm psf} &> 9.125 + 0.54 z_{\rm mod}. \end{aligned} \tag{2.19}$$

Furthermore, objects in patches of the sky with high galactic extinction (cf. Sec. 5.6) of $E_{\rm B-V,SFD} > 0.08$, as determined using the dust maps of Schlegel, Finkbeiner and Davis [128], are excluded.

A crucial issue in the analysis of galaxy clustering data is the mitigation of systematic errors. This is especially true in the analysis of primordial non-Gaussianity. That is why two chapters 4 and 5 are solely dedicated to this important issue, where I introduce an alternative error mitigation technique and describe its application to BOSS data. To compare our novel technique with the procedure commonly used in the BOSS collaboration, I briefly introduce their weight-based method here. When downloading BOSS data, the data file contains weights assigned to each galaxy that account for various systematic effects. As will be discussed in Sec. 2.4, clustering analyses are commonly based on counting the number of objects in spatial bins. To account for systematic effects, instead of just counting the objects, one sums over the total weights

$$w_{\rm tot} = (w_{\rm cp} + w_{\rm noz} - 1)\, w_{\rm star} w_{\rm see} \text{ [129]} \tag{2.20}$$

assigned to each galaxy (assuming that they are normalised such that $\langle w_{\rm tot}\rangle = 1$), where each of the individual weights are explained in the following paragraphs:

Fibre collisions ($w_{\rm cp}$) As mentioned above, the survey area is split into 3° circular tiles that each correspond to an aluminium plate containing holes to plug the fibres connected to the spectrograph. The number of fibres per tile is restricted to 1000 and the finite size of the fibre ferrules prohibit placing fibres closer than $62''$ apart from each other on the same plate. Because of that, a 5.8 per cent fraction of CMASS targets and 3.3 per cent of LOWZ targets cannot be assigned a spectroscopic fibre, even though they fulfil the selection criteria [25]. Not accounting for this would lead to artefacts in the clustering statistics because galaxies in high density regions are less likely to be spectroscopically observed. For instance, Reid *et al.* [24] have shown that this introduces a bias in redshift, because targets at redshifts close to where CMASS's angular target density is largest are more likely to be fibre collided (cf. Fig. 2.6). Furthermore, as the tiles are circular, their edges overlap with other tiles, such that, in overlapping regions, galaxy spectra can be obtained even though they are subtend angles that are closer than $62''$. This can cause a spurious clustering signal, because some regions are revisited while others are not. Therefore, if there is at least one other object of the

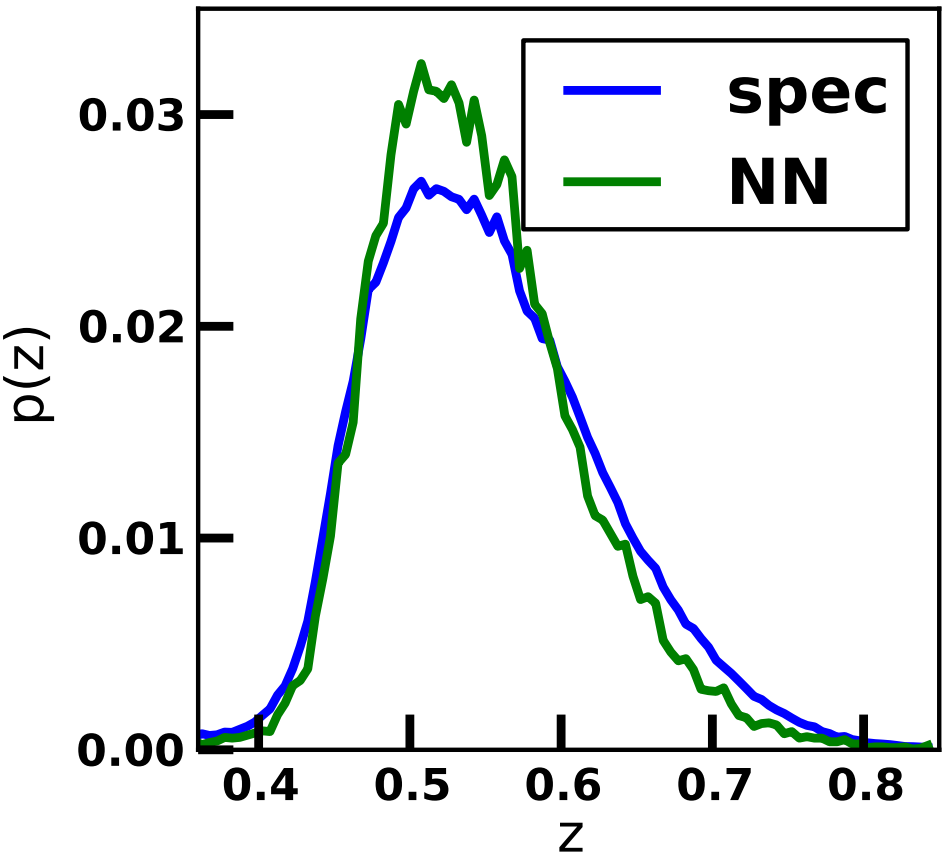

Figure 2.6: The normalised redshift probability distribution for CMASS targets that were assigned fibres (blue) and fibre collided galaxies (green). For collided galaxies, the nearest neighbour redshifts are used as a proxy; since the galaxy in a fibre collision pair that receives the fibre is randomly chosen, this is an unbiased estimate of the redshift distribution for objects without a fibre due to a fibre collision. Figure taken from [24].

same target class within the same fibre collision group, for each target without a fibre assigned to it, its nearest neighbour is up-weighted, thus the nearest neighbour's weight $w_{\rm cp}$ is increased by 1. However, if no other target of the same class obtained a fibre in the same fibre collision group, be it because there are no other ones targeted or because every other target happened to not have been assigned a fibre either, the galaxy is treated as a random incompleteness and no up-weighting takes place [25]. This nearest-neighbour weighting scheme assumes that all targets in a fibre collision group are equally likely to be fibre collided. A problem with this assumption is that the tiling was optimised to obtain as many redshifts as possible. For instance, in a triplet of galaxies, the pair with the largest angular separation is more likely to have fibres assigned to them. As the standard approach of measuring the two point correlation function is based on counting pairs of galaxies with a certain separation rather than individual galaxies, Bianchi & Percival [130] suggest a pairwise-inverse-probability weighting scheme, where the probability of observing each pair of targets is established by repeating the selection algorithm. Each pair of targets is then assigned the inverse of this probability as a weight. As a Fourier counterpart of the pairwise-inverse-probability weighting scheme has not been established yet, I use the nearest neighbour weighting scheme as described by Reid *et al.* [25] in my analyses in Chapter 5.

Redshift Failures ($w_{\rm noz}$) Even if a fibre is assigned to a target, there is a chance of the spectroscopic pipeline failing to measure its redshift robustly. This happens for 1.8 per cent of all CMASS targets and 0.5 per cent of all LOWZ targets [25]. Redshift failures do not occur randomly, but there is some dependence, e.g., on the $i_{\rm fib2}$ magnitude and the distance of the fibre from the plate centre. For instance, Fig. 2.7 shows that redshift measurements of fainter galaxies are more likely to fail. Fig. 2.8 shows that redshift failures also occur more often towards the edges of the spectroscopic tiles. Similar to the effect of fibre collisions, redshift failures are mitigated by up-weighting the nearest neighbour of a galaxy where obtaining a redshift failed [25]. The main difference compared to $w_{\rm cp}$ is that the nearest neighbour chosen for which the value of $w_{\rm noz}$ is increased is restricted to have been observed using the same plate on the same date.

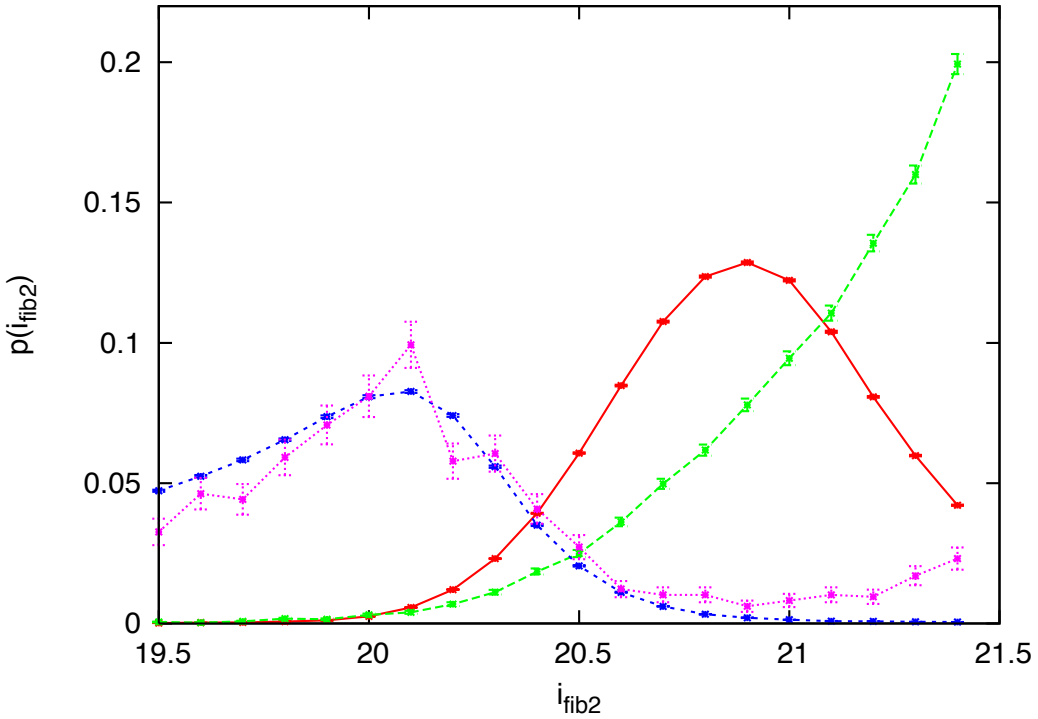

Figure 2.7: Normalised $i_{\rm fib2}$ distributions of redshift failures (green) and redshift successes (red) for the CMASS sample. Redshift failures constitute 1.8 per cent of the CMASS targets observed by SDSS-III BOSS. These are contrasted against normalised distributions for the LOWZ sample of redshift failures (pink, dotted) and redshift successes (blue, dashed). Error bars were calculated assuming Poisson statistics. Figure taken from [25].

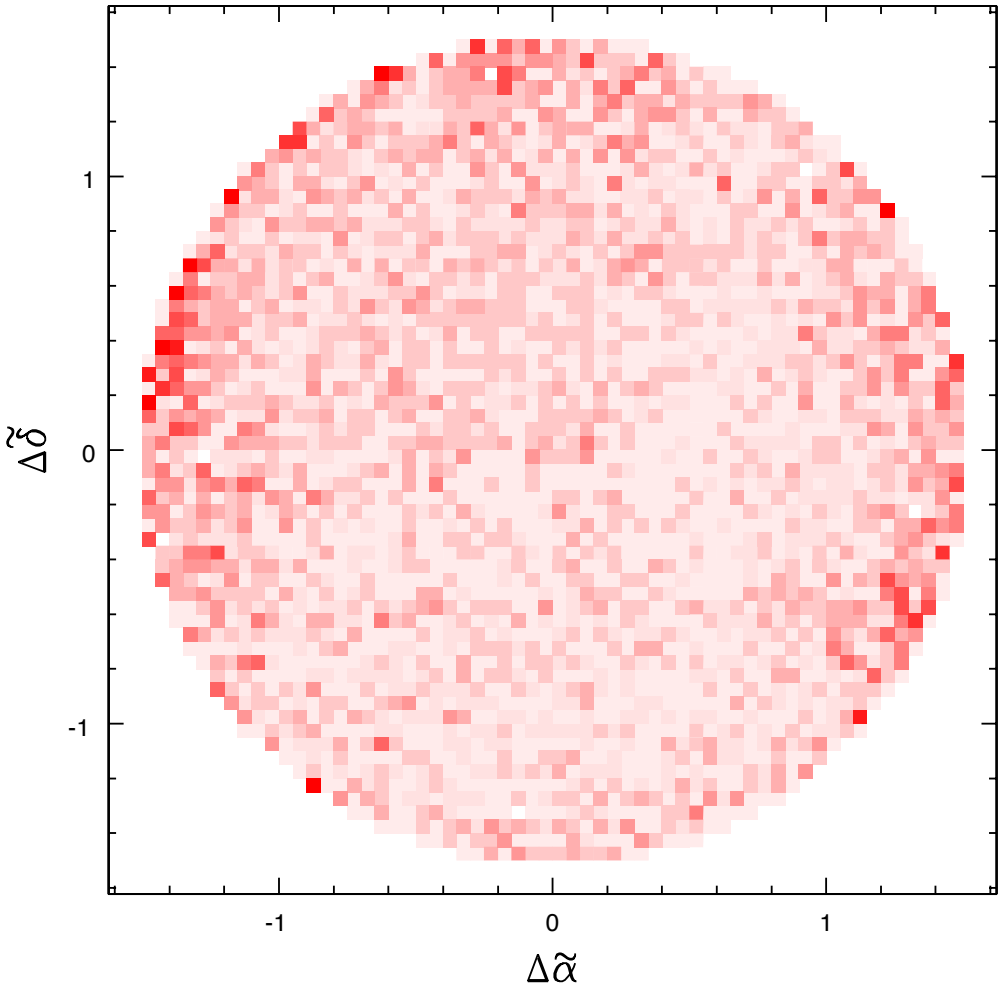

Figure 2.8: The percentage of failed CMASS redshifts as a function of the position on the tile, averaged over 817 DR9 tiles. The lightest regions are 0 per cent and the darkest regions are 12 per cent. $\Delta\tilde{\alpha}$ is the distance along the right ascension direction and $\Delta\tilde{\delta}$ is the distance along the declination direction (both given in degrees and transformed so that the true angular separations are represented). Figure taken from [26].

Stellar Density ($w_{\rm star}$) Ross *et al.* found in their BOSS DR9 $f_{\rm NL}$ analysis that regions that contain many foreground stars have as much as 20 per cent less galaxies than expected using a random catalogue to quantify the mask (cf. Sec. 2.4) [6]. While it is not understood in detail how the presence of stars affects the observed distribution of galaxies, we know a number of effects that can cause stars to contaminate a galaxy catalogue:

- **Obscuration**, i.e. galaxies that are dimmer than a bright foreground star cannot be observed if their angular positions are too close to that star.

- A second effect arises from the target selection algorithm. The selection is based on the targets' magnitudes in previous photometric observations, which might appear brighter in brighter regions, exceeding the magnitude threshold of the selection criterion. The objects in brighter regions are therefore less likely to have a spectrograph pointed at them. I refer to this effect as **selection bias**.

- Another effect is **confusion**, i.e. a star is mistaken as a galaxy. However, this effect is negligible for spectroscopic surveys, as the spectra of stars and galaxies are easily distinguishable.

This is the most important systematic effect in analyses of the largest scale clustering needed for $f_{\rm NL}$ measurements. The effect of the stellar density and its associated weight $w_{\rm star}$ is discussed in more detail in Sec. 5.2.

Seeing ($w_{\rm see}$) The light travelling to ground based telescopes has to travel through the Earth's atmosphere. Due to turbulence in the atmosphere, its refractive index changes on short time scales. This blurs the image of an astronomical object and the flux of the object is spread out. This causes an increased magnitude error, and hence, makes it problematic to distinguish between galaxies and stars in the target selection process [131], because the star-galaxy separation cut relies entirely on magnitudes (cf. Eq. 2.19). This can cause spurious fluctuations in the observed density field of galaxies [131]. The seeing can be quantified by measuring the apparent diameter of a point source. The effect of seeing in BOSS is accounted for by introducing the weight $w_{\rm see}$. This will also be discussed in more detail in Sec. 5.4.

Total Angular Systematic Weight ($w_{\rm sys}$) The data files from the 12th data release (DR12) of BOSS combine the last two weights into the weight $w_{\rm sys} \equiv w_{\rm star} w_{\rm see}$.

To test the impact of each of these weights, I have estimated the power spectrum from the BOSS DR12 CMASS North Galactic Cap (NGC) sample "switching off" one weight at a time (i.e. setting its value equal to one for all galaxies). The resulting power spectra are plotted in Fig. 2.9. Similar to [6], it can be found that $w_{\rm sys}$ has the greatest impact on the power spectrum measurements at large scales. By comparing Fig. 2.9 with Fig. 1.9 it becomes apparent that the angular systematics could even mimic a false $f_{\rm NL}$ signal. As a consequence, Chapters 4 and 5 are dedicated to this important issue. But first, I continue with introducing the successor of BOSS, which is the extended BOSS and other ongoing, and future surveys.

2.3 Ongoing and Future Surveys

2.3.1 Extended Baryon Oscillation Spectroscopic Survey

The extended BOSS (eBOSS) is part of the fourth stage of SDSS (SDSS-IV) and continues the success of BOSS, extending to higher redshifts that correspond to eighty percent of cosmic history. It will observe approximately 375,000 LRGs (in the redshift range $0.6 < z < 0.8$), 260,000 ELGs ($0.6 < z < 1$) and 740,000 quasi-stellar objects (quasars or QSOs, $0.9 < z < 3.5$) over a range of distances that have not been extensively explored by three dimensional maps of large-scale structure so far. This range includes the transition from deceleration to acceleration in the current standard ΛCDM model of cosmology (cf. Sec. 1.1.3). As the number density of QSOs above $z > 2.2$ is too low to trace the large scale structure of the Universe (cf. Fig. 2.10), only those with $z < 2.2$ are used for clustering statistics, whereas QSOs with $z > 2.1$ are used as backlights for Lyman α (Lyα) absorption [27], which probes the density of neutral gas along the line of sight. The number of QSOs useful for clustering analyses therefore amounts to about $\sim$500,000 objects. eBOSS observations started in July 2014, and the first data set, labelled DR13 in the tradition

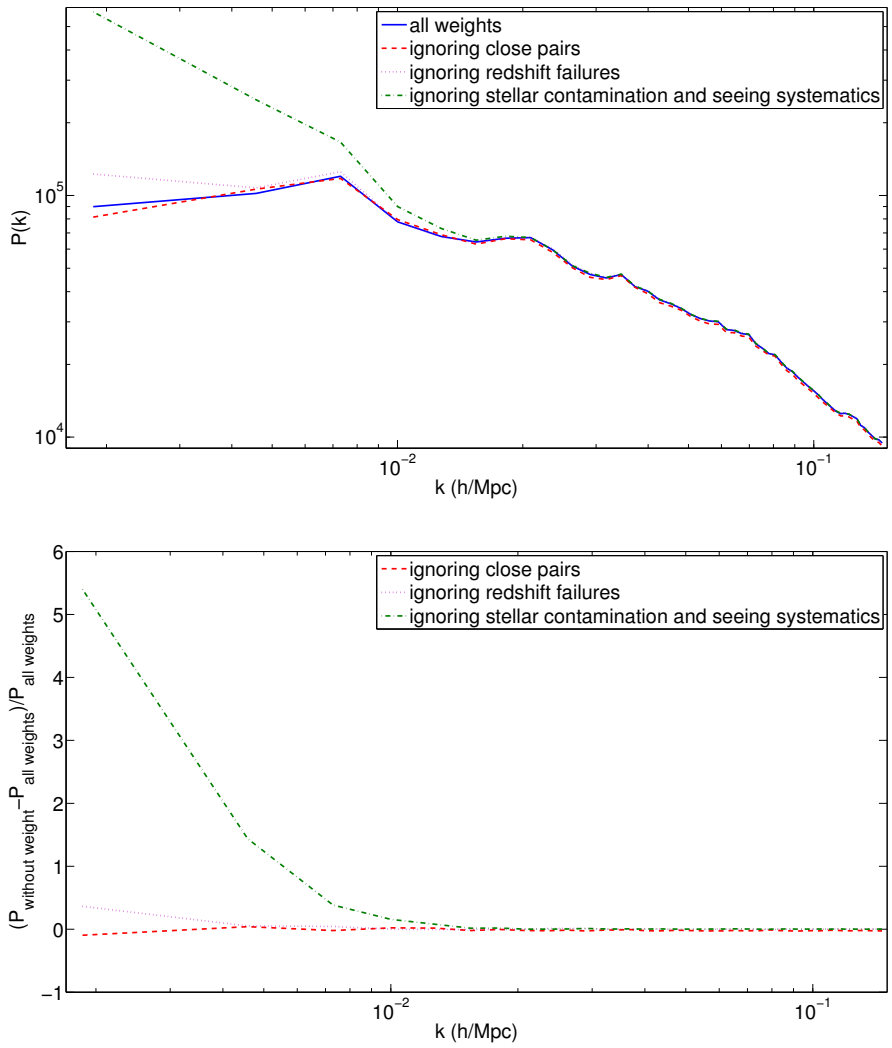

Figure 2.9: The BOSS DR12 CMASS NGC power spectrum using all weights (solid blue line), setting $w_{\rm cp} = 1$ for all galaxies (dashed red line), setting $w_{\rm zf} = 1$ (dotted purple line), and setting $w_{\rm sys} = 1$ (dot-dashed green line). The lower panel shows the fraction of the power spectra omitting one of the weights and the power obtained by using all weights.

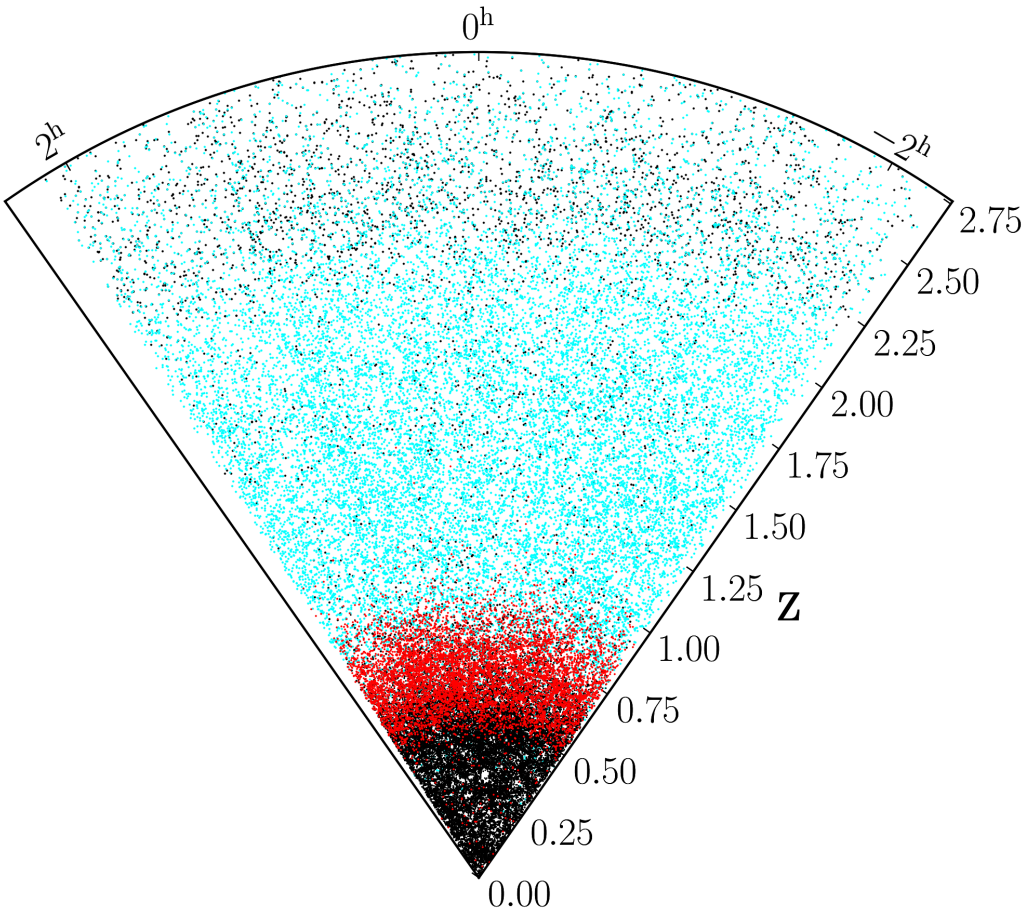

Figure 2.10: Slice along right ascension through the eBOSS redshift sample, 5° wide in declination and centred at $\delta = +22°$. Black points indicate previously known redshifts from SDSS-I through SDSS-III. Cyan points show eBOSS quasars and red points represent eBOSS LRGs. Figure taken from [27].

of SDSS, is already publicly available, also from http://www.sdss.org/data/. It contains spectra of 147,000 QSOs that have been used for the first BAO measurement between redshifts 0.8 and 2.2 that is in good agreement with ΛCDM [11]. After 6 years of taking data, it will cover the area plotted in Fig. 2.12 and provide with 6.6 Gpc^3 the largest volume that so far has been spectroscopically sampled. Such a large volume brings the predicted 68 % confidence level (CL) error on f_{NL} down to $\sigma(f_{\mathrm{NL}}) = 15.74$, only using the QSOs [132]. As eBOSS observes emission line galaxies (ELGs), LRGs and QSOs in overlapping volumes, tracers that have different biasing properties, one can apply the multi tracer technique (cf. Sec. 1.5.2) to beat cosmic variance and further reduce the error on f_{NL} [103, 133, 134]. Zhao *et al.* [132] predict that the error will decrease to $\sigma(f_{\mathrm{NL}}) = 10.5$. A joint analysis of the eBOSS power and bi-spectrum could provide f_{NL}-constraints of $\sigma(f_{\mathrm{NL}}) \sim 1$ (cf. Tab. 2.1, [10]) that are more than competitive with Planck, and are particularly interesting as they can disfavour spectator field models (cf. Sec. 1.3.4) and rule out most alternatives to inflation.

2.3.2 Hobby-Eberly Telescope Dark Energy Experiment

The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX, [135]) at the McDonald Observatory in Texas is observing about a million Lyα emitters within a similarly large volume as BOSS, but at considerably higher redshifts ($1.9 < z < 3.5$). Data from HETDEX will become available within the next two years. HETDEX is the first "blind" spectroscopic survey, i.e. it does not pre-select its target objects. Instead it scans the survey volume and selects objects according to their emission lines. The density of tracers is therefore expected to be high compared to eBOSS. On the other hand, not using a target sample raises new questions for the systematic contaminant mitigation, e.g. the treatment of close pairs. Nonetheless, due to the high density of tracers and high redshifts, it will be an interesting probe for f_{NL}-studies with $\sigma(f_{\mathrm{NL}}) \sim 6$.

2.3.3 Dark Energy Spectroscopic Instrument

The Dark Energy Spectroscopic Instrument (DESI, [136]) will operate between 2018 and 2022 at the Mayall 4-metre Telescope on top of Kitt Peak in Arizona. It will observe spectra of 18 million ELGs in the ($0.1 < z < 1.8$) redshift range, 4 million LRGs between $0.1 < z < 1.1$ and 3 million QSOs at redshifts $0.1 < z < 1.9$. The expected number densities are unprecedently large, especially at low redshifts. This will allow a sub two per mille measurement of the BAO (cf. Fig. 2.11). The linear bias parameters of DESI's tracers are assumed to be $b_0^{(\mathrm{ELG})} = 0.84/D_1(z)$, $b_0^{(\mathrm{LRG})} = 1.7/D_1(z)$ and $b_0^{(\mathrm{QSO})} = 1.2/D_1(z)$ [9]. Thus, there are three tracers whose bias properties are different enough to apply the multi tracer technique to measure the scale dependence of the bias and hence constrain f_{NL}. DESI will also map the Lyman-alpha forest up to redshift $z < 3.5$. A high throughput spectrograph will be fitted onto the telescope, allowing to take thousands of spectra simultaneously. Therefore, DESI will provide a dense, deep and wide survey, covering with at least 14,000 deg^2 a much larger fraction of the sky than eBOSS and HETDEX. This will provide also long angular modes for f_{NL} measurements that allow constraints of $\sigma(f_{\mathrm{NL}}) = 4.37$ from the power spectrum alone and $\sigma(f_{\mathrm{NL}}) \sim 0.2$ including the bi-spectrum (cf. Tab. 2.1, [10]). However, this bispectrum forecast is too optimistic, as the authors of [10] did not take the covariance between different triangles into account. Yet, even though this might degrade the signal by a factor of about 5, as suggested in the same article, DESI has the potential for ground-breaking f_{NL} results.

2.3.4 Euclid

Euclid [137] will take large scale structure observations to outer space. This does not only allow for observations across the whole sky, but also avoids atmospheric contamination of the data. A 1.2-metre telescope will be mounted on a satellite under development by the European Space Agency (ESA). Its launch is planned for 2020. It is expected to provide spectra of 35 million galaxies in optical and near infrared wavelengths, covering the redshift range $0.7 < z < 1.8$ and 15,000 deg^2. Moreover, Euclid will observe about 50 billion objects photometrically. This will provide the most accurate BAO measurement in the

foreseeable future (cf. Fig. 2.11). $f_{\rm NL}$-constraints from Euclid alone are expected to be $\sigma(f_{\rm NL}) = 4.71$ (power spectrum) and $\sigma(f_{\rm NL}) \sim 0.3$ (power and bi-spectra combined, cf. Tab. 2.1, [10]). Even though Euclid will observe more objects than DESI on the full sky, its $f_{\rm NL}$ predictions are less stringent because it observes only a single tracer. However, it can improve the knowledge we will have on $f_{\rm NL}$ at the time of when the data becomes available, if one combines Euclid with other surveys. For instance, when combined with Integrated Sachs Wolfe measurements and Planck data, the Euclid Red Book [137] states $\sigma(f_{\rm NL}) = 2$ as a scientific objective. Fedeli *et al* [138] made a forecast of $\sigma(f_{\rm NL}) = 0.87$ from the scale dependent bias alone, combining the galaxy-galaxy and cluster-cluster power spectra with the galaxy-cluster cross power spectrum. However, their analysis is based on the settings and specifications described in the Euclid Assessment Study Report [139], rather than the Euclid Red Book [137]. Giannantonio *et al* [140] have shown that $f_{\rm NL}$ constraints from photometric Euclid data degrade when updating the specifications from the Euclid Assessment Study Report to the Red Book, whereas those from spectroscopic data improve. As the clusters are measured photometrically, the forecasts of Fideli *et al* are likely to be too low. The Giannantonio *et al* $f_{\rm NL}$ constraints are with $\sigma(f_{\rm NL}) = 4.1$ a bit better than the ones from Tellarini *et al* [10] mentioned above.

2.3.5 Dark Energy Survey

The Dark Energy Survey (DES, [28]) is a photometric near-infrared survey in five optical filters undertaken since 2013 using the 4-meter Victor M. Blanco Telescope at the Cerro Tololo Inter-American Observatory in Chile. It covers the 5000 square degree area in the Southern sky plotted in Fig. 2.12, observing 300 million galaxies with shapes, photometric redshifts and angular positions, as well as 100 thousand galaxy clusters and 1000 type Ia supernovae. Combining weak lensing and galaxy clustering with priors from the Planck temperature power spectrum, Giannantonio *et al* [140] predict $f_{\rm NL}$ measurements with an accuracy of $\sigma(f_{\rm NL}) = 8$. Fonseca *et al* [141] predict $\sigma(f_{\rm NL}) = 11.9$ from galaxy clustering only, using 8 redshift bins between $0 < z < 2$. However, the DES Year 1 (DES Y1, [142]) galaxy catalogue only contains galaxies with redshifts $0.15 < z < 0.9$. Furthermore, both Giannantonio *et al.* [102] and

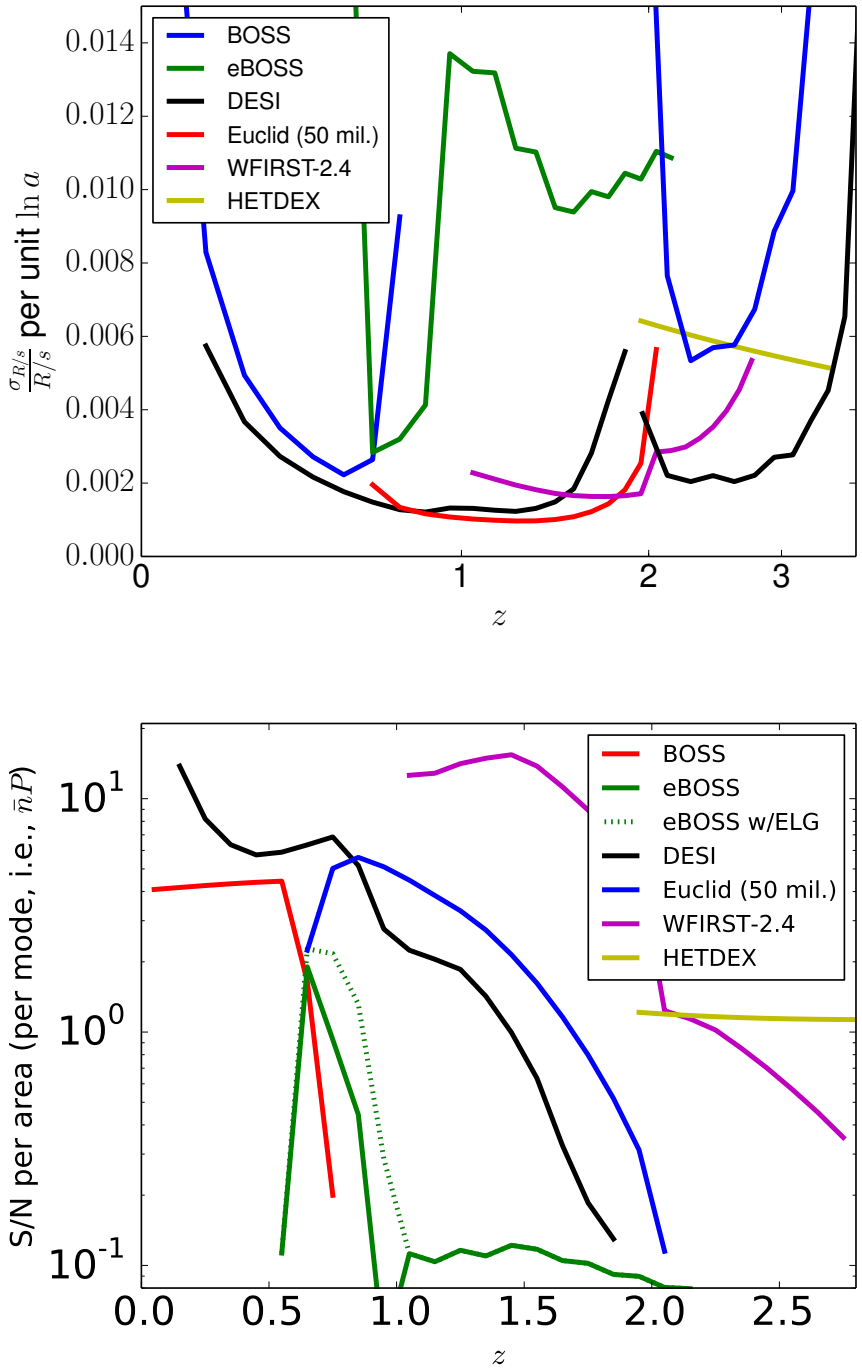

Figure 2.11: *Top:* Fractional error density on the dilation factor as a function of redshift, per unit $\ln(a)$, thus, the effect of the width Δz is removed in this plot. *Bottom:* Comparison of the galaxy number density in terms of $\bar{n}P(k = 0.14h^{-1}\text{Mpc}, \mu = 0.6)$. Figures taken from [9].

Table 2.1: Forecasts for $\sigma_{f_{\rm NL}}$ from different redshift surveys using measurements of the power spectrum (P) or a combination of the power and bispectrum (B). A "+b_0" in the table header indicates a scenario where the linear bias is marginalised over during the power or bispectrum measurements, while those without indicate that the linear bias is fixed. The forecasts are taken from Font-Ribera *et al.* [9] and Tellarini *et al.* [10]. Where there are results for more than one fiducial bias reported by Tellarini *et al.*, only the most conservative forecast is included in this table.

Survey	P+b_0 [9]	P+b_0 [10]	P [10]	B+b_0 [10]	B [10]
BOSS	23	21.30	13.28	2.47	1.48
eBOSS	-	14.21	11.12	2.02	1.29
Euclid	-	6.00	4.71	0.71	0.35
DESI	3.8	5.43	4.37	0.48	0.37
BOSS+eBOSS	11	-	-	-	-
BOSS+Euclid	6.7	5.64	4.44	0.59	0.35

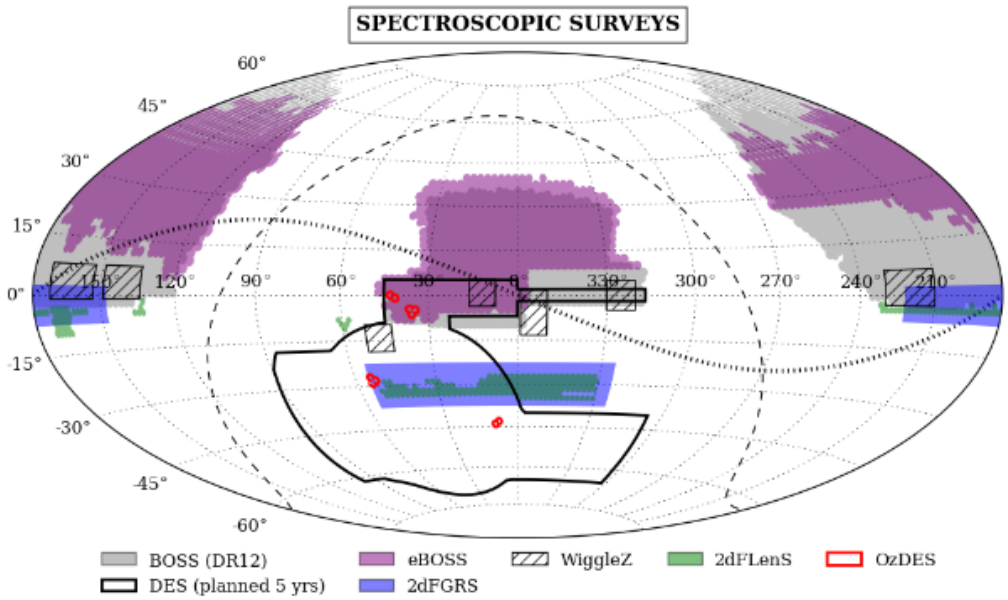

Figure 2.12: Footprints of DES and some selected completed or ongoing spectroscopic surveys (as of December 2015). This is a Hammer projection in equatorial coordinates, with the dashed and dotted lines indicating the Galactic plane and the ecliptic plane, respectively. Figure taken from [28].

Fonseca *et al.* assume that the number of galaxies is constant in all redshift bins, whereas the actual DES Y1 redshift distribution varies significantly with redshift. Thus, it is unlikely that DES will provide more insight into the primordial density distribution than eBOSS, but we see in the following section, that it is a good candidate to apply the multi tracer technique when combined with the Meer Karoo Array Telescope (MeerKAT) Large Area Synoptic Survey (MeerKLASS), as its fiducial galaxy bias of $b_0 = 1 + 0.84z$ is substantially different from the bias of MeerKLASS observations.

2.3.6 Radio Surveys

Like optical surveys, one can distinguish two categories of radio surveys: continuum radio surveys and spectral-line radio surveys. Continuum radio surveys can be regarded as an equivalent to photometric surveys. They scan the sky and just record the radio emission coming from every direction the telescope visits. They cover large cosmic volumes at higher redshifts than modern optical

surveys. However, they do not provide redshift information about the objects they observe, thus one can only measure 2-dimensional clustering statistics. As a consequence, they alone do not provide tight constraints on $f_{\rm NL}$. Yet, the first constraints on $f_{\rm NL}$ from radio data alone have been obtained using data of the National Radio Astronomy Observatory (NRAO) Very Large Array (VLA) Sky Survey (**NVSS**). The VLA in New Mexico surveyed 82 per cent between 1993 and 1996. It found with 3σ confidence that $f_{\rm NL} > 0$ and a best fitting value of $f_{\rm NL} = 62 \pm 27$ [143], but a reanalysis showed that $f_{\rm NL}$ constraints from radio data are prone to systematics such as sidelobe effects of bright sources, galactic foreground and the cosmic radio dipole. After taking these effects into account, the NVSS data is consistent with Gaussianity [144].

A lot more interesting for non-Gaussianity studies are **spectral-line radio catalogues**, which can be regarded as a radio equivalent to redshift surveys. They map the intensity of radio emission from well defined emission lines, such as the 21 centimetre line that is due to atomic transitions of electrons between the two hyperfine levels of the hydrogen 1s ground state. Comparing the frequency of the received signal with the one in our rest-frame, one can accurately infer the redshift of the emitting source. Due to the very different biasing properties of these radio sources compared to optical sources, the combination of an optical and a radio survey is a powerful tool to constrain $f_{\rm NL}$ using the multi-tracer technique (cf. Sec. 1.5.2). Especially useful for non-Gaussianity measurements is **intensity mapping**, where, instead of resolving individual galaxies, the integrated emission of 21 centimetre emission lines is used to reconstruct the underlying power spectrum of fluctuations in the matter distribution. The angular resolution of this technique is about a factor of 1000 worse than that of galaxy surveys. However, it is a good method to obtain tight $f_{\rm NL}$ constraints, because it allows a quick reconstruction of the ultra-large scale power spectrum out to redshift $z \sim 4$ [145].

The **MeerKAT** telescope will be composed of 64 dishes with 13.5 metre diameter. It will become fully operational in April 2018. A cosmological survey **MeerKLASS** has been proposed. The proposal is not yet publicly available, but preliminary specifications can be found in e.g. [141, 146]. It will observe in the redshift range of $0 < z < 1.45$ with a sky coverage of at least 1000 square degrees, $0 < z < 0.58$ in the L band ($900 < \nu < 1670$ MHz) and $0.4 < z < 1.45$

at ultra high frequencies (UHF, $580 < \nu < 1015$ MHz). It will observe for 3 to 15 weeks. The $f_{\rm NL}$ constraints from MeerKAT are forecast as $\sigma(f_{\rm NL}) = 43.8$ and $\sigma(f_{\rm NL}) = 56.5$ for the ultra high frequency (UHF) and low frequency bands, respectively. However, with a linear bias of up to $b_0 = 3$ [147], it can provide interesting constraints on $f_{\rm NL}$ when combined with DES. Fonseca *et al* predict for the combination of the L-band with DES data $f_{\rm NL}$ constraints of $\sigma(f_{\rm NL}) = 3.6$, and for the UHF-band they expect even $\sigma(f_{\rm NL}) = 2.3$.

In 2024, MeerKAT is planned to be absorbed into the Square Kilometre Array (**SKA**). SKA will allow to conduct a radio continuum survey covering 30,000 deg^2. There are two stages of SKA planned. Phase 1 will start in 2024, when MeerKAT is absorbed into SKA, and provides 10 per cent of the total collecting area. As the name suggests, the second stage will have a total collecting area of one square kilometre, which considerably improves the survey power. Its completion is estimated for 2030. Yamauchi, Takahashi and Oguri [133] made forecasts for $f_{\rm NL}$ measurements from a radio continuum survey carried out with either stages of SKA. Their constraints for stage 1 is $\sigma(f_{\rm NL}) = 1.64$ and $\sigma(f_{\rm NL}) = 0.66$ for stage 2. SKA will also provide a spectral line catalogue comprising tens of hundreds of millions of galaxies out to redshift $z \sim 2$. Camera, Santos and Maartens [148] predict that such a survey will provide $f_{\rm NL}$ constraints of $\sigma(f_{\rm NL}) = 1.54$ for the completed survey, thus stage 2. This is at odds with the forecast by Yamauchi, Takahashi and Oguri [133], as it is based on a radio continuum survey, which is expected to be less accurate. But then Camera, Santos and Maartens use a different definition of $f_{\rm NL}$ where the redshift dependence of the critical density $\delta_{\rm c}$ is turned into a redshift dependence of $f_{\rm NL}$, thus the Camera, Santos and Maartens forecast is actually $\sigma(f_{\rm NL}) = 1.18$, and furthermore, they take several relativistic effects into account, such that the forecasts by Yamauchi, Takahashi and Oguri [133] can be considered over-optimistic.

On the other hand, Yamauchi, Takahashi and Oguri [133] also considered combining SKA with Euclid using the multi-tracer technique. They expect an improved $f_{\rm NL}$ constraint of $\sigma(f_{\rm NL}) = 0.61$ for stage 1 combined with Euclid. Even though their initial forecast is too optimistic, it still shows that it is a great improvement that can potentially bring forth more interesting insights into the physics of the early Universe.

2.4 Galaxy Clustering Analysis Techniques

Having introduced a selection of cosmological surveys, I now briefly review how the galaxy density field and its corresponding clustering statistics are obtained from the positions and redshifts observed by these surveys.

2.4.1 Random Catalogues

When the over-density $\delta = \frac{\rho}{\bar{\rho}} - 1$ (cf. Eq. 1.61) is measured from a galaxy survey, understanding where galaxies could have been observed is as important as actually observing them. The completeness and purity of observations varies along the angular components of the survey mask. The radial component depends on the galaxy selection criteria and can also change with angular position. This is commonly addressed by generating random catalogues, i.e. a catalogue of Poisson sampled positions from the expected background density $\bar{\rho}$ (that is without clustering) under the same spatial selection function as the actual galaxy catalogue. The generation of such random catalogues is done in several steps. The first step in the creation of a random catalogue is to pick random angular positions distributed according to the completeness of the survey. For the example of BOSS, the completeness in a given sector i is estimated as

$$C_{\mathrm{BOSS},i} \equiv \frac{N_{\mathrm{star},i} + N_{\mathrm{gal},i} + N_{\mathrm{fail},i} + N_{\mathrm{cp},i}}{N_{\mathrm{star},i} + N_{\mathrm{gal},i} + N_{\mathrm{fail},i} + N_{\mathrm{cp},i} + N_{\mathrm{missed},i}}, \tag{2.21}$$

where $N_{\mathrm{star},i}$, $N_{\mathrm{gal},i}$, $N_{\mathrm{fail},i}$, $N_{\mathrm{cp},i}$, $N_{\mathrm{missed},i}$ are the numbers of objects spectroscopically confirmed to be stars, objects that were spectroscopically confirmed a galaxies, objects whose classification failed, close-pair objects of which no spectra could be taken due to fibre collision (cf. Sec. 2.2) but with at least one more object in the same target class, and all other objects without spectra [25]. In this context, a sector is defined by the area covered by the same (intersection of) spectroscopic tiles. The completeness of the BOSS CMASS sample is plotted in Fig. 2.13. As said earlier, we use the distribution of Eq. (2.21) to pick random angular positions. However, another ingredient to the overall mask are so called veto masks. Veto masks account for regions where spectra cannot be taken for various reasons. For instance, the area around the central bolts of the tiles are vetoed, because the plates are secured at that point to

the focal plane of the telescope and consequently no fibres can be placed there. Veto masks are also placed around targets outside the galaxy catalogue that have higher targeting priority, such as Lyman-α quasars. Areas around bright stars are also vetoed, whereas the areas around stars with magnitudes larger than 11.5 are kept in both the galaxy catalogue and the random catalogue and are dealt with in Sec. 5.2. For these reasons, a fraction of the random positions mimicking the completeness of the survey is then rejected if they fall into a veto mask.

Finally, each object that is still in the random catalogue after vetoing is assigned a random redshift that follows the distribution of the (weighted) galaxies, similar to Fig. 2.5. The random catalogues provided by the BOSS collaboration contain 50 times more random galaxies than there are in the galaxy catalogue. For the angular weight fitting (cf. Sec. 2.2, 5.2 and 5.4), Ross *et al.* [113] used random catalogues containing only angular positions of 10 times the number of galaxies. We use the full "50 times" random catalogues in our analyses in Chapter 5.

These random catalogues are a crucial part in all clustering analyses techniques presented in this section.

2.4.2 Power Spectrum Estimation

With a galaxy and a random catalogue at hand, one can make measurements of the galaxy over-density field and hence of its power spectrum. A classic power spectrum estimation method was suggested by Feldman, Kaiser & Peacock [109] and is introduced in this subsection as the **FKP estimator**. We shall see in Sec. 4.2 that this is a simplification of the more optimal, but computationally more expensive, Quadratic Maximum Likelihood (QML) estimator.

The first step to measure the power spectrum consists of dividing the survey volume into a cubic grid. The galaxy and random catalogues are folded into this grid and we count the number of galaxies $n_{\mathrm{g}}(\mathbf{x})$ and random catalogue objects $n_{\mathrm{r}}(\mathbf{x})$ in each grid cell around position $\mathbf{x}$. The simplest grid assignment algorithm, the Nearest-Grid-Point algorithm, counts each galaxy solely at its nearest grid point. This can introduce large truncation or aliasing errors, and therefore the Cloud-In-Cell algorithm is more often applied. A multi-linear

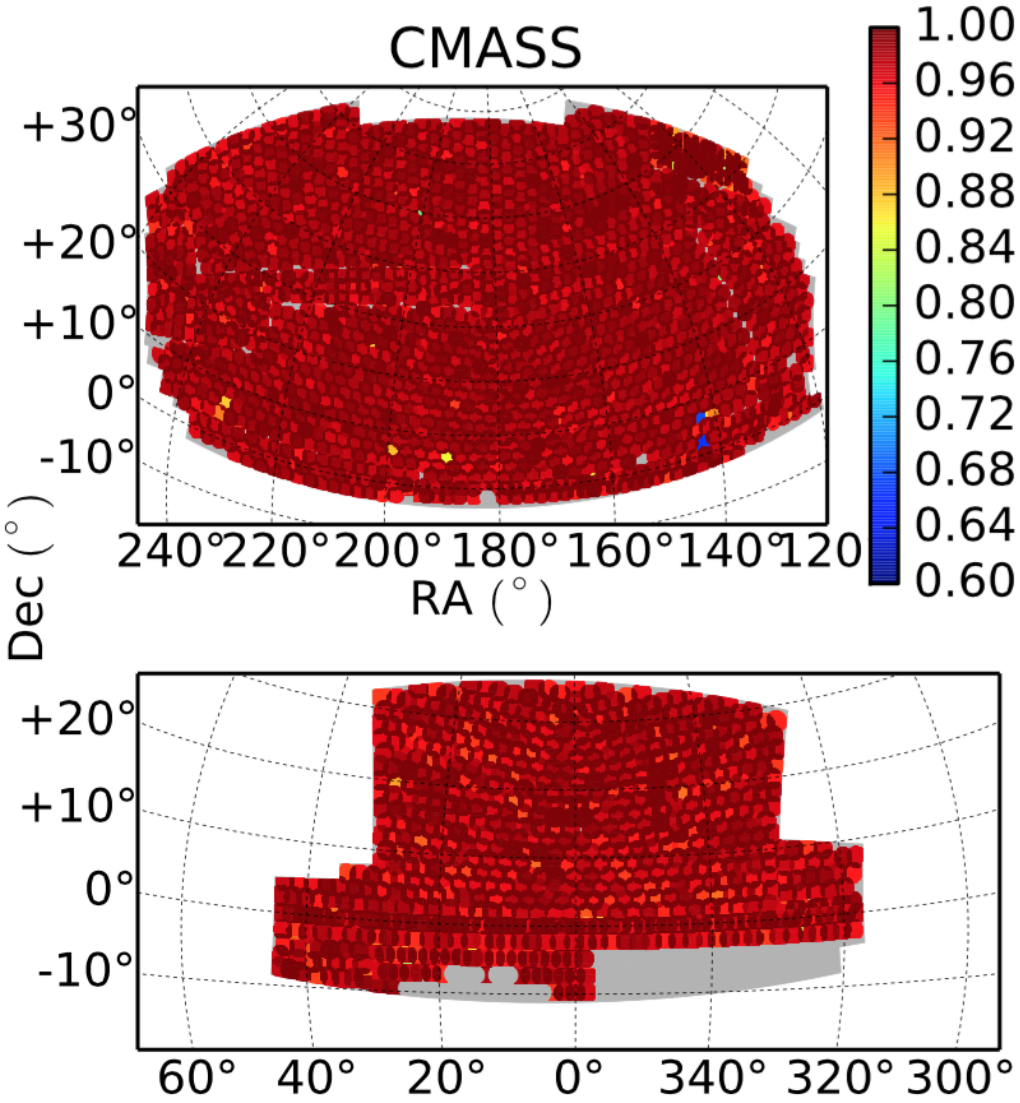

Figure 2.13: Completeness maps for the BOSS CMASS DR12 sample in the north and south Galactic caps. The mean completeness is 98.8% for the CMASS sample. Figure taken from [25].

interpolation determines fractions of each galaxy to be added to eight grid points in its neighbourhood. This effectively treats each galaxy as a cubical cloud with uniform density. For example, the particle in Fig. 2.14 is entirely assigned to grid point 1 using the Nearest-Grid-Point algorithm. Using Cloud-In-Cell, the particle is assigned mostly to grid point 1, but also by a smaller fraction to grid points 2 and 3, and with an even smaller weight to grid point 4, relative to the area covered by the particle cloud. To account for the systematic effects discussed in Sec. 2.2, each galaxy is also weighted according to Eq. (2.20) while they are being assigned to the grid.

As the noise varies as the number density varies across the sky, Feldman, Kaiser & Peacock [109] introduced a weight

$$w_{\mathrm{FKP}}(\mathbf{x}) \equiv \frac{1}{1 + \langle n_{\mathrm{g}} \rangle(\mathbf{x}) P_{\mathrm{est}}(k)} \tag{2.22}$$

into their estimator of the over-density field

$$D(\mathbf{x}) = w_{\mathrm{FKP}}(\mathbf{x}) \frac{n_{\mathrm{g}}(\mathbf{x}) - \alpha n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}}, \tag{2.23}$$

where

$$I_2 = \int \mathrm{d}^3\mathbf{x} \langle n_{\mathrm{g}} \rangle^2(\mathbf{x}) w_{\mathrm{FKP}}^2(\mathbf{x}) \text{ [109]} \tag{2.24}$$

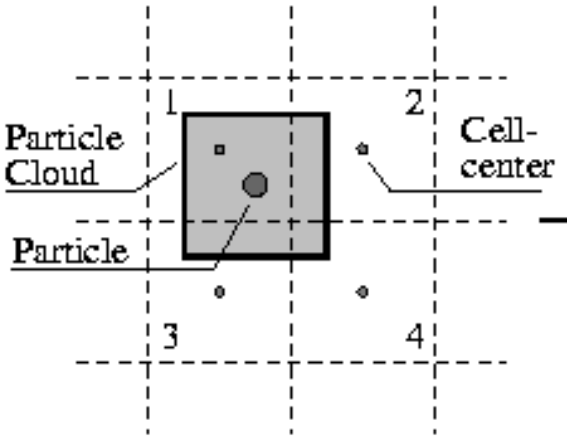

Figure 2.14: 2-dimensional visualisation of the Cloud-In-Cell mass assignment algorithm. Figure taken from [29].

acts both as a normalisation for the FKP weights and ensures that $D(\mathbf{x})$ is a density rather than just a number count. The FKP weights optimise the power spectrum measurement if the fluctuations are Gaussian with a preliminary power spectrum estimate $P_{\mathrm{est}}(k)$. In BOSS DR12 analyses, $P_{\mathrm{est}}(k) = 10000 \frac{\mathrm{Mpc}^3}{h^3}$ is assumed at all scales. α is a parameter that corrects for the higher number of random objects compared to galaxies and ensures that the density is on average

$$\sum_{\mathbf{x}} D(\mathbf{x}) = 0. \tag{2.25}$$

The next step in the power spectrum estimation is to perform a Fast Fourier Transform (FFT) of $D(\mathbf{x})$, which is written as $D(\mathbf{k})$. Taking the ensemble

average of its absolute value squared, we find that, due to the discrete sampling of the density field using galaxies, the power spectrum of each mode is overestimated by a constant shot noise term

$$\left\langle |D(\mathbf{k})|^2 \right\rangle = P(\mathbf{k}) + P_{\text{shot}} \text{ [109]}. \tag{2.26}$$

This shot noise term can be estimated by summing over the FKP weights of all objects r in the random catalogue:

$$P_{\text{shot}} = \frac{1+\alpha}{I_2} \sum_r w_{\text{FKP}}^2(\mathbf{x}_r). \tag{2.27}$$

The final FKP estimator of the power spectrum is then the average value of $|D(\mathbf{k})|^2$ in a k-bin, with the shot noise term subtracted:

$$\widehat{P}_{\text{FKP}}(\mathbb{k}) = \frac{1}{N_{\mathbb{k}}} \sum_{\mathbf{k}_i \in \mathbb{k}}^{N_{\mathbb{k}}} |D(\mathbf{k}_i)|^2 - P_{\text{shot}} \text{ [109]}, \tag{2.28}$$

where a k-bin is denoted as $\mathbb{k}$ and $N_{\mathbb{k}}$ is the number of modes within the k-bin in question.

As motivated in Sec. 2.1, the anisotropic power spectrum provides valuable information about the Universe, especially RSDs. For such applications, ignoring the noise term, one can extend the FKP estimator to an estimator of the power spectrum multipoles

$$\widehat{P}_\ell(\mathbf{k}) = \frac{2\ell+1}{I_2} \int \frac{\mathrm{d}\Omega_{\mathbf{k}}}{4\pi} \int \mathrm{d}\mathbf{x}_1 \int \mathrm{d}\mathbf{x}_2 D(\mathbf{x}_1) D(\mathbf{x}_2) e^{i\mathbf{k}(\mathbf{x}_1 - \mathbf{x}_2)} \mathcal{L}_\ell(\mu) \left(\frac{\mathbf{k}\mathbf{x}_h}{|\mathbf{k}||\mathbf{x}_h|} \right) \tag{2.29}$$

using the same multipole notation as in Eq. (2.6) and writing a solid angle element in Fourier space as $\mathrm{d}\Omega_{\mathbf{k}}$. $\mathbf{x}_h \equiv \frac{x_1+\mathbf{x}_2}{2}$ is the line-of-sight of a pair of galaxies, which can be approximated as the vector $\mathbf{x}_h \approx \mathbf{x}_2$ to one of the galaxies to obtain the separable **Yamamoto estimator**:

$$\begin{aligned} \widehat{P}_0^{\text{山本}}(k) &= I_2^{-1} \int \frac{\mathrm{d}\Omega_{\mathbf{k}}}{4\pi} |A_0(\mathbf{k})|^2, \\ \widehat{P}_2^{\text{山本}}(k) &= \frac{5}{2I_2} \int \frac{\mathrm{d}\Omega_{\mathbf{k}}}{4\pi} A_0(\mathbf{k}) \left[3A_2^*(\mathbf{k}) - A_0^*(\mathbf{k}) \right], \text{ etc.}, \end{aligned} \tag{2.30}$$

where

$$A_n(\mathbf{k}) \equiv \int \mathrm{d}\mathbf{x} \left(\frac{\mathbf{k}\mathbf{x}_h}{|\mathbf{k}||\mathbf{x}_h|} \right)^n D(\mathbf{x}) e^{i\mathbf{k}\mathbf{x}} \tag{2.31}$$

is a multipole generalisation of $D(\mathbf{k})$ [149,150]. However, the $\mathbf{k}$ dependence of the integrand prohibits the direct FFT application. This can be circumvented by the decomposition of Bianchi *et al.* [151], where, e.g.,

$$A_2(\mathbf{k}) = \frac{1}{|\mathbf{k}|^2} \sum_{i\in\{x,y,z\}} \sum_{j\in\{x,y,z\}} \mathbf{k}_i \mathbf{k}_j B^{ij}(\mathbf{k}) \tag{2.32}$$

can be written in terms of the Fast-Fourier-Transformable

$$B^{ij}(\mathbf{k}) \equiv \int \mathrm{d}\mathbf{x} \frac{\mathbf{x}_i \mathbf{x}_j}{\mathbf{x}^2} D(\mathbf{x}) e^{i\mathbf{k}\mathbf{x}}, \tag{2.33}$$

where $\mathbf{x}_i$, $\mathbf{k}_i$, etc. are vector components of $\mathbf{x}$ and $\mathbf{k}$.

2.4.3 Correlation Function Estimation

The correlation function and the power spectrum form a Fourier pair, thus, their information content is the same. However, either of them can be more suitable to solve different problems. When inferring scale dependent quantities such as the f_{NL} dependent bias, measuring the power spectrum is a more convenient choice, but, e.g., anisotropic measurements are easier, as correlation function measurements are based on pair counting, which can trivially be extended by including the angle between the pair orientation and the line-of-sight μ.

To measure the two point correlation function, one can count all pairs of (weighted) galaxies that are separated by a distance in the range $r - \delta r/2 < r < r + \delta r/2$, where r is a bin centre and δr the bin length. The number of pairs falling into the bin around r is written as $DD(r)$. Doing the same for the random catalogue and calling the normalised binned pair count of the random objects $RR(r)$, one can estimate the two point correlation function using the simple **Peebles-Hauser estimator**

$$\widehat{\xi}_{\mathrm{PH}}(r) = \frac{DD(r)}{RR(r)} - 1. \tag{2.34}$$

This estimator is however biased and the **Landy-Szalay estimator**

$$\widehat{\xi}_{\mathrm{LS}}(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} \text{ [152]} \tag{2.35}$$

is usually used instead. $DR(r)$ is the number of data-random pairs, analogous to $DD(r)$ and $RR(r)$. As the random catalogue is uncorrelated with the data,

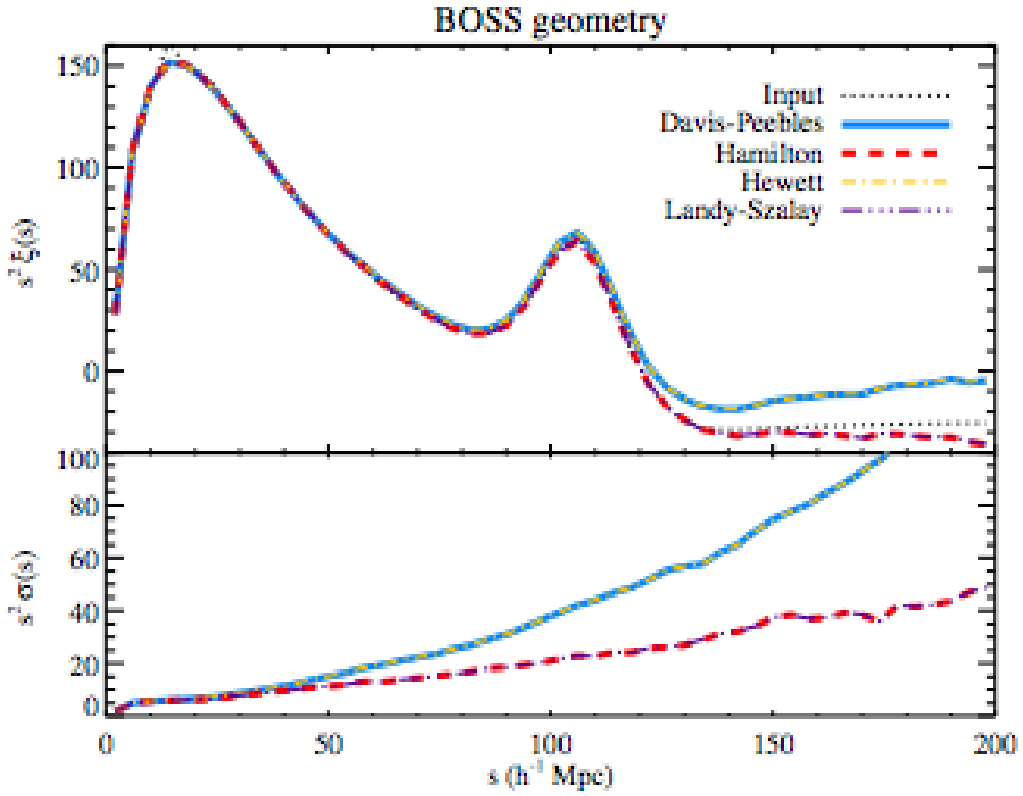

Figure 2.15: Input and reconstructed two-point correlation functions obtained from 120 log-normal mock catalogues using the various estimators available in the literature for the BOSS DR9 survey volume. The bottom panel show the root mean square of each estimator with corresponding colour and linestyle. In each case, the Hamilton and Landy-Szalay lines are exactly superposed as well as the Davis-Peebles and Hewett lines. Figure taken from [30].

the normalised number of data-random pairs is on average the same as the normalised number of random-random pairs: $\langle DR(r)\rangle = \langle RR(r)\rangle$. For a random catalogue with infinitely many objects and for a simple survey geometry, the Landy-Szalay estimator is unbiased and optimal. For a finite number of randoms, the Landy-Szalay estimator is biased, but by a much smaller amount as the Peebles-Hauser estimator, as Vargas-Magaña *et al.* [30] have tested using mock catalogues for BOSS DR9 (cf. Fig. 2.15). They also found that the lesser used **Hamilton estimator**

$$\widehat{\xi}_{\mathrm{H}}(r) = \frac{DD(r)RR(r)}{DR^2(r)} - 1 \tag{2.36}$$

performs as well as the Landy-Szalay estimator. As the bias becomes larger at large scales, one should rather apply mock-based iterative estimation technique (e.g. [30]), or use the power spectrum, as we do in our f_{NL} analyses.

2.4.4 The Window Function

Another important caveat to make before finalising this chapter is that the power spectrum estimators introduced here provide "windowed" quantities, because there are no full sky galaxy surveys available. The estimated density field

$$\widehat{\delta}(\mathbf{x}) = W(\mathbf{x})\delta(\mathbf{x}) \tag{2.37}$$

can therefore be regarded as the product of the true over-density field $\delta(\mathbf{x})$ and a window function $W(\mathbf{x})$, which is zero outside of the survey boundaries. In Fourier space, the observed over-density field

$$\widehat{\delta}(\mathbf{k}) = \int \frac{\mathrm{d}^3\mathbf{q}}{\sqrt{2\pi}^3} W(\mathbf{k}-\mathbf{q})\delta(\mathbf{q}) \tag{2.38}$$

is then actually convolved with a window function. A power spectrum measured from such an over-density field is therefore not directly comparable to a theoretical model prediction thereof. As convolutions are a much easier mathematically procedure than deconvolutions, it is common practice to estimate the window function from the survey mask expressed by a random catalogue and convolve the model power spectrum $P_{\mathrm{mod}}(\mathbf{k})$ when data is fitted to it:

$$P_{\mathrm{win}}(\mathbf{k}) = \int \frac{\mathrm{d}^3\mathbf{q}}{\sqrt{2\pi}^3} P_{\mathrm{mod}}(\mathbf{q})|W(\mathbf{k}-\mathbf{q})|^2. \tag{2.39}$$

As the random catalogue has been generated such that its number density is finite only within the survey volume and proportional to the completeness, the window function can be estimated as the Fourier transform of the number density of the random catalogue:

$$W(\mathbf{k}) = \frac{\alpha}{\sqrt{I_2}} \int \frac{\mathrm{d}^3\mathbf{r}}{\sqrt{2\pi}^3} n_\mathrm{r}(\mathbf{r}) e^{i\mathbf{k}\mathbf{r}}. \tag{2.40}$$

For an anisotropic power spectrum, one can apply the formalism of Wilson *et al.* [153]. The original version of this formalism has been derived using the global plane parallel approximation, which makes it invalid for the wide angle surveys needed for f_NL studies. However, an adoption using the local plane parallel formalism can be found in [154]. In this formalism, the effect of the window function is included into the model power spectrum in three main steps:

1. The multipoles of the model power spectrum are Fourier transformed to provide the correlation function multipole model.

2. The window function monopoles are estimated from the random catalogue as

$$W_\ell^2(r) \propto \sum RR(r,\mu)\mathcal{L}_\ell(\mu) \tag{2.41}$$

and then applied to the correlation function multipole model ξ_ℓ:

$$\begin{aligned}\xi_0^\mathrm{win} &= \xi_0 W_0^2 + \frac{1}{5}\xi_2 W_2^2 + \mathcal{O}(\xi_4) \\ \xi_2^\mathrm{win} &= \xi_0 W_2^2 + \xi_2\left[W_0^2 + \frac{2}{7}W_2^2 + \frac{2}{7}W_4^2\right] + \mathcal{O}(\xi_4).\end{aligned} \tag{2.42}$$

3. The convolved model power spectrum multipoles can than be obtained by a Hankel transformation of the windowed correlation function multipoles:

$$P_\ell^\mathrm{win} = 4\pi(-i)^\ell \int \mathrm{d}r r^2 \xi_\ell^\mathrm{win}(r) j_\ell(rk), \tag{2.43}$$

where $j_\ell(rk)$ is a Bessel function.

Now that we have a method to estimate power spectra using Eq. (2.28) and a window-convolved model power spectrum from Eq. (2.39), we can focus on inferring cosmological parameters by comparing the two in the next chapter.

Chapter 3

Cosmological parameter inference from the galaxy power spectrum

This chapter explains how one can infer information about cosmological parameters given a measurement of the galaxy power spectrum $P_{\rm gg}(k)$. I introduce the basics of the statistical framework needed for cosmological inference in Sec. 3.1 before focussing on the power spectrum likelihood and posterior distributions in the following sections. These sections are based on my article "Cosmological parameter inference from galaxy clustering: The effect of the posterior distribution of the power spectrum" [4] which I have written in collaboration with Will Percival and Lado Samushia. In order to make any inference from the galaxy power spectrum, one has to know its likelihood or posterior distribution, which for simple cases can be calculated analytically. For general cases, one usually assumes the likelihood or posterior to be a multi-variate Gaussian with a covariance matrix. The estimation of the covariance matrix is a critical step in the analysis of data, which is the scope of Sec. 3.2. In Sec. 3.3, the standard assumption of the posterior shape being Gaussian is first tested mode-by-mode on a toy example where we measure the power spectrum itself, and then by estimating the impact of using different posterior distributions on a real survey, i.e. BOSS (cf. Sec. 2.2). As the variance of a Gaussian distribution is not Gaussian distributed itself, the power spectrum of a Gaussian density field is not either. A derivation of the "true" distribution of the power

spectrum, i.e. the distribution of the variance of a Gaussian density field, is also presented in Sec. 3.3. This posterior does not contain the covariance of the power spectrum, which is fine for a perfect measurement. In real applications, we have to take into account that the survey does not cover the whole sky, or that there might be additional noise. It is common practice to include these effects into the power covariance matrix. Therefore, we study alternative posterior shapes for power spectrum estimates inspired by CMB analyses that contain covariance matrices in Sec. 3.4. The inverse cubic normal (ICN) distribution agrees with the true posterior remarkably well and in a Taylor expansion of the two distribution functions (Sec. 3.5), one can see that the ICN distribution approximates the true distribution well up to the 17th order in parameter space. We further test the ICN distribution by postdicting a $f_{\rm NL}$-measurement for a data sample like the data release 9 (DR9) of BOSS and we make predictions of Euclid $f_{\rm NL}$-measurements in Sec. 3.6. The conclusion of [4] is reproduced in Sec. 3.7.

3.1 Introduction to the Statistics of Inference

In pure logic, a hypothesis H can either be true or false, often represented by 1 or 0. Cosmology, however, is an empirical science. Cosmologists can formulate hypotheses about the Universe, these can be tested against observations, but then one cannot say with certainty, whether the hypothesis is true or false. However, one can extend the notion of true or false to a spectrum of degrees of plausibility of the hypothesis, represented with continuous values between 0 and 1. In the Bayesian interpretation of statistics, this degree of plausibility is called the **probability** $\mathscr{P}(H)$ of the hypothesis H. This is an extension to Aristotelian logic, because a hypothesis with $\mathscr{P}(H) = 0$ is certainly false, one with $\mathscr{P}(H) = 1$ is certainly true; the smaller the value of $\mathscr{P}(H)$ the more we believe H is false, and, vice versa, the larger its value, the more we believe H is true. This is equivalent to the first of Cox's postulates, formalised by Jaynes [155]. One can further demand that H is certainly either true or false, i.e.

$$\mathscr{P}(H) + \mathscr{P}(\bar{H}) = 1. \tag{3.1}$$

Finally, if there is more than one way of reasoning, then every possible way must lead to the same conclusion [155]. Considering three propositions A, B and C, this means that we assign the same plausibility to the conjunction of A and B given C no matter whether we first learnt about A given C and then about B given our knowledge about A and C, i.e.

$$\mathscr{P}(AB|C) = \mathscr{P}(A|C)\mathscr{P}(B|AC), \tag{3.2}$$

or whether we first learnt about B given C and then about A, i.e.

$$\mathscr{P}(AB|C) = \mathscr{P}(B|C)\mathscr{P}(A|BC). \tag{3.3}$$

From equating Eq. (3.2) and (3.3) one immediately obtains Bayes' theorem

$$\mathscr{P}(A)\mathscr{P}(B|A) = \mathscr{P}(B)\mathscr{P}(A|B). \tag{3.4}$$

This is the basis of Bayesian analysis, where we are interested in the probability of a certain hypothesis (be it a model, a parameter, or a set of parameters) H given some data D. As this is the probability assigned to H before taking the relevant evidence into account, it is commonly just referred to as the **posterior** (or posterior distribution function) and is denoted by another typeface

$$\mathcal{P} \equiv \mathscr{P}(H|D) \tag{3.5}$$

throughout this thesis. The Bayesian analysis is then a way to update our **prior** knowledge about H, which will be denoted as

$$\Pi \equiv \mathscr{P}(H), \tag{3.6}$$

taking into account the probability of the data before it is known, given the hypothesis to be tested, which is called the **likelihood** and from now on will be written as

$$L \equiv \mathscr{P}(D|H). \tag{3.7}$$

The remaining probability function entering Bayes' equation is called the **evidence**

$$\mathcal{Z} \equiv \mathscr{P}(D). \tag{3.8}$$

It describes the probability of the data without assuming anything about the theory to be tested, and, in practice, is often just treated as a normalisation

factor to ensure integrating over the posterior for all possible hypotheses equals one. In our notation, we then have

$$\mathcal{P} \propto L\Pi. \tag{3.9}$$

This equation also holds for probability density functions (pdf) $f(x)$ of continuous probability functions that can be written as

$$\mathscr{P}(x_1 < X < x_2) = \int_{x_1}^{x_2} \mathrm{d}x f(x). \tag{3.10}$$

One advantage of Bayesian analysis is that if there is new data available after the first analysis, one can update one's knowledge about H by applying Bayes' theorem iteratively. The posterior of the first experiment then serves as prior of the second:

$$\begin{aligned} \mathcal{P}(H|D_1D_2) &\propto L(D_2|H)\mathcal{P}(H|D_1) \\ &\propto L(D_2|H)L(D_1|H)\Pi(H). \end{aligned} \tag{3.11}$$

Thus, Bayesian analysis is a good way to update one's knowledge every time new data is available. What still remains to be addressed is the question of the initial prior $\Pi(H)$. This can be problematic, because, in principle, two observers might interpret the same observation, using exactly the same data, differently, just because their initial guess of the probability of the hypothesis was different. One can argue, that this is actually a positive feature of Bayesian statistics, as it provides a systematic way to quantify one's assumptions about the problem before looking at the data. However, every observer should be aware of **Cromwell's rule**: "I beseech you, in the bowels of Christ, think it possible that you may be mistaken." [156], i.e. a prior probability of 0 should only be assigned to statements that are logically false. This ensures, after applying enough informative likelihoods, convergence to a unique posterior distribution (**Bernstein-von Mises theorem** [31]). On the other hand, it is useful to include domain knowledge in the prior, e.g. in a measurement of the mass m, one can exclude unphysical negative masses a priori by setting $\Pi(m < 0) = 0$, possibly making the posterior more predictive. Inspired by statistical mechanics, one can introduce a systematic way to build a prior, namely by the **principle of maximum entropy**. The entropy

$$H = -\int_{-\infty}^{\infty} \mathrm{d}x \ln(f(x)) f(x), \tag{3.12}$$

of a probability distribution with pdf $f(x)$ describes the "ignorance" of the system. The pdf that maximises the entropy is the one with the least encoded information and the principle of maximum entropy therefore states that one should use it as a prior. Using the technique of Lagrange multipliers, one can impose constraints on the pdf. One constraint that can always be included is $\int_{-\infty}^{\infty} \mathrm{d}x f(x) = 1$. In the absence of any further prior knowledge, one has to maximise

$$-\int_{-\infty}^{\infty} \mathrm{d}x \ln(f(x)) f(x) + \lambda \left[\int_{-\infty}^{\infty} \mathrm{d}x f(x) - 1 \right], \tag{3.13}$$

which yields a flat prior

$$f(x) = e^{-1-\lambda} = \text{const.} \tag{3.14}$$

For the flat prior to be finite, its domain has to be finite, so that the second term in the bracket in Eq. (3.13) becomes $\int_{-x_{\min}}^{x_{\max}} \mathrm{d}x f(x) - 1$. Another prominent example is when the mean μ and the variance σ^2 is given. The pdf maximising

$$\begin{aligned} -\int_{-\infty}^{\infty} \mathrm{d}x \ln(f(x)) f(x) &+ \lambda_1 \left[\int_{-\infty}^{\infty} \mathrm{d}x f(x) - 1 \right] \\ &+ \lambda_2 \left[\int_{-\infty}^{\infty} \mathrm{d}x f(x) x - \mu \right] \\ &+ \lambda_3 \left[\int_{-\infty}^{\infty} \mathrm{d}x f(x)(x-\mu)^2 - \sigma^2 \right] \end{aligned} \tag{3.15}$$

is the Gaussian pdf

$$f(x) = \frac{e^{\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}. \tag{3.16}$$

Once we have applied Bayes' Theorem to obtain the posterior based on our prior and a measurement that is summarised in the likelihood, we have to "explore" the posterior distribution to infer the cosmological parameters we are interested in (and even more if there are other parameters that require to be marginalised over). If there are only one or two parameters, one can just visualise the posterior by directly plotting their posterior distributions and/or plot a 2-dimensional posterior contour plot. One can also calculate summary statistics analytically such as the mean or standard variation, at least if the data is sufficiently close to Gaussian to give these quantities any meaning. If the number of parameters is larger, but still sufficiently small, one can calculate the posterior numerically on a grid of possible parameter values. It

is still useful to visualise the posterior by marginalising over all parameters bar one and plot the marginalised posterior for the remaining parameter, or a 2-dimensional posterior contour plot, that is particularly useful to explore degeneracies between a pair of parameters. To obtain these marginalised plots, one can simply sum over the grid in the dimensions in question if the parameter space is low dimensional. For most cosmological analyses, the number of parameters is usually too large for grid based methods to be feasible. Therefore, the parameter space is commonly sampled using a Markovian random walk. A process is Markovian if the probability of going from a state X_n to X_{n+1} depends only on the current state X_n. According to the Markov Chain Convergence Theorem, an irreducible and aperiodic Markov chain, that is a Markov chain where it is possible to get from each state to any other in finite time and where the greatest common divisor of all possible steps n after which the chain can return to its initial state equals 1, converges in total variation to its unique stationary distribution [157]. In practice, this means that after a few steps, the so called "burn in", the Markov chain samples high density regions of the target distribution.

A popular **Markov Chain Monte Carlo** (MCMC) algorithm is the **Metropolis-Hastings** algorithm [158, 159]:

1. Choose an arbitrary starting point θ_0 and an arbitrary, but symmetric, *proposal* or *jumping distribution* $q(\theta_n|\theta') = q(\theta'|\theta_n)$. This can be for example a Gaussian centred around θ_n.

2. Pick a candidate θ' for the next sampling point from the *proposal distribution* $q(\theta|\theta')$.

3. Calculate the acceptance ratio
$$a = \frac{\mathcal{P}(\theta')}{\mathcal{P}(\theta_n)}. \tag{3.17}$$

4. If $a > 1$, then the posterior of the candidate is larger than at θ_n, thus, the candidate is automatically accepted as the next sampling point $\theta_{n+1} = \theta'$. If $a < 1$, θ' is accepted with probability a, otherwise $\theta_{n+1} = \theta_n$. Accepting less probable sampling points avoids that the chain gets stuck around a local maximum of the posterior distribution, failing to sample more probable areas in parameter space.

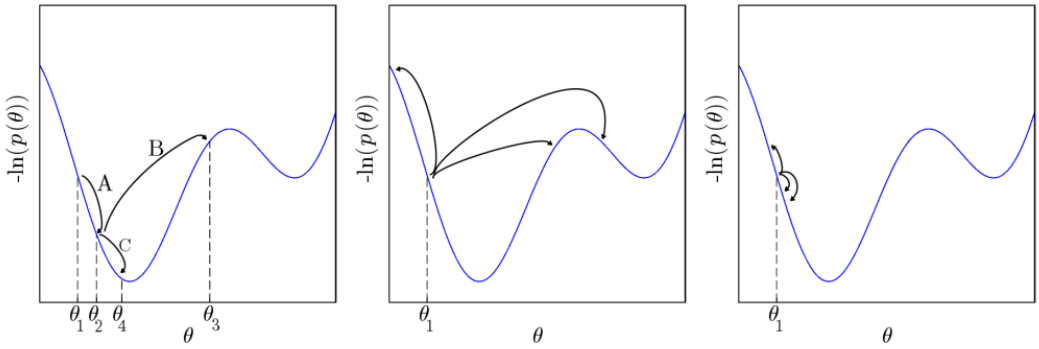

Figure 3.1: *Left panel*: An example of a Markov chain constructed by the Metropolis-Hastings algorithm: starting at θ_1, θ_2 is proposed and accepted (step A), θ_3 is proposed and refused (step B), θ_4 is proposed and accepted (step C). The resulting chain is $\{\theta_1, \theta_2, \theta_2, \theta_4, ...\}$. *Central panel*: An example of what happens with too broad a jump size: the chain lacks mobility because all the proposals are unlikely. *Right panel*: An example of what happens with too narrow a jump size: the chain samples the parameter space very slowly. Figure taken from [31].

5. Repeat 2. to 5. until the chain seems to have converged.

This is the algorithm implemented in the Cosmological Monte Carlo (CosmoMC) code [160, 161]. There is no intrinsic check of whether the chain has converged. Therefore, several chains with different starting points θ_0 usually run in parallel and they are stopped when they are all in the same area of parameter space. Another issue is the choice of the *jumping distribution*, which is demonstrated in Fig. 3.1. If it is chosen to be too narrow, the chain is very inefficient and it might get stuck in a local posterior maximum (which corresponds to a minimum in the plot). Theoretically, the chain would converge to the global maximum eventually, but that is likely to take much longer than the computing time available. If the *proposal distribution* is chosen too wide, jumps might be longer than the features of the posterior surface and the proposals are rejected most of the time.

An efficient way around that problem is hybrid or **Hamiltonian Monte Carlo** [162], where the negative logarithm of the posterior is interpreted as a "potential" in Hamiltonian mechanics. Instead of choosing a random candidate

for the next sampling point, each parameter obtains a random "momentum" and the proposal is deterministically chosen by solving the Hamiltonian equations of motion in parameter space. Due to energy conservation, the acceptance rate is theoretically equal to 1, although numerical errors can lower this value a bit.

The techniques introduced in this section will be applied to inference of cosmological parameters from the galaxy power spectrum in the remainder of this chapter, with a special focus on measuring $f_{\rm NL}$.

3.2 The Covariance Matrix of the Galaxy Power Spectrum

In order to infer information on cosmological parameters from a measured power spectrum as described in the previous section, one has to know the likelihood of power-spectra, which, for simple cases, can be calculated analytically. For general cases, one usually assumes the likelihood or posterior to be a multi-variate Gaussian with a covariance matrix

$$C_{ij} \equiv \langle P(k_i)P(k_j)\rangle - \langle P(k_i)\rangle \langle P(k_j)\rangle \,. \tag{3.18}$$

The estimation of the covariance matrix is a critical step in the analysis of data. Internal methods such as the sub-sample, jackknife and bootstrap methods have been widely used in the past, but Norberg *et al.* [32] have shown that they are not able to faithfully reproduce variances:

- **Sub-sample Method**: The data is split into $N_{\rm sub}$ sub-samples and the clustering statistics are estimated for each sub-sample separately, using Eq. (3.18). If the sub-samples are independent, one obtains the correct covariance for an $N_{\rm sub}$-times smaller sample volume, which can easily be accounted for, because the covariance is proportional to the volume. However, due to the presence of long-range modes, which we are actually also interested in for $f_{\rm NL}$ analyses, the sub-samples are not independent and the covariances thus obtained are not reliable [32].

- **Jackknife** [163]: The data is again split into $N_{\rm sub}$ sub-samples, but now the clustering measurement is repeated omitting a different sub-sample

each time. The overall power spectrum $\widehat{P}$ is estimated as the mean of the power spectra of each jackknife P_i:

$$\widehat{P}(k) = \frac{1}{N_{\rm sub}} \sum_{m=1}^{N_{\rm sub}} P_m(k). \tag{3.19}$$

The power covariance matrix is then estimated as

$$\widehat{C}_{ij} = \frac{N_{\rm sub}-1}{N_{\rm sub}} \sum_{m=1}^{N_{\rm sub}} [P_m(k_i) - \widehat{P}(k_i)][P_m(k_j) - \widehat{P}(k_j)], \tag{3.20}$$

where the $\frac{N_{\rm sub}-1}{N_{\rm sub}}$-prefactor takes the lack of independence between each resampling into account [164]. The jackknife technique provides fairly accurate covariances on large scales, but fails to reproduce its correct scale dependence. On smaller scales, the jackknife estimate of the covariance depends on the number of sub-samples $N_{\rm sub}$. In a study by Norberg *et al.*, increasing $N_{\rm sub}$ actually lead to a worse estimate of the covariance of the correlation function on small scales [32]. One of the main plots of [32] has been included as Fig. 3.2 to clarify the previous statements. The jackknife covariances are represented as blue and cyan lines. They agree well with the black line representing "mock"-based covariances at large scales, but the jackknife error estimates are too large at small scales. Splitting the data into 27 sub-samples (blue line) over-predicts the error less than splitting into 64 sub-samples (cyan line). Especially for $f_{\rm NL}$-analyses one should also keep in mind that internal methods in general are limited by the sampling volume and therefore fail to include the full cosmic variance [32].

- **Bootstrapping** [165]: The original data is still divided into $N_{\rm sub}$ sub-samples, but all no sub-samples are omitted in bootstrapping. Each sub-sample is assigned a randomly chosen weight. Contrary to the previous two techniques, there is no limit on the number of boot resamplings $N_{\rm boot}$. After $N_{\rm boot}$ resamplings, one can estimate the power as

$$\widehat{P}(k) = \frac{1}{N_{\rm boot}} \sum_{m=1}^{N_{\rm boot}} P_m(k), \tag{3.21}$$

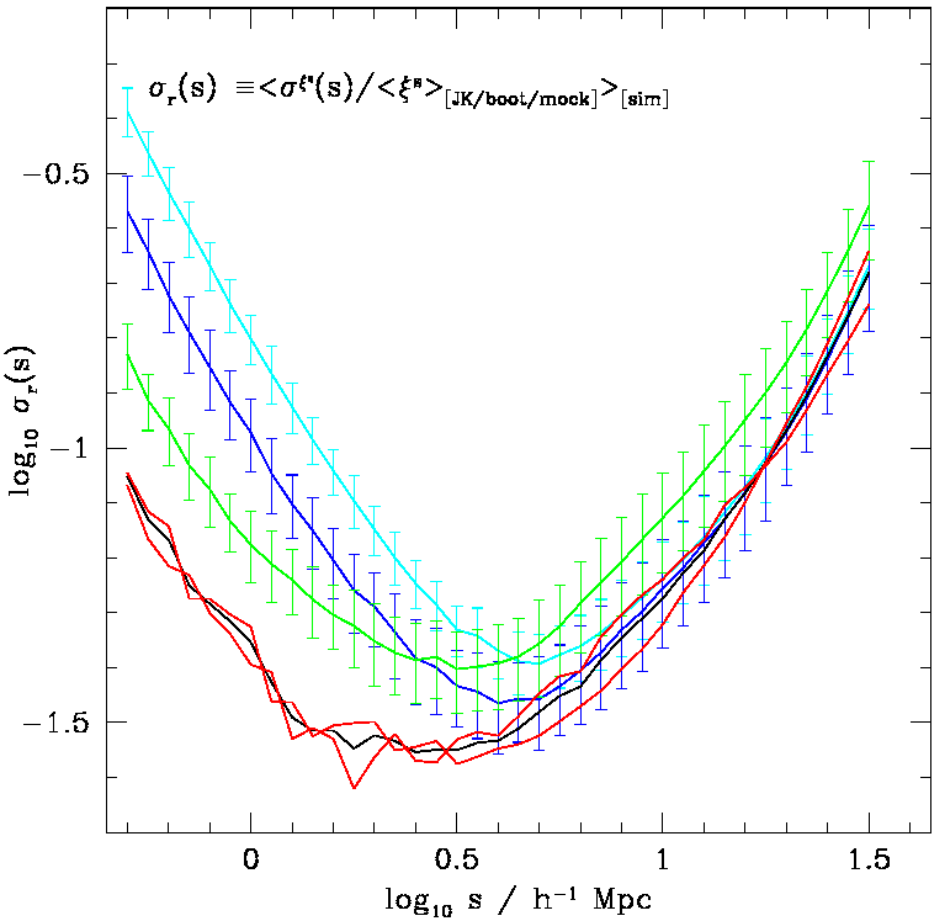

Figure 3.2: The mean relative variance of ξ_s as a function of scale for different error estimators: black – 100 "mock" data sets; red – two different samples of the mocks (the first 50 & the last 50); green – bootstrap errors after splitting each data set into 27 sub-samples; blue and cyan – jackknife errors measured after splitting each data set into 27 and 64 sub-samples respectively (i.e. Jack-27 and Jack-64). Figure taken from [32].

and the covariance matrix as

$$\widehat{C}_{ij} = \frac{1}{N_{\rm boot}-1}\sum_{m=1}^{N_{\rm boot}}[P_m(k_i)-\widehat{P}(k_i)][P_m(k_j)-\widehat{P}(k_j)]. \tag{3.22}$$

This technique reproduces the scale dependence of the covariance, but overestimates the error on all scales (see the green line in Fig. 3.2). As already mentioned for the jackknife technique, bootstrapping, as an internal method, fails to include full cosmic variance in the covariance matrix, due to the sampling volume limit [32].

Robust estimates are often instead obtained from mock galaxy catalogues. In recent analyses of the Baryon Oscillation Spectroscopic Survey (BOSS, [166]), these were generated from second order Lagrangian Perturbation Theory (2LPT, cf. 1.2.3) matter fields using a friends-of-friends group finder [167] to find haloes [168, 169]. Their masses were calibrated by comparisons with N-body simulations. A Halo Occupation Distribution, which can be obtained from small scale pilot N-body simulations, then prescribed how to populate these haloes with mock galaxies, and the geometry and the efficiency of the survey were sampled. Methods for producing mocks include

- **N-body simulations**: The continuous dark matter distribution is discretised into dark matter 'particles', each with a mass, velocity and position. These are updated using a discrete version of Eq. (1.80). A popular discretisation scheme is the Kick-Drift-Kick or Leapfrog method, where the position of each particle is updated after full timesteps according to the velocity at half the time between each position update:

$$\begin{aligned}\mathbf{r}(t_{i+1}) &= \mathbf{r}(t_i) + \mathbf{v}(t_{i+1/2})H_0\int_{a(t_i)}^{a(t_{i+1})}\frac{{\rm d}a}{a^3H(a)}\\ \mathbf{v}(t_{i+1/2}) &= \mathbf{v}(t_{i-1/2}) - \nabla\Phi(t_i)\frac{H_0}{a(t_i)}\int_{a(t_{i-1/2})}^{a(t_{i+1/2})}\frac{{\rm d}a}{a^2H(a)}\ [170].\end{aligned} \tag{3.23}$$

This requires the calculation of $\nabla\Phi$ at each timestep, which scales with the square of the number of particles, and is therefore generally a time consuming and thus expensive procedure.

- Comoving Lagrangian Acceleration (**COLA**, [171]): The idea behind COLA is to evolve the position $\mathbf{r}_{\rm LPT}$ using second order Lagrangian perturbation theory (cf. Sec. 1.2.3) at large scales, and then discretise only

the time evolution of the third order residual $\mathbf{r}_{\rm res} \equiv \mathbf{r} - \mathbf{r}_{\rm LPT}$. This allows the use of wider timesteps where only linear and mildly non-linear scales are of importance, such as in the generation of mock galaxy catalogues.

- Pinpointing Orbit-Crossing Collapsed Hierarchical Objects (**PINOCCHIO**, [172]) avoids N-body simulations completely by making use of the Zel'dovich approximation (1.97). Dark matter haloes can be formed where orbit-crossing occurs, i.e. where the Jacobian of the transformation from Eulerian to Lagrangian coordinates vanishes.

- Effective Zel'dovich Approximation Mocks (**EZmocks**, [173]), as the name suggests, also rely on the Zel'dovich approximation. Non-linear effects, such as tidal fields, non-linear and non-local stochastic bias, etc. are included through effective models and by iteratively matching the probability distribution to N-body simulations.

- Perturbation Theory Catalog Generator of Halo and Galaxy Distributions (**PATCHY**, [174]) extends second order Lagrangian perturbation theory with a small scale spherical collapse model. It includes a non-linear and non-local stochastic bias similarly to EZmocks.

The covariance matrix is then the sample variance and cross correlation of the power spectra from the different mock realisations [169, 175, 176]. The covariance matrix computed from the mocks will depend on the cosmological model that was used to generate them. It is computationally costly to produce mock catalogues for each possible cosmological model and set of parameters to be tested, so one usually chooses a cosmological model which will produce a P(k) reasonably close to the measured one and uses the covariance matrix computed from the mocks created assuming that model. This approximation does not hold in general, especially at large scales. In [4], we studied other ways of approaching this problem, which I go on to discuss in Sec. 3.4, including using approximations to the true posterior distribution to obtain accurate inferences without requiring a covariance matrix for each cosmological model. We apply the most suitable of these approximations and the true distribution to provide a pdf for measurements of the non-Gaussianity parameter $f_{\rm NL}$. Our result provides a complementary method to analysing $f_{\rm NL}$ directly from $\delta_{\rm g}(\mathbf{x})$, as described in [177].

3.3 The Power Spectrum Likelihood

In this section, we elaborate the analytic likelihood and posterior functions of the galaxy clustering power spectrum assuming a Gaussian density field. We consider this posterior function as "truth" and compare it to commonly used approximations of the galaxy power spectrum posterior function for single modes, which we shall introduce in Sec. 3.3.3.

3.3.1 The True Distribution of the Density Contrast Under the Assumption of a Gaussian Density Field

The positions of the galaxies in a survey can be transformed into the galaxy over-density field (cf. Eq. (1.61))

$$\delta_{\mathrm{g}}(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}(\mathbf{x})}{\bar{\rho}(\mathbf{x})} = \frac{n(\mathbf{x}) - \bar{n}(\mathbf{x})}{\bar{n}(\mathbf{x})}, \tag{3.24}$$

where $n(\mathbf{x})$ is the measured galaxy number density and $\bar{n}(\mathbf{x})$ the expected value. Fourier transforming $\delta(\mathbf{x})$ yields

$$\delta_{\mathbf{k}} \equiv \frac{1}{V} \int \mathrm{d}^3\mathbf{x} \delta(\mathbf{x}) \exp(i\mathbf{k}\mathbf{x}) \tag{3.25}$$

whose covariance matrix

$$\left\langle \delta_{\mathbf{k}_1} \delta^*_{\mathbf{k}_2} \right\rangle = \frac{(2\pi)^3}{V} \delta_D(\mathbf{k_1} - \mathbf{k_2}) P(\mathbf{k_1}) \tag{3.26}$$

is given by the power spectrum $P(\mathbf{k})$, as defined in Sec. 1.2.4. In the derivations of [4] and of this chapter, we make the assumption that $\delta_{\mathbf{k}}$ forms a Gaussian random field. Intuitively, any result based on that assumption seems unsuitable for tests of primordial non-Gaussianity, but, as I show in Appendix A, the error we make due to that assumption is a lot smaller than approximating the posterior of the power spectrum as Gaussian. Therefore, the probability of measuring a particular value of the real and imaginary parts $\left(\widehat{\delta_u}, \widehat{\delta_v}\right)$ of a single $\widehat{\delta_{\mathbf{k}}} = \widehat{\delta_u} + i\widehat{\delta_v}$ is assumed to obey a zero centred Gaussian distribution with standard deviation half the true power $\frac{1}{2}P_T(\mathbf{k})$. Thus the respective distributions of the real and imaginary parts of the density field

read

$$\mathcal{Z}\left(\widehat{\delta_{\mathbf{u}}}\right) = \frac{1}{\sqrt{\pi P_T(\mathbf{k})}} \exp\left(-\frac{\widehat{\delta_u}^2}{P_T(\mathbf{k})}\right), \text{ and}$$
$$\mathcal{Z}\left(\widehat{\delta_{\mathbf{v}}}\right) = \frac{1}{\sqrt{\pi P_T(\mathbf{k})}} \exp\left(-\frac{\widehat{\delta_v}^2}{P_T(\mathbf{k})}\right). \tag{3.27}$$

We use the letter $\mathcal{Z}$ here, because we have assumed that the true power is known, i.e. the distribution only depends on the data, thus, it is formally a Bayesian evidence. As the power spectrum is estimated from the absolute value of the density field, I derive its distribution in the following way. The cumulative distribution function (CDF) of the absolute value $\widehat{|\delta_{\mathbf{k}}|} = \sqrt{\widehat{\delta_{\mathbf{u}}}^2 + \widehat{\delta_{\mathbf{v}}}^2}$ is the joint cumulative distribution of finding a set of $\widehat{\delta_u}$ and $\widehat{\delta_v}$ within a radius $\widehat{|\delta_{\mathbf{k}}|}$:

$$\mathrm{CDF}(\widehat{|\delta_{\mathbf{k}}|}) = \int \mathrm{d}\widehat{\delta_{\mathbf{u}}} \int \mathrm{d}\widehat{\delta_{\mathbf{v}}} \mathcal{Z}\left(\widehat{\delta_{\mathbf{u}}}\right) \mathcal{Z}\left(\widehat{\delta_{\mathbf{v}}}\right) \Theta\left(\widehat{|\delta_{\mathbf{k}}|} - \sqrt{\widehat{\delta_{\mathbf{u}}}^2 + \widehat{\delta_{\mathbf{v}}}^2}\right). \tag{3.28}$$

The pdf of $\widehat{|\delta_{\mathbf{k}}|}$ is then the derivative of the CDF with respect to $\widehat{|\delta_{\mathbf{k}}|}$, where the derivative of the Heaviside Theta function is defined as the Dirac delta distribution:

$$\mathcal{Z}_R\left(\widehat{|\delta_{\mathbf{k}}|}\right) = \frac{\partial\,\mathrm{CDF}(\widehat{|\delta_{\mathbf{k}}|})}{\partial\widehat{|\delta_{\mathbf{k}}|}} = \int \mathrm{d}\widehat{\delta_{\mathbf{u}}} \int \mathrm{d}\widehat{\delta_{\mathbf{v}}} \mathcal{Z}\left(\widehat{\delta_{\mathbf{u}}}\right) \mathcal{Z}\left(\widehat{\delta_{\mathbf{v}}}\right) \delta_D\left(\widehat{|\delta_{\mathbf{k}}|} - \sqrt{\widehat{\delta_{\mathbf{u}}}^2 + \widehat{\delta_{\mathbf{v}}}^2}\right). \tag{3.29}$$

The integration becomes easy after inserting Eq. (3.27) and converting to polar coordinates. One can identify the distribution of $\widehat{|\delta_{\mathbf{k}}|}$ as a Rayleigh distribution that scales with the true power spectrum $P_T(\mathbf{k})$:

$$\begin{aligned}\mathcal{Z}_R\left(\widehat{|\delta_{\mathbf{k}}|}\right) &= \int \mathrm{d}\widehat{|\delta_{\mathbf{k}}|} \int_0^{2\pi} \mathrm{d}\varphi \frac{\exp\left(-\frac{\widehat{\delta_{\mathbf{u}}}^2 + \widehat{\delta_{\mathbf{v}}}^2}{P_T(\mathbf{k})}\right)}{\pi P_T(\mathbf{k})} \delta_D\left(\widehat{|\delta_{\mathbf{k}}|} - \sqrt{\widehat{\delta_{\mathbf{u}}}^2 + \widehat{\delta_{\mathbf{v}}}^2}\right) \\ &= \frac{2\widehat{|\delta_{\mathbf{k}}|}}{P_T(\mathbf{k})} \exp\left(-\frac{\widehat{|\delta_{\mathbf{k}}|}^2}{P_T(\mathbf{k})}\right).\end{aligned} \tag{3.30}$$

Throughout this chapter, Eq. (3.30) is regarded as the "true" distribution of $\widehat{|\delta_{\mathbf{k}}|}$ to which several approximations will be compared to later.

As $\langle\delta_{\mathbf{k}}\rangle = 0$ for all models by definition, any model dependence enters the Rayleigh distribution only in the covariance of the density field, which is equal to the true power spectrum $P_T(\mathbf{k})$. A Rayleigh distribution peaks

at the value of its scale parameter which is in our case the square root of the true power spectrum $P_T(\mathbf{k})$. Measurements of $\widehat{\delta_{\mathbf{k}}}$ have been used to make cosmological inferences when they have been further decomposed into spherical harmonics and spherical Bessel functions, because radial and angular modes can be distinguished, allowing an easy analysis of redshift-space distortions (cf. Sec. 2.1). However, this method is rather complex and computationally expensive [178,179]. It is difficult to linearly compress $\widehat{\delta_{\mathbf{k}}}$ efficiently maximally retaining information. We are therefore considering this distribution in terms of the power spectrum in the following section.

3.3.2 The Posterior in Terms of the Power

To obtain the "true" distribution of the power spectrum, we rewrite the Rayleigh distribution in terms of the power. We replace the true power spectrum $P_T(\mathbf{k})$ with a hypothesis, i.e. a model power spectrum, $P_H(\mathbf{k})$, and $\widehat{\delta_{\mathbf{k}}}$ with $\sqrt{\widehat{P}(\mathbf{k})}$ in equation (3.30) which in this way depends on both data and model, and hence becomes a likelihood:

$$L_R\left(\widehat{P}(\mathbf{k})\middle|\, P_H(\mathbf{k})\right) = \frac{2\sqrt{\widehat{P}(\mathbf{k})}}{P_H(\mathbf{k})} \exp\left(-\frac{\widehat{P}(\mathbf{k})}{P_H(\mathbf{k})}\right). \tag{3.31}$$

We can use Bayes' theorem (cf. Eq. (3.9)) to find the posterior. In line with the principle of maximum entropy (cf. Sec. 3.1) we assume a uniform prior

$$\Pi\left(P_H(\mathbf{k})\right) = \begin{cases} \frac{1}{P_{\max}(\mathbf{k})} & \text{, if } 0 \leq P_H(\mathbf{k}) \leq P_{\max}(\mathbf{k}), \\ 0 & \text{otherwise,} \end{cases} \tag{3.32}$$

which requires an arbitrary choice of $P_{\max}(\mathbf{k})$. We assume that $P_{\max}(\mathbf{k})$ is far in the right tail of the likelihood such that $\frac{\Pi(P_H(\mathbf{k}))}{\mathcal{Z}\left(\widehat{P}(\mathbf{k})\right)}$ is effectively constant and hence acts only as a normalisation factor. Alternatively, one could argue that all orders of magnitude of the power spectrum between arbitrary $P_{\min}(\mathbf{k}) > 0$ and $P_{\max}(\mathbf{k}) > P_{\min}(\mathbf{k})$ are equally likely, which is equivalent to choosing a logarithmic prior

$$\Pi\left(P_H(\mathbf{k})\right) = \begin{cases} \frac{\ln(P_{\max}(\mathbf{k})/P_{\min}(\mathbf{k}))}{P_H(\mathbf{k})} & \text{, if } P_{\min}(\mathbf{k}) \leq P_H(\mathbf{k}) \leq P_{\max}(\mathbf{k}), \\ 0 & \text{otherwise.} \end{cases} \tag{3.33}$$

However, with Bernstein-von Mises theorem, both Eq. (3.32) and (3.33) converge to a unique prior if the range between $P_{\min}(\mathbf{k})$) and $P_{\max}(\mathbf{k})$ is chosen wide enough. The uniform prior has the advantage of imposing only one condition that is in disagreement with Cromwell's rule (by setting $\Pi(P_H) = 0$ for $P_H > P_{\max}$ to normalise the distribution rather than due to logical exclusion), whereas for the logarithmic prior, we also have to set $\Pi(P_H) = 0$ for $P_H < P_{\min}$ for pure normalisation reasons. Therefore, we choose a uniform prior and, thus, for the "true" posterior we have

$$\begin{aligned}\mathcal{P}_R\left(P_H(\mathbf{k})\left|\widehat{P}(\mathbf{k})\right.\right) &= \frac{L_R\left(\widehat{P}(\mathbf{k})\middle|\, P_H(\mathbf{k})\right)}{\int \mathrm{d}P_H\; L_R\left(\widehat{P}(\mathbf{k})\middle|\, P_H(\mathbf{k})\right)} \\ &\propto \frac{2\sqrt{\widehat{P}(\mathbf{k})}}{P_H(\mathbf{k})}\exp\left(-\frac{\widehat{P}(\mathbf{k})}{P_H(\mathbf{k})}\right). \end{aligned} \tag{3.34}$$

As $\widehat{P}$ is a constant in the posterior, one can rewrite Eq. (3.34) such that the log-posterior only depends on the ratio $\widehat{P}(\mathbf{k})/P_H(\mathbf{k})$:

$$-2\ln(\mathcal{P}_R) = 2M\ln\left(\frac{P_H(\mathbf{k})}{\widehat{P}(\mathbf{k})}\right) + 2M\frac{\widehat{P}(\mathbf{k})}{P_H(\mathbf{k})} + \text{const.} \tag{3.35}$$

One can follow the method of [180] to write $\mathcal{P}_R$ in a Gaussian-like way and introduce the function

$$\gamma(x) \equiv \sqrt{-\ln(x) + x} \tag{3.36}$$

to make Eq. (3.35) look more quadratic:

$$-2\ln(\mathcal{P}_R) = 2M\left[\gamma\left(\frac{\widehat{P}(\mathbf{k})}{P_H(\mathbf{k})}\right)\right]^2 + \text{const.} \tag{3.37}$$

We can also define

$$P_\gamma(\mathbf{k}) \equiv P_f(\mathbf{k})\gamma\left(\frac{\widehat{P}(\mathbf{k})}{P_H(\mathbf{k})}\right) \tag{3.38}$$

for some fiducial model with power P_f. P_γ has then a symmetric Gaussian posterior with a fixed variance $\tilde{C}_k = \frac{2P_f^2(\mathbf{k})}{M}$ evaluated for our fiducial model:

$$-2\ln(\mathcal{P}_R) = 4P_\gamma\tilde{C}_k^{-1}P_\gamma + \text{const.} \tag{3.39}$$

In general, things are more complicated than this simple picture. For example, the survey geometry leads to a convolution of $\delta_{\mathbf{k}}$, and non-linear effects

distort the small scale mode distribution. These cause mode coupling and hence introduce off-diagonal entries into the power covariance matrix. There is no obvious way to include these effects in Eq. (3.35), because it does not explicitly contain the power covariance. Ideally, we would still like to use a single distribution which should be matched to simulations (e.g. [181]). Therefore, I introduce and test approximations to the "true" that explicitly contain the power covariance in the following subsections. First, I look at the common Gaussian approximations before broadening the choice of distributions by considering a number of forms for the likelihood inspired by CMB analyses.

3.3.3 Common Approximations of the Likelihood/Posterior of the Power Spectrum

Often, the power-spectrum is directly analysed, incorrectly assuming that it follows a Gaussian distribution, thus the distribution of a finite empirical realisation of the power spectrum $\widehat{P}(\mathbf{k})$ would read

$$\mathcal{Z}\left(\widehat{P}(\mathbf{k})\right) = \frac{\exp\left(-\frac{1}{2}\frac{\left[\widehat{P}(\mathbf{k})-P_T(\mathbf{k})\right]^2}{C_\mathbf{k}}\right)}{\sqrt{2\pi C_\mathbf{k}}}, \tag{3.40}$$

where $C_\mathbf{k} \equiv \langle P_T^2(\mathbf{k})\rangle = \frac{2P_T^2(\mathbf{k})}{M}$ is the variance of the true power spectrum P_T at a bin centred around $\mathbf{k}$ comprising M independent modes. Note that we assume that the widths and positions of the $\mathbf{k}$-bins are such that window effects are negligible [109] and different modes are independent.

As in Sec. 3.3.2, we replace $P_T(\mathbf{k})$ with $P_H(\mathbf{k})$ in equation (3.40) making it a likelihood:

$$L\left(\widehat{P}(\mathbf{k})\middle|\, P_H(\mathbf{k})\right) = \frac{\exp\left(-\frac{1}{2}\frac{\left[\widehat{P}(\mathbf{k})-P_H(\mathbf{k})\right]^2}{C_\mathbf{k}^H}\right)}{\sqrt{2\pi C_\mathbf{k}^H}}, \tag{3.41}$$

where $C_\mathbf{k}^H \equiv \langle P_H^2(\mathbf{k})\rangle$ is the variance for the hypothetical power spectrum $P_H(\mathbf{k})$. Note that the histogram of the Gaussian likelihood $L\left(\widehat{P}(\mathbf{k})\middle|\, P_H(\mathbf{k})\right)$ is not Gaussian, because the variance of each $\mathbf{k}$-mode is different, thus the histogram is a sample of a mixture of independent Gaussians (cf. e.g. [31]).

However, in practice one chooses a fiducial model with power spectrum

$\tilde{P}(\mathbf{k})$ and estimates the variance $\tilde{C}_{\mathbf{k}} \equiv \left\langle \tilde{P}^2(\mathbf{k}) \right\rangle$ for this particular choice:

$$L\left(\widehat{P}(\mathbf{k}) \middle| P_H(\mathbf{k}), \tilde{C}_{\mathbf{k}}\right) = \frac{\exp\left(-\frac{1}{2}\frac{\left[\widehat{P}(\mathbf{k})-P_H(\mathbf{k})\right]^2}{\tilde{C}_{\mathbf{k}}}\right)}{\sqrt{2\pi\tilde{C}_{\mathbf{k}}}}. \tag{3.42}$$

For mock based variance calculations, $\tilde{P}(\mathbf{k})$ is the cosmology of the mocks used in their analysis.

We can again use Bayes' theorem (cf. Eq. 3.9) and assume the same uniform prior as before to find the posterior. For the posterior assuming a Gaussian distribution in $\widehat{P}(\mathbf{k})$ with model-dependent covariance we have

$$\begin{aligned}\mathcal{P}_D\left(P_H(\mathbf{k}) \middle| \widehat{P}(\mathbf{k})\right) &= \frac{L\left(\widehat{P}(\mathbf{k}) \middle| P_H(\mathbf{k})\right)}{\int \mathrm{d}P_H \; L\left(\widehat{P}(\mathbf{k}) \middle| P_H(\mathbf{k})\right)} \\ &\propto \frac{\exp\left(-\frac{1}{2}\frac{\left[\widehat{P}(\mathbf{k})-P_H(\mathbf{k})\right]^2}{C^H_{\mathbf{k}}}\right)}{\sqrt{2\pi C^H_{\mathbf{k}}}},\end{aligned} \tag{3.43}$$

where we adopt the subscript notation $\mathcal{P}_D$ of [180]. Note that both the exponential and the covariance matrix $C^H_{\mathbf{k}}$ depend on $P_H(\mathbf{k})$.

If a fixed covariance is assumed, we have to apply the Bayesian Eq. (3.9) to Eq. (3.42) giving

$$\mathcal{P}_f\left(P_H(\mathbf{k}) \middle| \widehat{P}(\mathbf{k}), \tilde{C}_{\mathbf{k}}\right) \propto \frac{\exp\left(-\frac{1}{2}\frac{\left[\widehat{P}(\mathbf{k})-P_H(\mathbf{k})\right]^2}{\tilde{C}_{\mathbf{k}}}\right)}{\sqrt{2\pi\tilde{C}_{\mathbf{k}}}}. \tag{3.44}$$

This is the most commonly assumed posterior distribution in parameter inferences from the galaxy power spectrum. The problem with this approximation is that one has to assume the parameters one wants to measure a priori, as the variance $C_{\mathbf{k}}$ depends on the model and its parameters. The posterior distribution $\mathcal{P}_D$ is more accurate in this respect, as the covariance is updated for every possible value of the power spectrum, which depends on the model to be tested and its parameters. However, this is not feasible in practice, because, for each possible set of parameters, the covariance matrix has to be estimated from a few thousand mock realisations of the observations. The next section shows further shortcomings.

3.3.4 A Simple Test of the Posterior Shapes for the isotropically averaged power spectrum

In this subsection we combine the single mode posterior functions of the last section to posterior functions of the band-power. The band-power is the power spectrum averaged over a k-range around the k-value in question to increase the signal-to-noise ratio. We do not take any anisotropic effects, such as redshift space distortions, into account. This is conservative because the effective volume for higher multipole moments (cf. Eq. 3.51) is smaller, therefore containing fewer independent modes and thence amplifying the effect of choosing different posterior shapes.

In Gaussian cases, we suppose that our volume is large enough to accommodate M independent complex Gaussian distributed samples of $\delta_{\mathbf{k}}$. Estimating the power spectrum as

$$\widehat{P}(\mathbf{k}) = \frac{1}{M}\sum_{m=1}^{M}\delta(\mathbf{k}_m)\delta^*(\mathbf{k}_m), \tag{3.45}$$

one can calculate the covariance matrices at higher numbers of modes M:

$$\begin{aligned} C_{ab} =& \langle\widehat{P}(\mathbf{k}_a)\widehat{P}(\mathbf{k}_b)\rangle - \langle\widehat{P}(\mathbf{k}_a)\rangle\langle\widehat{P}(\mathbf{k}_b)\rangle \\ =& \frac{1}{M_a M_b}\sum_{m=1}^{M_a}\sum_{n=1}^{M_b}\left[\langle\delta(\mathbf{k}_m)\delta^*(\mathbf{k}_m)\delta(\mathbf{k}_n)\delta^*(\mathbf{k}_n)\rangle\right. \\ & \left. - \langle\delta(\mathbf{k}_m)\delta^*(\mathbf{k}_m)\rangle\,\langle\delta(\mathbf{k}_n)\delta^*(\mathbf{k}_n)\rangle\right]. \end{aligned}$$

Using Wick's theorem (cf. Sec. 1.2.4), the 4-point correlator can be written as the sum of three products of 2-point correlators, of which one is cancelled by the second term in the square brackets. The remaining 2-point correlators can be replaced by a discrete version of Eq. (1.104):

$$\begin{aligned} C_{ab} =& \frac{1}{M_a M_b}\sum_{m=1}^{M_a}\sum_{n=1}^{M_b}\left[\langle\delta(\mathbf{k}_m)\delta(\mathbf{k}_n)\rangle\,\langle\delta^*(\mathbf{k}_m)\delta^*(\mathbf{k}_n)\rangle\right. \\ & \left. + \langle\delta(\mathbf{k}_m)\delta^*(\mathbf{k}_n)\rangle\,\langle\delta^*(\mathbf{k}_m)\delta(\mathbf{k}_n)\rangle\right] \\ =& \frac{1}{M_a M_b}\sum_{m=1}^{M_a}\sum_{n=1}^{M_b}\left[\delta_D(\mathbf{k}_m+\mathbf{k}_n)P^2(\mathbf{k}_m) + \delta_D(\mathbf{k}_m-\mathbf{k}_n)P^2(\mathbf{k}_m)\right]. \end{aligned}$$

Due to the hermiticity of $\delta(\mathbf{k})$, the power spectrum is anti-symmetric, so the two terms in the remaining square brackets are equal to each other, and the

covariance is

$$C_{ab} = \frac{2}{M}\delta_D(\mathbf{k}_a - \mathbf{k}_b)P^2(\mathbf{k}_a). \tag{3.46}$$

We can obtain $\mathcal{P}_R\left(P_H(\mathbf{k})\left|\widehat{|\delta_{\mathbf{k}}|}\right.\right)$ for the band-power by multiplying together the single mode expressions for all modes that fall into the k-bin around $\mathbf{k}$ in the band-powering process, which is the averaging over power spectrum estimates of all values of $\mathbf{k}$ in the respective k-bin.

The three different posterior shapes $\mathcal{P}_R$, $\mathcal{P}_D$ and $\mathcal{P}_f$ of P_H are plotted in Fig. 3.3. In the top panel of Fig. 3.3, we plot single mode posterior distributions for which we adopt $\widehat{|\delta_{\mathbf{k}}|} = 100$ and $P_T(\mathbf{k}) = \widehat{P}(\mathbf{k}) = \widehat{|\delta_{\mathbf{k}}|}^2 = 10000$. Note that a different choice would shift the peak positions and normalisation factor, but preserve the shapes. We make two different choices for the fixed covariance to see the effect of making the wrong assumption. For the dotted red line, we choose the covariance matrix which corresponds to the true power spectrum $P_T(\mathbf{k})$, i.e. $\tilde{C}_{\mathbf{k}} = 2P_T^2(\mathbf{k}) = 50000000$, and for the dashed-dotted line, we consider that our guess of the power spectrum is 5 per cent lower than the actual power spectrum, i.e. $\tilde{C}_{\mathbf{k}} = 45125000$. The panels in the middle and at the bottom of Fig. 3.3 show the posterior distributions for 10 and 100 independent modes respectively.

Fig. 3.3 and 3.4 show that different choices of the covariance matrix provide very different posterior distributions for a small number of modes, but if we can increase the number of independent modes, we see the effect of the central limit theorem and the posterior distribution functions become more and more similar. This means, that for many applications, Gaussian approximations work, but in the case of an $f_{\rm NL}$-analysis, where one has to rely on a small number of modes, assuming a Gaussian posterior distribution can lead to wrong confidence intervals.

We observe that the maximum of the fixed-covariance Gaussian posterior always agrees with the true value, even if the wrong fiducial model has been chosen. However, if we choose the wrong covariance matrix, we over- or underestimate the error of our measurements. If we do not fix the covariance, the best fit, i.e. the maximum of the posterior, has an offset with regard to the true value, which decreases as the number of modes increases. We also notice the long right tails of the varying-covariance Gaussian and the posterior

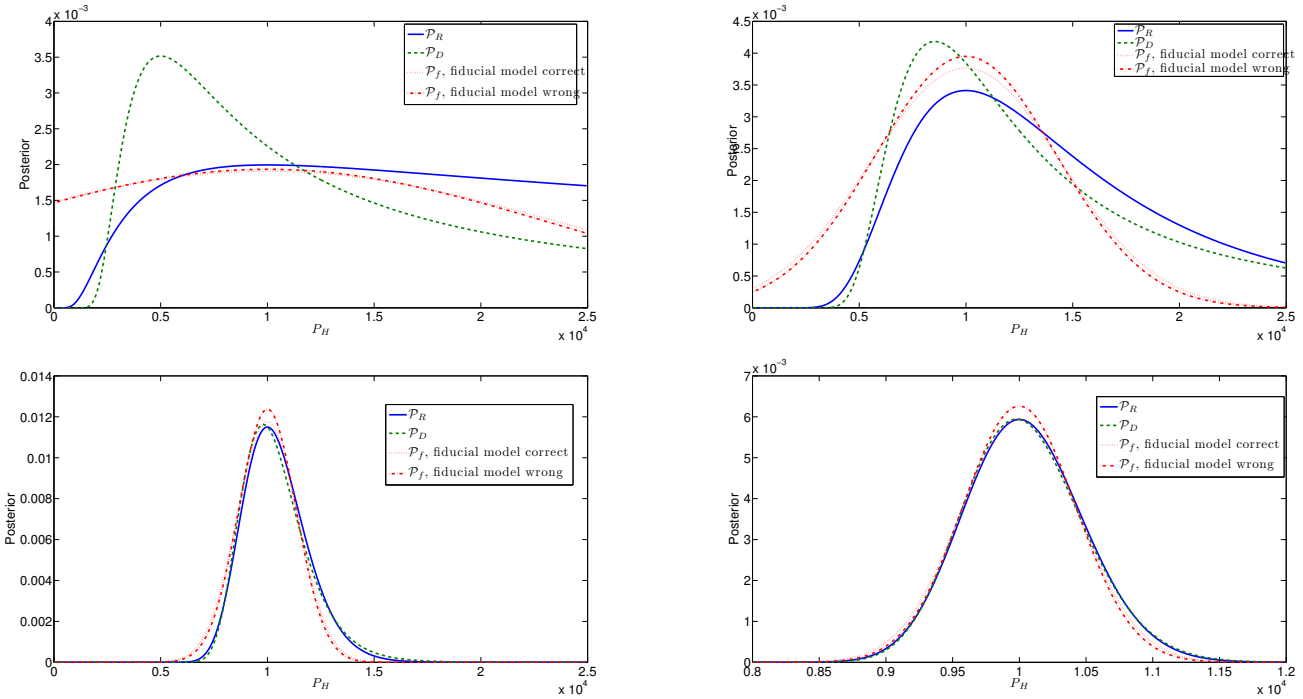

Figure 3.3: Comparison of different posterior distribution functions for 1, 10, 100 and 1000 independent modes (from top left to bottom right). The blue line represents the product of single Rayleigh distributed modes (true posterior distribution) and some of the approximations, such as the Gaussian posterior distribution with a model-dependent covariance (green), and the Gaussian posterior where the covariance is estimated for a fixed fiducial model (red). The posterior takes the form of the dotted red line if the fiducial and the true power spectra agree, the dashed-dotted line shows the effect of choosing a fiducial model of which the power spectrum is wrong by 5 per cent.

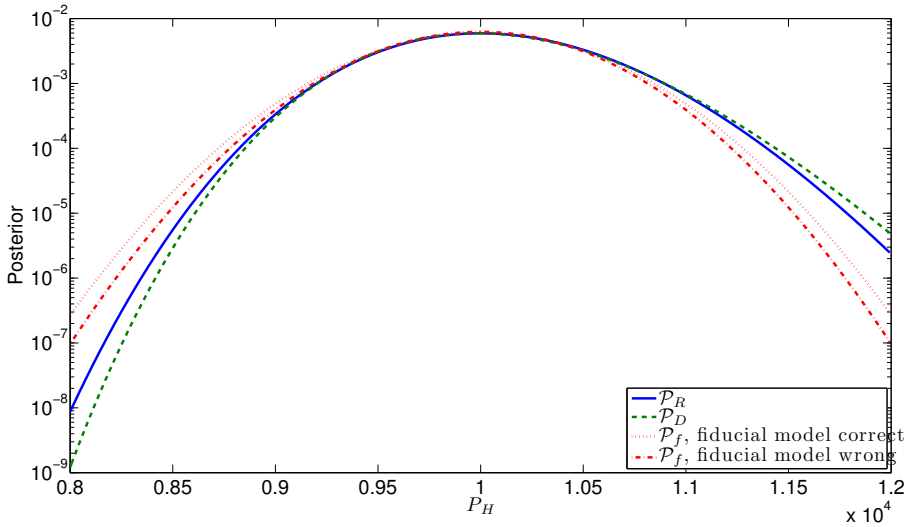

Figure 3.4: Same as the bottom right panel of Fig. 3.3, but with a logarithmic ordinate.

measured from $\widehat{|\delta_{\mathbf{k}}|}$. The logarithmic plot in Fig. 3.4 shows that the tails of all approximations disagree with the true posterior distribution. However, $\mathcal{P}_D$ is the Gaussian approximation that is closest to the truth. The next section shows how this affects inference from BOSS data.

3.3.5 Application to a Real Survey

We have seen that a Gaussian distribution for $P_H(\mathbf{k})$ is not a good approximation to the true Rayleigh distribution if the number of modes is small. In this section, we study whether this has an impact on a real survey. We will base our analysis on an analytic linear error for the power spectrum and errors, but use survey parameters for the data release 11 (DR11) of BOSS. For a real survey, we have to take into account that the discrete positions of the galaxies in a given survey are sampled from a continuous random field by a Poisson point process [109]. To take this sampling process into account, Eq. (3.26) becomes

$$\left\langle \delta_{\mathbf{k}_1} \delta^*_{\mathbf{k}_2} \right\rangle = \frac{(2\pi)^3}{V} \delta_D(\mathbf{k_1} - \mathbf{k_2}) \left[P(\mathbf{k_1}) + \bar{n}^{-1} \right], \tag{3.47}$$

and hence also

$$\mathcal{P}\left(P_H(\mathbf{k}) \left| \widehat{|\delta_{\mathbf{k}}|} \right. \right) \propto \frac{\widehat{|\delta_{\mathbf{k}}|} \exp\left(-\frac{\widehat{|\delta_{\mathbf{k}}|}^2}{P_H(\mathbf{k}) + \bar{n}^{-1}} \right)}{P_H(\mathbf{k}) + \bar{n}^{-1}}. \tag{3.48}$$

The average number density $\bar{n} = 2 \times 10^{-4}\ \frac{h^3}{\mathrm{Mpc}^3}$ was matched to the number of galaxies contained in the BOSS DR11 CMASS sample (690,826) and we also match its survey volume $V_S = 10\ \mathrm{Gpc}^3$ [129] assuming $h = 0.7$. For the covariance matrices of the Gaussians, we need to know the number of modes [109, 182]

$$M = V_n V_{\mathrm{eff}}(\mathbf{k}), \tag{3.49}$$

where

$$V_n \equiv \frac{k_n^2 \triangle k_n}{2\pi^2} \tag{3.50}$$

is the k-space-"volume" of the nth k-bin centred at k_n with width $\triangle k_n$, and

$$V_{\mathrm{eff}}(\mathbf{k}) \equiv V_S \left[\frac{\bar{n}P(\mathbf{k})}{1 + \bar{n}P(\mathbf{k})}\right]^2 \tag{3.51}$$

is the effective volume. The BOSS collaboration [129] calculated the power spectrum in Fourier modes averaged over bin widths of $\triangle k = 0.008h\ \mathrm{Mpc}^{-1}$. The values of the k-bin centres and their corresponding number of modes M are M=18, 180 and 500 in the three lowest k-bins centred at $k = 0.004, 0.012$ and $0.02\ \mathrm{Mpc}\ h^{-1}$. We model the measured power spectrum as $\widehat{P}(\mathbf{k}) = b^2 P_{\mathrm{lin}}(\mathbf{k})$, where $b = 1.87$ is the large-scale bias and $P_{\mathrm{lin}}(\mathbf{k})$ is a linear power spectrum produced by CAMB [22]. For the other measurement we take $\widehat{|\delta_{\mathbf{k}}|} = \sqrt{\widehat{P}(\mathbf{k}) + \bar{n}^{-1}}$. The resulting posterior distributions for the three lowest k-bins are plotted in Fig. 3.5. At the largest scales, i.e. $k = 0.004h\ \mathrm{Mpc}^{-1}$, neither $\mathcal{P}_D$ or $\mathcal{P}_f$ match $\mathcal{P}_R$. At $k = 0.012h\ \mathrm{Mpc}^{-1}$ and $k = 0.02h\ \mathrm{Mpc}^{-1}$ $\mathcal{P}_f$ and $\mathcal{P}_D$ become more similar, but neither of them features the asymmetric shape of $\mathcal{P}_R$. Additionally, $\mathcal{P}_f$ and $\mathcal{P}_D$ produce smaller error bars compared to $\mathcal{P}_R$. We can also numerically compare the distributions if we introduce the Kullback-Leibler (KL) divergence [183]. A distribution $\mathcal{P}_1$ is "better" than $\mathcal{P}_2$, if the loss of information due to approximating the true distribution with $\mathcal{P}_1$ is less than the same loss caused by using $\mathcal{P}_2$ as an approximation. If we use a pdf g to approximate another pdf f, a measure of the loss of information is given by the KL divergence

$$D_{\mathrm{KL}}\left(g||f\right) \equiv \int_{-\infty}^{\infty} \mathrm{d}x f(x) \ln\left(\frac{f(x)}{g(x)}\right), \tag{3.52}$$

which is, to some extent, a relative entropy (cf. Sec. 3.1). The KL divergences given in Tab. 3.1 tell us the same story as Fig. 3.5. The KL divergences of

Table 3.1: Kullback-Leibler divergences of the different approximations with respect to the true $\mathcal{P}_R$ at different scales k_n for BOSS DR11 CMASS.

$k_n \, \frac{\mathrm{Mpc}}{h}$	$D_{\mathrm{KL}}\left(\mathcal{P}_D \| \mathcal{P}_R\right)$	$D_{\mathrm{KL}}\left(\mathcal{P}_f \| \mathcal{P}_R\right)$	$D_{\mathrm{KL}}\left(\mathcal{P}_f^{\mathrm{wrong}} \| \mathcal{P}_R\right)$
0.004	0.021	0.383	0.336
0.012	0.005	0.034	0.037
0.02	0.003	0.014	0.020

the Gaussian approximation with a varying covariance $\mathcal{P}_D$ is at all scales less than the KL divergence of $\mathcal{P}_f$, i.e. $\mathcal{P}_D$ is a better approximation to the true $\mathcal{P}_R$. On the downside, its best fit has an offset with respect to $\mathcal{P}_R$. We will therefore investigate alternative posterior shapes in the next section, which are not Gaussian and more accurately approximate the true posterior distribution.

3.4 Studying Alternative Posterior Shapes

We have seen in the previous sections that the true posterior distribution $\mathcal{P}_R$ is not well approximated by either $\mathcal{P}_f$ or $\mathcal{P}_D$ if the number of independent modes is low, which is the case at large scales, i.e. the small values of k where most of the f_{NL}-signal comes from. A similar problem arises when cosmological models are fitted to cosmic microwave background (CMB) power spectra, which are Wishart distributed. Bond, Jaffe and Knox [184]; Smith, Challinor and Rocha [185]; Percival and Brown [176]; as well as Hamimeche and Lewis [180] have studied alternative distribution shapes that approximate the Wishart distribution. Taking a similar approach to Verde *et al.* [177] and Percival and Brown [176], we expand the natural logarithm of Eq. (3.34) around the maximum $P_H(\mathbf{k}) \equiv (1+\varepsilon)\widehat{|\delta_{\mathbf{k}}|}^2$:[1]

$$-2\ln\left(\mathcal{P}_R\right) = 2M\left(\frac{\varepsilon^2}{2} - \frac{2\varepsilon^3}{3} + \frac{3\varepsilon^4}{4} + \mathcal{O}\left(\varepsilon^5\right)\right) + \text{const.} \tag{3.53}$$

This equation agrees to third order with the Taylor expansions of the logarithms of the following distributions:

[1]For realistic, noisy measurements of $\widehat{|\delta_{\mathbf{k}}|}$ and $\widehat{P}(\mathbf{k})$, $P_H(\mathbf{k})$ has to be replaced by $P_H(\mathbf{k}) + \bar{n}^{-1}$ everywhere in this section. For simplicity, we do not write the noise explicitly.

- the inverse cubic normal (ICN) distribution [185]

$$-2\ln(\mathcal{P}_{\mathrm{ICN}}) = 18\tilde{C}_{\mathbf{k}}^{-1}\left[\widehat{P}(\mathbf{k}) - \widehat{P}(\mathbf{k})^{4/3}P_H(\mathbf{k})^{-1/3}\right]^2, \tag{3.54}$$

- the offset log-normal (OLN) distribution

$$-2\ln(\mathcal{P}_{\mathrm{OLN}}) = 2(1+a)\tilde{C}_{\mathbf{k}}^{-1}\left[\widehat{P}(\mathbf{k})\ln\left(\frac{P_H(\mathbf{k})+a\widehat{P}(\mathbf{k})}{\widehat{P}(\mathbf{k})+a\widehat{P}(\mathbf{k})}\right)\right]^2 \tag{3.55}$$

  if $a = -1/4$,

- and combinations of any of the distributions given in chapter 5.1 of [176].

We can see from Fig. 3.6 that the 3rd order diverges for large values of the model power spectrum P_H. Hence the optimal free parameter a might differ from $a = -1/4$. Therefore, we use the KL divergence to optimise a in the offset log-normal distribution $\mathcal{P}_{\mathrm{OLN}}$. It can be found to be $a = -0.201$ at $k = 0.004\frac{\mathrm{Mpc}}{h}$, $a = -0.240$ at $k = 0.012\frac{\mathrm{Mpc}}{h}$ and $a = -0.242$ at higher values of k. $\mathcal{P}_{\mathrm{OLN}}$ peaks at the maximum of the true distribution $\mathcal{P}_R$ and it approximates the tails of the true distribution a bit better than the Gaussian approximations, but as Fig. 3.5 shows, it is still obviously different from $\mathcal{P}_R$.

The ICN distribution [185] fits the true distribution better. Fig. 3.6 shows a remarkable agreement between $\mathcal{P}_R$ and $\mathcal{P}_{\mathrm{ICN}}$. Writing both $-2\ln(\mathcal{P}_R)$ and $-2\ln(\mathcal{P}_{\mathrm{ICN}})$ as Taylor series, we see that their Taylor coefficients are equal for $k \leq 3$ and approximately equal for much higher orders, as shown in the next Section 3.5.

3.5 Comparison of "true" and ICN Posterior Taylor Series

In this section, we compare the Taylor Series of $\mathcal{P}_R$ and $\mathcal{P}_{\mathrm{ICN}}$ to explain why they are so similar. We write the hypothetical power spectrum $P_H \equiv (1+\varepsilon)\widehat{|\delta_{\mathbf{k}}|}^2$ as a perturbation around the measured power. The Rayleigh posterior hence becomes

$$-2\ln(\mathcal{P}_R) = 2\ln(1+\varepsilon) + \frac{2}{1+\varepsilon}. \tag{3.56}$$

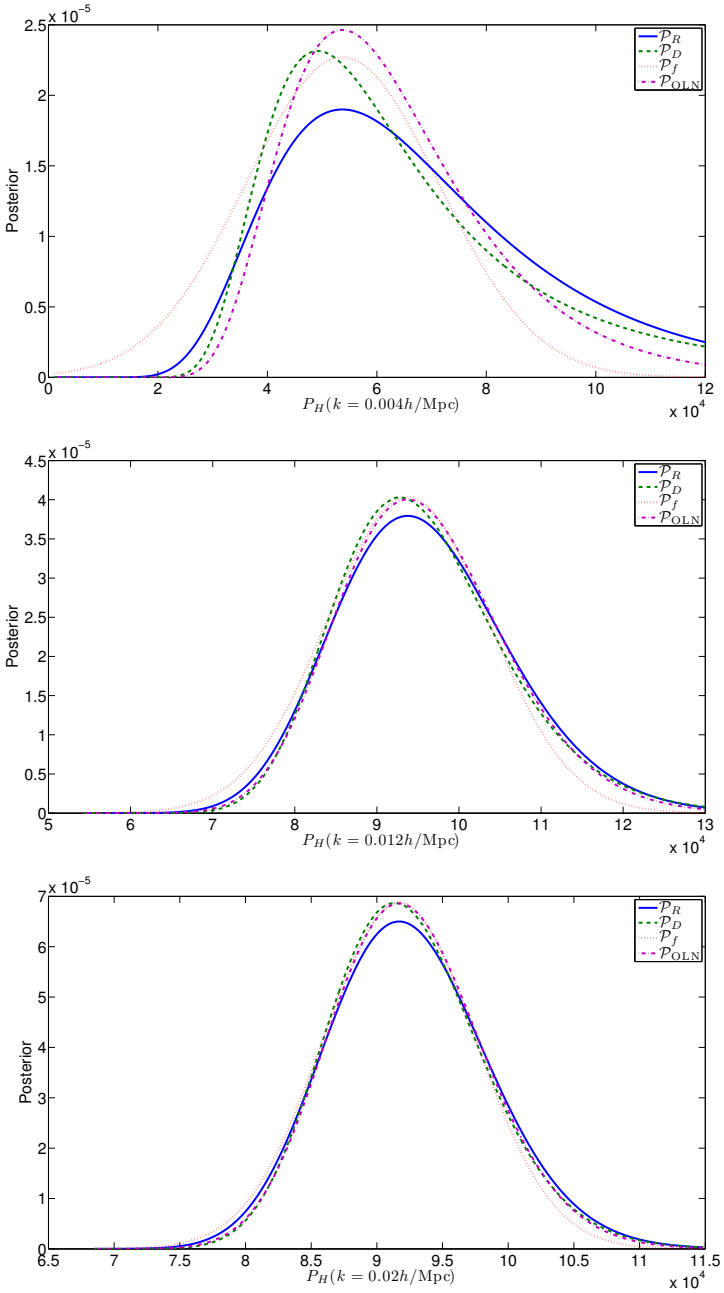

Figure 3.5: Posterior distribution functions of the hypothetical power spectrum $P_H(\mathbf{k})$ for the three lowest $\mathbf{k}$-bins of BOSS DR11 CMASS. The colour coding is the same as in Fig. 3.3, with the addition of the offset log-normal (OLN) posterior distribution plotted in magenta.

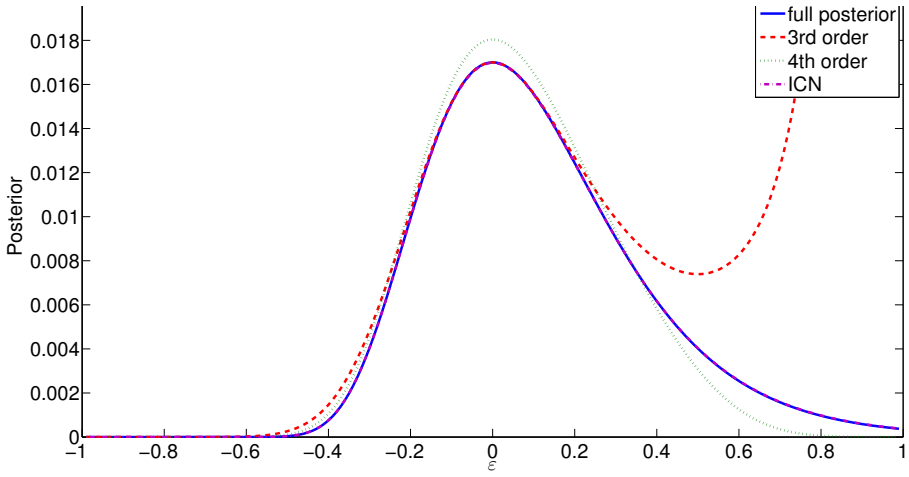

Figure 3.6: Third and fourth order Taylor expansion to the true posterior shape $\mathcal{P}_R$ with $M = 20$ modes. The x-axis is a perturbation $\varepsilon \equiv P_H/\widehat{P} - 1$ of the model power spectrum P_H around the average recovered best-fit value $\widehat{P}$. As the third order approximation is not normalisable, the normalisation has been chosen such that it agrees with the 4^{th} order at the maximum. The true posterior shape agrees very well with the inverse cubic normal posterior shape.

Ignoring the irrelevant zero order contribution, the Taylor series reads

$$-2\ln(\mathcal{P}_R) = 2\sum_{\kappa=1}^{\infty}(-1)^{\kappa}\varepsilon^{\kappa}\frac{\kappa-1}{\kappa}. \tag{3.57}$$

Similarly, the ICN distribution in terms of ε is given by

$$\begin{aligned}-2\ln(\mathcal{P}_{\rm ICN}) &= 9\left[1-(1+\varepsilon)^{-1/3}\right]^2 \\ &= 9\left[1-2(1+\varepsilon)^{-1/3}+(1+\varepsilon)^{-2/3}\right].\end{aligned} \tag{3.58}$$

We make use of the generalised binomial series $(1+\varepsilon)^{\alpha} = \sum_{\kappa=0}^{\infty}\binom{\alpha}{\kappa}\varepsilon^{\kappa}$, where $\binom{\alpha}{\kappa} \equiv \frac{\Gamma(\alpha+1)}{\Gamma(\kappa+1)\Gamma(\alpha-\kappa+1)}$ is the generalised binomial coefficient, and obtain the series

$$-2\ln(\mathcal{P}_{\rm ICN}) = 9\sum_{\kappa=1}^{\infty}\varepsilon^{\kappa}\left[\binom{-\frac{2}{3}}{\kappa}-2\binom{-\frac{1}{3}}{\kappa}\right]. \tag{3.59}$$

Again, we have ignored irrelevant constant terms. The negative entries in the binomial coefficients can be removed using $\binom{\alpha}{\kappa} = (-1)^{\kappa}\binom{\kappa-\alpha-1}{\kappa}$:

$$-2\ln(\mathcal{P}_{\rm ICN}) = 9\sum_{\kappa=1}^{\infty}(-1)^{\kappa}\varepsilon^{\kappa}\left[\binom{\kappa-\frac{1}{3}}{\kappa}-2\binom{\kappa-\frac{2}{3}}{\kappa}\right]. \tag{3.60}$$

If we insert values for $\kappa \leq 3$, we find the equality

$$2\frac{\kappa-1}{\kappa} = 9\left[\binom{\kappa-\frac{1}{3}}{\kappa}-2\binom{\kappa-\frac{2}{3}}{\kappa}\right]. \tag{3.61}$$

Thus, $\mathcal{P}_R$ and $\mathcal{P}_{\rm ICN}$ are the same to third order. What is even more striking is that for larger κ,the approximation

$$2\frac{\kappa-1}{\kappa} \approx 9\left[\binom{\kappa-\frac{1}{3}}{\kappa}-2\binom{\kappa-\frac{2}{3}}{\kappa}\right]. \tag{3.62}$$

still holds. The numerical values of the first 20 coefficients can be compared in Tab. 3.2. For $\kappa < 17$, the two sides differ by less than 20%. Therefore, the agreement between $\mathcal{P}_R$ and $\mathcal{P}_{\rm ICN}$ is high. Due to this good agreement, we study in the next section how the ICN distribution would perform in $f_{\rm NL}$-analyses using BOSS DR11 and Euclid data.

3.6 The Effect on Primordial Non-Gaussianity Measurements

In this section, we test the effect of using different posterior distribution shapes on the inference of a real observable. The largest deviations between the posteriors are at small k and we would therefore expect the largest effects for

Table 3.2: Comparison of the numerical values of the Taylor coefficients of the "true" pdf $\mathcal{P}_R$ and the ICN approximation.

κ	$2\frac{\kappa-1}{\kappa}$	$9\left[\binom{\kappa-\frac{1}{3}}{\kappa}-2\binom{\kappa-\frac{2}{3}}{\kappa}\right]$
1	0.	0.
2	1.	1.
3	1.33	1.33
4	1.5	1.48
5	1.6	1.56
6	1.67	1.59
7	1.71	1.61
8	1.75	1.62
9	1.78	1.62
10	1.8	1.62
11	1.82	1.61
12	1.83	1.61
13	1.85	1.60
14	1.86	1.59
15	1.87	1.58
16	1.88	1.57
17	1.88	1.56
18	1.89	1.55
19	1.89	1.54
20	1.9	1.53

parameters dependent on these modes. At these scales, (local) primordial non-Gaussianity alters the biasing law between dark-matter halos and the underlying mass-density field (c.f. Sec. 1.5.2, [1, 2, 100, 101, 102, 186, 187, 188]) The resulting alteration of the bias is given by Eq. (1.189).

Here we work to first order in $\delta_{\mathbf{k}}$, so that we can continue to assume that $\delta_{\mathbf{k}}$ is drawn from a Gaussian distribution, with an altered $P(\mathbf{k})$, i.e. the first order effect of non-Gaussianity is to $P_H(\mathbf{k})$, keeping the distribution the same. Furthermore, we do not alter V_n (Eq. 3.50) to include any coupling between modes from the non-Gaussian signal. Where k is very small, higher order corrections to $\delta_{\mathbf{k}}$ will become increasingly important (e.g. [189]), suggesting that the Gaussian limit for $\delta_{\mathbf{k}}$ will break down here.

3.6.1 Comparison to BOSS Results

We use BOSS DR9 parameters and the same CAMB linear matter power spectrum as [6]. We also assume $\delta_c = \frac{1.686}{D(z)}$ as expected from the spherical collapse model in an Einstein-de Sitter universe and a flat prior for $f_{\rm NL}$. We plot $f_{\rm NL}$ posterior functions in Fig. 3.7, assuming a measurement of a power spectrum with underlying $f_{\rm NL} = 0$. $\mathcal{P}_f$ is not symmetric, as both a linear and a quadratic term of $f_{\rm NL}$ enter the power spectrum. The inverse cubic normal distribution agrees again very well with $\mathcal{P}_R$. $\mathcal{P}_R$, $\mathcal{P}_f$ and $\mathcal{P}_{\rm ICN}$ reproduce the true value as their best fit estimate. Using $\mathcal{P}_D$, the most likely value of $f_{\rm NL}$ is $f_{\rm NL} = -25.5$ considering the same k-bins as [6] in their analysis of DR9 BOSS data, i.e. $0.004\frac{h}{\rm Mpc} \leq k \leq 0.05\frac{h}{\rm Mpc}$.

One has to keep in mind that there are different definitions of the measured value. The commonly published value is the posterior mean $\langle f_{\rm NL}\rangle$, due to the fact that if $f_{\rm NL}$ is fitted as part of a longer list of cosmological parameters, one has to rely on Markov chain Monte Carlo techniques (e.g. [160], cf. Sec. 3.1). In general, such techniques cannot provide accurate estimates of the best-fit value. Hence, data analysis papers more often present $\langle f_{\rm NL}\rangle$ as their results. If the posterior is asymmetric, the best fit and posterior mean do not agree. Given a flat $f_{\rm NL}$-prior, we expect $f_{\rm NL} = 11.4$ using $\mathcal{P}_R$. Based on our arguments in Sec. 3.3.1 and 3.6, we think of the mean of $\mathcal{P}_R$ as the correct estimate of $f_{\rm NL}$. This seems counter-intuitive because our input was that we measure a power spectrum which corresponds to $f_{\rm NL} = 0$. There are two different explanations

Table 3.3: $f_{\rm NL}$-postdictions of the best fit $f_{\rm NL}^{\rm (BF)}$ and marginalised best fits $\langle f_{\rm NL}\rangle$, as well as its 95% confidence interval, for BOSS DR9 using different shapes of the posterior distribution.

posterior	$f_{\rm NL}^{\rm (BF)}$	$\langle f_{\rm NL}\rangle$	95% confidence interval
$\mathcal{P}_R$	0	11.4	-71.5< $f_{\rm NL}$ <100.7
$\mathcal{P}_D$	-25.5	-11.9	-68.2< $f_{\rm NL}$ <53.4
$\mathcal{P}_f$	0	-7.7	-90.9< $f_{\rm NL}$ <71.0
$\mathcal{P}_{\rm ICN}$	0	11.4	-71.9< $f_{\rm NL}$ <101.2

for that, one for the "true" distribution and good approximations to it, such as ICN, and another explanation for the Gaussian approximations. The power spectrum is strictly non-negative, to wit there is a hard lower limit on the probability distribution of the power spectrum, but no upper one, so there is a higher chance of measuring a larger power spectrum value, and hence $f_{\rm NL}$-value, than a lower one. Therefore, one finds a positive mean (cf. Tab. 3.3) when marginalising over all possible values of $f_{\rm NL}$. In the presence of more informative likelihoods, this mean gets closer to zero, as we shall see in the next subsection where we do the same analysis using Euclid. Contrarily, in the case of a symmetric distribution of the power, one has to keep in mind that this does not translate into a symmetric distribution of $f_{\rm NL}$. Due to primordial non-Gaussianity altering the power spectrum both linearly and quadratically, the fixed Gaussian log-likelihood $-\ln(L_f) \propto 2b_0(b_0-1)f_{\rm NL}A(k)+(b_0-1)^2 f_{\rm NL}^2 A^2(k)$ contains these terms as well, making negative values of $f_{\rm NL}$ more likely than positive ones and yielding a negative mean of $\langle f_{\rm NL}\rangle = -7.7$ (cf. Tab. 3.3). This effect is even amplified by the model-dependent covariance matrix in $\mathcal{P}_D$, which brings the mean down to $\langle f_{\rm NL}\rangle = -11.9$. This is another point against using Gaussian posterior approximations in $f_{\rm NL}$ inferences.

The choice of the posterior distribution also affects the error estimation. If we use $\mathcal{P}_R$ or $\mathcal{P}_{\rm ICN}$, the length of our postdicted 95% $f_{\rm NL}$-confidence interval (C.I., cf. Tab. 3.3) is similar to the length of Ross *et al.*'s most naïve case ii 95% C.I., i.e. $32 < f_{\rm NL} < 198$ [6]. Approximating the "true" posterior using a Gaussian distribution leads to an underestimation of the error.

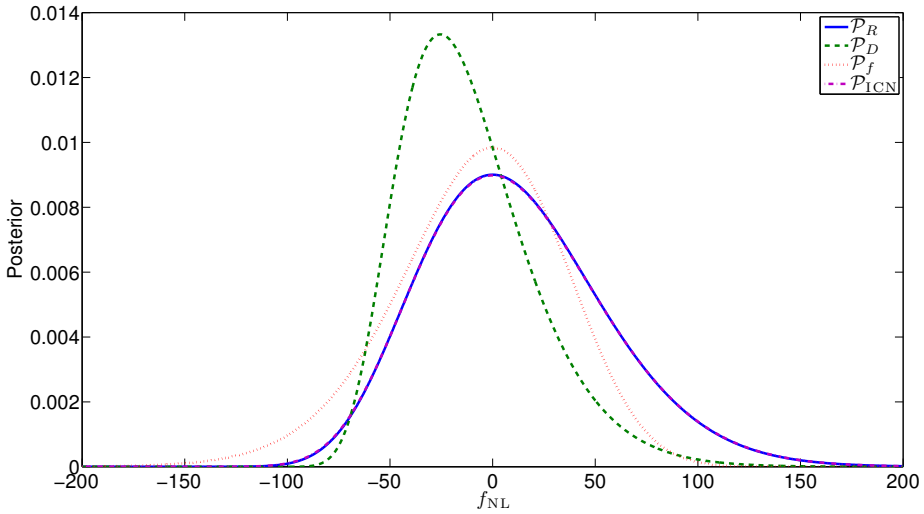

Figure 3.7: Analytic $f_{\rm NL}$-posterior functions for a BOSS like survey combining all k-bins.

3.6.2 Euclid Predictions

We make similar predictions for the Euclid survey [137]. We assume bias values $b(z) = \sqrt{1+z}$, matched to simulations of [190] and also assumed in [191], and number densities $\bar{n}(z)$ predicted for Euclid by Pozzetti *et al.* [192], and a survey covering 15000 square degrees. We generate CAMB matter power spectra $P(\mathbf{k}, z)$ for the redshift range $0.9 < z < 1.74$. Note that the aim of [4], and hence this chapter, is to test how the use of different posterior shapes influences cosmological measurements, but not primarily to make $f_{\rm NL}$-predictions. More rigorous predictions for Euclid $f_{\rm NL}$-measurements can be found e.g. in [137, 138, 140] or [133]. These studies also include 3-point statistics, weak lensing tomography, measurements of the integrated Sachs-Wolfe effect and/or the use of the multitracer technique. Their constraints are therefore tighter than ours.

As Euclid will probe a much larger volume, it will accommodate many more k-modes and hence we see good agreement of $\mathcal{P}_f$ with $\mathcal{P}_R$ in Fig. 3.8. As against our results in Sec. 3.3.5, fixing the covariance provides better $f_{\rm NL}$ results than the inferences from a posterior with varying covariance. However, $\mathcal{P}_{\rm ICN}$ is still the best approximation and accurately reproduces the marginalised $f_{\rm NL}$-value of $\mathcal{P}_R$ and its 95% C.I., whereas using $\mathcal{P}_f$ yields the correct width of the 95% C.I., but its position and the marginalised value have an offset of 0.38

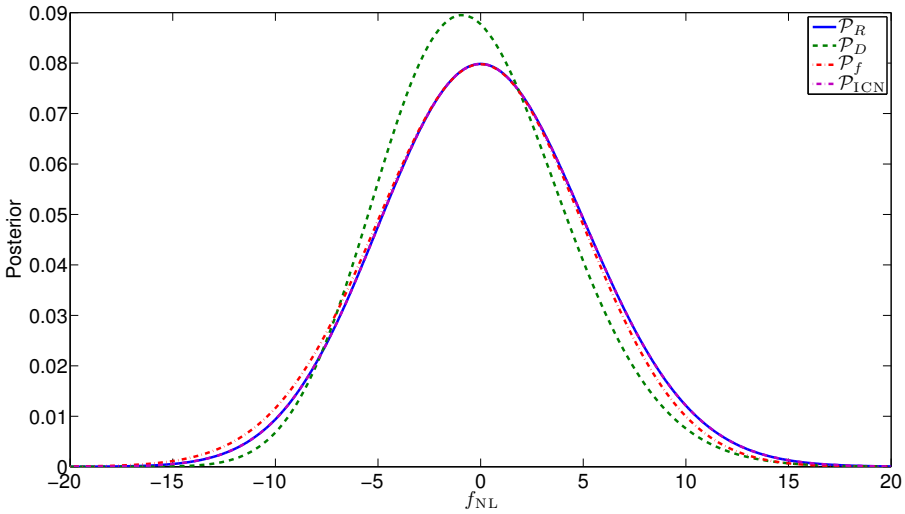

Figure 3.8: Analytic $f_{\rm NL}$-posterior functions for a Euclid like survey combining all k-bins.

Table 3.4: $f_{\rm NL}$-predictions similar to Tab. 3.3, but for Euclid.

posterior	$f_{\rm NL}^{\rm (BF)}$	$\langle f_{\rm NL} \rangle$	95% confidence interval
$\mathcal{P}_R$	0	0.24	-9.0< $f_{\rm NL}$ <9.4
$\mathcal{P}_D$	-1.0	-0.30	-8.4< $f_{\rm NL}$ <7.8
$\mathcal{P}_f$	0	-0.14	-9.4< $f_{\rm NL}$ <9.0
$\mathcal{P}_{\rm ICN}$	0	0.24	-9.0< $f_{\rm NL}$ <9.4

(cf. Tab. 3.4). We therefore still recommend either using $\mathcal{P}_{\rm ICN}$ or $\mathcal{P}_R$ when cosmological models are fitted to power spectra from galaxy surveys even as large as Euclid.

3.7 Conclusions

We have studied different posterior shapes that can be used in the fitting process of cosmological models to power spectra from galaxy surveys. As the underlying matter density field is at least approximately Gaussian, we assume that the true posterior distribution $\mathcal{P}_R$ is based on a Rayleigh likelihood distribution in δ. Assuming Gaussian posteriors in $P(\mathbf{k})$, be it with a fixed or a varying covariance matrix, does not approximate $\mathcal{P}_R$ well and yields biased

best-fit values and wrong error estimates especially on large scales where statistics are not good enough to make use of the central limit theorem.

If one confines oneself to use Gaussian posterior shapes, it depends on the parameter one wants to constrain whether a fixed or varying covariance matrix provides more accurate results. We found that the posterior shape $\mathcal{P}_D$ with varying covariance follows $\mathcal{P}_R$ closer than $\mathcal{P}_f$ with a fixed covariance when the power spectrum P_H (or any parameter linear in the power spectrum) is fitted to the power spectrum $\widehat{P}$, but when $f_{\rm NL}$ is fitted to $\widehat{P}$ it is the other way round.

Due to these reasons, we advise against using Gaussian posterior distributions. Instead, we have found that posterior distributions, such as the inverse cubic normal distribution $\mathcal{P}_{\rm ICN}$ (cf. Eq. 3.54) or applying Hamimeche and Lewis's method [180] to $\mathcal{P}_R$ (cf. Eq. 3.39), provide simple, more accurate alternatives. They confidently reproduce the correct width of the 95 % confidence intervals in our simplified predictions of $f_{\rm NL}$-measurements. However, in our analysis we have not accounted for the effect of the survey window function, or the effect of non-Gaussianity of the density field on the posterior shape. Therefore, the final decision about which posterior is the best to use should be done after testing these methods against simulations which account for the non-linear effects that we have ignored for simplicity in our analytic calculations. We leave this for future work.

A major advantage of the non-Gaussian posteriors presented in this chapter, is the fact that their covariance matrices do not depend on the power spectrum of the model to be tested. The estimation of covariance matrices is a critical and computationally expensive step in the data analysis.

Chapter 4

Unbiased Contaminant Removal for 3D Galaxy Power Spectrum Measurements

The other, and more crucial, big issue when measuring $f_{\rm NL}$ from the galaxy power spectrum, is that the power spectrum at large scales is contaminated by non-cosmic sources, such as foreground stars and variations of the performance of the telescope across the sky. The traditional way of using weights to ease the effect of such contaminants has been described in Sec. 2.2 and can result in biased estimates of the power spectrum [193]. An unbiased and optimal power spectrum estimator is computationally unfeasible. In the article [5], Will Percival, David Bacon, Lado Samushia and I introduced a technique that can mitigate the contamination quickly, taking a slight sub-optimality as a trade-off.

Our model and assumptions are introduced in Sec. 4.1. In Sec. 2.4.2, I introduced the FKP estimator. In Sec. 4.2, I present the Quadratic Maximum Likelihood (QML) estimator and show how the FKP estimator is an approximation to this lossless power spectrum estimator. I introduce the systematics removal techniques, *mode deprojection* and *mode subtraction*, in sections 4.3 and 4.4, respectively, and show that before normalisation their resulting power spectra are the same, also when extended to multiple contaminants. I introduce a new normalisation factor in Sec. 4.5 for a single contaminant and compare it to the normalisation of the quadratic maximum likelihood (QML)

estimator of [194]. This derivation is also extended to allow for a non-diagonal covariance. I show that one can apply our methodology also to multiple contaminants in Sec. 4.6 and test the different methods on simulations in Sec. 4.7. The chapter is summarised in Sec. 4.8.

4.1 A Mathematical Model of Contamination

The observed galaxy field can be contaminated with fluctuations of non-cosmological origin, such as variations due to the galactic extinction and the stellar density. Often the contaminants are not known exactly (e.g. we may know the shape of the spurious mode but may not know its exact amplitude) which makes their exact removal impossible. These modes have the potential to strongly bias cosmological constraints derived from the clustering measurements, so we need to correct or suppress these misleading modes in a responsible way.

I start with introducing the basic mathematical problem that we wish to solve and introduce the main methods of removing contaminants discussed in literature. We assume that we have measured the galaxy density field as real numbers in configuration space, which we (fast) Fourier transform to obtain a density field $F(\mathbf{k})$, which is Hermitian because it is the Fourier transform of a real field. Furthermore, we assume that the contamination can be described by another Hermitian field $f(\mathbf{k})$, such that the true density field is given by

$$D(\mathbf{k}) = F(\mathbf{k}) - \varepsilon_{\rm true} f(\mathbf{k}), \tag{4.1}$$

with the true amplitude of the contaminant $\varepsilon_{\rm true}$ unknown. In cases with multiple contaminants (which we label with capital Latin indices), we extend Eq. (4.1) to

$$D(\mathbf{k}) = F(\mathbf{k}) - \sum_A \boldsymbol{\varepsilon}_A^{(\rm true)} f_A(\mathbf{k}). \tag{4.2}$$

Furthermore, we assume that $F(\mathbf{k})$ and $f(\mathbf{k})$ are uncorrelated, which is a valid assumption for most sources of systematics since they originate from our Galaxy or due to telescope effects. Large scale surveys will reduce the current sample variance limitation on the power spectrum on scales where the systematic errors have a significant impact. As a consequence, having control

of these systematics is a key requirement to provide accurate cosmological measurements.

In order to investigate techniques for estimating the power spectrum in the presence of contaminants, we separate the process into three separate stages: (i) removing the contaminant signal, (ii) estimating the uncontaminated cosmological power spectrum, (iii) debiasing the resulting estimates. Two techniques are in common usage for removing the contaminant signal (i): The first is *mode subtraction* (cf. Sec. 4.4 and Sec. 4.5), where contaminants are removed by fitting the amplitude ε of the contaminant field $f(\mathbf{k})$ to the data and simply subtracted off from $F(\mathbf{k})$. The second is *mode deprojection* (cf. [195, 196] and Sec. 4.3), which is based on assigning infinitely large covariances to contaminated modes, thus removing them from any analysis. In our nomenclature, a mode is a linear combination of Fourier modes rather than a single $\mathbf{k}$-mode. This is reflected in the naming of *mode subtraction* and *mode deprojection*. This choice of names shall distinguish the *mode subtraction* technique from a third technique for removing the contaminant signal, called *template subtraction*, where the observed power spectra are corrected using best-fit amplitudes derived via cross-correlations between the data and the templates. This is done on the power spectrum level. As not all modes entering the power spectrum for a given multipole ℓ or wave number $\mathbf{k}$ are affected equally by a certain contaminant, this is a sub-optimal way of mitigating the contaminant. Furthermore, Elsner *et al.* [193] have shown that this method provides a biased estimate of the power and I do not consider it further. For (ii), the power spectrum P(k) is commonly estimated by the FKP estimator [109], which is an approximation to the Quadratic Maximum Likelihood (QML) estimator [194]. As well as being optimal, the QML estimator has the advantage of producing unbiased power spectrum estimates. However, when applying this methodology to data with N_{mode} modes, one has to calculate and invert an $N_{\mathrm{mode}} \times N_{\mathrm{mode}}$ matrix. The complexity of the inversion of such a matrix scales with $\mathcal{O}(N_{\mathrm{mode}}^3)$ for straightforward inversion algorithms and it can be shown that there is a theoretical lower bound for the complexity of $\mathcal{O}(N_{\mathrm{mode}}^2 \log(N_{\mathrm{mode}}))$ [197]. On top of that an overall $N_{\mathrm{bin}} \times N_{\mathrm{bin}}$ normalisation matrix has to be computed after binning the data into N_{bin} bins, which makes the application of this methodology unfeasible for future surveys with increased number of modes N_{mode}. In [5], we

suggested a modified FKP-style *mode subtraction* approach. We show that this technique can be made unbiased and, on a mode-by-mode basis, is mathematically identical to *mode deprojection*. The FKP estimator with debiased *mode subtraction* is not optimal in that it discards more information than the full QML estimator, but we expect that, in realistic cases, this loss of information will be small.

4.2 Power spectrum estimators

In this section, I review the *quadratic maximum likelihood* (QML) estimator [194] and show how it reduces to the simpler FKP estimator [109] in the limit of uncorrelated modes with equal noise per mode in each bin. Even without considering any contaminants, the FKP estimator is easier to implement and is used in most recent analyses of large-scale structure, while the QML estimator is optimal but difficult to implement especially on smaller scales.

To introduce the QML concept separately from the problem of error mitigation, I first consider $F(\mathbf{k})$ in this subsection to be uncontaminated. Assuming, as in the previous chapter, that the density field is Gaussian, its log-likelihood reads

$$-2\ln\mathcal{L} = \ln\left(\det\mathbf{C}\right) + \sum_{\alpha\beta} F^*(\mathbf{k}_\alpha)\mathbf{C}^{-1}_{\alpha\beta}F(\mathbf{k}_\beta). \tag{4.3}$$

Note that $\mathbf{C}$ is the mode-by-mode covariance matrix of the over-density field here. The binned power spectrum is given by the average of the diagonal entries of the full mode-by-mode density covariance that fall into the bin $\mathbb{k}_i$:

$$P(k_i) \equiv \frac{1}{N_{\mathbb{k}_i}}\sum_{\alpha\beta}\delta_{\alpha\beta}\Theta(\mathbf{k}_\alpha \in \mathbb{k}_i)\mathbf{C}_{\alpha\beta} + P_{\rm shot} \equiv \frac{1}{N_{\mathbb{k}_i}}\sum_{\alpha\beta}\delta_{\alpha\beta}\Theta_{\alpha i}\mathbf{C}_{\alpha\beta} + P_{\rm shot}. \tag{4.4}$$

Throughout this chapter I use the following notation: Fourier modes $\mathbf{k}_\alpha$ are written in bold and are labelled with Greek indices, bins are denoted in blackboard bold $\mathbb{k}_i$ and have lower case Latin indices. The effective k-value of a bin is written in italic with the same index as the bin. In the remainder of this Chapter, I ignore the shot noise term $P_{\rm shot}$ (cf. Eq. 2.27) to keep equations simpler. As it is only a constant offset, the equations discussed here are still valid also for real applications if this constant offset is added back, as will be done in Chapter 5. The covariance is quadratic in the over-density field,

hence, it seems reasonable to make a quadratic ansatz for a power spectrum estimator

$$\widehat{P}(k_i) = \sum_{\alpha\beta} F^*(\mathbf{k}_\alpha)\bar{\mathbf{E}}_{\alpha\beta}(k_i)F(\mathbf{k}_\beta), \tag{4.5}$$

where $\bar{\mathbf{E}}$ is a matrix that weighs each combination of modes such that the power spectrum estimate is optimal, i.e. its variance is minimal. As the estimator should also be **unbiased**, it is convenient to split

$$\bar{\mathbf{E}}_{\alpha\beta}(k_i) \equiv \sum_j \mathbf{N}^{-1}_{ij}\mathbf{E}_{\alpha\beta}(k_j) \tag{4.6}$$

into a renormalised set of weights $\mathbf{E}_{\alpha\beta}(k_j)$ and another matrix $\mathbf{N}_{ij}$ that will ensure that power spectrum estimates are unbiased. In this new notation, the power spectrum estimator reads

$$\widehat{P}(k_i) = \sum_j \mathbf{N}^{-1}_{ij} \sum_{\alpha\beta} F^*(\mathbf{k}_\alpha)\mathbf{E}_{\alpha\beta}(k_j)F(\mathbf{k}_\beta) \equiv \sum_j \mathbf{N}^{-1}_{ij}\mathbf{p}_j, \tag{4.7}$$

where I have defined

$$\mathbf{p}_j \equiv \sum_{\alpha,\beta} F^*(\mathbf{k}_\alpha)\mathbf{E}_{\alpha\beta}(k_j)F(\mathbf{k}_\beta). \tag{4.8}$$

With $\langle F^*(\mathbf{k}_\alpha)F(\mathbf{k}_\beta)\rangle = \mathbf{C}_{\alpha\beta}$ the expectation value of Eq. (4.7) is given by

$$\left\langle \widehat{P}(k_i) \right\rangle = \sum_j \mathbf{N}^{-1}_{ij} \sum_{\alpha\beta} \mathbf{E}_{\alpha\beta}(k_j)\mathbf{C}_{\alpha\beta}. \tag{4.9}$$

Thus, in order for the estimator to be unbiased, i.e. $\left\langle \widehat{P}(k_i) \right\rangle = P(k_i)$, we need

$$\sum_j \mathbf{N}^{-1}_{ij}\mathbf{E}_{\alpha\beta}(k_j) = \frac{1}{N_{\mathbb{k}_i}}\delta_{\alpha\beta}\Theta_{\alpha i} = \frac{\partial P(k_i)}{\partial \mathbf{C}_{\alpha\beta}}, \tag{4.10}$$

where the last equality is the inverse derivative of Eq. (4.4) that allows using matrix notation. This equation is solved by

$$\mathbf{N}_{ij} = \mathrm{tr}\left\{ \mathbf{E}(k_i)\frac{\partial \mathbf{C}}{\partial P(k_j)} \right\}. \tag{4.11}$$

Note that, in order for this equation to make sense, one has to make an assumption of a prior covariance matrix $\mathbf{C}$.

To obtain the **optimal** quadratic maximum likelihood (QML) estimator [182], one has to minimise the covariance of the power spectrum estimates,

which is denoted by the letter $\mathbf{V}$ to avoid confusion with the covariance matrix of the over-density field:

$$\mathbf{V}_{ij} \equiv \left\langle \widehat{P}(k_i)\widehat{P}(k_j) \right\rangle - \left\langle \widehat{P}(k_i) \right\rangle \left\langle \widehat{P}(k_j) \right\rangle . \tag{4.12}$$

After inserting Eq. (4.5), one can make use of Wick's theorem to reduce the 4-point correlator of F in the power covariance. As $\langle F^*(\mathbf{k}_\alpha)F(\mathbf{k}_\beta)\rangle = \mathbf{C}_{\alpha\beta}$, the power covariance reads

$$\mathbf{V}_{ij} = \bar{\mathbf{E}}_{\alpha\beta}(k_i)\bar{\mathbf{E}}_{\gamma\delta}(k_j) \left[\mathbf{C}_{\alpha\gamma}\mathbf{C}_{\beta\delta} + \mathbf{C}_{\alpha\delta}\mathbf{C}_{\beta\gamma}\right] . \tag{4.13}$$

Both $\bar{\mathbf{E}}$ and $\mathbf{C}$ are symmetric matrices, so after relabeling $\alpha \leftrightarrow \beta$ in the first sum, one can see that both terms are actually equal and identify the trace

$$\mathbf{V}_{ij} = 2\,\mathrm{tr}\left[\mathbf{C}\bar{\mathbf{E}}(k_i)\mathbf{C}\bar{\mathbf{E}}(k_j)\right] . \tag{4.14}$$

To minimise this variance subject to the unbiasedness condition of Eq. (4.11), one can introduce the Lagrange multiplier λ and minimise

$$\mathrm{tr}\left[\mathbf{C}\bar{\mathbf{E}}(k_i)\mathbf{C}\bar{\mathbf{E}}(k_i) - \lambda\left(\bar{\mathbf{E}}(k_i)\frac{\partial \mathbf{C}}{\partial P(k_i)} - 1\right)\right] , \tag{4.15}$$

which yields

$$2\mathbf{C}\bar{\mathbf{E}}(k_i)\mathbf{C} - \lambda\frac{\partial \mathbf{C}}{\partial P(k_i)} = 0. \tag{4.16}$$

Thus, the estimator matrix

$$\bar{\mathbf{E}}(k_i) = \frac{\lambda}{2}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial P(k_i)}\mathbf{C}^{-1} \tag{4.17}$$

optimises the power spectrum estimation. Going back to the unbarred matrix, we then have

$$\mathbf{E}(k_i) = \mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial P(k_i)}\mathbf{C}^{-1} = -\frac{\partial \mathbf{C}^{-1}}{\partial P(k_i)}, \tag{4.18}$$

which provides an unbiased estimate if

$$\mathbf{N}_{ij} = \mathrm{tr}\left\{\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial P(k_i)}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial P(k_j)}\right\} . \tag{4.19}$$

The estimator is **lossless**, if its covariance is equal to the Fisher information matrix

$$\mathbf{F}_{ij} \equiv \left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial P(k_i)\partial P(k_j)} \right\rangle , \tag{4.20}$$

which is defined as the expected Hessian of the log-likelihood. Inserting Eq. (4.3), one finds

$$\mathbf{F}_{ij} = \frac{1}{2}\,\mathrm{tr}\left\{\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial P(k_i)}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial P(k_j)}\right\} = \frac{\mathbf{N}_{ij}}{2}. \tag{4.21}$$

This Fisher matrix can also be obtained by substituting Eq. (4.18) into Eq. (4.14), thus

$$\mathbf{V}_{ij} = \mathbf{F}_{ij}^{-1}, \tag{4.22}$$

which means that the QML estimator is lossless.

As we have seen, the contribution of each pair of Fourier modes is weighted by how the inverse of the density field covariance matrix $\mathbf{C}$ changes with respect to the prior of the power spectrum of the respective bin:

$$\mathbf{E}(k_j) = -\frac{\partial \mathbf{C}^{-1}}{\partial P(k_j)}, \tag{4.23}$$

This is the equation that makes the QML estimator prohibitively expensive. However, assuming a Gaussian density field, the QML estimator provides an unbiased estimate of the power spectrum with minimal errors.

Under the assumption that all modes are independent, the covariance of the density field is given by the power spectrum (and the Kronecker delta $\delta_{\mu\nu}$):

$$\mathbf{C}_{\mu\nu} = \delta_{\mu\nu} P(k_\mu). \tag{4.24}$$

For the derivative of $\mathbf{C}$ with respect to $P(k_i)$ we found earlier that it is unity if the modes $\mathbf{k}_\alpha$ and $\mathbf{k}_\beta$ are equal and contained in the bin $\mathbb{k}_i$, and zero otherwise, which can be written using the Heaviside function Θ as:

$$\frac{\partial \mathbf{C}_{\alpha\beta}}{\partial P(k_i)} = \delta_{\alpha\beta}\Theta(\mathbf{k}_\alpha \in \mathbb{k}_i) \equiv \delta_{\alpha\beta}\Theta_{\alpha i}. \tag{4.25}$$

Given Eq. (4.24) and (4.25), we find

$$\mathbf{E}_{\alpha\beta}(k_j) = \frac{\delta_{\alpha\beta}}{P^2(k_\alpha)}\Theta_{\alpha j} \tag{4.26}$$

and

$$\mathbf{N}_{ij} = \frac{N_{\mathbb{k}_i}}{P^2(k_i)}\delta_{ij}, \tag{4.27}$$

where $N_{\mathbb{k}_i}$ is the total number of modes in a given bin $\mathbb{k}_i$. Hence the QML estimator of Eq. (4.7) reduces to the FKP estimator (cf. Sec. 2.4.2, [109])

under the assumption that the covariance is constant within the k-bin, where several modes are combined into bins and the absolute values squared of the density field of each bin are summed:

$$\widehat{P}(k_i) = \frac{1}{N_{\mathbb{k}_i}} \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} |F(\mathbf{k}_\alpha)|^2 \,. \tag{4.28}$$

The difference here is that the QML estimator uses a prior of the power spectrum $P(k_\alpha)$ to weight contributions from each mode optimally, which means that the covariance of the power spectrum is minimal. The FKP estimator is commonly applied even when the assumptions of Eq. (4.24) to (4.27) are not valid. Note that in practice, both power spectrum estimators include a shot noise term (cf. Sec. 2.4.2), because the continuous density field is estimated using a discrete sample of galaxies. However, as the shot noise is not affected by systematic contaminants, it is ignored throughout this chapter. It will be introduced in Sec. 5.1 where I describe how the techniques discussed in this chapter can be applied to real data.

4.3 Removing Contaminants: mode deprojection

In this subsection, I describe how *mode deprojection* can be applied to estimate the 3D galaxy power spectrum. The method was first suggested in [195] in the context of noisy, irregularly sampled data. Applications and extensions to angular power spectra can be found for WMAP data in [198], for photometric SDSS-III data in [199], for photometric quasars of the XDQSOz catalogue in [196] and [108] and for 2D galaxy clustering in general in [193]. We use the notation of [193] for consistency.

Suppose we estimate the power spectrum using QML and that there is only a single contaminant. Then one can suppress contaminated modes in the covariance matrix updating the covariance matrix as [193]

$$\mathbf{C}_{\alpha\beta} \to \tilde{\mathbf{C}}_{\alpha\beta} = \mathbf{C}_{\alpha\beta} + \lim_{\sigma\to\infty} \sigma f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta), \tag{4.29}$$

i.e. letting the covariances of contaminated modes tend to infinity. Making use of the Sherman-Morrison matrix inversion lemma that states that the inverse

of the sum of an invertible matrix $\mathbf{A}$ and the outer product of two vectors $u_\alpha v_\beta$ can be calculated as

$$(\mathbf{A}_{\alpha\beta} + u_\alpha v_\beta)^{-1} = \mathbf{A}^{-1}_{\alpha\beta} - \frac{\mathbf{A}^{-1}_{\alpha\gamma} u_\gamma v_\delta \mathbf{A}^{-1}_{\delta\beta}}{1 + v_\gamma \mathbf{A}^{-1}_{\gamma\delta} u_\delta} \text{ [200]}, \tag{4.30}$$

one can see that (if $f(\mathbf{k}) \neq 0 \ \forall \mathbf{k}$) the inverse updated covariance matrix converges to

$$\tilde{\mathbf{C}}^{-1}_{\alpha\beta} = \mathbf{C}^{-1}_{\alpha\beta} - \frac{\sum_{\mu\nu} \mathbf{C}^{-1}_{\alpha\mu} f(\mathbf{k}_\mu) f^*(\mathbf{k}_\nu) \mathbf{C}^{-1}_{\nu\beta}}{\sum_{\mu\nu} f^*(\mathbf{k}_\mu) \mathbf{C}^{-1}_{\mu\nu} f(\mathbf{k}_\nu)}. \tag{4.31}$$

Now supposing that the modes are independent, i.e. Eq. (4.24) holds, we can insert it into Eq. (4.31) so that

$$\tilde{\mathbf{C}}^{-1}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P(k_\alpha)} - \frac{1}{R_P} \frac{f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta)}{P(k_\alpha) P(k_\beta)} \tag{4.32}$$

where we have defined

$$R_P \equiv \sum_\mu \frac{|f(\mathbf{k}_\mu)|^2}{P(k_\mu)}, \tag{4.33}$$

for simplicity. Taking the derivative of Eq. (4.32) with respect to $P(k_i)$, we obtain the updated estimator matrix[1]

$$\tilde{\mathbf{E}}_{\alpha\beta}(k_j) = \frac{\delta_{\alpha\beta}}{P^2_\alpha} \Theta_{\alpha j} - \frac{1}{R_P} \frac{f_\alpha f^*_\beta}{P_\alpha P_\beta} \left(\frac{\Theta_{\alpha j}}{P_\alpha} + \frac{\Theta_{\beta j}}{P_\beta} - \frac{t_j}{R_P} \right), \tag{4.34}$$

where

$$t_i \equiv \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \frac{|f(\mathbf{k}_\alpha)|^2}{P^2(k_\alpha)}. \tag{4.35}$$

After inserting Eq. (4.34) into Eq. (4.8), we obtain for the two point function

$$\begin{aligned} \mathbf{p}_i &= \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \left\{ \frac{|F(\mathbf{k}_\alpha)|^2}{P^2(k_\alpha)} - \frac{2}{R_P} \operatorname{Re} \left[S_P \frac{F^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P^2(k_\alpha)} \right] \right. \\ &\quad \left. + \frac{|S_P|^2}{R^2_P} \frac{|f(\mathbf{k}_\alpha)|^2}{P^2(k_\alpha)} \right\} \\ &= \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \frac{\left| F(\mathbf{k}_\alpha) - \frac{S_P}{R_P} f(\mathbf{k}_\alpha) \right|^2}{P^2(k_\alpha)}, \end{aligned} \tag{4.36}$$

where we have defined

$$S_P \equiv \sum_{\mathbf{k}_\alpha} \frac{F^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P_\alpha}. \tag{4.37}$$

[1] writing $f_\alpha \equiv f(\mathbf{k}_\alpha)$ and $P_\alpha \equiv P(k_\alpha)$ to save space

S_P is real, because $F(\mathbf{k})$ and $f(\mathbf{k})$ are Hermitian fields with real Fourier transforms.

Eq. (4.36) is in a considerably simpler form than Eq. (4.8) and does not require calculating many matrix elements of the estimator matrix $\mathbf{E}$. We show in the next section that we can consider this equation as a best-fit of the contaminants in the data.

We can normalise the updated mode deprojected QML estimator by replacing $\mathbf{C}$ by $\tilde{\mathbf{C}}$ in Eq. (4.19). As the term that suppresses contaminated modes from the covariance matrix in Eq. (4.29) does not depend on the power $P(k)$, we have $\frac{\partial \tilde{\mathbf{C}}_{\alpha\beta}}{\partial P(k_i)} = \frac{\partial \mathbf{C}_{\alpha\beta}}{\partial P(k_i)}$ and hence the normalisation is

$$\begin{aligned}
\tilde{\mathbf{N}}_{ij} =& \sum_{\alpha\mu\nu\rho} \tilde{\mathbf{C}}^{-1}_{\alpha\mu}\delta_{\mu\nu}\Theta_{\mu i}\tilde{\mathbf{C}}^{-1}_{\nu\rho}\delta_{\rho\alpha}\Theta_{\alpha j} \\
=& \sum_{\alpha\mu} |\tilde{\mathbf{C}}^{-1}_{\alpha\mu}|^2\Theta_{\mu i}\Theta_{\alpha j} \\
=& \sum_{\alpha\mu} \Theta_{\alpha i}\Theta_{\mu j}\left[\frac{\delta_{\alpha\mu}}{P^2(k_\alpha)}\left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)}\right) + \frac{1}{R_P^2}\frac{|f(\mathbf{k}_\alpha)f(\mathbf{k}_\mu)|^2}{P^2(k_\alpha)P^2(k_\mu)}\right]
\end{aligned} \tag{4.38}$$

where we have used the Hermitian property of $\tilde{\mathbf{C}}^{-1}$ in the third equality. In the first term in the square brackets, $\mathbf{k}_\alpha$ has to be in both $\mathbb{k}_i$ and $\mathbb{k}_j$, hence we can replace one Θ with δ_{ij}, such that Eq. (4.38) can be written as a diagonal matrix with diagonal elements $\tilde{n}$ and the outer product of a vector with itself:

$$\begin{aligned}
\tilde{\mathbf{N}}_{ij} =& \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \frac{\delta_{ij}}{P^2(k_\alpha)}\left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)}\right) + \frac{t_i t_j}{R_P^2} \\
\equiv& \tilde{n}_i\delta_{ij} + \frac{t_i t_j}{R_P^2},
\end{aligned} \tag{4.39}$$

This means that we can apply the Sherman-Morrison matrix inversion lemma [200]:

$$\tilde{\mathbf{N}}^{-1}_{ij} = \frac{\delta_{ij}}{\tilde{n}_i} - \frac{1}{R_P^2 + \sum_\ell \frac{t_\ell^2}{\tilde{n}_\ell}}\frac{t_i}{\tilde{n}_i}\frac{t_j}{\tilde{n}_j}. \tag{4.40}$$

As $\tilde{\mathbf{N}}^{-1}$ is not diagonal, it does not reduce to a simple FKP style estimator, i.e. if we have $N_{\rm bin}$ bins, we have to calculate for each bin the $N_{\rm mode} \times N_{\rm mode}$ estimator matrix $\mathbf{E}$ and we have to invert the $N_{\rm bin} \times N_{\rm bin}$ normalisation matrix. This is not feasible for 3D clustering, because of the large number of modes to be considered, especially if we want to choose narrow bins. Including several

contaminants makes it even more costly. The way of functioning of this method is illustrated using minimalist examples in Appendix B.

One way around this is a new framework introduced by Leistedt & Peiris [196] which they call *extended mode projection* and that selectively removes modes based on cross correlations with the data. However, this procedure reintroduces a small bias [193].

Another possibility is using the methodology of the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS)-collaboration, which is similar to that described in Sec. 4.4, but applied at the power spectrum level. However, this method is also biased [193]. Although [113] show that, for the Completed SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS DR12), the bias is much smaller than the statistical uncertainty, it was shown in the appendix of [26] that the bias is significant when one attempts to correct for many systematics. Furthermore, we expect smaller statistical uncertainties with future surveys, so in Sec. 4.4 and 4.5, we consider a computationally cheaper way of removing this small bias.

4.4 Removing Contaminants: mode subtraction

Here I consider *mode subtraction* and its link to *mode deprojection*. In order to remove contaminants we start by treating the true, but unknown, amplitude of the contamination $\varepsilon_{\rm true}$ in Eq. (4.1) as a free parameter ε, so that an estimate of the true density field $D(\mathbf{k})$ reads

$$\widehat{D}(\mathbf{k}) = F(\mathbf{k}) - \varepsilon f(\mathbf{k}). \tag{4.41}$$

Note that this is different to the *template subtraction* method introduced by [199], which is used by the BOSS collaboration and works entirely at the level of power spectra, whereas Eq. (4.41) works at the map level, thus it does not affect uncontaminated modes that are at the same multipole ℓ or wave number k as contaminated modes.

We can write a simplified model of the Gaussian likelihood whose maximum is given by the QML (cf. Eq. (4.7) and [182]) in the approximation of a

diagonal covariance matrix, with a small contaminant that does not affect the covariance. This is given by

$$-2\ln\mathcal{L} = \ln(2\pi) + \ln\left(\prod_{\alpha} P(k_\alpha)\right) + \sum_{\alpha} \frac{|F(\mathbf{k}_\alpha) - \varepsilon f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}. \tag{4.42}$$

We can therefore find ε by minimising Eq. (4.42), which is equivalent to simultaneously fitting ε and the model parameters entering the model power spectrum. The derivative of $\ln\mathcal{L}$ with respect to ε reads

$$\frac{\partial \ln\mathcal{L}}{\partial\varepsilon} = \sum_{\alpha} \frac{\mathrm{Re}\left[f(\mathbf{k}_\alpha)F^*(\mathbf{k}_\alpha)\right] - \varepsilon|f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}. \tag{4.43}$$

This expression is equal to zero if the likelihood maximised, i.e.

$$\varepsilon^{(\mathrm{BF})} = \frac{S_P}{R_P}. \tag{4.44}$$

The uncontaminated estimate of the density field is hence given by

$$\widehat{D}(\mathbf{k}_\alpha) = F(\mathbf{k}_\alpha) - \frac{S_P}{R_P} f(\mathbf{k}_\alpha), \tag{4.45}$$

and we can estimate the power as

$$\widehat{P}(k_i) = \frac{1}{N_{\mathbb{k}_i}} \sum_{\mathbf{k}_\alpha} \left| F(\mathbf{k}_\alpha) - \frac{S_P}{R_P} f(\mathbf{k}_\alpha) \right|^2. \tag{4.46}$$

This is similar to the *mode deprojection* result of Eq. (4.36) with a bias, missing the inverse noise matrix convolution of Eq. (4.7). The bias of this estimate comes about because S_P is correlated with the true density field $D(\mathbf{k})$. This correlation is similar to that created by the internal linear combination (ILC) method (e.g. [201]) for the analysis of cosmic microwave background (CMB) data. Based on this knowledge, an unbiased FKP-style estimator is built in the next section.

4.5 An Unbiased FKP-Style Estimator

In this section a simple, although sub-optimal, way to remove the bias on the power spectrum estimate resulting from imperfectly removing systematics using either Eq. (4.36) or (4.46), is presented. A straightforward way to remove the bias consists of calculating the expectation value of the power from each

mode analytically, assuming Eq. (4.1), and divide out the bias. We start with calculating some useful expectations which we need for the final result, the first one being the expectation of the correlation of the uncorrected density field of the data $\langle F(\mathbf{k}_\alpha)F^*(\mathbf{k}_\beta)\rangle$. Eq. (4.1) relates the uncorrected field with the true density field and the contamination template, allowing us to write

$$\begin{aligned}\langle F(\mathbf{k}_\alpha)F^*(\mathbf{k}_\beta)\rangle &= \langle (D(\mathbf{k}_\alpha) + \varepsilon_{\rm true} f(\mathbf{k}_\alpha))\,(D^*(\mathbf{k}_\beta) + \varepsilon_{\rm true} f^*(\mathbf{k}_\beta))\rangle \\ &= \langle D(\mathbf{k}_\alpha)D^*(\mathbf{k}_\beta)\rangle + \langle \varepsilon_{\rm true}\,(f(\mathbf{k}_\alpha)D^*(\mathbf{k}_\beta) + f^*(\mathbf{k}_\beta)D(\mathbf{k}_\alpha))\rangle \\ &\quad + \left\langle \varepsilon^2_{\rm true} f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta)\right\rangle . \end{aligned} \tag{4.47}$$

As D is the true density field, for $\alpha = \beta$, the first term equals the true power spectrum (cf. Eq. 1.104)

$$\langle D(\mathbf{k}_\alpha)D^*(\mathbf{k}_\beta)\rangle = \delta_{\alpha\beta}P(k_\alpha). \tag{4.48}$$

The second term of Eq. (4.47) contains only one random variable, which is again the true density field that is a zero-centred field with $\langle D(\mathbf{k}_\alpha)\rangle = 0$, hence the second term is zero. The third term is not a random term, so we can just drop the angle brackets and write the auto-correlation of the uncorrected observed field as

$$\langle F(\mathbf{k}_\alpha)F^*(\mathbf{k}_\beta)\rangle = \delta_{\alpha\beta}P(k_\alpha) + \varepsilon^2_{\rm true} f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta). \tag{4.49}$$

The next expectation value that we need in the derivation of the unbiased estimator is the one of the best-fitting amplitude $\varepsilon_{\rm BF}$. The normalisation R_P is a constant term, so the expectation of the amplitude is proportional to the one of S_P:

$$\langle \varepsilon_{\rm BF}\rangle = \frac{\langle S_P\rangle}{R_P} = \frac{1}{R_P}\sum_\alpha \left\langle \frac{F^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P(k_\alpha)}\right\rangle , \tag{4.50}$$

where in the second equation I have just inserted the definition of S_P from Eq. (4.37). One can again substitute F with Eq. (4.1), such that

$$\langle \varepsilon_{\rm BF}\rangle = \frac{1}{R_P}\sum_\alpha \left\langle \frac{D^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P(k_\alpha)} + \varepsilon_{\rm true}\frac{f^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P(k_\alpha)}\right\rangle . \tag{4.51}$$

The first term vanishes because of $\langle D\rangle = 0$, and the second term is known, so we can drop the brackets:

$$\langle \varepsilon_{\rm BF}\rangle = \varepsilon_{\rm true}\frac{1}{R_P}\sum_\alpha \frac{f^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P(k_\alpha)}. \tag{4.52}$$

The sum is exactly the definition of R_P (cf. Eq. 4.33) and we find that our best fit is unbiased, i.e.

$$\langle \varepsilon_{\rm BF} \rangle = \varepsilon_{\rm true}. \tag{4.53}$$

However, the expectation of the square of $\varepsilon_{\rm BF}$ is less trivial:

$$\begin{aligned}\langle \varepsilon_{\rm BF}^2 \rangle =& \frac{\langle S_P^2 \rangle}{R_P^2} = \frac{1}{R_P^2} \sum_{\alpha\beta} \left\langle \frac{F^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha) F(\mathbf{k}_\beta) f^*(\mathbf{k}_\beta)}{P(k_\alpha) P(k_\alpha)} \right\rangle \\ =& \frac{1}{R_P^2} \sum_{\alpha\beta} \langle F^*(\mathbf{k}_\alpha) F(\mathbf{k}_\beta) \rangle \frac{f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta)}{P(k_\alpha) P(k_\alpha)},\end{aligned} \tag{4.54}$$

where, in the first line, I used the fact that the complex conjugate of the real quantity $S_P^* = S_P$. The expectation in the second line can be replaced with the previous result from Eq. (4.49):

$$\begin{aligned}\langle \varepsilon_{\rm BF}^2 \rangle =& \frac{1}{R_P^2} \sum_{\alpha\beta} \left(\delta_{\alpha\beta} P(k_\alpha) + \varepsilon_{\rm true}^2 f^*(\mathbf{k}_\alpha) f(\mathbf{k}_\beta) \right) \frac{f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta)}{P(k_\alpha) P(k_\alpha)} \\ =& \frac{1}{R_P^2} \sum_{\alpha} \frac{|f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)} + \varepsilon_{\rm true}^2 \frac{1}{R_P^2} \sum_{\alpha\beta} \frac{|f(\mathbf{k}_\alpha)|^2 |f(\mathbf{k}_\beta)|^2}{P(k_\alpha) P(k_\beta)},\end{aligned} \tag{4.55}$$

and the sums are again equal to R_P. Hence, the inverse of R_P is the variance of $\varepsilon_{\rm BF}$:

$$\langle \varepsilon_{\rm BF}^2 \rangle = \varepsilon_{\rm true}^2 + \frac{1}{R_P}. \tag{4.56}$$

The last remaining expectation needed is the cross correlation between the contaminant's amplitude and the measured density field

$$\langle \varepsilon_{\rm BF} F(\mathbf{k}_\alpha) \rangle = \frac{1}{R_P} \sum_{\beta} \left\langle \frac{F^*(\mathbf{k}_\beta) f(\mathbf{k}_\beta)}{P(k_\beta)} F(\mathbf{k}_\alpha) \right\rangle = \frac{1}{R_P} \sum_{\beta} \frac{f(\mathbf{k}_\beta)}{P(k_\beta)} \langle F^*(\mathbf{k}_\beta) F(\mathbf{k}_\alpha) \rangle \,. \tag{4.57}$$

Again using Eq. (4.49), one obtains

$$\begin{aligned}\langle \varepsilon_{\rm BF} F(\mathbf{k}_\alpha) \rangle =& \frac{1}{R_P} \sum_{\beta} \frac{f(\mathbf{k}_\beta)}{P(k_\beta)} \left(\delta_{\alpha\beta} P(k_\alpha) + \varepsilon_{\rm true}^2 f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta) \right) \\ =& \left(\varepsilon_{\rm true}^2 + \frac{1}{R_P} \right) f(\mathbf{k}_\alpha).\end{aligned} \tag{4.58}$$

The most important intermediate results of this section so far are summarised in Table 4.1. With these equations at hand, we can calculate the expectation of the biased power spectra in Eq. (4.36) and (4.46), which are

Table 4.1: Expectation values of quantities entering Eq. (4.60).

$\langle F(\mathbf{k}_\alpha)F^*(\mathbf{k}_\beta)\rangle$	$\delta_{\alpha\beta}P(k_\alpha) + \varepsilon^2_{\rm true}f(\mathbf{k}_\alpha)f^*(\mathbf{k}_\beta)$
$\langle\varepsilon_{\rm BF}\rangle$	$\varepsilon_{\rm true}$
$\langle\varepsilon^2_{\rm BF}\rangle$	$\varepsilon^2_{\rm true} + \frac{1}{R_P}$
$\langle\varepsilon_{\rm BF}F(\mathbf{k}_\alpha)\rangle$	$\left(\varepsilon^2_{\rm true} + \frac{1}{R_P}\right) f(\mathbf{k}_\alpha)$

the two-point function of the estimator for the corrected density field $\widehat{D}$ in Eq. (4.1):

$$\begin{aligned}\langle|\widehat{D}(\mathbf{k}_\alpha)|^2\rangle &= \langle|F(\mathbf{k}_\alpha) - \varepsilon_{\rm BF}f(\mathbf{k}_\alpha)|^2\rangle \\ &= \langle|F(\mathbf{k}_\alpha)|^2\rangle - 2\,\mathrm{Re}\left[\langle\varepsilon_{\rm BF}F(\mathbf{k}_\alpha)\rangle f^*(\mathbf{k}_\alpha)\right] + \langle\varepsilon^2_{\rm BF}\rangle|f(\mathbf{k}_\alpha)|^2 \\ &= P(k_\alpha) + \varepsilon_{\rm true}|f(\mathbf{k}_\alpha)|^2 - 2\left(\varepsilon^2_{\rm true} + \frac{1}{R_P}\right)|f(\mathbf{k}_\alpha)|^2 + \left(\varepsilon^2_{\rm true} + \frac{1}{R_P}\right)|f(\mathbf{k}_\alpha)|^2 \\ &= P(k_\alpha) - \frac{|f(\mathbf{k}_\alpha)|^2}{R_P}, \end{aligned} \tag{4.59}$$

hence, we see that the naïve power spectrum estimators of Eq. (4.36) and (4.46) is biased with a factor

$$1 - \frac{1}{R_P}\frac{|f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}, \tag{4.60}$$

and we can build an unbiased estimator of the power by dividing each mode in Eq. (4.36) and (4.46) by Eq. (4.60). If we want to debias the two-point function using Eq. (4.60), we have to assume a prior power spectrum. Note that the QML approach also requires the prior knowledge of the power spectrum. We will see in Sec. 4.7 and 5.2 that the impact of adopting a slightly wrong prior power spectrum is indeed small. Our final estimator of the power spectrum is then

$$\widehat{P}(k_i) = \frac{1}{N_{\mathbb{k}_i}}\sum_{\mathbf{k}_\alpha}\frac{\left|F(\mathbf{k}_\alpha) - \frac{S_P}{R_P}f(\mathbf{k}_\alpha)\right|^2}{1 - \frac{1}{R_P}\frac{|f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}}. \tag{4.61}$$

Eq. (4.61) is one of the key results of [5]: this is an extension of the FKP estimator that removes potential contaminants from the data in an unbiased way, without the need for large matrices. Moreover, as it is in the same form as the well established FKP estimator, this can easily be folded into estimators for redshift-space clustering such as those by Bianchi *et al.* and Scoccimarro [151, 202].

One can interpret the same debiasing factor also in the framework of the QML Fisher information matrix $\mathbf{N}$, which in the QML approach performs both the debiasing and optimisation effects, as it can also be derived from $\tilde{\mathbf{N}}$. Without binning, i.e. each bin contains only one Fourier mode, Eq. (4.39) simplifies to

$$\tilde{\mathbf{N}}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P^2(k_\alpha)}\left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)}\right) + \frac{1}{R_P^2}\frac{|f(\mathbf{k}_\alpha)|^2}{P^2(k_\alpha)}\frac{|f(\mathbf{k}_\beta)|^2}{P^2(k_\beta)}. \tag{4.62}$$

The difference between the two approaches is that QML provides an unbiased optimal power estimate, whereas Eq. (4.61) has been constructed such that it is only unbiased, i.e. the powers in the denominators of Eq. (4.62) act as optimal weights to each mode. If we allow for some information loss within bins, by assuming the expected power is constant within each bin, we can replace $P^2(k_\beta)$ by $P(k_\alpha)P(k_\beta)$, such that

$$\tilde{\mathbf{N}}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P^2(k_\alpha)}\left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)}\right) + \frac{1}{R_P^2}\frac{|f(\mathbf{k}_\alpha)|^2}{P^3(k_\alpha)}\frac{|f(\mathbf{k}_\beta)|^2}{P(k_\beta)}. \tag{4.63}$$

This normalisation is proportional to the Fisher information matrix [182], from which we marginalise out contributions from other modes by summing over all modes $\mathbf{k}_\beta$:

$$\begin{aligned}\sum_\beta \tilde{\mathbf{N}}_{\alpha\beta} =& \frac{1}{P^2(k_\alpha)}\left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)}\right) + \frac{1}{R_P}\frac{|f(\mathbf{k}_\alpha)|^2}{P^3(k_\alpha)} \\ =& \frac{1}{P^2(k_\alpha)}\left(1 - \frac{|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)}\right).\end{aligned} \tag{4.64}$$

This is exactly Eq. (4.60) with a factor of $\frac{1}{P^2(k_\alpha)}$ that cancels out the difference between Eq. (4.36) and Eq. (4.46). We have therefore shown that Eq. (4.61) is a non-optimal, but unbiased, approximation to using the QML normalisation with *mode deprojection*. In the limit of narrow bins, when the power spectrum does not change significantly within the bin, Eq. (4.61) is mathematically identical to the QML estimate. The main inequivalence between the two techniques lies in the fact that with mode deprojection, one effectively marginalises over ε, whereas using mode subtraction, we find a best fitting value $\varepsilon_{\rm BF}$ which has a variance of $\mathrm{Var}(\varepsilon_{\rm BF}) = \frac{1}{R_P}$, as can be read from Tab. 4.1. Mode deprojection therefore retains the optimality of the QML estimator, whilst mode subtraction slightly increases the measurement uncertainty by fitting an additional parameter. The impact of this sub-optimality is studied

later in examples in Sec. 4.7. In fact, we will argue later that this is actually a weaker effect than many common approximations applied when using the FKP estimator, such as ignoring large-scale window effects, when averaging large scale modes.

4.5.1 Notes on assuming a diagonal covariance

Note that, even in the absence of systematics, we have assumed a diagonal covariance matrix in the derivation of both the *mode subtraction* and the debiasing step. In practice the covariance matrix has off-diagonal terms due to the effect of the survey window. However, this is usually not included when calculating the data power spectrum but, instead, it is included as a convolution in the model power spectrum. One can show that Eq. (4.61) still holds in the general case of having a non-diagonal covariance matrix, as long as R_P is generalised as in Eq. (4.33). I present a generalisation of the derivation in the main section, defining the covariance matrix of the true density

$$\mathbf{C}_{\alpha\beta} \equiv \left\langle D_\alpha D_\beta^* \right\rangle . \tag{4.65}$$

Still assuming that the true signal and the contaminant are uncorrelated, Eq. (4.49) can be generalised to

$$\left\langle F_\alpha F_\beta^* \right\rangle = \mathbf{C}_{\alpha\beta} + \varepsilon_{\rm true}^2 f_\alpha f_\beta^* . \tag{4.66}$$

As done in Sec. 4.4, we introduce a free parameter ε, such that

$$\widehat{D}_\alpha \equiv F_\alpha - \varepsilon f_\alpha . \tag{4.67}$$

Assuming that the true density field is Gaussian with Gaussian auto-correlation, its log-likelihood reads

$$-2\ln\mathcal{L} = \sum_{\alpha\beta} \left(F_\alpha - \varepsilon f_\alpha\right)^* \mathbf{C}^{-1}_{\alpha\beta} \left(F_\beta - \varepsilon f_\beta\right) + \text{const.} \tag{4.68}$$

To find the best-fitting $\varepsilon^{(\rm BF)}$, we take the derivative of the log-likelihood with respect to ε:

$$\begin{aligned} -2\frac{\partial \ln\mathcal{L}}{\partial\varepsilon} =& -\sum_{\alpha\beta} f_\alpha^* \mathbf{C}^{-1}_{\alpha\beta} \left(F_\beta - \varepsilon f_\beta\right) - \sum_{\alpha\beta} \left(F_\alpha - \varepsilon f_\alpha\right)^* \mathbf{C}^{-1}_{\alpha\beta} f_\beta \\ =& 2\varepsilon \sum_{\alpha\beta} f_\alpha^* \mathbf{C}^{-1}_{\alpha\beta} f_\beta - \sum_{\alpha\beta} \left[f_\alpha^* \mathbf{C}^{-1}_{\alpha\beta} F_\beta + F_\alpha^* \mathbf{C}^{-1}_{\alpha\beta} f_\beta \right] . \end{aligned} \tag{4.69}$$

As $\mathbf{C}$ is a covariance matrix of complex random variables, it is Hermitian positive-semidefinite, such that the second sum can be written as

$$\sum_{\alpha\beta}\left[f_\alpha^*\mathbf{C}_{\alpha\beta}^{-1}F_\beta + F_\alpha^*\mathbf{C}_{\alpha\beta}^{-1}f_\beta\right] = 2\,\mathrm{Re}\left[\sum_{\alpha\beta} f_\alpha^*\mathbf{C}_{\alpha\beta}^{-1}F_\beta\right]. \tag{4.70}$$

For shortness and in analogy to Sec. 4.4, let me call this sum

$$S_P \equiv \sum_{\alpha\beta}\mathrm{Re}\left[f_\alpha^*\mathbf{C}_{\alpha\beta}^{-1}F_\beta\right] \tag{4.71}$$

and the first sum in Eq. (4.69) shall be called

$$R_P \equiv \sum_{\alpha\beta} f_\alpha^*\mathbf{C}_{\alpha\beta}^{-1}f_\beta. \tag{4.72}$$

We obtain the best-fitting, i.e. maximum likelihood, value

$$\varepsilon^{(\mathrm{BF})} = \frac{S_P}{R_P} \tag{4.73}$$

by equating Eq. (4.69) to zero.

Now I want to calculate the expectation value of the biased power spectrum estimator

$$\begin{aligned}\left\langle \widehat{D}_\alpha^2\right\rangle =& \left\langle \left| F_\alpha - \frac{S_P}{R_P}f_\alpha\right|^2\right\rangle \\ =& \langle |F_\alpha|^2\rangle - 2\,\mathrm{Re}\left[\left\langle\frac{S_P}{R_P}F_\alpha\right\rangle f_\alpha^*\right] + \left\langle\frac{S_P^2}{R_P^2}\right\rangle |f_\alpha|^2.\end{aligned} \tag{4.74}$$

I calculate each term separately and in analogy to the diagonal case:

1. The first term $\langle|F_\alpha|^2\rangle = \mathbf{C}_{\alpha\alpha} + \varepsilon_{\mathrm{true}}^2|f_\alpha|^2$ is a special case of Eq. (4.66).

2. To calculate the second term, we reexpand S_P and use the fact that $\mathrm{Re}\,[F_\alpha f_\alpha^*] = \mathrm{Re}\,[F_\alpha^* f_\alpha]$:

$$2\,\mathrm{Re}\,[\langle S_P F_\alpha^* f_\alpha\rangle] = 2\,\mathrm{Re}\left[\sum_{\gamma\beta} f_\gamma^*\mathbf{C}_{\gamma\beta}^{-1}\left\langle F_\beta F_\alpha^*\right\rangle f_\alpha\right] \tag{4.75}$$

   After reinserting Eq. (4.66), I get

$$2\,\mathrm{Re}\,[\langle S_P F_\alpha^* f_\alpha\rangle] = 2\,\mathrm{Re}\left[\sum_{\gamma\beta} f_\gamma^*\mathbf{C}_{\gamma\beta}^{-1}\mathbf{C}_{\beta\alpha}f_\alpha + \varepsilon_{\mathrm{true}}^2\sum_{\gamma\beta} f_\gamma^*\mathbf{C}_{\gamma\beta}^{-1}f_\beta f_\alpha^* f_\alpha\right]. \tag{4.76}$$

In the first term we have $\sum_\beta \mathbf{C}^{-1}_{\gamma\beta}\mathbf{C}_{\beta\alpha} = \delta_{\gamma\alpha}$, and in the second term we find the definition of R_P. Thus, the second term of Eq. (4.74) is

$$2\,\mathrm{Re}\left[\langle S_P F^*_\alpha f_\alpha\rangle\right] = 2|f_\alpha|^2\left(1+\varepsilon^2_{\mathrm{true}}R_P\right). \tag{4.77}$$

3. In the third term, we can again make use of Eq. (4.66):

$$\begin{aligned}\left\langle S_P^2\right\rangle &= \sum_{\alpha\beta\gamma\delta}\mathrm{Re}\left[f^*_\alpha\mathbf{C}^{-1}_{\alpha\beta}f_\gamma\mathbf{C}^{-1}_{\gamma\delta}\left\langle F_\beta F^*_\delta\right\rangle\right]\\ &= \sum_{\alpha\beta\gamma\delta}\mathrm{Re}\left[f^*_\alpha\mathbf{C}^{-1}_{\alpha\beta}f_\gamma\mathbf{C}^{-1}_{\gamma\delta}\mathbf{C}_{\beta\delta} + \varepsilon^2_{\mathrm{true}}f^*_\alpha\mathbf{C}^{-1}_{\alpha\beta}f_\gamma\mathbf{C}^{-1}_{\gamma\delta}f_\beta f^*_\delta\right]\end{aligned} \tag{4.78}$$

In the first term, we have again $\sum_\beta \mathbf{C}^{-1}_{\alpha\beta}\mathbf{C}_{\beta\delta} = \delta_{\alpha\delta}$, and the second term is equal to R_P^2, such that

$$\left\langle S_P^2\right\rangle = \sum_{\alpha\gamma}\mathrm{Re}\left[f^*_\alpha f_\gamma\mathbf{C}^{-1}_{\gamma\alpha}\right] + \varepsilon^2_{\mathrm{true}}R_P^2 = R_P + \varepsilon^2_{\mathrm{true}}R_P^2 \tag{4.79}$$

Recollecting 1.-3. and inserting into Eq. (4.74) yields

$$\begin{aligned}\left\langle\left|F_\alpha - \frac{S_P}{R_P}f_\alpha\right|^2\right\rangle =& \mathbf{C}_{\alpha\alpha} + \varepsilon^2_{\mathrm{true}}|f_\alpha|^2 - 2|f_\alpha|^2\left(\varepsilon^2_{\mathrm{true}} + \frac{1}{R_P}\right) + \left(\varepsilon^2_{\mathrm{true}} + \frac{1}{R_P}\right)|f_\alpha|^2\\ =& \mathbf{C}_{\alpha\alpha} - \frac{|f_\alpha|^2}{R_P}.\end{aligned} \tag{4.80}$$

As the power spectrum $P(k_\alpha) = \mathbf{C}_{\alpha\alpha}$ is defined as the diagonal elements of the covariance matrix, the debiasing step is the same for a non-diagonal covariance matrix as for a diagonal one (cf. Sec. 4.5), we just have to use the generalised definition of R_P as in Eq. (4.33). However, to compute the generalised R_P, one has to invert the full $N_{\mathrm{mode}} \times N_{\mathrm{mode}}$ covariance matrix, which makes this approach computationally almost as expensive as using the QML estimator. We will argue that, in most cases, Eq. (4.61) provides a good estimate of the power, even in the presence of covariant modes, and we will provide a further correction term that corrects for using Eq. (4.61) when off-diagonal covariances are important.

Suppose we apply Eq. (4.61) assuming a diagonal covariance matrix, even though there are covariances between different modes. Then, we find a best fitting

$$\varepsilon'_{\mathrm{BF}} = \frac{\sum_\alpha \frac{F^*_\alpha f_\alpha}{P_\alpha}}{\sum_\mu \frac{|f_\mu|^2}{P_\mu}} \tag{4.81}$$

instead of the true

$$\varepsilon_{\rm BF} = \frac{\sum_{\alpha\beta} f^*_\alpha \mathbf{C}^{-1}_{\alpha\beta} F_\beta}{\sum_{\alpha\beta} f^*_\alpha \mathbf{C}^{-1}_{\alpha\beta} f_\beta}. \tag{4.82}$$

The expectations are the same $\langle \varepsilon'_{\rm BF} \rangle = \langle \varepsilon_{\rm BF} \rangle = \varepsilon_{\rm true}$, but their variances are different. For the approximate estimate we have

$$\begin{aligned}
\left\langle \varepsilon'^2_{\rm BF} \right\rangle =& \frac{\left\langle \sum_{\alpha\beta} \frac{F^*_\alpha f_\alpha F_\beta f^*_\beta}{P_\alpha P_\beta} \right\rangle}{R'^2_P} \\
=& \frac{\sum_{\alpha\beta} \frac{f_\alpha \mathbf{C}_{\alpha\beta} f^*_\beta}{P_\alpha P_\beta}}{R'^2_P} + \frac{\sum_{\alpha\beta} \frac{\varepsilon^2_{\rm true} |f_\alpha|^2 |f_\beta|^2}{P_\alpha P_\beta}}{R'^2_P} \\
=& \frac{1}{R'^2_P} \sum_{\alpha\beta} \frac{f_\alpha \mathbf{C}_{\alpha\beta} f^*_\beta}{P_\alpha P_\beta} + \varepsilon^2_{\rm true}.
\end{aligned} \tag{4.83}$$

Unlike in the previous estimates, the covariance matrix does not cancel in the first term. Similarly,

$$\begin{aligned}
\langle \varepsilon'_{\rm BF} F^*_\alpha f_\alpha \rangle =& \frac{1}{R'_P} \left\langle \sum_\beta \frac{f^*_\beta F_\beta F^*_\alpha f_\alpha}{P_\beta} \right\rangle \\
=& \frac{1}{R'_P} \sum_\beta \frac{f_\alpha \mathbf{C}_{\alpha\beta} f^*_\beta}{P_\beta} + \varepsilon^2_{\rm true} |f_\alpha|^2 .
\end{aligned} \tag{4.84}$$

Combining the previous two equations, we obtain

$$\left\langle |F_\alpha - \varepsilon'_{\rm BF} f_\alpha|^2 \right\rangle = C_{\alpha\alpha} - \frac{2}{R'_P} \sum_\beta \frac{f_\alpha \mathbf{C}_{\alpha\beta} f^*_\beta}{P_\beta} + \frac{|f_\alpha|^2}{R'^2_P} \sum_{\gamma\beta} \frac{f_\gamma \mathbf{C}_{\gamma\beta} f^*_\beta}{P_\gamma P_\beta}. \tag{4.85}$$

Splitting the covariance matrix

$$C_{\alpha\beta} = P_\beta \left(\delta_{\alpha\beta} + \Delta_{\alpha\beta} \right) \tag{4.86}$$

into a diagonal and off-diagonal elements yields

$$\left\langle |F_\alpha - \varepsilon'_{\rm BF} f_\alpha|^2 \right\rangle = P_\alpha - \frac{|f_\alpha|^2}{R'_P} \left[1 + \sum_{\gamma\beta} f_\gamma \Delta_{\gamma\beta} f^*_\beta \left(\frac{2\delta_{\alpha\gamma}}{|f_\alpha|^2} - \frac{1}{R'_P P_\gamma} \right) \right]. \tag{4.87}$$

Hence, one can perform mode subtraction assuming a diagonal covariance matrix and then apply another correction term which is linear in its off-diagonal elements. The advantage of this procedure is that it does not require any inversion of the $N^2_{\rm mode}$ covariance matrix. If the off-diagonal elements are small, then the bias correction reverts back to the form of Eq. (4.60).

4.6 Removing multiple Contaminants

We have shown the equivalence of expected power spectrum values after applying *mode deprojection* and *debiased mode subtraction* for one contaminant. A realistic survey has several sources of potential contaminants, so I show here that this equivalence holds for an arbitrary number of templates. For *mode deprojection*, we have to update the covariance matrix with a sum over all templates, and thus we have to replace Eq. (4.29) with

$$\tilde{\mathbf{C}}_{\alpha\beta} = \mathbf{C}_{\alpha\beta} + \lim_{\sigma\to\infty} \sigma \sum_{A=1}^{N_{\rm sys}} f_A(\mathbf{k}_\alpha) f_A^*(\mathbf{k}_\beta). \tag{4.88}$$

Starting from Eq. (4.88), I present the derivation of the unbinned *mode deprojection* power spectrum

$$\widehat{P}(k_\alpha) = \left| F(\mathbf{k}_\alpha) - \sum_{AB} \mathbf{S}_A \mathbf{R}^{-1}_{AB} f_B(\mathbf{k}_\alpha) \right|^2, \tag{4.89}$$

below. Let me start by rewriting Eq. (4.88) in matrix notation

$$\tilde{\mathbf{C}} = \mathbf{C} + \lim_{\sigma\to\infty} \sigma \mathbf{f} \mathbf{I}_{N_{\rm sys}} \mathbf{f}^\dagger, \tag{4.90}$$

defining an $N_{\rm mode} \times N_{\rm sys}$ matrix $\mathbf{f}_{\alpha A} \equiv f_A(\mathbf{k}_\alpha)$ and writing the $N_{\rm sys} \times N_{\rm sys}$ identity matrix as $\mathbf{I}_{N_{\rm sys}}$, such that I can invert $\tilde{\mathbf{C}}$ using the Woodbury matrix identity

$$\begin{aligned}
\tilde{\mathbf{C}}^{-1} &= \mathbf{C}^{-1} - \mathbf{C}^{-1} \lim_{\sigma\to\infty} \sigma \mathbf{f} \left(\mathbf{I}^{-1}_{N_{\rm sys}} + \mathbf{f}^\dagger \mathbf{C}^{-1} \sigma \mathbf{f} \right)^{-1} \mathbf{f}^\dagger \mathbf{C}^{-1} \\
&= \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{f} \left(\mathbf{f}^\dagger \mathbf{C}^{-1} \mathbf{f} \right)^{-1} \mathbf{f}^\dagger \mathbf{C}^{-1} \\
&\equiv \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{f} \mathbf{R}^{-1} \mathbf{f}^\dagger \mathbf{C}^{-1},
\end{aligned} \tag{4.91}$$

where, in the last line, $\mathbf{R} \equiv \mathbf{f}^\dagger \mathbf{C}^{-1} \mathbf{f}$ is a matrix equivalent to the factor R_P in previous sections. If we again assume $\mathbf{C}_{\alpha\beta} = \delta_{\alpha\beta} P(k_\alpha)$, it reads

$$\mathbf{R}_{AB} = \sum_{\mu\nu} f_A^*(\mathbf{k}_\mu) \frac{\delta_{\mu\nu}}{P(k_\mu)} f_B(k_\nu) = \sum_\mu \frac{f_A^*(\mathbf{k}_\mu) f_B(k_\mu)}{P(k_\mu)}. \tag{4.92}$$

The inverse updated covariance matrix then reads

$$\tilde{\mathbf{C}}^{-1}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P(k_\alpha)} - \sum_{AB} \frac{f_A(\mathbf{k}_\alpha) \mathbf{R}^{-1}_{AB} f_B^*(\mathbf{k}_\beta)}{P(k_\alpha) P(k_\beta)}. \tag{4.93}$$

If we do not bin, but apply mode deprojection to each mode separately, the matrix $\tilde{\mathbf{E}}$ simplifies to

$$\tilde{\mathbf{E}}_{\alpha\beta}(k_j) = \sum_{\mu\nu} \tilde{\mathbf{C}}^{-1}_{\alpha\mu}\delta_{\mu j}\delta_{\mu\nu}\tilde{\mathbf{C}}^{-1}_{\nu\beta} = \tilde{\mathbf{C}}^{-1}_{\alpha j}\tilde{\mathbf{C}}^{-1}_{j\beta}. \tag{4.94}$$

After inserting Eq. (4.93) into Eq. (4.94), one can derive a generalisation of Eq. (4.36):

$$\begin{aligned}
\mathbf{p}_j =& \sum_{\alpha\beta} F^*(\mathbf{k}_\alpha)\tilde{\mathbf{E}}_{\alpha\beta}(k_j)F(\mathbf{k}_\beta) \\
=& \frac{1}{P^2(k_j)}\left(|F(\mathbf{k}_j)|^2 - \sum_{AB\alpha}\frac{F^*(\mathbf{k}_\alpha)f_A(\mathbf{k}_\alpha)}{P(k_\alpha)}\mathbf{R}^{-1}_{AB}f^*_B(\mathbf{k}_j)F(\mathbf{k}_j)\right. \\
& - \sum_{AB\beta} F^*(\mathbf{k}_j)f_A(\mathbf{k}_j)\mathbf{R}^{-1}_{AB}\frac{f^*_B(\mathbf{k}_\beta)F(\mathbf{k}_\beta)}{P(k_\beta)} \\
& \left. + \sum_{ABCD\alpha\beta}\frac{F^*(\mathbf{k}_\alpha)f_A(\mathbf{k}_\alpha)}{P(k_\alpha)}\mathrm{R}^{-1}_{AB}f^*_B(\mathbf{k}_j)f_C(\mathbf{k}_j)\mathrm{R}^{-1}_{CD}\frac{f^*_D(\mathbf{k}_\beta)F(\mathbf{k}_\beta)}{P(k_\beta)}\right) \\
=& \frac{1}{P^2(k_j)}\left(|F(\mathbf{k}_j)|^2 - 2\,\mathrm{Re}\left[\sum_{AB}\mathbf{S}_A\mathbf{R}^{-1}_{AB}f^*_B(\mathbf{k}_j)F(\mathbf{k}_j)\right] + \left|\sum_{AB}\mathbf{S}_A\mathbf{R}^{-1}_{AB}f_B(\mathbf{k}_j)\right|^2\right) \\
=& \frac{1}{P^2(k_j)}\left|F(\mathbf{k}_j) - \sum_{AB}\mathbf{S}_A\mathbf{R}^{-1}_{AB}f_B(\mathbf{k}_j)\right|^2,
\end{aligned} \tag{4.95}$$

where I defined $\mathbf{S}_A \equiv \sum_\alpha \frac{f_A(\mathbf{k}_\alpha)F^*(\mathbf{k}_\alpha)}{P(k_\alpha)}$ analogously to S_P. To show the equivalence of power spectrum estimates after *mode deprojection* and *mode subtraction*, I also derive an equivalent expression in the *mode subtraction* framework.

To apply mode subtraction to multiple contaminants, one has to extend the likelihood given in Eq. (4.42) to

$$-2\ln\mathcal{L} = \ln(2\pi) + \ln\left(\prod_\alpha P(k_\alpha)\right) + \sum_\alpha \frac{|F(\mathbf{k}_\alpha) - \sum_A \boldsymbol{\varepsilon}_A f_A(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}. \tag{4.96}$$

Writing $\boldsymbol{\varepsilon}$ as a vector, and taking the derivative with respect to $\boldsymbol{\varepsilon}_B$ yields

$$\frac{\partial\chi^2}{\partial\boldsymbol{\varepsilon}_B} = -2\boldsymbol{\varepsilon}_B\sum_\alpha\frac{f_B(\mathbf{k}_\alpha)F^*(\mathbf{k}_\alpha) - \sum_A\boldsymbol{\varepsilon}_A f_B(\mathbf{k}_\alpha)f^*_A(\mathbf{k}_\alpha)}{P(k_\alpha)}. \tag{4.97}$$

This derivative is zero if

$$\sum_\alpha\frac{f_B(\mathbf{k}_\alpha)F^*(\mathbf{k}_\alpha)}{P(k_\alpha)} = \sum_{A\alpha}\frac{\boldsymbol{\varepsilon}_A f_B(\mathbf{k}_\alpha)f^*_A(\mathbf{k}_\alpha)}{P(k_\alpha)}, \tag{4.98}$$

which reads

$$\mathbf{S} = \mathbf{R}\boldsymbol{\varepsilon} \tag{4.99}$$

in matrix notation. The best fitting $\boldsymbol{\varepsilon}$ value is therefore given by

$$\boldsymbol{\varepsilon}^{(\mathrm{BF})} = \mathbf{R}^{-1}\mathbf{S}. \tag{4.100}$$

Note that this would require fitting the amplitude of all contaminants simultaneously. The absolute value squared of the best fitting signal is hence equal to Eq. (4.89). Hence, we also do not need large $N_{\mathrm{mode}} \times N_{\mathrm{mode}}$ matrices when we have to remove several potential contaminants.

We can calculate the debiasing factor

$$\sum_j \tilde{\mathbf{N}}_{ij} P^2(k_i) = 1 - \sum_{AB} \frac{f_A(\mathbf{k}_i)\mathbf{R}^{-1}_{AB} f^*_B(\mathbf{k}_i)}{P(k_i)} \tag{4.101}$$

analogously to Sec. 4.4 from the *mode deprojection* normalisation matrix without binning. Our unbiased estimator of the power spectrum thus reads

$$\widehat{P}(k_\alpha) = \frac{\left|F(\mathbf{k}_\alpha) - \sum_{AB} \mathbf{S}_A \mathbf{R}^{-1}_{AB} f_B(\mathbf{k}_\alpha)\right|^2}{1 - \sum_{AB} \frac{f_A(\mathbf{k}_i)\mathbf{R}^{-1}_{AB} f^*_B(\mathbf{k}_i)}{P(k_i)}} \tag{4.102}$$

when we mitigate the effect of several contaminants, or if we use more than one template to describe a single contaminant.

4.7 Testing Contaminant Removal

After deriving the equations needed to mitigate contaminants when measuring the 3D galaxy power spectrum, it is time to test these using simple toy examples. In this section, simple artificial contaminants are removed in power spectrum measurements from simulated density fields, using the hitherto described methodologies.

4.7.1 Gaussian Spike Contaminant

As a first test, we generate 3-dimensional Gaussian random fields according to an input power spectrum that we calculate using CAMB [22]. Each of these fields consists of a $16 \times 16 \times 16$ grid, in a box of length $3136h^{-1}$ Mpc. An example of such a field is shown in the top panel of Fig. 4.1. We contaminate

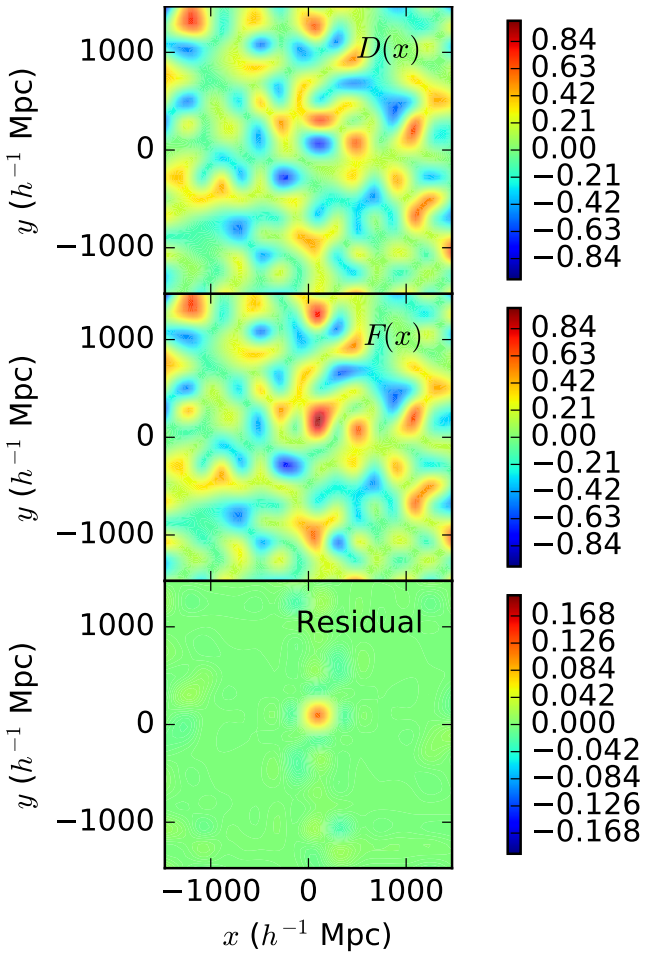

Figure 4.1: A slice through a realisation of a Gaussian random field contaminated with a Gaussian spike used in Sec. 4.7.1. The top panel shows the "clean" Gaussian random field (corresponding to $D(\mathbf{k})$ through Fourier transform) in configuration space. In the central panel, we have plotted the contaminated field (Fourier pair of $F(\mathbf{k})$) with an obvious Gaussian overdensity in the centre. The bottom panel shows the residual, i.e. the difference of the field after *mode subtraction* (i.e. the Fourier transform of $F(\mathbf{k}) - \varepsilon^{(\mathrm{BF})} f(\mathbf{k})$, cf. Eq. (4.46)) and the input field. The best-fitting $\varepsilon^{(\mathrm{BF})}$ for this particular realisation amounts to 1.078.

these Gaussian random fields by adding a real Gaussian spike in k-space with width $\sigma^2 = 10^{-5}h$ Mpc, centred around $k = 0.01h$ Mpc^{-1}, such that its maximum lies within a bin with sufficiently good statistics. The Fourier transform of this contaminant field is again a Gaussian spike in the centre of the box with some long wavelength fluctuations around it. The amplitude of the real part over-density in $\mathbf{k}$-space is 100, thus having the same order of magnitude as the "true" density field. An example of this setup can be seen in the central panel of Fig. 4.1. We calculate four different power spectra:

1. We do not account for the contaminants and just average the absolute values squared of the density field in each bin (cf. Eq. (4.28)).

2. We perform a naïve *mode subtraction*, i.e. we subtract off the template, but do not debias the two-point function (cf. Eq. (4.46)).

3. We debias the previous power spectrum by applying Eq. (4.60).

4. We use the full QML estimator with *mode deprojection*.

In the cases (ii) to (iv), we have to assume a prior power spectrum, which we take as equal to the input power. We shall test the effect of this assumption with the next example. As each bin contains modes with a range of different $\mathbf{k}$-values, we have to clarify what we mean by the prior power spectrum $P(k_i)$ for a specific bin. We find that the power spectrum measurements are closest to the input values, when we assume that the input power spectrum $P(k_i)$ is given by the average of the prior power spectrum values for each mode in the respective bin, i.e.

$$P(k_i) \equiv \frac{1}{N_{\mathbb{k}_i}} \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} P(k_\alpha). \tag{4.103}$$

In Fig. 4.2, we can clearly see an increase of power in the bins around $k = 0.01$ in case (i). Subtracting off the template in the naïve way (method (ii)) is biased in the bins affected by the spike. However, this bias is only a 1 part in a thousand effect. Methods (iii) and (iv) both reproduce the input power spectrum well, removing the bias. A significant difference between their error bars cannot be observed. It is therefore sufficient in this case to use the FKP-style estimator we introduced in Sec. 4.5.

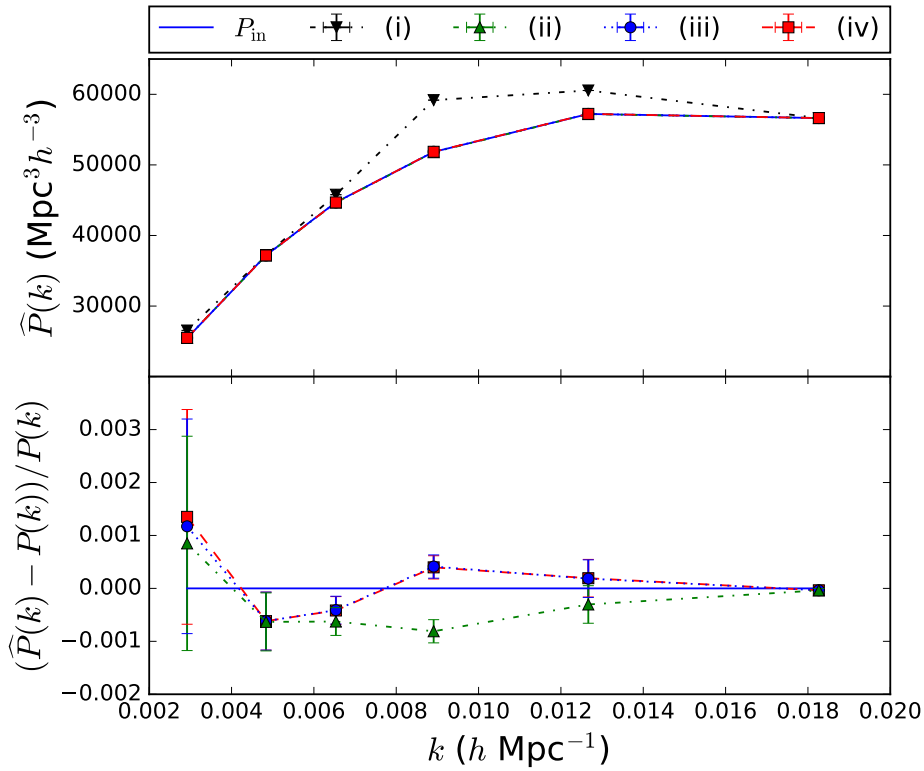

Figure 4.2: Means and standard deviations of the power spectra of 70,000 realisations of Gaussian random fields contaminated with a real Gaussian spike. The top panel shows the input power spectrum as a solid blue line, as well as the power spectra obtained with methods (i)-(iv) as described in Sec. 4.7.1. In the lower panel, we plot fractional errors for methods (ii)-(iv).

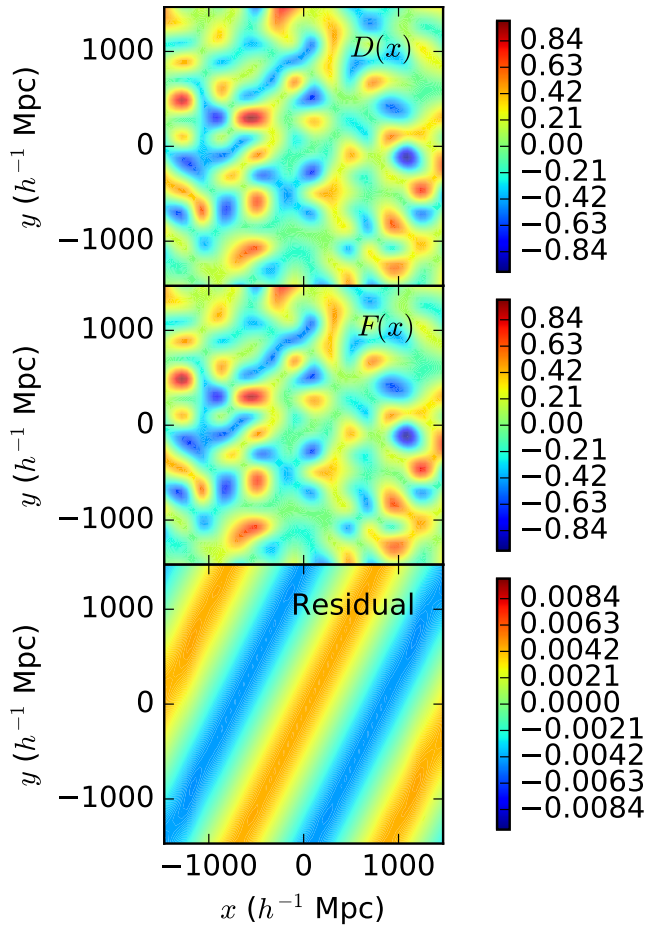

Figure 4.3: This plot is similar to Fig. 4.1, but shows a slice through a field with a single contaminated mode as described in Sec. 4.7.2. The best-fitting $\varepsilon^{(\mathrm{BF})}$ for this particular realisation amounts to 1.005. All panels appear very similar; this is quantified in Fig. 4.4.

4.7.2 Single Contaminated Mode

As a second example we use Eq. (4.60) to construct a contaminant that would lead to a strong bias in the recovered $P(k)$ without the debiasing step. Eq. (4.60) only contains positive quantities and is normalised such that the bias is a value between 0 and 1. 1 corresponds to an unbiased estimate, hence 0 is the maximal bias. This extreme case would be fulfilled if f is large for one mode and 0 otherwise. Therefore, we construct a contaminant that is a large number at the modes corresponding to $\mathbf{k} = \pm(0.003, 0.003, 0.003)h$ Mpc^{-1}. An example of this setup can be found in Fig. 4.3. The top panel again shows an uncontaminated Gaussian random field, the central panel shows the same

field with the contaminant added. The contaminant itself is not as prominent as the one in Fig. 4.1, because this single contaminated mode just adds a long wavelength contribution in real space. The bottom panel shows the residual of the field after subtracting the template.

We measure the same cases (i)-(iv) as in the previous subsection, which we plot in Fig. 4.4. The prior power is again the input power. If we were to apply this to a real survey, we would not know the true power, so we perform a few runs, where we first assume a flat prior power spectrum $P(k) = 1\ \forall k$, and then iteratively compute the power with the power from the previous run as the prior power spectrum. The effect of the prior power spectrum is negligible, because the result in the first step provides the same result as assuming the input power as prior.

The data points for all cases (i)-(iv) are close to the input power in all bins but the second. In the second bin, the power spectrum for case (i) extends beyond the plotted range, chosen to highlight differences between the other approaches. In case (ii), the power is significantly underestimated. The bias amounts to about 2 per cent, i.e. it highly affects measurements where small-$\mathbf{k}$ modes are crucial, such as $f_{\rm NL}$-measurements. The difference between the cases (iii) and (iv) is much smaller, even in this extreme example.

4.8 Summary

In [5], we have considered methods to remove contaminants when measuring the 3D galaxy power spectrum from a given density field, focussing on *mode deprojection* and *mode subtraction*. In order to understand how these are related, we have decomposed the problem into separate steps. In particular we have separated *mode deprojection* from power spectrum estimation - they are often considered together - arguing that this split makes sense given the mathematical similarity of *mode deprojection* and *mode subtraction*. We argue that the QML estimation is not practical for modern surveys with large numbers of observed modes, but that we can apply *mode subtraction* to the FKP-estimator, using the mathematical similarity of *mode deprojection* and *mode subtraction*, thus avoiding having to create large estimator and covariance matrices for all modes. The resulting estimate is biased, but can easily

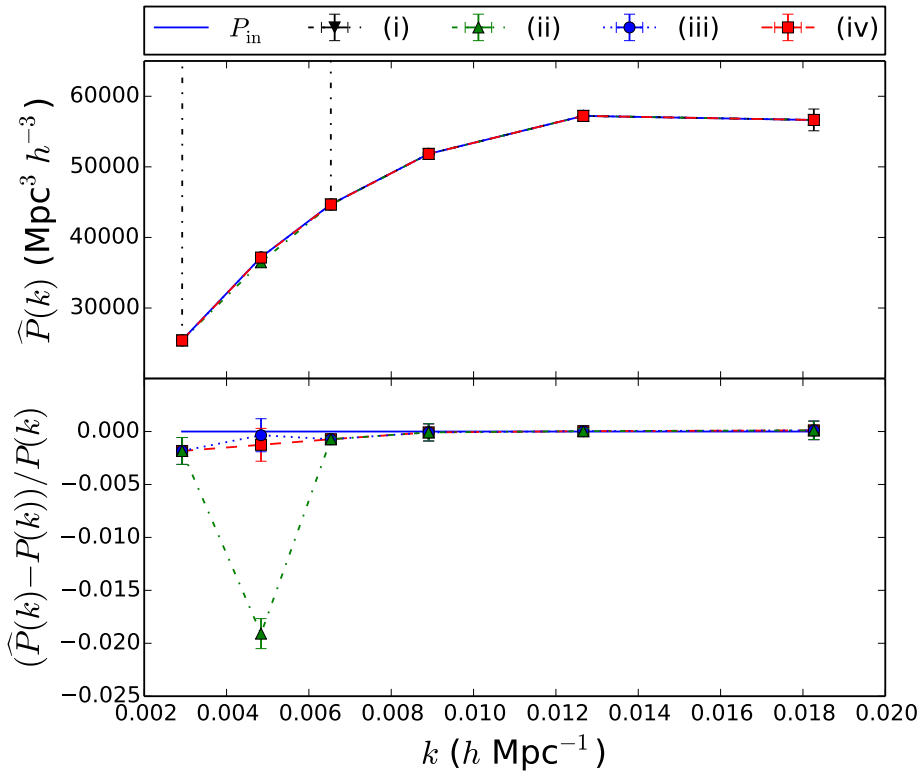

Figure 4.4: Means and standard deviations of the power spectra of 1000 realisations of Gaussian random fields contaminated with Hermitian Gaussian spikes. The red dots represent measurements, where the contamination has not been taken into account. For the blue dots, *mode deprojection* has been used to remove the spikes. For the green dots, we used *debiased mode subtraction*. The solid blue line shows the input power spectrum.

be made unbiased with a simple correction, again that can be implemented without the inversion of large matrices. This correction is easily extended to the case of multiple contaminants and is not affected if the modes are correlated even without the effects of contaminants. The final result of this chapter and [5] is the suggestion that 3D galaxy power spectrum should be estimated using Eq. (4.61),

$$\widehat{P}(k_i) = \frac{1}{N_{\mathbb{k}_i}} \sum_{\mathbf{k}_\alpha} \frac{\left| F(\mathbf{k}_\alpha) - \frac{S_P}{R_P} f(\mathbf{k}_\alpha) \right|^2}{1 - \frac{1}{R_P} \frac{|f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}}. \tag{4.104}$$

While theoretically it is sub-optimal, in practice the degradation of signal is expected to be less than ignoring window effects in the optimisation of mode averaging when using the standard FKP estimator.

In the next chapter we shall see that a rigorous treatment of systematics is important especially at the large scales that are important for $f_{\rm NL}$ measurements. Ross *et al.* [6] found the large scale galaxy power spectrum heavily affected by foreground stars in their $f_{\rm NL}$ analysis using BOSS DR9. However, removing the effect of stars using the standard weighting technique resulted in a power spectrum measurement that is not in good agreement with a model power spectrum for any value of $f_{\rm NL}$. Similarly, the angular power spectrum of photometric SDSS quasars showed excess power at large scales due to systematics [203], which could be removed by a contaminant mitigation technique that is similar to the one presented in this chapter [196]. It is therefore worth applying this technique in $f_{\rm NL}$ measurements from BOSS to see whether the previously found offset can also be explained by systematic contamination.

Chapter 5

Sources of Data Contamination when Constraining Primordial Non-Gaussianity using BOSS DR12 data

In this chapter, I illustrate how the mode subtraction technique can be applied to real BOSS DR12 CMASS data. I consider several possible sources of data contamination and check whether they are affecting the power spectrum measurements at the scales that are interesting when measuring $f_{\rm NL}$. There are two tasks left to be done before the mode subtraction technique described in Sec. 4.4 can be applied to a real galaxy catalogue: The first one is to analyse how each contaminant affects the number of targets that can be observed. This will be described in Sec. 5.2 to 5.9 for each contaminant separately. The second step is to transform this knowledge into a template field f that can be used to mitigate against these contaminants in large-scale clustering measurements. This procedure is the same for every contaminant that affects the observed target density multiplicatively, and I start by developing the general framework in Sec. 5.1. This chapter shows that the mode subtraction technique works well to remove the effect of stars on $f_{\rm NL}$ measurements. I apply the same technique to other known contaminants, such as varying seeing conditions, the airmass and extinction. As the power spectrum does not meet our model power spectrum even after accounting for all of these effects, I describe the work I have

done so far to identify possible additional contaminants in Sec. 5.10, which would enable robust large-scale clustering measurements from these data. This work is still unfinished and I will present what remains to be done in Chapter 6.

5.1 Ingredients for Mode Subtraction in Applications to Real Surveys

In [5] and in the previous chapter, I have shown how an uncorrected density field F can be corrected for several systematic foreground contaminants using the mode subtraction technique with templates f_A. As before, the corrected density field is denoted as D. In Sec. 2.4.2, I have discussed the classical FKP framework. Unlike in Sec. 2.4.2, I explicitly write those weights $w(\mathbf{x})$ that are position dependent, e.g. like the stellar density and seeing weights described in Sec. 2.2. The corrected density field then reads

$$D(\mathbf{x}) = w_{\mathrm{FKP}}(\mathbf{x}) \frac{w(\mathbf{x}) n_{\mathrm{g}}(\mathbf{x}) - \alpha n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}}, \tag{5.1}$$

where the random number density $n_{\mathrm{r}}(\mathbf{x})$, the normalisation I_2 and the FKP weight $w_{\mathrm{FKP}}(\mathbf{x})$ are the same as in Eq. 2.23. Likewise, α is a parameter that ensures that the density is on average

$$\sum_{\mathbf{x}} D(\mathbf{x}) = 0. \tag{5.2}$$

Only $n_{\mathrm{g}}(\mathbf{x})$ has to be understood differently to Sec. 2.4.2 as the position dependent weights are written explicitly. Note that the galaxies are still weighted for fibre collisions and redshift failures that are not position dependent effects. The ensemble average of the absolute value squared of the Fourier transform $D(\mathbf{k})$ provides an estimate of the power spectrum of each mode, but is overestimated by a constant shot noise term

$$\left\langle |D(\mathbf{k})|^2 \right\rangle = P(\mathbf{k}) + P_{\mathrm{shot}} \text{ [109]}. \tag{5.3}$$

This shot noise term is given in Eq. (2.27).

The aim of this section is to translate Eq. (2.23), or in fact Eq. (5.1), into the debiased mode subtraction framework, i.e. writing D in terms of F and f,

and thence, identifying how F and f are related to the weight $w(\mathbf{x})$ and the observed galaxy and random counts, n_g and n_r, respectively. Eq. (5.1) reflects the fact that most known contaminants affect the observed galaxy density multiplicatively. To give an example, a bright star obscures a fraction of the targets in its angular vicinity, i.e. the number of targets that are not observed depends on the number of targets that actually exist. In spite of that, the underlying assumption behind mode subtraction is given by Eq. (4.1), i.e. that a template of the contaminant f can be subtracted from the observation F to obtain a "clean" density field D. This apparent contradiction can be lifted by moving the weights from acting on the observed galaxy density to letting their inverse act on the random catalogue. Mathematically speaking, we divide both the numerator and denominator of the fraction in Eq. (2.23) by the weights and obtain

$$D(\mathbf{x}) = w'_\mathrm{FKP}(\mathbf{x}) \frac{n_\mathrm{g}(\mathbf{x}) - \alpha w^{-1}(\mathbf{x}) n_\mathrm{r}(\mathbf{x})}{\sqrt{I_2}}, \tag{5.4}$$

where $w'_\mathrm{FKP}(\mathbf{x}) \equiv w_\mathrm{FKP}(\mathbf{x}) w(\mathbf{x})$ is an updated FKP weight. We know the parameter α well as it has been chosen to generate the random catalogue. The amplitude of the contaminant is less known, so we split the second term into a part without weights and into another one with the weights, such that we can introduce another free parameter $\varepsilon \approx \alpha$ that we can marginalise over:

$$D(\mathbf{x}) = w'_\mathrm{FKP}(\mathbf{x}) \frac{n_\mathrm{g}(\mathbf{x}) - \alpha n_\mathrm{r}(\mathbf{x}) - \varepsilon [w^{-1}(\mathbf{x}) - 1] n_\mathrm{r}(\mathbf{x})}{\sqrt{I_2}}. \tag{5.5}$$

In order to fulfil Eq. (2.25), we can constrain

$$\alpha = \frac{\sum_\mathbf{x} [n_\mathrm{g}(\mathbf{x}) - \varepsilon [w^{-1}(\mathbf{x}) - 1] n_\mathrm{r}(\mathbf{x})]}{\sum_\mathbf{x} n_\mathrm{r}(\mathbf{x})}. \tag{5.6}$$

Recalling that in the case of not including weights, we have

$$\alpha_\mathrm{FKP} = \frac{\sum_\mathbf{x} n_\mathrm{g}(\mathbf{x})}{\sum_\mathbf{x} n_\mathrm{r}(\mathbf{x})}, \tag{5.7}$$

we can split α into two terms: one independent and one proportional to ε:

$$\begin{aligned} \alpha &= \alpha_\mathrm{FKP} - \varepsilon \frac{\sum_\mathbf{x} [w^{-1}(\mathbf{x}) - 1] n_\mathrm{r}(\mathbf{x})}{\sum_\mathbf{x} n_\mathrm{r}(\mathbf{x})} \\ &= \alpha_\mathrm{FKP} - \varepsilon \left[\frac{\sum_\mathbf{x} w^{-1}(\mathbf{x}) n_\mathrm{r}(\mathbf{x})}{\sum_\mathbf{x} n_\mathrm{r}(\mathbf{x})} - 1 \right]. \end{aligned} \tag{5.8}$$

Thus, we can write Eq. (5.5) as

$$\begin{aligned} D(\mathbf{x}) =& w'_{\mathrm{FKP}}(\mathbf{x}) \frac{n_{\mathrm{g}}(\mathbf{x}) - \alpha_{\mathrm{FKP}} n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}} \\ & - \varepsilon w'_{\mathrm{FKP}}(\mathbf{x}) \frac{[\frac{\sum_{\mathbf{x}} w^{-1}(\mathbf{x}) n_{\mathrm{r}}(\mathbf{x})}{\sum_{\mathbf{x}} n_{\mathrm{r}}(\mathbf{x})} + w^{-1}(\mathbf{x})] n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}}. \end{aligned} \tag{5.9}$$

In the framework of [5] and the previous chapter, we can identify

$$F(\mathbf{x}) = w'_{\mathrm{FKP}}(\mathbf{x}) \frac{n_{\mathrm{g}}(\mathbf{x}) - \alpha_{\mathrm{FKP}} n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}} \tag{5.10}$$

and

$$f(\mathbf{x}) = \frac{w'_{\mathrm{FKP}}(\mathbf{x}) n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}} \left[\frac{\sum_{\mathbf{x}} w^{-1}(\mathbf{x}) n_{\mathrm{r}}(\mathbf{x})}{\sum_{\mathbf{x}} n_{\mathrm{r}}(\mathbf{x})} + w^{-1}(\mathbf{x}) \right]. \tag{5.11}$$

Thus, the uncorrected field F is similar to the FKP field without systematic weights, but with a modified FKP weight, and the template is the expected correction that has to be subtracted based on expectation on the galaxy number density from the random catalogue and the systematic weight.

One big advantage of the mode subtraction framework is that it can be easily extended to N_{cont} different contaminant templates. Different contaminants can be included in the traditional weighting scheme by just multiplying n_{g} with a weight for each contaminant one can imagine. However, to do so, $w(\mathbf{x})$ has to be known exactly. If the functional form of the weight is not exactly known, the mode subtraction framework allows to include more than one template for each contaminant. Having a free parameter for each template then naturally mitigates the templates that are supported by the data, and that in the right amount. When dealing with more than one template, it is convenient to model the total weight in terms of contributions from each contaminant $Y_A(\mathbf{x})$:

$$w'_{\mathrm{FKP}}(\mathbf{x}) = \frac{w_{\mathrm{FKP}}(\mathbf{x})}{1 + \sum_{A=1}^{N_{\mathrm{cont}}} Y_A(\mathbf{x})}, \tag{5.12}$$

where $\langle Y_A(\mathbf{x}) \rangle = 0$ for all contaminants A. Eq. (2.23) then reads

$$\begin{aligned} D(\mathbf{x}) =& w_{\mathrm{FKP}}(\mathbf{x}) \frac{\prod_A w_A(\mathbf{x}) n_{\mathrm{g}}(\mathbf{x}) - \alpha n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}} \\ =& w'_{\mathrm{FKP}}(\mathbf{x}) \frac{n_{\mathrm{g}}(\mathbf{x}) - \alpha \left(1 + \sum_{A=1}^{N_{\mathrm{cont}}} Y_A(\mathbf{x})\right) n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}}. \end{aligned} \tag{5.13}$$

As said, we introduce free parameters ε_A for each contaminant to take the uncertainties of each of their amplitudes into account. We then have

$$D(\mathbf{x}) = w'_{\mathrm{FKP}}(\mathbf{x}) \frac{n_{\mathrm{g}}(\mathbf{x}) - \alpha n_{\mathrm{r}}(\mathbf{x}) - \sum_{A=1}^{N_{\mathrm{cont}}} \varepsilon_A Y_A(\mathbf{x}) n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}}. \tag{5.14}$$

To ensure again that the density field fulfils $\langle D(\mathbf{x}) \rangle = 0$, we need

$$\alpha = \alpha_{\mathrm{FKP}} - \sum_{A=1}^{N_{\mathrm{cont}}} \varepsilon_A \frac{\langle Y_A n_{\mathrm{r}} \rangle}{\langle n_{\mathrm{r}} \rangle}. \tag{5.15}$$

Recollecting all ε_A terms yields

$$D(\mathbf{x}) = w'_{\mathrm{FKP}}(\mathbf{x}) \frac{n_{\mathrm{g}}(\mathbf{x}) - \alpha_{\mathrm{FKP}} n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}} - \sum_{A=1}^{N_{\mathrm{cont}}} \varepsilon_A \frac{Y_A(\mathbf{x}) - \frac{\langle Y_A n_{\mathrm{r}} \rangle}{\langle n_{\mathrm{r}} \rangle}}{\sqrt{I_2}} w'_{\mathrm{FKP}}(\mathbf{x}) n_{\mathrm{r}}(\mathbf{x}), \tag{5.16}$$

where we can read off

$$F(\mathbf{x}) = w'_{\mathrm{FKP}}(\mathbf{x}) \frac{n_{\mathrm{g}}(\mathbf{x}) - \alpha_{\mathrm{FKP}} n_{\mathrm{r}}(\mathbf{x})}{\sqrt{I_2}} \tag{5.17}$$

and

$$f_A(\mathbf{x}) = w'_{\mathrm{FKP}}(\mathbf{x}) \frac{Y_A(\mathbf{x}) - \frac{\langle Y_A n_{\mathrm{r}} \rangle}{\langle n_{\mathrm{r}} \rangle}}{\sqrt{I_2}} n_{\mathrm{r}}(\mathbf{x}) \tag{5.18}$$

in the same way as we did to obtain Eq. (5.10) and (5.11). The field F is again similar to the FKP field. Each $Y_A(\mathbf{x})$ describes how contaminant A affects the number of galaxies in a certain region around the point $\mathbf{x}$. Although the effect of most contaminants is expected to be relative to F, this section has shown how absolute templates f can be constructed using the expected number of galaxies from the random catalogue n_{r}. Each template is an estimate of the absolute number density that has to be added or subtracted to correct for the contaminant in question. The following sections will show how the Y_A are obtained in practice for specific contaminants.

5.2 Stellar Density Counts

As already mentioned in Sec. 2.2, the stars affect galaxy clustering measurement through obscuration, selection bias and confusion. For spectroscopic surveys, we expect confusion to be negligible. Hence, the higher the stellar density is, the lower is the number of galaxies we observe, as found by Ross *et*

al. [6]. It has been confirmed in [113] that foreground stars cause the strongest systematic error in BOSS CMASS data. The foreground stars are mostly within our own Galaxy, which can be described as a half-sky mode in Fourier space. Thus, the foreground stars add large-scale power to the actual galaxy power spectrum in a very similar way as a positive $f_{\rm NL}$ signal [6]. As stellar densities have been reported as the main source of systematic error in BOSS CMASS data [6], it is the first systematic we want to confront using the mode subtraction technique.

In order to be able to compare the results using mode subtraction to the results using the weights as in [113], we generate the first set of templates in a similar way as Ross *et al*'s weights. We start with the SDSS DR8 star catalogue in Hierarchical Equal Area isoLatitude Pixelization of a sphere (HEALPix, cf. Fig. 5.1). Both the BOSS data and the catalogue of stars that we use thus come from the same survey, hence having similar footprints (there are additional stripes running through the Milky Way in the catalogue of stars that are masked out in Fig. 5.1) and instrumental systematics. The advantage of using HEALPix is that the number count in each pixel is proportional to the angular density of stars, because all pixels cover equally sized areas. Another advantage of HEALPix in general is that the resolution of a map can be easily changed due to the hierarchical ordering of the cells. The resolution can be identified by the "number of pixels per side" ($N_{\rm side}$), which is related to the total number of pixels on the sphere $N_{\rm pix}$ by $N_{\rm pix} = 12N_{\rm side}^2$. The resolution in Fig. 5.1 is $N_{\rm side} = 256$. Ross *et al.* reduced this to $N_{\rm side} = 128$ to reduce the shot noise in the stellar data. We reduce the resolution to $N_{\rm side} = 64$, because our FKP grid has only 128^3 grid points and there is no need to resolve the stars within a grid cell.

The number count and the number density of stars are only proportional to each other in cells that are entirely within the survey footprint. In the original map of stars we could see prominent edge effects, because HEALPix cells on the edges are only partially filled. We reduce this effect by assuming the following for the completeness of each HEALPix-pixel: pixels in the $N_{\rm side} = 256$-map that have only non-zero neighbours are complete, and for every neighbour that is zero, we assume that the pixel in question is 25 per cent less complete, such that cells, whose neighbours are all empty, are also empty. We generate a

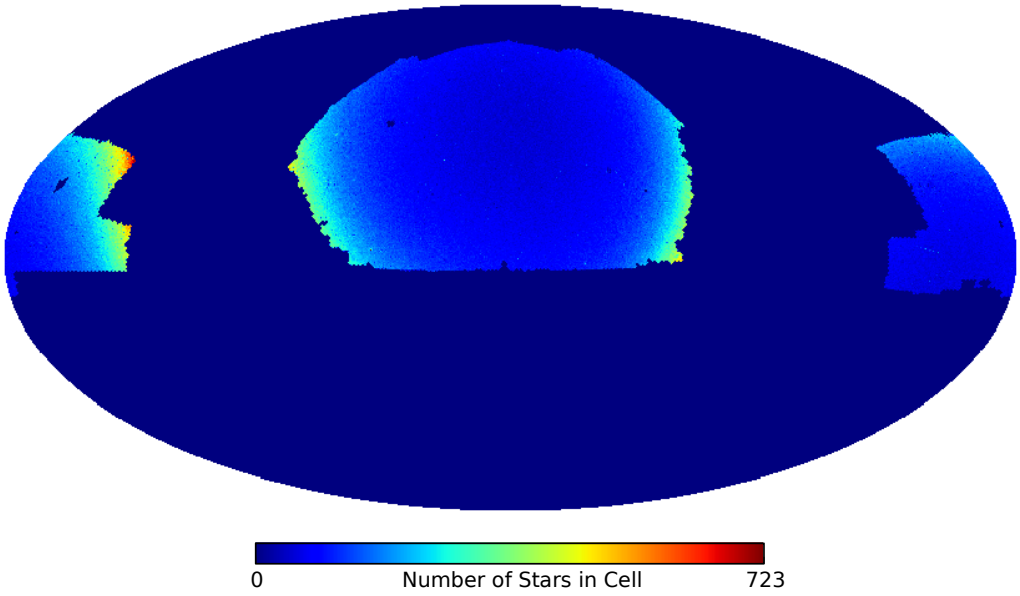

Figure 5.1: The distribution of stars in the 8$^{\text{th}}$ SDSS data release in HEALPix. The map is presented in Mollweide projection, equatorial coordinates, astronomical orientation, i.e. east is left, and it has been rotated by 180° to show the NGC in the centre. The catalogue includes stars in areas that were not targeted by BOSS. These are masked out by setting the number of stars equal to zero in the respective cells.

HEALPix map with these completeness values and reduce its resolution to $N_{\text{side}} = 64$ in the same way as the map of stars. We divide the number count of each partially filled pixel by this resulting completeness map, such that we obtain a map whose entries are proportional to the number density of stars.

To obtain a template to mitigate stellar contamination, we pixelise the BOSS DR12 CMASS galaxy and random catalogues in the same way as the stars. We then bin pixels by their number count of stars and average $\frac{n_{\text{g}}}{\alpha n_{\text{r}}}$ in the cells corresponding to the cells in each stellar number count bin. In Fig. 5.2, we observe a similar trend as in [113]: In pixels containing less than 1500 stars, we observe more galaxies than we expect from the random catalogue, whereas in pixels with more than 2000 stars, we seem to miss galaxies in the observations.

For the effect of obscuration it is reasonable to assume that galaxies with

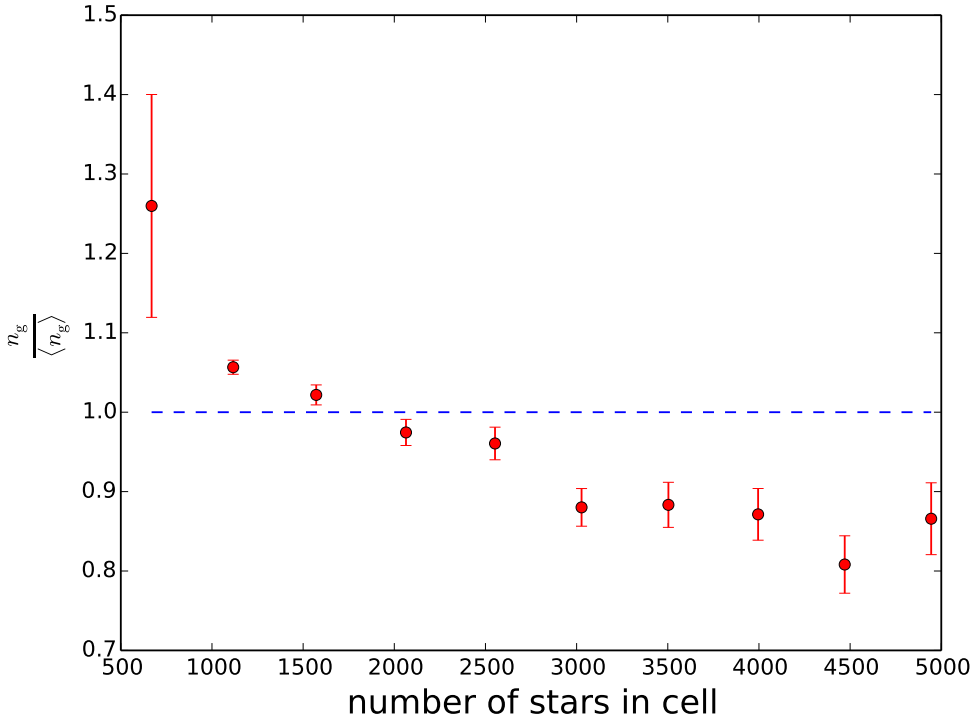

Figure 5.2: The average fraction of observed galaxies to expected galaxies from the random catalogue in cells with given numbers of stars.

different magnitudes are affected differently by foreground stars. Ross *et al.* therefore made plots similar to Fig. 5.2, but with the galaxies split into subsamples by their i-band magnitudes within a $2''$ aperture radius i_{fib2}. We also follow that procedure to generate Fig. 5.3, where we can see that galaxy observations are affected very differently according to their brightness: galaxies with $i_{\mathrm{fib2}} < 20.6$ do not cause a significant deviation between the expected and observed number of galaxies (the best-fitting $\frac{n_{\mathrm{g}}}{\langle n_{\mathrm{g}}\rangle}$-line for $i_{\mathrm{fib2}} < 20.3$ even goes up as the number of stars increases, but the error bars of all data points are so large that they all agree with $\frac{n_{\mathrm{g}}}{\langle n_{\mathrm{g}}\rangle} = 1$). For fainter galaxies, the best-fitting $\frac{n_{\mathrm{g}}}{\langle n_{\mathrm{g}}\rangle}$-lines are negative and are steeper the larger the galaxies' magnitudes (i.e. the fainter they are). This meets our expectation, because part of the stellar contamination effect is due to obscuration. To obtain the Ross *et al.* weights for a given galaxy, one finds a linear regression line

$$\frac{n_{\mathrm{g}}}{\langle n_{\mathrm{g}}\rangle}^{(1)}(n_{\mathrm{stars}}, i_{\mathrm{fib2}}) \equiv C_0(i_{\mathrm{fib2}}) + C_1(i_{\mathrm{fib2}}) n_{\mathrm{stars}} \tag{5.19}$$

through the data points for the i_{fib2}- range into which the galaxy's i_{fib2}- magnitude falls. This function is then evaluated at the number of stars in the pixel which contains the galaxy. The weight for the ith galaxy in the survey with magnitude i_{fib2} at right ascension α_i and δ_i is then given by

$$w_i = \frac{1}{\frac{n_{\mathrm{g}}}{\langle n_{\mathrm{g}}\rangle}^{(1)}(n_{\mathrm{stars}}(\alpha_i, \delta_i), i_{\mathrm{fib2}})}. \tag{5.20}$$

The mode subtraction technique, however, requires a template field in configuration space or Fourier space, so one cannot generate the template using the i_{fib2}-values of individual galaxies. Instead, we average i_{fib2} in redshift slices (cf. Fig. 5.4), because we expect galaxies to be fainter the further away they are, and we assign the averages to template grid cells according to their redshifts. Apart from this, the weights entering Eq. (5.11) are obtained in the same way as Ross *et al.*'s weights:

$$w(\mathbf{x}) = w(z, \alpha, \delta) = \frac{1}{\frac{n_{\mathrm{g}}}{\langle n_{\mathrm{g}}\rangle}^{(1)}(n_{\mathrm{stars}}(\alpha, \delta), \langle i_{\mathrm{fib2}}\rangle(z))}. \tag{5.21}$$

We compute the BOSS DR12 CMASS NGC power spectrum using the FKP-style estimator with debiased mode subtraction as described in Sec. 4.5. The resulting power spectrum is shown in Fig. 5.5. We compute the power

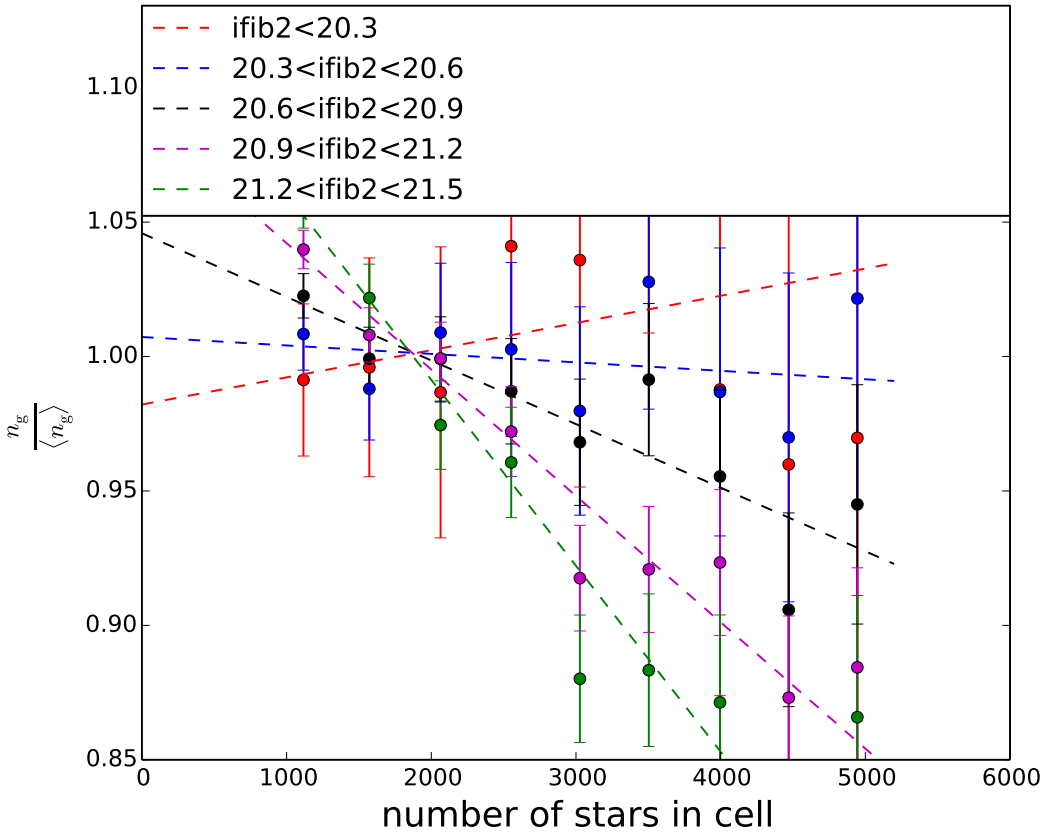

Figure 5.3: Plot similar to Fig. 5.2, but points in different colours are for different subsamples of galaxies with different $i_{\rm fib2}$ ranges. The dashed lines are the best-fitting lines through the data points.

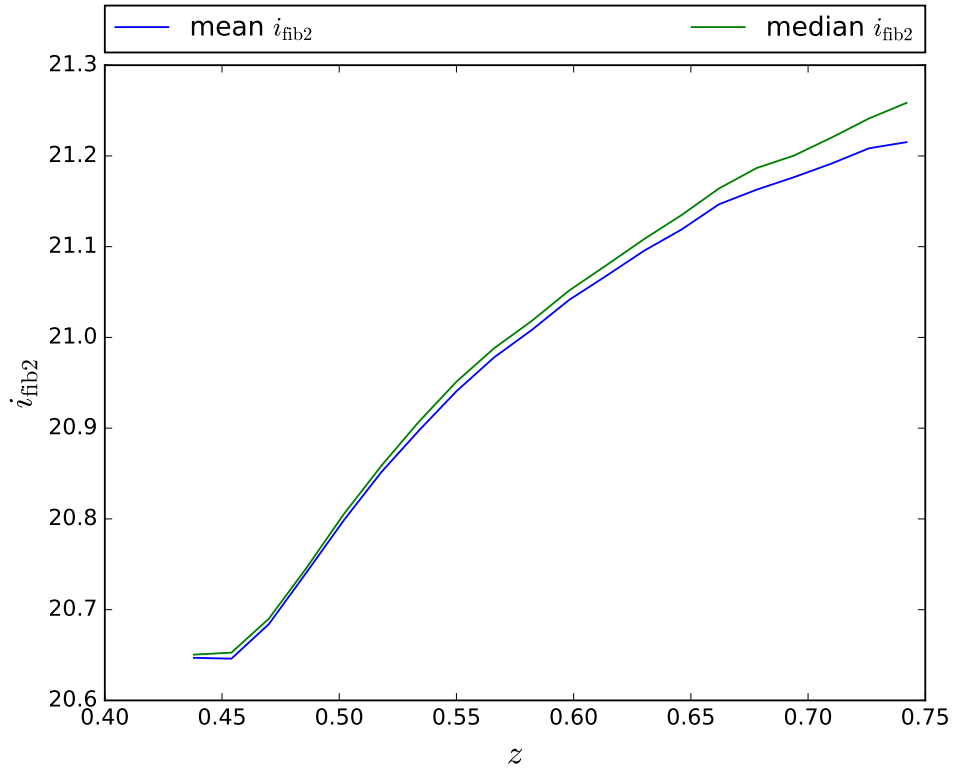

Figure 5.4: Mean and median values of the i_{fib2}-magnitudes of BOSS DR12 CMASS galaxies at given redshifts z.

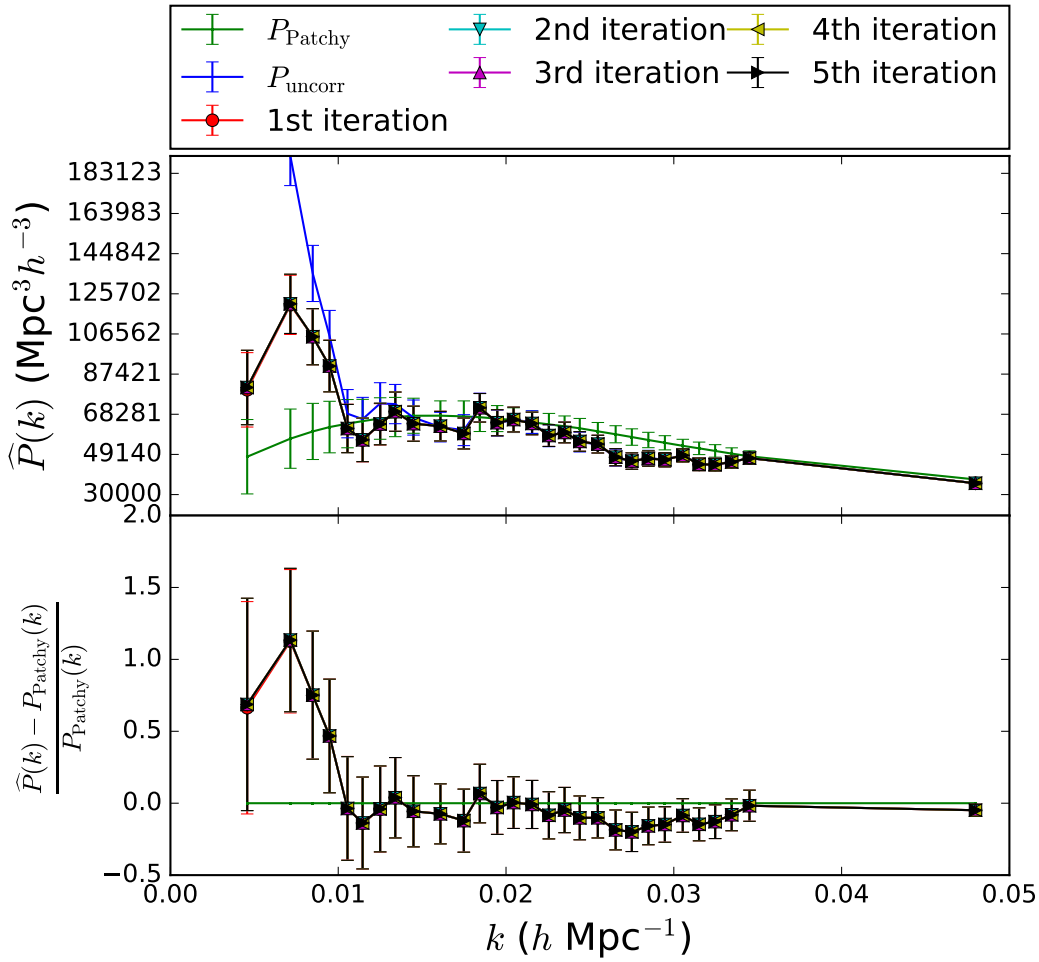

Figure 5.5: The power spectra of the uncorrected BOSS DR12 CMASS NGC galaxies $P_{\rm uncorr}$ and after 1 to 5 iterations of the debiased mode subtraction procedure, compared to the average power spectrum of the PATCHY mocks $P_{\rm Patchy}$.

spectra of 2048 PATCHY mock catalogues (cf. Sec. 3.2), not only to estimate the variance of the power spectrum, but also to use the average of the mock power spectra as the prior power spectrum needed in the debiasing step. To check that also for real data the choice of the prior power spectrum for the debiasing step does not matter, we use the average of the PATCHY power spectra as the input prior power spectrum for the first run of our error mitigation procedure, and then, we iterate by rerunning the programme with the previous output power spectrum as the prior for the next run. We cannot see any significant difference between the power spectra of the five runs plotted in Fig. 5.5. Furthermore, all of the five spectra agree well with the power spectrum obtained by mitigating the stellar density contamination using Ross *et al.*'s weights. This shows that our method can successfully remove the stellar contamination to first order. On the other hand, we also observe a significant discrepancy between the average PATCHY power and our result. This means, that our findings do not favour the ΛCDM universe with $f_{\mathrm{NL}} = 0$ that has been used to generate the PATCHY mocks. However, comparing our result with the prediction of Fig. 1.9, one can observe that our power spectrum does not have the shape one would expect for a positive f_{NL}-signal, which would diverge towards $+\infty$ as $k \to 0$. A similar discrepancy has been identified in the BOSS DR9 f_{NL}-analysis [6]. The remainder of this chapter is concerned with attempts to explain this discrepancy. The advantage of using our template based error mitigation technique compared to using weights is that it is straight-forward to extend the error mitigation by including new templates. This will be done extensively in the next sections.

5.3 Higher Order Templates

We can see in Fig. 5.2 that there is a trend of seeing less galaxies in directions with higher numbers of stars. However, it is not obvious whether the relationship between not observed galaxies and the number of stars is linear, as it was assumed by Ross *et al.* and when we generated the templates in the previous section. As said at the end of the previous section, the big advantage of our template based method is that we can add more templates for any form of contamination we have a reason to include. To liberate ourselves from the

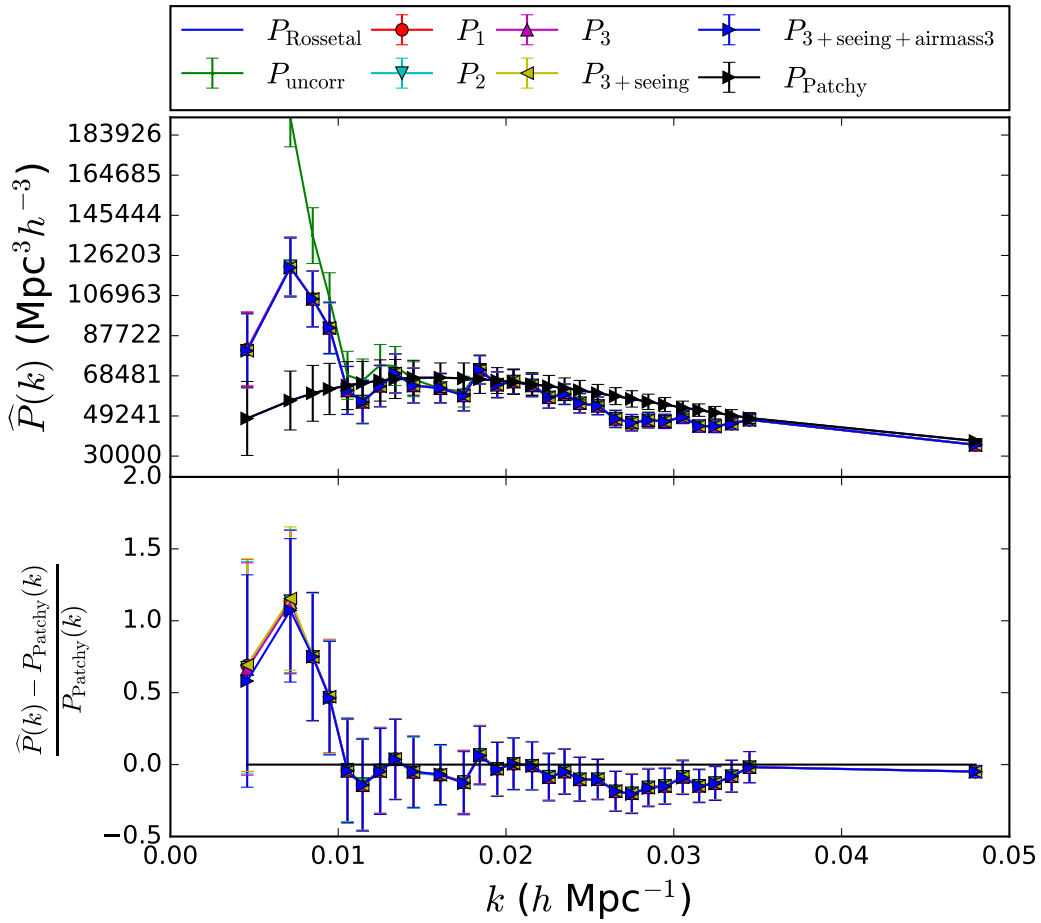

Figure 5.6: The power spectra of the uncorrected BOSS DR12 CMASS NGC galaxies and after applying debiased mode subtraction with a first order stellar template P_1, and additional second order template P_2, a third order template P_3, as well as also using a seeing template $P_{3+\rm seeing}$ and an 3 first, second and third order airmass templates $P_{3+\rm seeing+airmass3}$, compared to the average power spectrum of the PATCHY mocks.

linear assumption, we fit higher order polynomials

$$\frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(N)}(n_{\rm stars}, i_{\rm fib2}) \equiv \sum_{\ell=0}^{N} C_\iota(i_{\rm fib2}) n_{\rm stars}^{\ell} \tag{5.22}$$

and build new templates using Eq. (5.12)-(5.18), where

$$Y_1(z,\alpha,\delta) \equiv \frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(1)}(n_{\rm stars}(\alpha,\delta), \langle i_{\rm fib2}\rangle(z)) - 1 \tag{5.23}$$

is given by the same linear fits as plotted in Fig. 5.3,

$$Y_2(z,\alpha,\delta) \equiv \frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(2)}(n_{\rm stars}(\alpha,\delta), \langle i_{\rm fib2}\rangle(z)) - \frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(1)}(n_{\rm stars}(\alpha,\delta), \langle i_{\rm fib2}\rangle(z)) \tag{5.24}$$

is given by the difference between the quadratic and linear fits,

$$Y_3(z,\alpha,\delta) \equiv \frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(3)}(n_{\rm stars}(\alpha,\delta), \langle i_{\rm fib2}\rangle(z)) - \frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(2)}(n_{\rm stars}(\alpha,\delta), \langle i_{\rm fib2}\rangle(z)) \tag{5.25}$$

is the difference between the cubic and quadratic fits, etc. In this way, correlations between the Y_A are reduced. Therefore, templates that correspond to expansion orders that are actually not in the data obtain naturally negligible best-fitting values of $\varepsilon^{(\rm BF)}$. We use Eq. (4.100) to find the best-fitting amplitude of each template, which are listed in Tab. 5.1. The amplitude for the first and third order templates, $\varepsilon_1^{(\rm BF)}$ and $\varepsilon_3^{(\rm BF)}$, respectively, do not change significantly when other templates are fitted at the same time. The second order amplitude changes, but it is always at least one order of magnitude less than $\varepsilon_1^{(\rm BF)}$ and $\varepsilon_3^{(\rm BF)}$. The fourth order amplitude $\varepsilon_4^{(\rm BF)}$ is also much smaller, which suggests that the true relationship between observed number of stars and galaxies is odd. We compute the debiased mode subtracted power spectra, which we plot in Fig. 5.6. We observe that, even though $\varepsilon_3^{(\rm BF)}$ is almost as large as $\varepsilon_1^{(\rm BF)}$, including the third order stellar contamination template, or in fact any other higher order template, does not change the resulting power spectrum significantly.

5.4 Seeing

The effect of seeing has been addressed in Sec. 2.2. As the effect is purely angular and does not depend on intrinsic properties of the galaxies, we can

Table 5.1: Best-fitting contamination amplitudes $\varepsilon^{(\mathrm{BF})}$ for a power spectrum measurement using different numbers of stellar templates.

order	1 template	2 templates	3 templates	4 templates
1^{st}	0.0071	0.0071	0.0072	0.0073
2^{nd}		0.0008	0.0001	0.0008
3^{rd}			0.0055	0.0054
4^{th}				-0.0001

Table 5.2: Best-fitting contamination amplitudes for a power spectrum measurement using three stellar templates and seeing weights (left) and replacing the seeing weights by seeing templates (right).

template	3 stellar templates	3 star + 1 seeing template
stars 1^{st} order	0.00719	0.00739
stars 2^{nd} order	0.00009	-0.00002
stars 3^{rd} order	0.00552	0.00576
seeing		-0.00237

build our templates by directly inserting the Ross *et al.* weights into Eq. (5.11), which are provided in the galaxy catalogue file and which are mapped in Fig. 5.7. The yellow lines and left facing triangles in Fig. 5.6 represents the power spectrum I obtained after replacing the direct application of the seeing weights with seeing templates, and using three templates for the stellar contamination. The plot shows that there is no difference between the results obtained using the weights and those obtained using templates based on the same weights. This shows that our method works and that the discrepancy between the measured and the theoretical power spectra are not due to using different error mitigation techniques inconsistently.

5.5 Airmass

Another variation in astronomical observations due to the Earth's atmosphere arises because light coming from a source close to the horizon has to travel through more atmosphere than the light coming from a source close to the

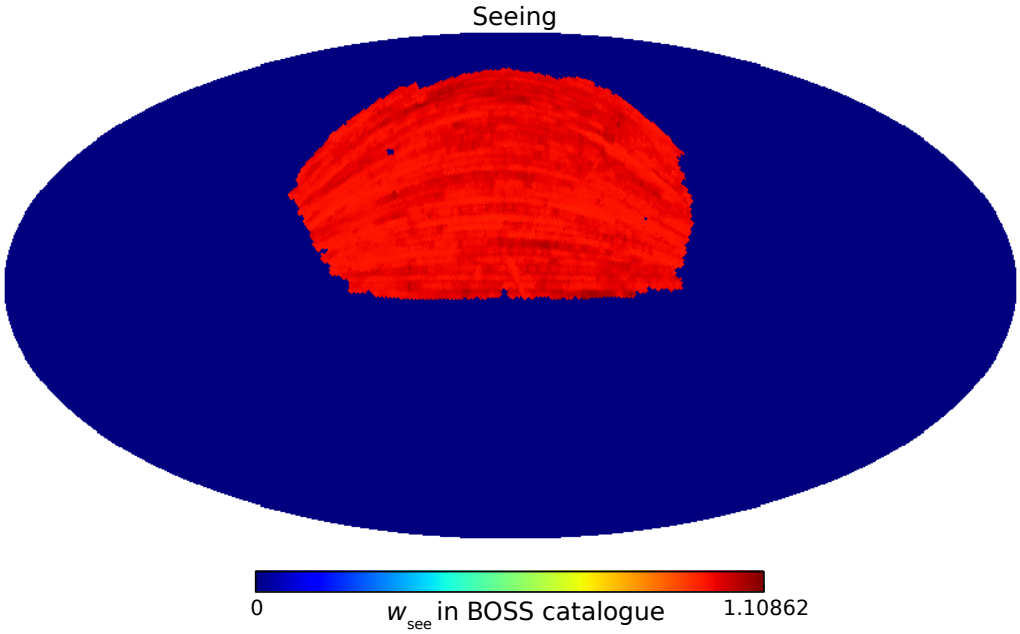

Figure 5.7: The seeing condition weights of BOSS DR12 CMASS NGC in HEALPix. The map is presented in Mollweide projection, equatorial coordinates and astronomical orientation, but it is rotated by 180° such that the region observed is in the centre of the map.

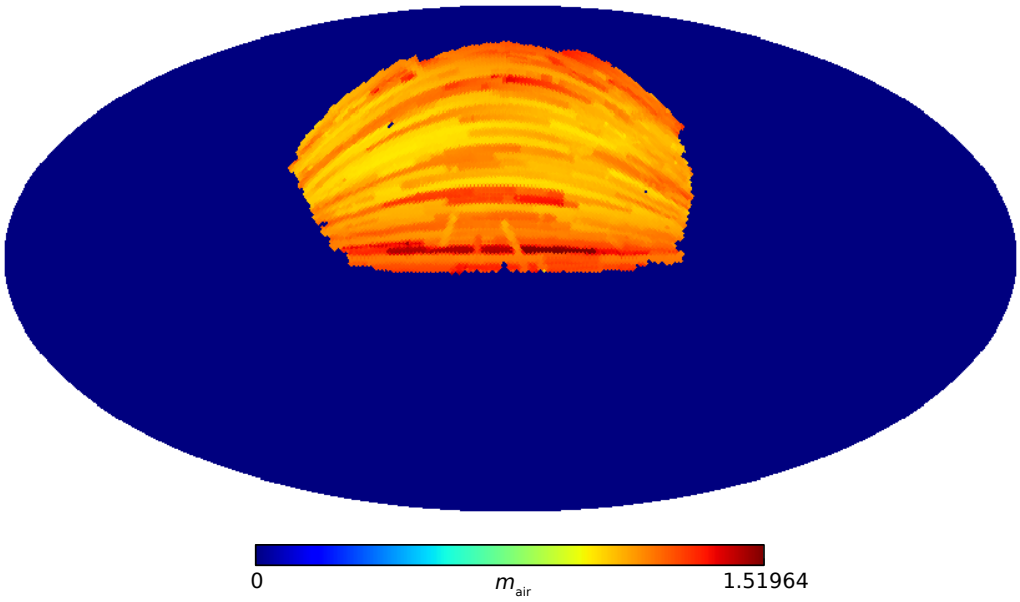

Figure 5.8: The airmass m_{air} (cf. Eq. 5.26) in the NGC sub-sample of BOSS DR12 CMASS in HEALPix. The map is presented in Mollweide projection, equatorial coordinates and astronomical orientation, but it is rotated by 180° such that the region observed is in the centre of the map.

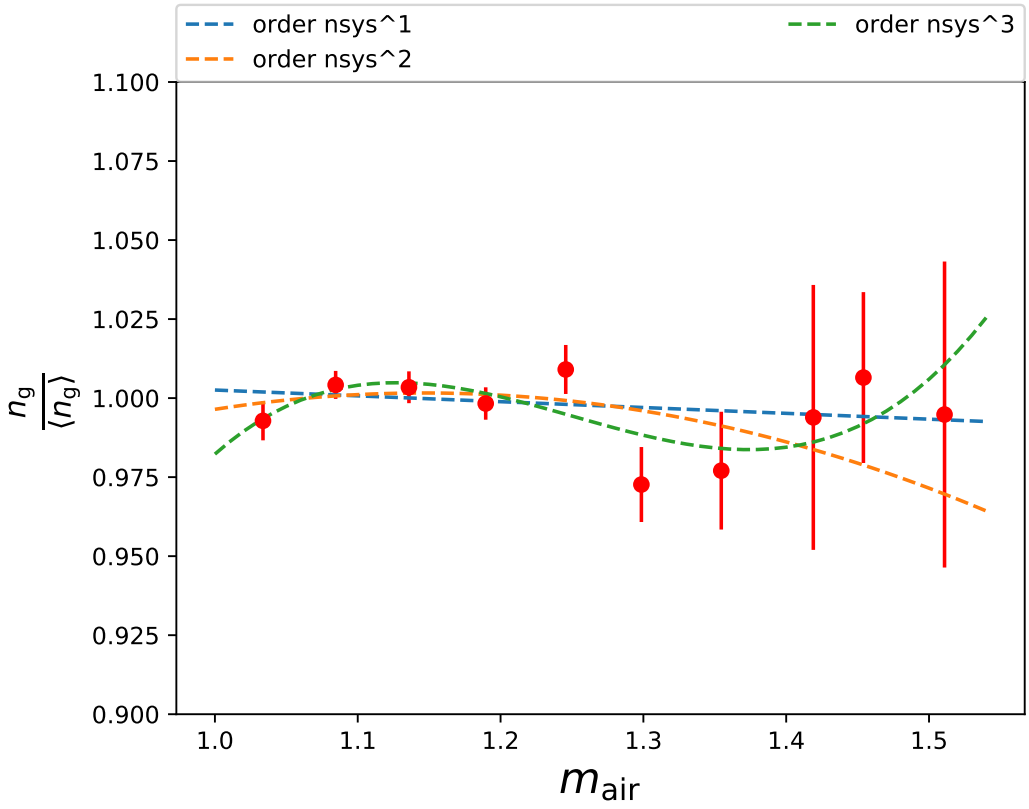

Figure 5.9: The relationship between observed galaxy density and airmass (cf. (5.26)).

zenith. The effect is quantified by the airmass

$$m_{\rm air} \equiv \frac{\int {\rm d}s\rho}{\int {\rm d}s_{\rm zenith}\rho}, \tag{5.26}$$

which is the column density, i.e. the integral over the mass density of the atmosphere ρ, along the line of sight s divided by the zenith column density. The mass density depends on time varying quantities such as the temperature and other weather phenomena. Furthermore, the angle between the zenith and the line of sight changes with the seasons. Hence, the amount of photons to be scattered or absorbed varies with both position and observing times, effectively varying the depth of the survey and the magnitude error. Information about the airmass is provided in the random catalogue. A map can be found in Fig. 5.8. It prominently shows the drift scanning strategy of SDSS. The airmass does not change much along SDSS scanning stripes, as the telescope remains stationary along a great circle, but there are sharp leaps from stripe to stripe, which can cause spurious fluctuations in the density field. A plot similar to Fig. 5.2 that relates $n_{\rm g}/\langle n_{\rm g}\rangle$ to the airmass is shown in Fig. 5.9, where the data points are consistent with $n_{\rm g}/\langle n_{\rm g}\rangle = 1$ for all values of airmass. The linear fit $n_{\rm g}/\langle n_{\rm g}\rangle^{(1)}$ through Fig. 5.9 is almost constantly equal to one. The quadratic fit $n_{\rm g}/\langle n_{\rm g}\rangle^{(2)}$ shows a slight negative trend at larger airmasses and the cubic fit $n_{\rm g}/\langle n_{\rm g}\rangle^{(3)}$ looks like an over-fit. Ross *et al.* [113] made a similar analysis including a χ^2 null test. Based on that test, they state that corrections for such a marginally significant effect are ill- advised. However, they recommend to reconsider this choice for any future studies of the clustering of BOSS galaxies at the largest scales, what this section is about.

We proceed in a similar way as for the stars. We fit the three polynomials

$$\frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(N)}(m_{\rm air}) = \sum_{i=0}^{N} C_i m_{\rm air}^i \tag{5.27}$$

to the data that we have plotted in Fig. 5.2. We define

$$Y_{{\rm am},N}(\alpha,\delta) \equiv \frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(N)}(m_{\rm air}(\alpha,\delta)) - \frac{n_{\rm g}}{\langle n_{\rm g}\rangle}^{(N-1)}(m_{\rm air}(\alpha,\delta)), \tag{5.28}$$

which we again insert into Eq. (5.12) and (5.18) to obtain templates to mitigate the effect of the airmass. To do so, we perform the mode subtraction method and find the best-fitting template amplitudes given in Tab. 5.3. The third

Table 5.3: Best-fitting contamination amplitudes for a power spectrum measurement using three stellar templates (left) and additionally three airmass templates (right).

template	stellar templates	stellar and airmass templates
stars 1^{st} order	0.0072	0.0061
stars 2^{nd} order	0.0001	-0.0013
stars 3^{rd} order	0.0055	0.0042
airmass 1^{st} order		-0.0014
airmass 2^{nd} order		0.0202
airmass 3^{rd} order		-0.0003

order template indeed is not favoured by the data and obtains a very small amplitude, suggesting that the third order describes noise rather than an actual effect of the airmass on the observed galaxy density. The first order is almost constant and equal to one, so it cannot be expected to significantly change the resulting power spectrum. The second order template, however, has the largest amplitude coefficient. Yet, including all templates into the power spectrum measurement does only lead to minor corrections in the result, as the blue line in Fig. 5.6 shows.

5.6 Galactic Extinction

Leaving Earth, the interstellar medium within our Galaxy has a similar effect to extragalactic light as the Earth's atmosphere, both cause the light to scatter. However, the interstellar medium changes on much larger time scales as the Earth's atmosphere, so this effect, called extinction, can be mapped. As blue light is more affected by scattering, extinction causes the light to become redder, and extinction is usually quantified as the difference between the observed and intrinsic B − V colour

$$E_{\mathrm{B-V}} = (\mathrm{B}-\mathrm{V})_{\mathrm{obs}} - (\mathrm{B}-\mathrm{V})_{\mathrm{int}}, \tag{5.29}$$

where B stands for the filter sensitive to blue light and V is sensitive to visible green-yellow light.

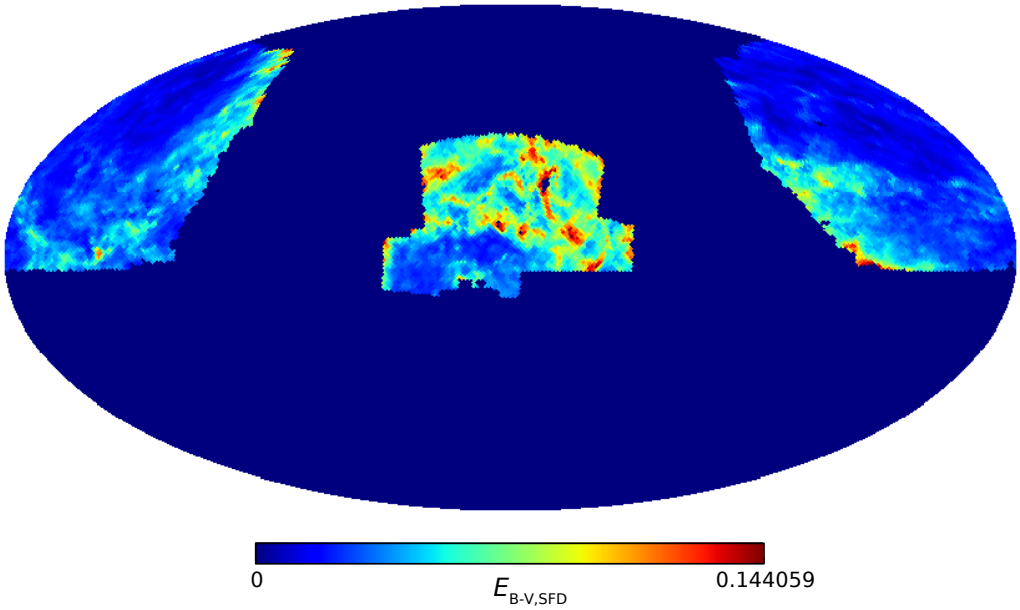

Figure 5.10: The values of $E_{\mathrm{E-V,SFD}}$ used to correct for extinction in the BOSS targetting in HEALPix. The map is presented in Mollweide projection, equatorial coordinates and astronomical orientation.

The photometric magnitudes used in the BOSS target selection were corrected using the dust map by Schlegel, Finkbeiner and Davis (SFD, [128]). Schlafly and Finkbeiner [204] found that, using a more accurate reddening law, the SFD map $E_{\mathrm{B-V,SFD}}$ has to be recalibrated such that

$$E_{\mathrm{B-V}} = 0.86 E_{\mathrm{B-V,SFD}} \text{ [204]}. \tag{5.30}$$

Due to the recalibration, there might be a colour-dependent shift in the target density. A similar χ^2 null-test by Ross *et al.* led to a similar conclusion as for the airmass test: Extinction weights do not significantly change the clustering statistics, but one should be prudent at large scales [113].

For that reason, we also test whether including extinction templates changes our power spectrum at large scales. The SFD values of $E_{\mathrm{E-V,SFD}}$ used in the BOSS targeting are mapped in Fig. 5.10. There, one can see that extinction mostly affects the SGC part of the BOSS footprint, which we have not analysed so far. Extinction in NGC occurs mostly in the regions close to the Galactic disk, similar to the stars in Fig. 5.1. We therefore might expect some correlation between the stellar and extinction templates, as their best-fitting

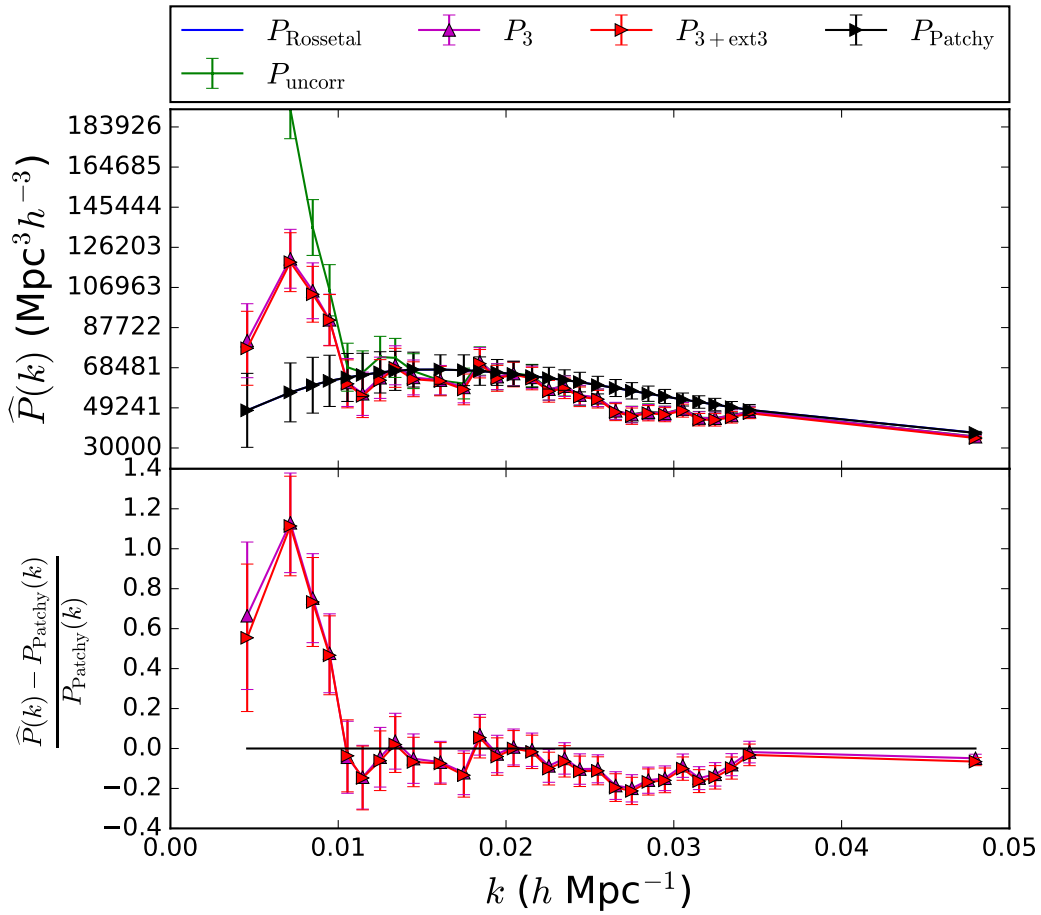

Figure 5.11: The power spectra of the uncorrected BOSS DR12 CMASS NGC galaxies (green), and those after mode subtraction using 3 stellar templates (magenta) and 3 stellar and 3 extinction templates (red), compared to the average power spectrum of the PATCHY mocks (black) and the power spectrum using the Ross *et al.* weights (blue).

Table 5.4: Best-fitting contamination amplitudes for a power spectrum measurement using three stellar templates (left) and additionally three extinction templates (right).

template	stellar templates	+ extinction templates
stars 1st order	0.0072	0.0070
stars 2nd order	0.0001	0.0023
stars 3rd order	0.0055	0.0043
extinction 1st order		-0.0009
extinction 2nd order		0.0016
extinction 3rd order		0.0020

amplitudes $\varepsilon^{(\mathrm{BF})}$, listed in Tab. 5.4, also suggest. The amplitudes of the first and third order stellar templates is slightly smaller when fitted at the same time as the extinction templates. The amplitudes of all extinction templates are less than all stellar template amplitudes, explaining why the power spectrum does not change much when extinction templates are included (cf. Fig. 5.11).

5.7 Scanning Stripes

Another possible source of data contamination is the instrument itself rather than astronomical or atmospheric foregrounds. For example, the telescope might have a calibration offset between different nights. Furthermore, we have already seen in Fig. 5.7 and 5.8 that time-varying systematics are mostly exposing the drift scanning strategy of SDSS. In fact, Fig. 5.12 shows that the observed number of galaxies in certain stripes can be significantly different from the number that is expected from the random catalogue.

We use Eq. (5.18) to build templates where

$$Y_A(\mathbf{x}) = \begin{cases} 1, \text{ if } \mathbf{x} \in \eta_A, \\ 0, \text{ else} \end{cases} \tag{5.31}$$

is either 1 or 0 depending on whether the position $\mathbf{x}$ falls into the scanning stripe η_A. The problem with this is that if we do this for all stripes, the matrix

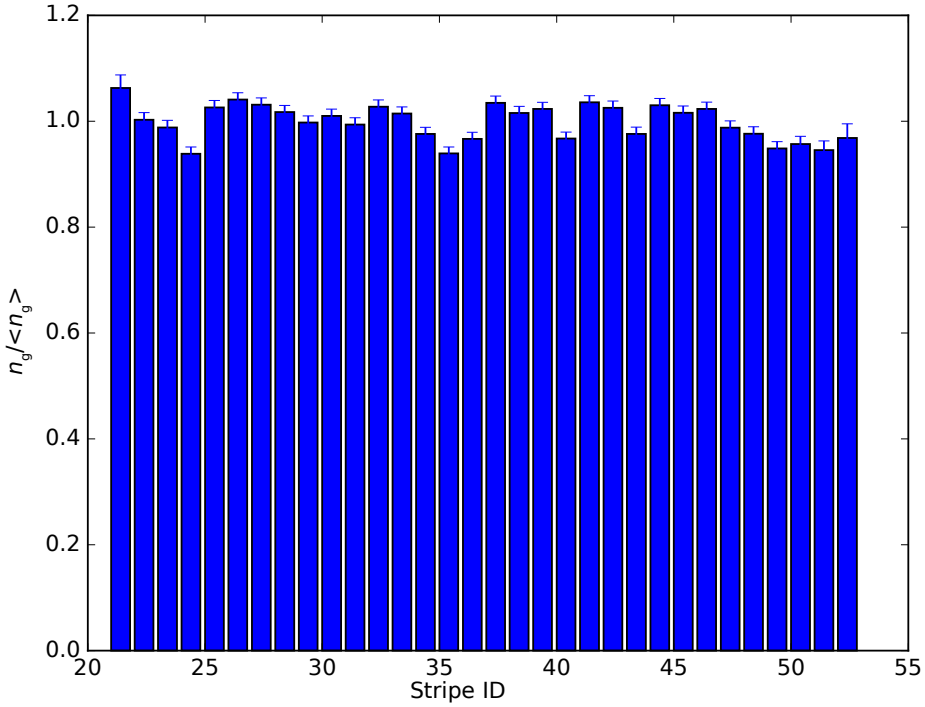

Figure 5.12: $n_{\mathrm{g}}/\langle n_{\mathrm{g}}\rangle$ in the different scanning stripes.

$\mathbf{R}$ (as defined in Eq. 4.92) is singular and therefore not invertible. From Eq. (4.100), we know that there is no unique set of best fitting amplitudes $\varepsilon^{(\mathrm{BF})}$ if $\mathbf{R}$ is singular. This can be explained by the fact that templates combined look very similar to the data and would allow for subtracting off all of the actual signal. The method works, however, when the templates of only a few of the stripes are being used. We start with stripes no. 34, 35 and 36, where there is a significant dip in the observed galaxy density compared to the density of the random galaxy catalogue in the same stripes (cf. Fig. 5.12). This causes the power spectrum to be slightly closer to the Patchy predictions, but it still remains significantly different, as can be seen in Fig. 5.13. Encouraged by this slight improvement, we include more templates symmetrically around the dip, until we reach the lowest possible stripe no. 21. We observe that the power spectrum moves back closer to the power spectrum that we obtained by only mitigating the effect of the foreground stars (cf. Fig. 5.13). Hence, our result depends on our choice of which stripes to include, but there is no well motivated way of telling us which templates we should include.

Alternatively, we test what happens if we just mask out stripes no. 34, 35 and 36. This masking causes an increase of power in the lowest k-bin. As the masking changes the survey window, we have to apply the same mask also to the Patchy mocks and compute their masked average power spectrum. We also plot these two masked power spectra in Fig. 5.13 and see that they agree even less than the unmasked power spectra.

As we do not have a well established reasonable suspicion of a systematic coming from the observational stripe pattern, we refrain from including any templates based on the scanning stripes alone, because it is not very clear to us, which of the stripes should be mitigated against, and our final result would depend on our not well motivated choice.

5.8 Sub-Sampling the Stars by Magnitude

As accounting for other known sources of contamination did not change the shape of the power spectrum significantly, we study possibilities to improve the stellar contamination mitigation procedure. The first such possibility that we study is to split the SDSS star catalogue into sub-samples according to

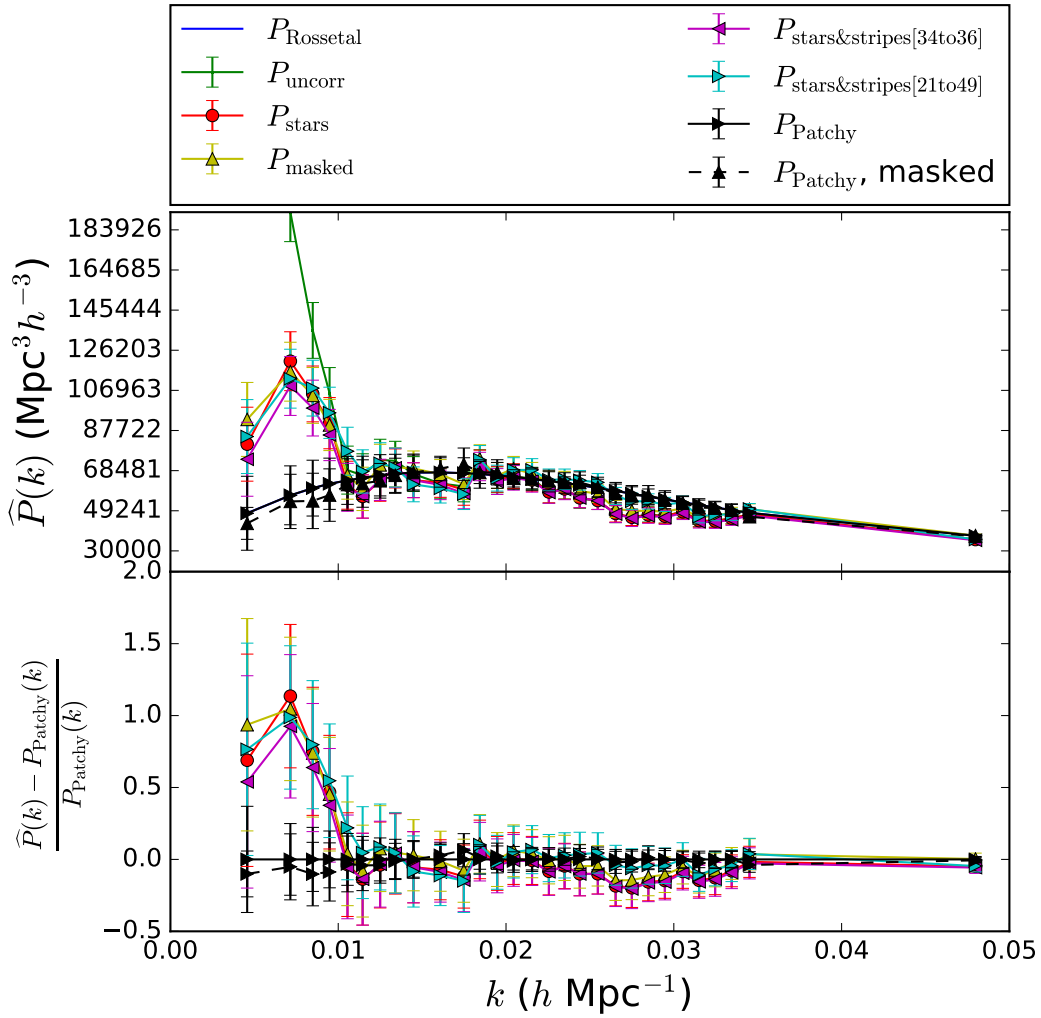

Figure 5.13: Power spectra of the uncorrected BOSS DR 12 CMASS NGC data (green), of the same data using mode subtraction to mitigate the effect of foreground stars (red), as well as stars and possible stripe dependent effects (magenta and cyan), compared to the power using Ross *et al.* stellar weights (blue) and to the Patchy power (black). The power spectrum of the data cleaned for stellar effects, but with stripes no. 34 to 36 masked out is plotted in yellow. The corresponding Patchy power with the same mask applied to it is plotted as a dashed black line for comparison.

Table 5.5: Best-fitting contamination amplitudes for a power spectrum measurement using five templates for different magnitude ranges of the stars. The values on the left hand side are obtained by fitting only one template at a time, whereas those on the right have been obtained in a simultaneous fit.

magnitude range	separate fit	simultaneous fit
$17.5 < i < 18.0$	0.007	0.013
$18.0 < i < 18.5$	0.006	-0.009
$18.5 < i < 19.0$	0.006	-0.004
$19.0 < i < 19.5$	0.007	0.009
$19.5 < i < 19.9$	0.006	-0.003

the stars magnitudes, as the angular radius within which faint stars affect the magnitudes of galaxies is smaller than the radius for bright stars, even though the effect is less dependent on the magnitude of stars than on the magnitude of the galaxies [131].

Our first attempt is splitting the star sample into two sub-samples at the central i-band magnitude value of $i = 18.7$. In Fig. 5.14 one can see that the two sub-samples also have different spatial distributions: Bright stars are more likely to be found close to the Galactic plane, whereas faint stars are more spread out. The $n_{\rm g}/\langle n_{\rm g}\rangle$ diagrams do not look very different, though. We refine the magnitude split of the stars and split them into five magnitude bins, each with a width of 0.5, except for the last bin with $19.5 < i < 19.9$. By comparing the masked maps of each sample (Fig. 5.17), one can see that the differences in the distribution of stars are only prominent in regions close to the galactic plane from where no galaxies enter BOSS. Therefore, the templates are all strongly correlated, which we can also see in Tab. 5.5. When fitted separately, all templates have roughly the same amplitude, and each template alone can remove the whole stellar contamination signal, suggesting that they contain mostly the same information. When combined, their amplitudes differ, but the resulting power spectrum does not change. The resulting power spectra are plotted in Fig. 5.16 We therefore conclude that the effect of stars at different magnitudes is less than we initially thought.

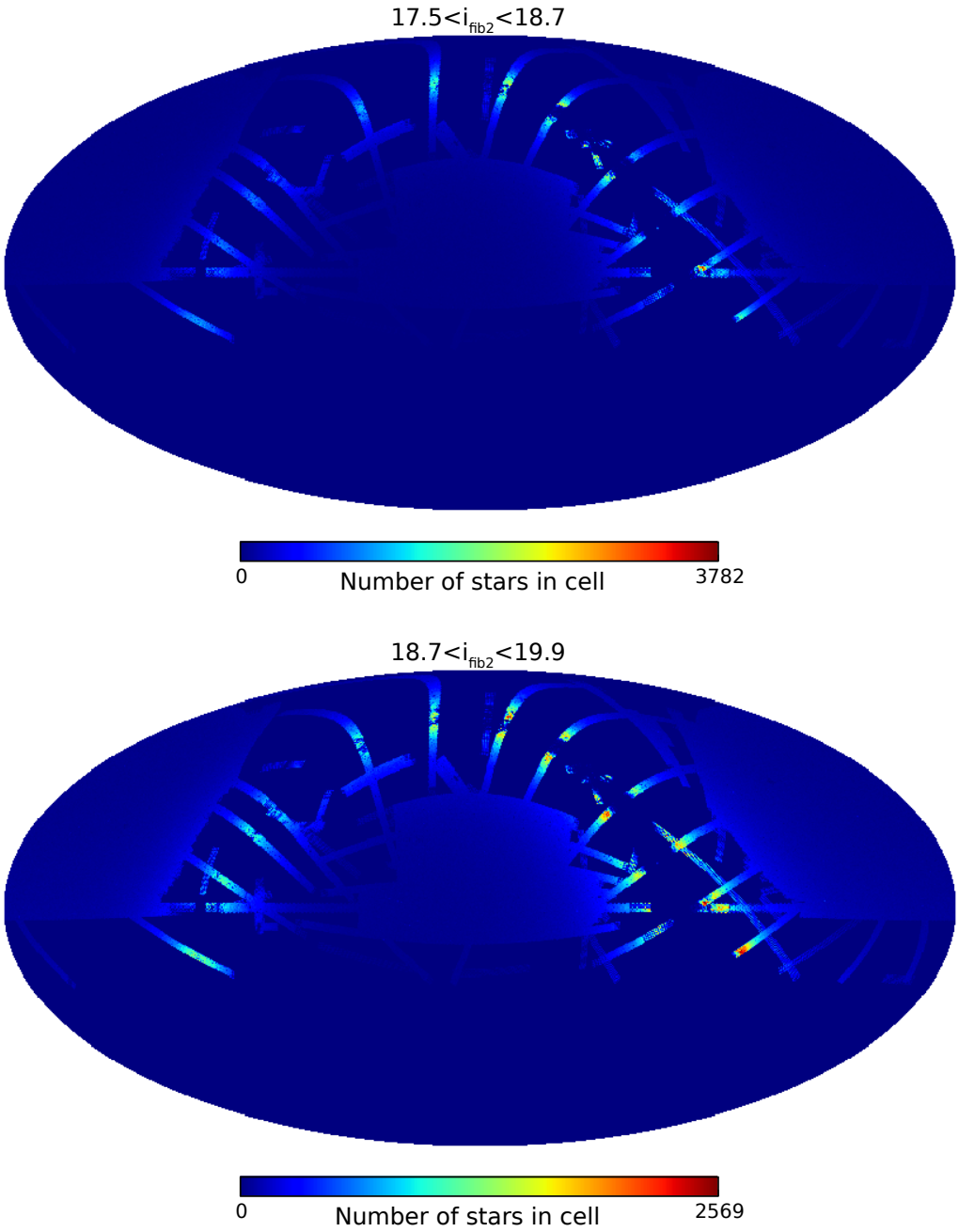

Figure 5.14: Maps of two sub-samples of the SDSS DR8 star catalogue. The upper panel shows the distribution of bright stars with $17.5 < i < 18.7$ and the lower one faint stars with $18.7 < i < 19.9$. The plot is in Mollweide projection and in equatorial coordinates with astronomical orientation.

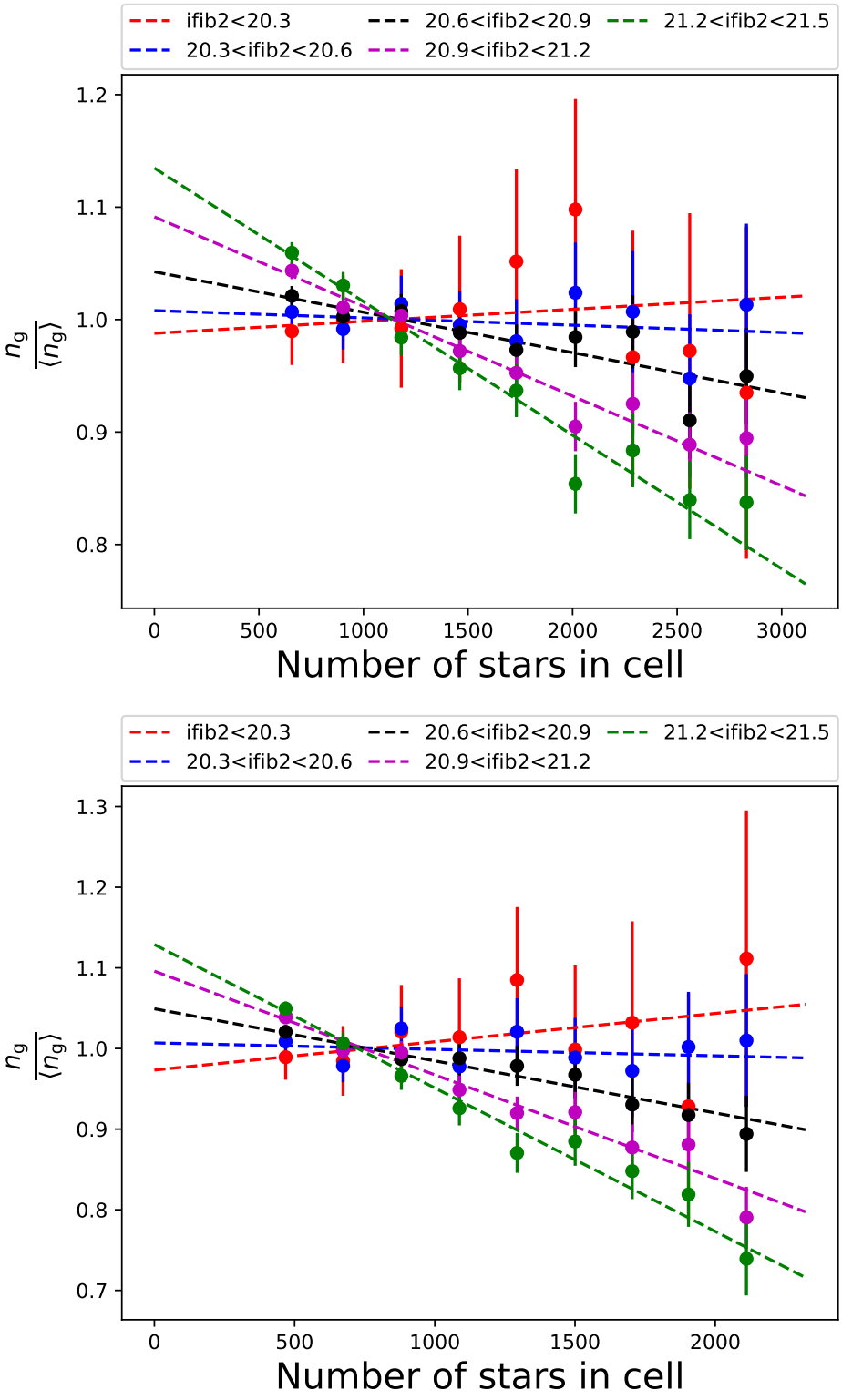

Figure 5.15: The relationship between observed galaxy density and the number of bright stars ($i < 18.7$, upper panel) and faint stars ($i > 18.7$, lower panel).

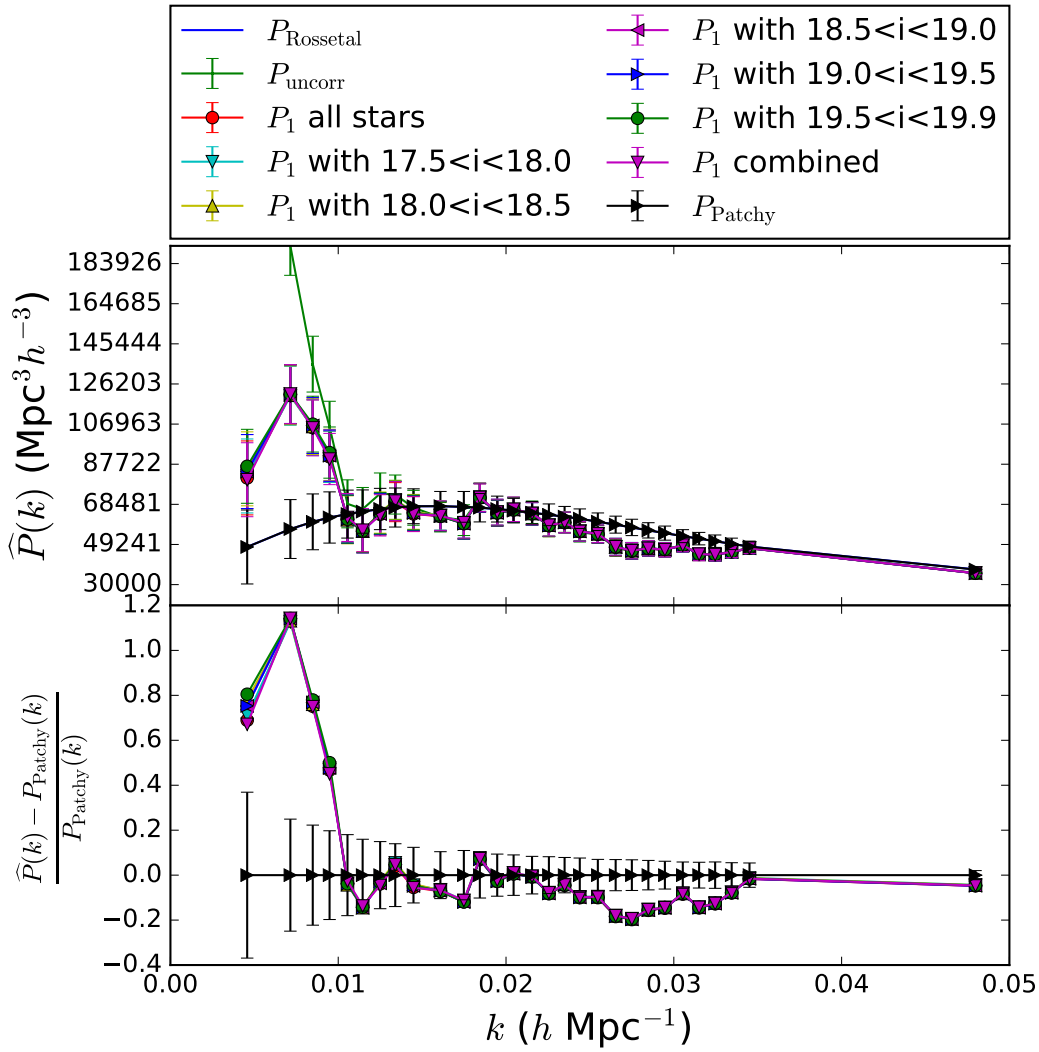

Figure 5.16: The power spectra of the uncorrected BOSS DR12 CMASS NGC galaxies (green), and those after mode subtraction using five different templates for stars with different magnitudes, compared to the average power spectrum of the PATCHY mocks (black) and the power spectrum using the Ross *et al.* weights (blue).

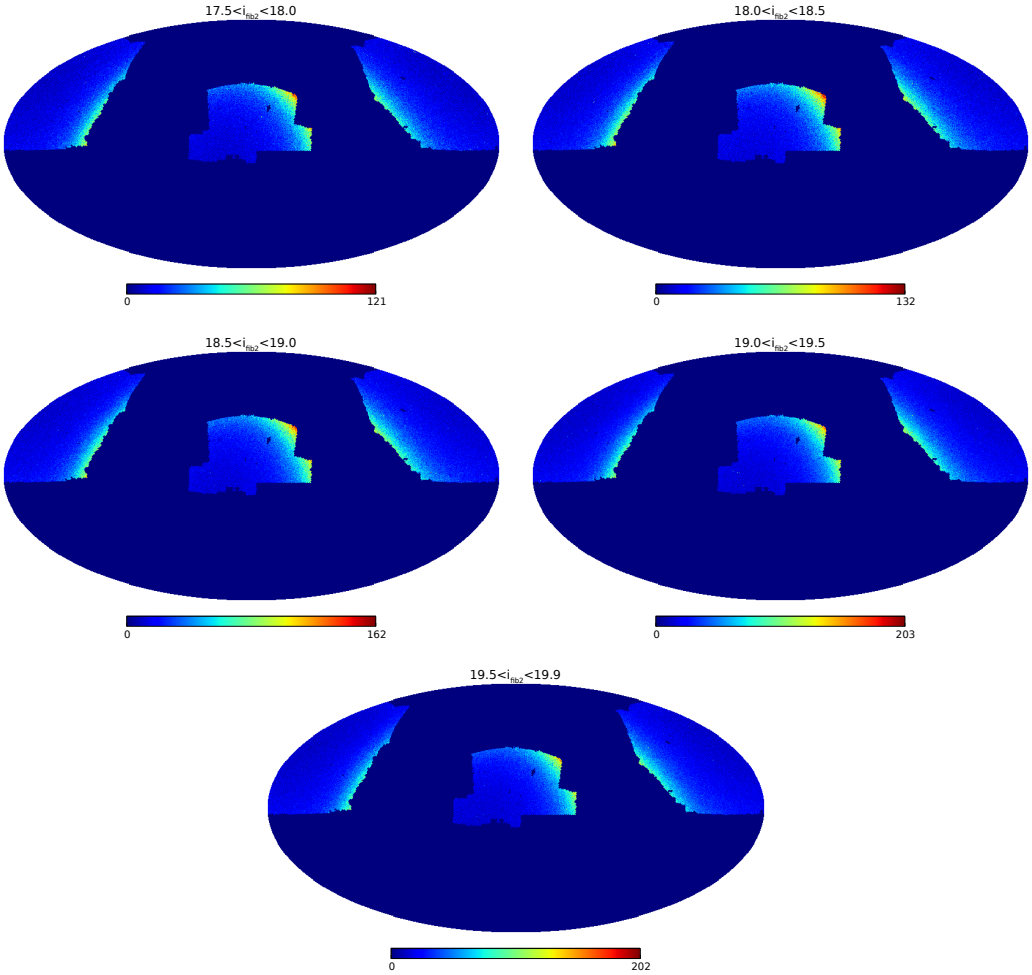

Figure 5.17: Maps of five sub-samples of the SDSS DR8 star catalogue. The panels show the distribution of stars with $17.5 < i < 18.0$ (top left), $18.0 < i < 18.5$ (top right), $18.5 < i < 19.0$ (centre left), $19.0 < i < 19.5$ (centre right), and $19.5 < i < 19.9$ (bottom). The plot is in Mollweide projection and in equatorial coordinates with astronomical orientation.

5.9 Number Count versus Integrated Magnitude

So far, all our stellar templates are based on the number count of stars in regions of the sky. As the effect of the stars is due to their light, this is not the only plausible way of doing so and we explore, as an alternative, to base the templates on the stellar foreground brightness $I(\alpha, \delta)$ in each HEALPix cell. The astronomical magnitude m of an object is defined through the decimal logarithm of its brightness I in units of the brightness of a reference object $I_{\rm ref}$:

$$m - m_{\rm ref} \equiv -2.5 \log_{10} \left(\frac{I}{I_{\rm ref}} \right) . \tag{5.32}$$

Given the i-band magnitudes provided in the star catalogue file and used in the previous section, we can get the stellar foreground brightness as the sum over the brightness of all stars in a HEALPix cell around the coordinates (α, δ):

$$I(\alpha, \delta) \propto \sum_{\rm stars \in cell} 10^{-i/2.5}. \tag{5.33}$$

The distribution of the stellar foreground brightness, mapped in Fig. 5.18, is very similar to the distribution of the number of foreground stars (cf. Fig. 5.1). However, in Fig. 5.19 we see that the relationship between observed galaxy density and the foreground brightness $n_{\rm g}/\langle n_{\rm g} \rangle (I(\alpha, \delta), i_{\rm fib2})$ looks very different compared to the same plot for the number counts (cf. Fig. 5.3), but this can be explained by the fact that the number count and the brightness are approximately logarithmically related. A linear fit does not agree well with the data, and the template based on it does worse in removing the contamination than the first order template based on the number counts, as the plot of the resulting power spectrum in Fig. 5.20 shows. Introducing higher order templates results in power spectra that are similar to the power spectra obtained in the sections before. It shows that the method of introducing templates based on a series expansion of the expected contaminant is working if more than one template are significant, and if they are uncorrelated. On the other hand, it also shows that there is no improvement by constructing the templates on the integrated brightness rather than on the number count of the foreground stars.

This concludes our list of known contaminants and we still obtain a power spectrum that does not agree with our expectation based on the PATCHY mocks. This means that we are left with the following possible conclusions:

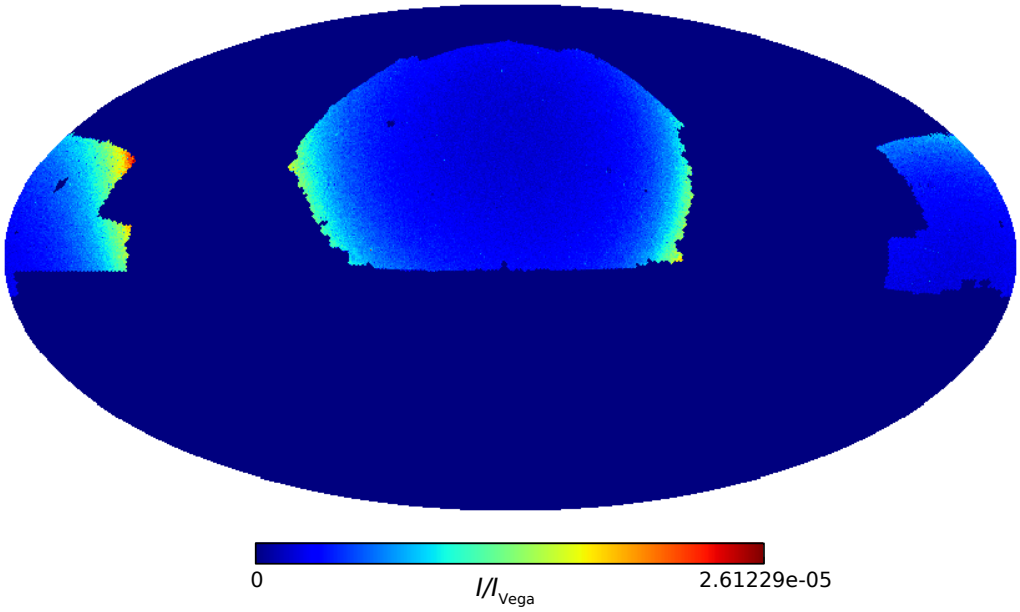

Figure 5.18: Map of the brightness distribution of the SDSS DR8 star catalogue. The plot is in Mollweide projection and in equatorial coordinates with astronomical orientation. The map is rotated by 180° to feature the NGC in the centre. The brightness is given in units of the brightness of the star Vega.

- There is unknown physics.
- We are modelling the power spectrum spectrum incorrectly at small k, e.g. we are using the wrong gauge.
- There is an unknown source of contamination.

The first two points in the list are beyond the scope of this thesis, and I focus on the last point. In the next section, I present my yet unfinished work about finding and identifying this potential unknown source of contamination.

5.10 Localisation of the Potentially Remaining Systematic

Having excluded known sources of data contamination, we explore the possibility of having one or more unknown systematics. We perform a few tests

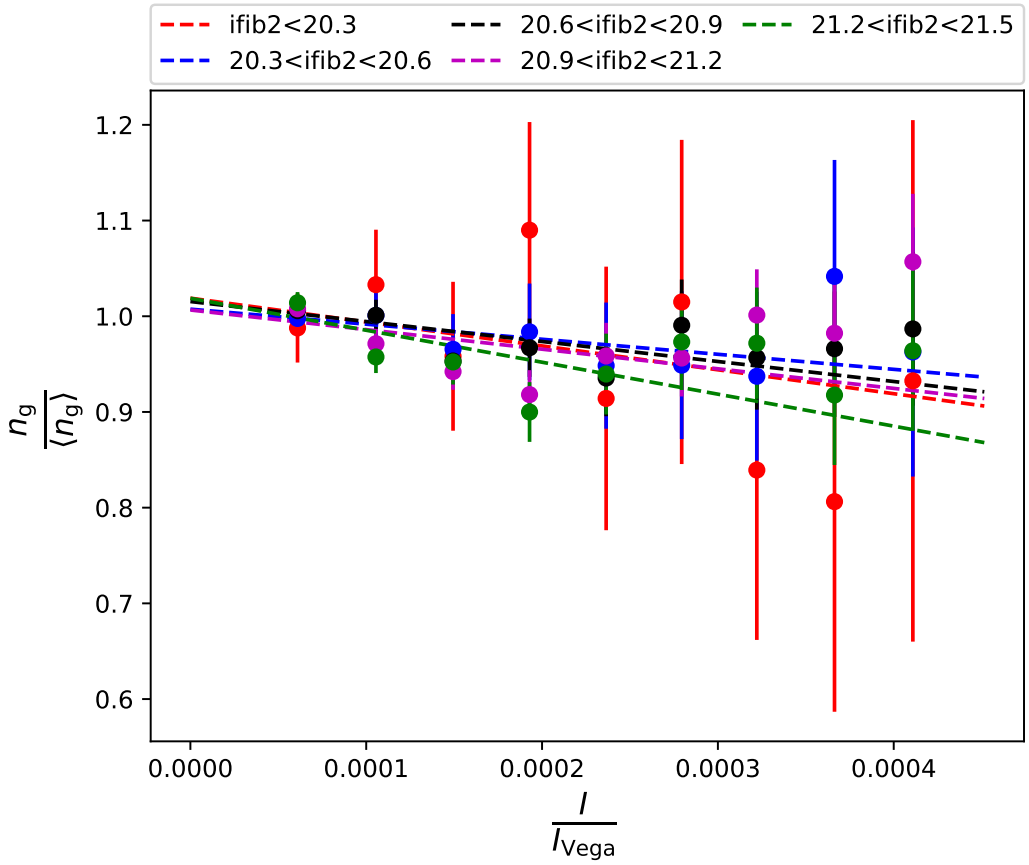

Figure 5.19: The relationship between observed galaxy density and the integrated stellar foreground brightness in units of the brightness of Vega.

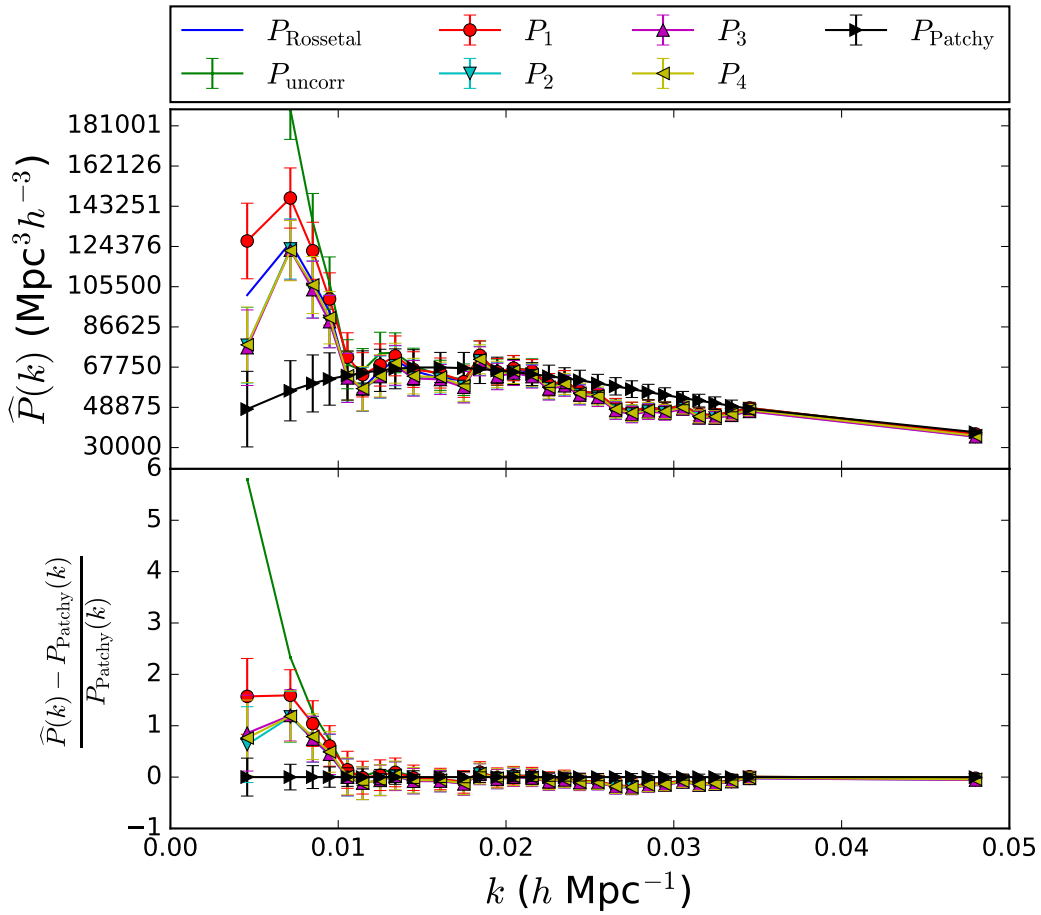

Figure 5.20: The power spectra of the uncorrected BOSS DR12 CMASS NGC galaxies and after 1 to 4 iterations of mitigating the effect of the foreground stellar brightness, compared to the average power spectrum of the PATCHY mocks.

to single out specific patterns or relationships that give us a clue about the nature of the remaining systematic.

We start with back transforming the galaxy density field into configuration space after mode subtraction to check whether there are any obvious features in the residual map (top panel of Fig. 5.21). The map shows an alignment of high density peaks that are within scanning stripes no. 34, 35 and 36, the same ones I already discussed in Sec. 5.7. These are also present in the galaxy density map before applying the stellar template (cf. top panel of Fig. 5.22). What is surprising is that, by eye, we see large over densities in the stripes that had less observed galaxies than expected. It remains to analyse the significance of these over densities for future work (cf. Sec. 6).

We also draw similar maps for the BOSS DR12 CMASS NGC data set split into two redshift bins at its effective redshift $z = 0.57$ to consider the possibility that galaxies in the lower redshift bin are lensing the light coming from the galaxies further away. This would let spurious large over-densities appear in the observed distribution of the galaxies further away. The two lower panels of Fig. 5.21 do not support this hypothesis, they actually show the opposite. This indicates that these suspicious looking over densities are not due to a systematic observational effect, because we expect systematics to affect the data equally in all redshift shells. A more thorough test based on cross-correlations between redshift shells will be presented in Sec. 6. Splitting the data into redshift shells before applying mode subtraction yields very similar looking maps (cf. centre and bottom panel of Fig. 5.22).

The power spectrum of the close sub-sample is larger than the one of the one further away, which is closer to the average Patchy power spectrum (cf. Fig. 5.23). Overall, they do not differ significantly, so we compare sub-sets from different regions in the sky instead.

We split the sky by angular coordinates, such as the declination δ. This split results in an intriguing result. When we split the sky into 20°-wide stripes, the power spectra of all stripes, bar the $(60° < \delta < 90°)$-stripe, are consistent with the Patchy power spectra on the same stripes (cf. Fig. 5.24). We reanalyse the Patchy catalogues for each stripe such that we only compare power spectra with the same underlying window functions. Note that the BOSS footprint does not extent above 80°, so the stripe is comparable to the other 20°-wide

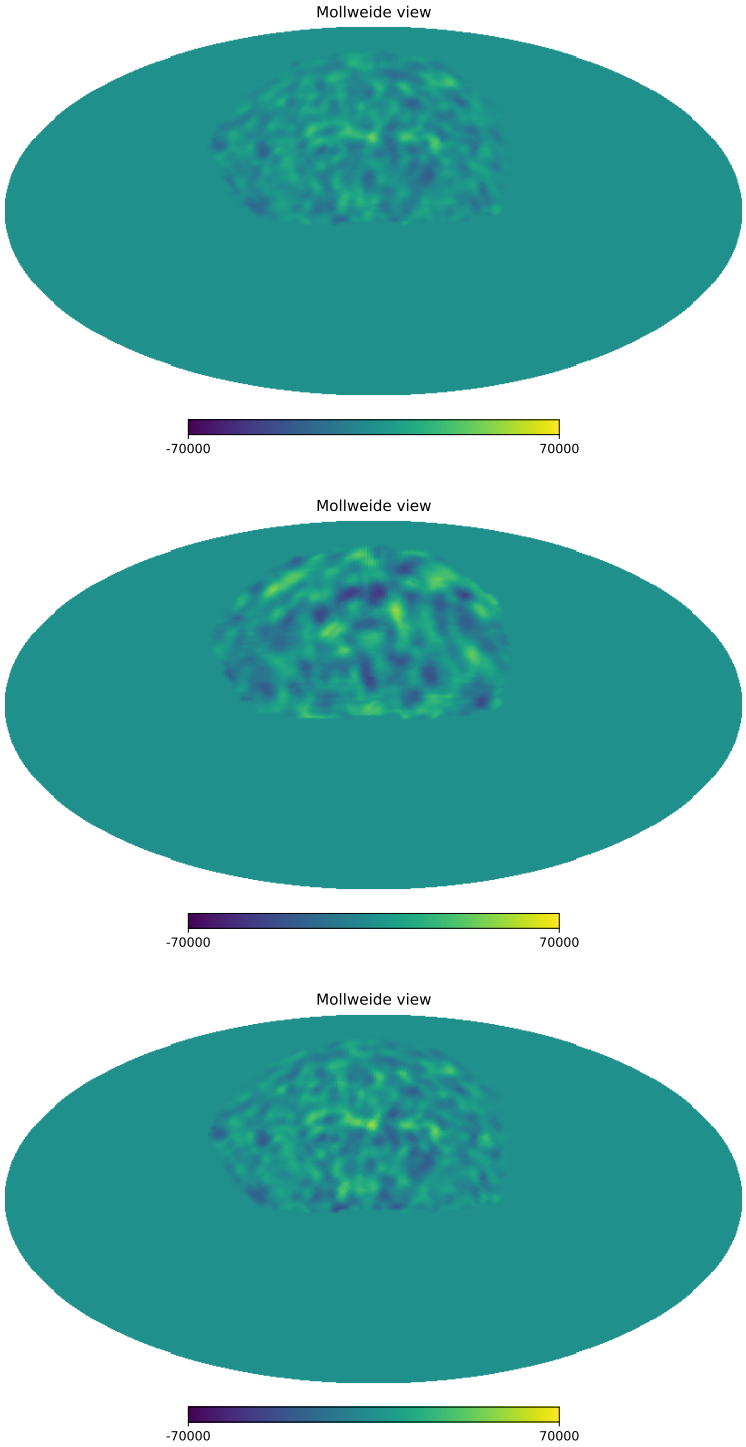

Figure 5.21: Maps of the cleaned galaxy density of the full BOSS DR12 CMASS NGC (top), and split into two redshift bins with $z < 0.57$ (centre) and $z > 0.57$ (bottom). The scale is arbitrary as the density field has not been normalised after performing the Fourier transform.

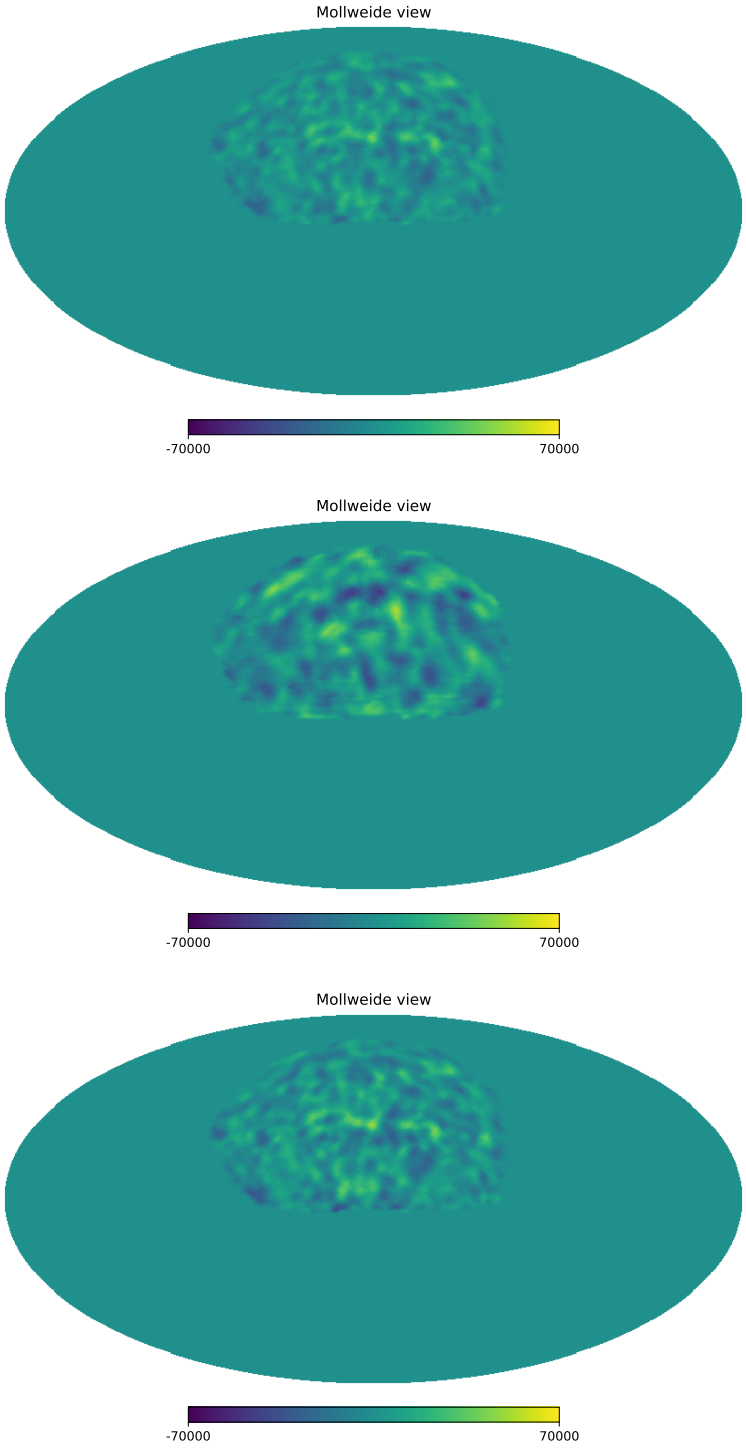

Figure 5.22: Maps of the galaxy density similar to Fig. 5.21 but before applying the stellar template.

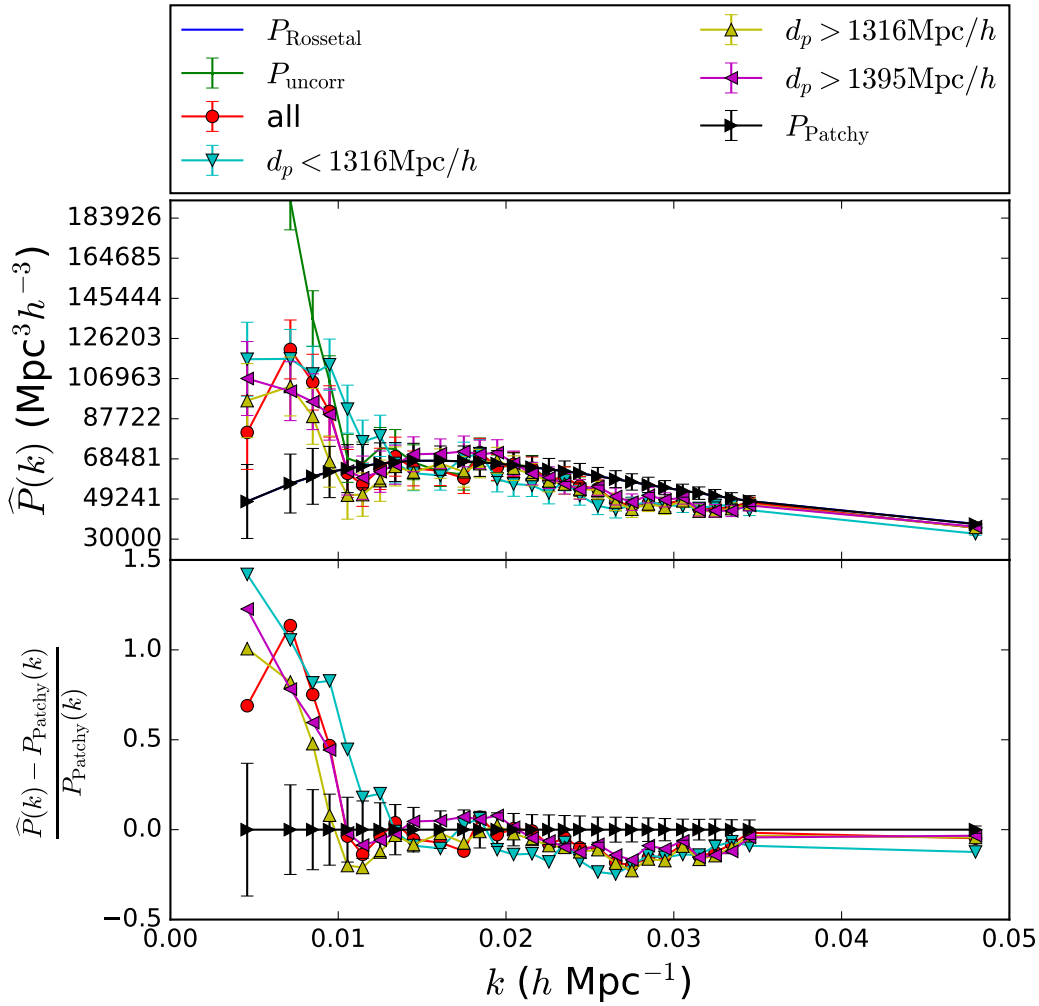

Figure 5.23: The power spectra of the uncorrected BOSS DR12 CMASS NGC galaxies and of its close ($z < 0.57 \sim d_p < 1316{\rm Mpc}/h$, cyan) and far ($z > 0.57 \sim d_p > 1316{\rm Mpc}/h$, yellow) sub-samples. The power spectrum of another far sub-sample split at $d_p > 1395{\rm Mpc}/h$ is plotted in magenta.

stripes. The surprise comes when we split the sky into 30°-wide stripes: The measured power spectra on all stripes do not agree with the corresponding power spectra measured from the Patchy mock catalogues on the same stripes. The deviations occur below $k \lesssim 0.01h/\mathrm{Mpc}$, which corresponds to scales of

$$\lambda = \frac{2\pi}{k} \gtrsim 628 \, \frac{\mathrm{Mpc}}{h}. \tag{5.34}$$

Modes of this size only fit into the declination stripe if it is aligned with the small circle defined by the declination, because, considering that the furthest galaxies in the BOSS CMASS sample are at distances of approximately 1700 $\frac{\mathrm{Mpc}}{h}$, the maximum distance between two objects in a 20°-wide stripe with the same right ascension and redshift is given by

$$1700 \, \frac{\mathrm{Mpc}}{h} \sin(20°) = 593 \, \frac{\mathrm{Mpc}}{h}. \tag{5.35}$$

In a 30°-wide stripe, this maximum distance increases to

$$1700 \, \frac{\mathrm{Mpc}}{h} \sin(30°) = 890 \, \frac{\mathrm{Mpc}}{h}, \tag{5.36}$$

providing enough volume to accommodate all $k \lesssim 0.01h/\mathrm{Mpc}$-modes regardless of their orientation. My hypothesis is therefore that we are left with two sources of contamination: One that affects our observation of galaxies at declinations greater than 60°, and another one that affects large scale modes that are aligned with hour circles, i.e. great circles through the celestial poles in equatorial coordinates. As the equatorial coordinate system is geocentric, I suspect that the remaining excess power at large scales is due to a systematic effect from astronomical foregrounds within our solar system or due to instrumental systematics. Due to time constraints, this hypothesis will be further tested after completion of this thesis, as I will describe in the following chapter.

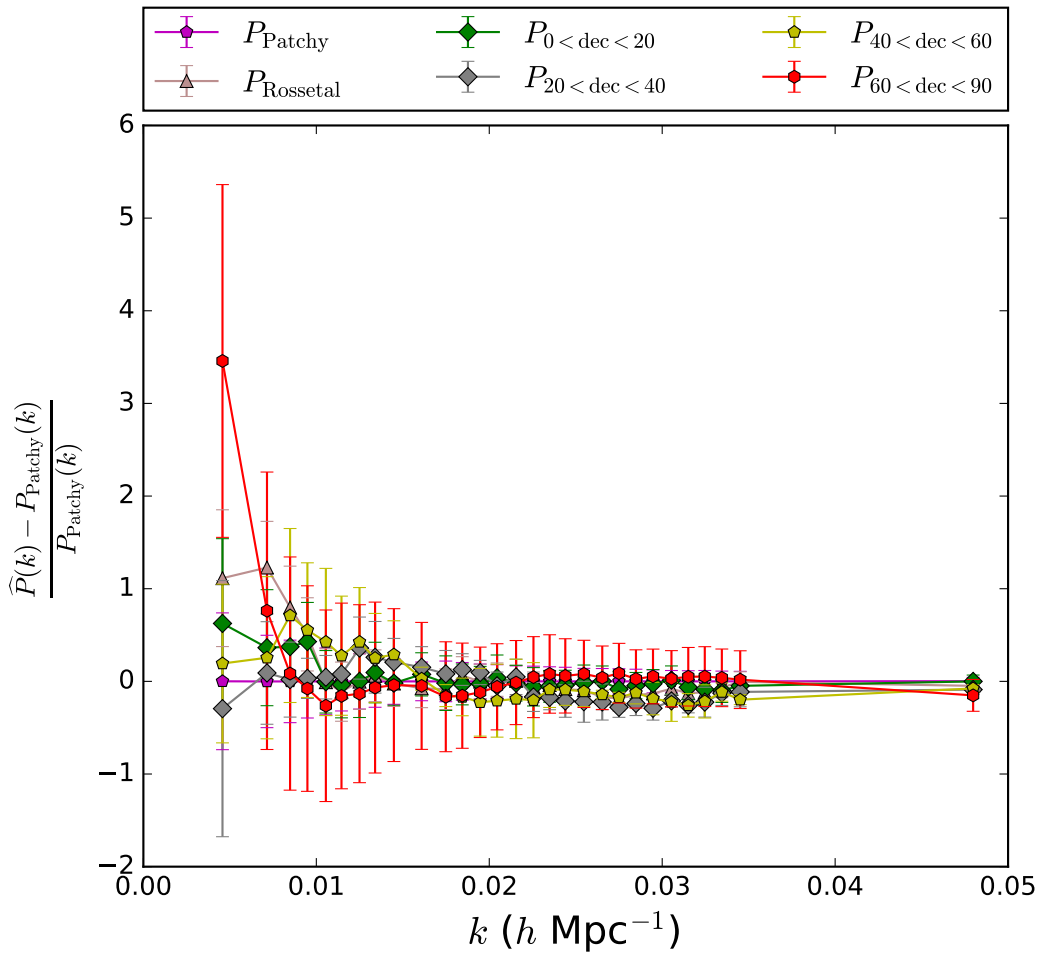

Figure 5.24: Fractional difference between the measured and average Patchy mock power spectra of the BOSS DR12 CMASS NGC galaxy sample on 20°-wide declination stripes.

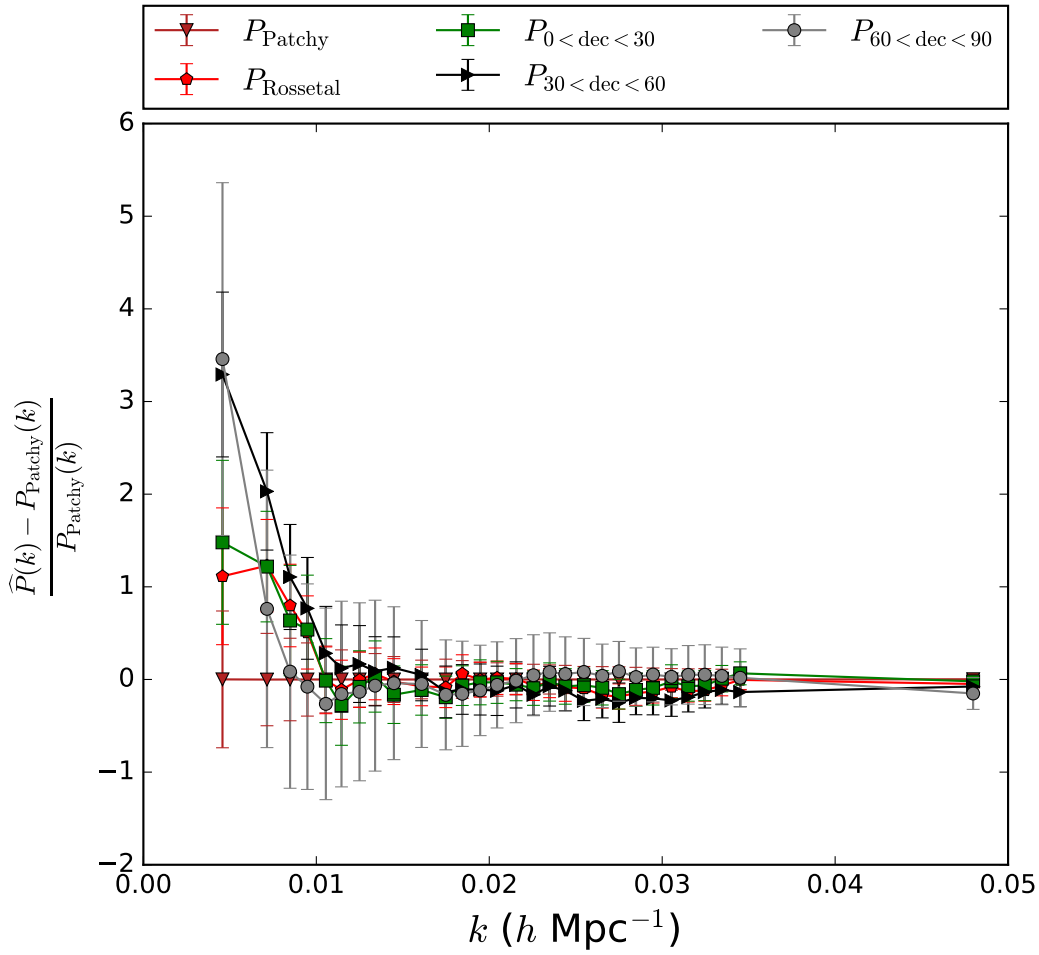

Figure 5.25: Similar plot to Fig. 5.24, but for 30°-wide declination stripes.

Chapter 6

Conclusion and Further Work

For almost 20 years, the concordance model of cosmology (cf. Sec. 1.1) has passed numerous observational tests. It is well supported by SN, BAO and CMB experiments, but there are also observations that are in tension with the concordance model. For instance, distance ladder measurements of the Hubble expansion rate may or may not be in tension with Hubble rate measurements from BAO and CMB data. The systematics of BAO and CMB experiments are well understood, but constraints on the value of the Hubble rate from these experiments are model dependent. On the other hand, measuring the relation between redshifts and distances to various local objects provides a model-independent value of H_0, but to obtain the distance to a specific standard candle, one has to anchor its distance scale by comparing the luminosity of that particular type of standard candle to the luminosity of another type of standard candle within the same galaxy. Depending on the anchors, the local value of the Hubble parameter is either consistent [20] or inconsistent [49] with the values obtained from CMB and BAO measurements. This might call for an extension of the basic ΛCDM concordance model with, e.g. more relativistic particles or dark energy that is not a cosmological constant (cf. Fig. 1.5). Still, the tension between the two different values of the Hubble parameter is consistent with fluctuations of the Hubble parameter within a ΛCDM cosmology [50, 51]. Weak lensing experiments also report results that are discrepant with BAO and CMB measurements, though, but at scales that are not well understood.

The concordance model is remarkably successful in terms of observations.

However, there are issues when some of the model's phenomena shall be linked to fundamental physics. One big question is why we observe an accelerated expansion. This is realised by a cosmological constant Λ that can so far not be linked to something we can observe in a laboratory.

This problem might be linked to another problem at earlier times. Observations of the CMB radiation are remarkably isotropic, even though only tiny regions of the sky that we can observe today were in causal contact at the time the CMB photons scattered for the last time. Furthermore, the Universe is extremely flat, such that its curvature must have been fine tuned during the Big Bang. We also need a mechanism to seed the fluctuations that grew into the structures we can observe today. The most successful paradigm to solve these problems is another epoch of accelerated expansion at early times called cosmic inflation.

Inflation (cf. Sec. 1.3) occurred at energy scales that are too high to be accessible with accelerator experiments, thus we have to rely entirely on cosmological observations to rule out classes of inflationary models and get insight into the physics of the early Universe. One way to distinguish these models is by measuring how close their predicted primordial fluctuations are to being Gaussian, described at first order by the parameter $f_{\rm NL}$. Local primordial non-Gaussianity alters the biasing law between dark-matter halos and the underlying mass-density field at the largest scales [1, 2]. Currently, the tightest constraints on the local $f_{\rm NL} = 0.8 \pm 5.0$ come from the cosmic microwave background (CMB) experiment Planck [3]. The next generation ground-based CMB experiment CMB-S4 will tighten the constraint to $\sigma(f_{\rm NL}) = 2.5$, or $\sigma(f_{\rm NL}) = 1.8$ if combined with Planck [106]. These are the strongest constraints achievable with ground-based CMB measurements due to cosmic variance, thus, we need a different approach to independently confirm these results, if not improve these constraints to further narrow down our understanding of the physics that governed the inflationary epoch. Galaxy clustering studies so far could not compete with the precision of the CMB $f_{\rm NL}$ results, but upcoming galaxy surveys, such as those I described in Sec. 2.3, will come close to independently confirm the Planck results. Combining future galaxy clustering and CMB data will improve $f_{\rm NL}$ constraints such that they will provide physically interesting results.

Before we can benefit from the improved data of future galaxy clustering surveys, we have to solve some challenges. With this thesis, I have contributed to two major issues of $f_{\rm NL}$ measurements from the galaxy power spectrum and then considered BOSS data and its possible contaminants:

The Posterior Distribution of the Galaxy Power Spectrum At the very large scales at which $f_{\rm NL}$ strongly affects the 3D galaxy power spectrum, Gaussian posterior distributions in the power do not approximate the posterior distribution $\mathcal{P}_R$ we expect for a Gaussian density field $\delta_{\mathbf{k}}$, even if we vary the covariance matrix according to the model to be tested. In Chapter 3, which is mostly based on [4], I compared alternative posterior distributions with $\mathcal{P}_R$, both mode-by-mode and in terms of expected measurements of $f_{\rm NL}$. Marginalising over a Gaussian posterior distribution $\mathcal{P}_f$ with fixed covariance matrix yields a posterior mean value of $f_{\rm NL}$ which, for a data set with the characteristics of Euclid, will be underestimated by $\triangle f_{\rm NL} = 0.4$, while for the SDSS-III BOSS DR9, it will be underestimated by $\triangle f_{\rm NL} = 19.1$. Adopting a different form of the posterior function, such as the inverse cubic normal distribution (ICN, [4, 185], cf. Figure 3.7), means that we do not necessarily require a different covariance matrix for each model to be tested: this dependence is absorbed into the functional form of the posterior. Thus, the computational burden of analysis is significantly reduced.

Systematic Contaminants Mitigation In Chapter 4 and [5], I assessed and developed techniques to remove contaminants when calculating the 3D galaxy power spectrum. The process was separated into three separate stages: (i) removing the contaminant signal, (ii) estimating the uncontaminated cosmological power spectrum, (iii) debiasing the resulting estimates. For (i), we showed that removing the best-fit contaminant (*mode subtraction*), and setting the contaminated components of the covariance to be infinite (*mode deprojection*) are mathematically equivalent. For (ii), performing a Quadratic Maximum Likelihood (QML, [182]) estimate after *mode deprojection* gives an optimal unbiased solution, although it requires the manipulation of large $N_{\rm mode}^2$ matrices ($N_{\rm mode}$ being the total number of modes), which is unfeasible for recent 3D galaxy surveys. Measuring a binned average of the modes for (ii) as

proposed in (FKP, [109]) is faster and simpler, but is sub-optimal and gives rise to a biased solution. We presented a method to debias the resulting FKP measurements that does not require any large matrix calculations, and argued that the sub-optimality of the FKP estimator compared with the QML estimator, caused by contaminants is less severe than that commonly ignored due to the survey window.

$f_{\rm NL}$ Constraints with BOSS In Chapters 3 and 4, as well as in the publications [4, 5], we have tested the ICN posterior shape and the unbiased FKP-style power spectrum estimator only on toy cases. We chose BOSS as the first real data test, because, as described in Sec. 2.4 and Chapter 3, the analysis of galaxy clustering data requires additional data such as mock galaxy catalogues. These are well tested and supported for seasoned surveys, such as the final BOSS CMASS sample. I have implemented the code to build the templates needed for the error mitigation from the SDSS stellar density catalogue. Using a set of templates, not only could I reproduce the same results as if I would apply the stellar density weights (see Fig. 5.5), but I do not have to assume that the relationship between the number of stars in one direction is linear to the effect they have on the galaxy density field. Instead, I was able to provide a test whether there are also higher order contributions.

Although the mode subtraction technique did a good job at removing a lot of signal due to contamination, our resulting power spectrum does not agree with our model power spectrum. We compare our measured power spectrum with the average power spectrum of Patchy mock realisations of the BOSS CMASS catalogue. This tests whether our findings are consistent with ΛCDM matching the cosmology of Patchy without significant $f_{\rm NL}$. For positive $f_{\rm NL}$, we expect to measure a power spectrum that is larger than the Patchy average and monotonous at the largest scales. For negative $f_{\rm NL}$, the power is suppressed at the lower k-range of our observed power spectrum, with a turn-over at very large scales depending on the value of $f_{\rm NL}$, but for values allowed by Planck [3], the turn-over happens below the minimal k value we can observe using BOSS data (cf. Fig. 1.9). Our measurement shows an excess power but also a turn over towards the lowest k-bin, thus it does not match our model of the $f_{\rm NL}$- and scale-dependent bias. This is similar to the $f_{\rm NL}$ measurement

from BOSS DR9 [6] where stellar density weights removed much of what looks like a positive $f_{\rm NL}$ signal, while leaving a residual that does not agree well with the model power spectrum. Particularly, the best fitting model power spectrum has $\chi^2 = 15.9$ with 10 degrees of freedom, which indicates that the model and the data are in tension. We therefore speculate that there is a further source of contamination. The remainder of Chapter 5 describes how the effects of seeing, airmass, galactic extinction and the observing strategy can be included into the analysis as additional contamination templates. As they do not bring our measurement closer to our expectation, we try different variations of building the stellar density templates. These also do not change our result. We split the survey data to localise where the excess power comes from to get clues that might help identifying the possibly remaining contaminant. We found that the excess power comes from modes that are separated by more than 20 degrees in declination and from modes that are further North than 60 degrees in declination. As the modes causing the excess signal seem to be aligned with equatorial coordinates, which has the Earth as a reference point, we suspect them to be a further unknown foreground contaminant, but a failure of the concordance ΛCDM model at large scales is also possible. Another possible explanation for our observations not meeting our model could lie in the fact that we are not taking all General Relativity effects into account. For example, the over-density field δ, and hence its power spectrum, is a gauge-dependent quantity and its meaning is ambiguous at large scales. Considering matter fluctuations at the hyper-surface defined by an observed redshift, their power spectrum resembles a power spectrum with non-zero $f_{\rm NL}$ in the standard Newtonian description except for a line of sight dependence: it is enhanced along the line of sight and suppressed in the transverse direction in a redshift dependent way. According to a study by Jaiyul Yoo [205], this becomes significant at scales larger than the ones we are observing.

To differentiate between these options is the next step in my future work, which I list in the following.

6.1 Cross-Correlation Based Tests of Contaminant Templates

In the photometric quasar analysis by Leistedt & Peiris [196], the decision of whether a template corresponds to a significant contaminant that should be mitigated against using extended mode projection is done based on the cross correlation between the template and the data. If the data is unaffected by a certain contaminant, the cross correlation between its corresponding template and the data is expected to be below the noise level. A strong cross correlation on the other hand is an imperative to mitigate against the contaminant. This procedure however relies on the ability to create such templates, which seems problematic for an unknown contaminant.

In contrast to the photometric quasar analysis, we can use the spectroscopic redshift information and split the BOSS CMASS catalogue similarly to Sec. 5.10, where I have found that the power spectra of the galaxy distribution in different redshift shells are consistent with each other. What I have not analysed so far, are the cross correlations between the redshift shells. By doing so, we can test whether the remaining offset in the power spectrum is a foreground angular contamination, as we suspect, or whether it is a cosmic signal. A foreground contaminant would affect all redshift slices and we would therefore see a strong correlation between different shells. If the excess power is caused by a cosmological signal, due to projection, the angular scale of the signal would change with redshift. To estimate by-chance correlations, one can do the same cross correlation studies to the mock catalogues. Assuming that cross correlations are due to an angular systematic, one could build templates based on the cross power spectrum and use mode subtraction to mitigate this angular systematic. The cross correlation signal could also provide us with more information to identify the possible contaminant. Identifying a potential contaminant would be an exciting contribution to the field.

Another possible way of obtaining templates for unknown contaminants is through cross correlating the over-density field in the same volume measured using different surveys. However, this only works for contaminants due to the instrument or due to its position on Earth, such as atmospheric contamination, because astronomical foregrounds are the same for all cosmological surveys.

Furthermore, if we only cross-correlate two surveys, we get a combination of the systematics of both surveys. To disentangle the contribution from each survey, we would have to cross correlate more surveys. It is therefore not interesting for our current BOSS analysis, since there are no other surveys that have large enough overlap with BOSS. This would only allow to create templates on small fractions of the survey area that are not useful for a full analysis.

6.2 eBOSS

As mentioned in Sec. 2.3.1, the extended BOSS (eBOSS) continues the success of BOSS, extending to higher redshifts of $z = 2.2$ that correspond to eighty percent of cosmic history. It observes galaxies and quasars in a range of distances that have not been explored by three dimensional maps of large-scale structure so far. There are big efforts to use this data to achieve constrains of $\sigma(f_{\rm NL}) = 15.74$ [132]. These constraints will only be achieved if we fully mitigate against systematics. After I will have tested the mode subtraction technique and hopefully understood the remaining excess signal, I want to contribute this knowledge to the eBOSS $f_{\rm NL}$ analysis.

6.3 Redshift weighting

Dealing with a survey that covers such a wide redshift range as eBOSS imposes another problem that I have not yet addressed in this thesis. As the galaxy bias evolves with time and thus with redshift, it is common practice to slice deep surveys into redshift shells. This removes long radial modes that are particularly interesting for $f_{\rm NL}$ analyses. Instead, Mueller, Percival & Ruggeri [206] developed a weighting scheme that takes the redshift evolution of the bias into account, allowing optimal exploitation of large scale structure data for $f_{\rm NL}$ measurements. Similar schemes have been elaborated for BAO and RSD measurements [207, 208]. The basic idea of these weighting schemes is to replace the FKP weights in Eq. (2.28) with a new weight

$$w = \sqrt{w_{\rm FKP} w_{f_{\rm NL}}} \tag{6.1}$$

such that it maximises the Fisher information matrix

$$F_{\rm ij} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle \tag{6.2}$$

for the parameters θ_i that are measured. Assuming a Gaussian likelihood for the weighted over-density squared $w\delta^2(\mathbf{k})$ with mean $P_w(\mathbf{k}) \equiv \int \mathrm{d}\mathcal{W}(z) w(z) P(k,z)$, where $\mathrm{d}\mathcal{W}(z)$ is a weighted volume element, and a covariance matrix $\mathbf{C}$ that is assumed to be independent of θ_i, the Fisher information for a single parameter reads

$$F_{ii} = \frac{\left(\sum_{\mathbf{k}} w\delta^2(\mathbf{k}) \frac{\partial P_w(\mathbf{k})}{\partial\theta_i}\right)^2}{\sum_{ab} w\delta^2(\mathbf{k}_a) \mathbf{C}_{ab} w\delta^2(\mathbf{k}_b)}. \tag{6.3}$$

In general, this can be maximised by a weight

$$w \propto \frac{\partial P_w(\mathbf{k})}{\partial\theta_i}. \tag{6.4}$$

As multiplying w with any constant factor does not change F_{ii}, it is useful to normalise w such that $\int \mathrm{d}\mathcal{W} w = 1$. In the case where we want to measure $\theta_i = f_{\rm NL}$, the weight for the monopole power spectrum is given by

$$w_{f_{\rm NL}} = \left(b + \frac{f}{3}\right)(b-p)\, D_1(z) \text{ [206]}. \tag{6.5}$$

These weights are straight-forward to apply, and will therefore be included in my final BOSS $f_{\rm NL}$ analysis. Furthermore, it might be worth to study if further improvement can be made by replacing the Gaussian likelihood assumed for Eq. (6.3) with the ICN likelihood of Chapter 3.

6.4 Future Surveys

Looking further into the future, there will be the upcoming surveys such as the Dark Energy Spectroscopic Instrument (DESI), and Euclid. As discussed in Sec. 2.3, they will pose tight constrains on $f_{\rm NL}$. The techniques that I have presented in this thesis will be very useful for achieving the goal of learning more about the early Universe.

Although my research is focussed on $f_{\rm NL}$, the contaminant mitigation introduced in Chapters 4 and 5 will be of interest for other clustering-based measurements, including baryon acoustic oscillations, redshift-space distortions, neutrino masses, dark energy equation of state and deviations from general relativity.

To conclude, this thesis introduces techniques that will be crucial in pushing the boundary of our understanding of cosmology, especially the inflationary epoch. In particular, mode subtraction has been shown to be able to mitigate the effect of foreground contaminants, if we are able to model the effect using templates. Understanding all possible contaminations is still work in progress.

Appendix A

Approximating the distribution of a non-Gaussian density field with a Normal distribution

In Chapter 3, I derived the distribution of the power spectrum of a Gaussian distributed density field. As the topic of this thesis is testing primordial non-Gaussianity, using the results of Chapter 3 for such tests needs some justification. In order to do so, I generate 100,000 zero-centred Gaussian random fields on a $4 \times 4 \times 4$-grid. Using Eq. (1.108), each Gaussian realisation can be transformed into a non-Gaussian field with a particular value of $f_{\rm NL}$. Here, I choose $f_{\rm NL} = \pm 5$, as this approximately corresponds to the minimum and maximum values of Planck polarisation $f_{\rm NL}$-measurements (cf. Sec. 1.6, [3]). A histogram of the power spectra of all realisations is plotted in Fig. A.1. The left tails of the power spectrum distribution seems unaffected by non-Gaussianity, but the power spectra of non-Gaussian fields are likely to be slightly larger than the the one for Gaussian fields.

I fit a Gaussian distribution to each histogram, which is also plotted in Fig. A.1. As in Chapter 3, one can use the Kullback-Leibler divergence as a measure of how well one probability distribution approximates another. In Tab. A.1, I compare the Kullback-Leibler divergences for approximating the distribution of the power of the non-Gaussian random fields with either the power of a Gaussian random field or a Gaussian power spectrum. The table shows that the effect of assuming a Gaussian density is two orders of magnitude smaller

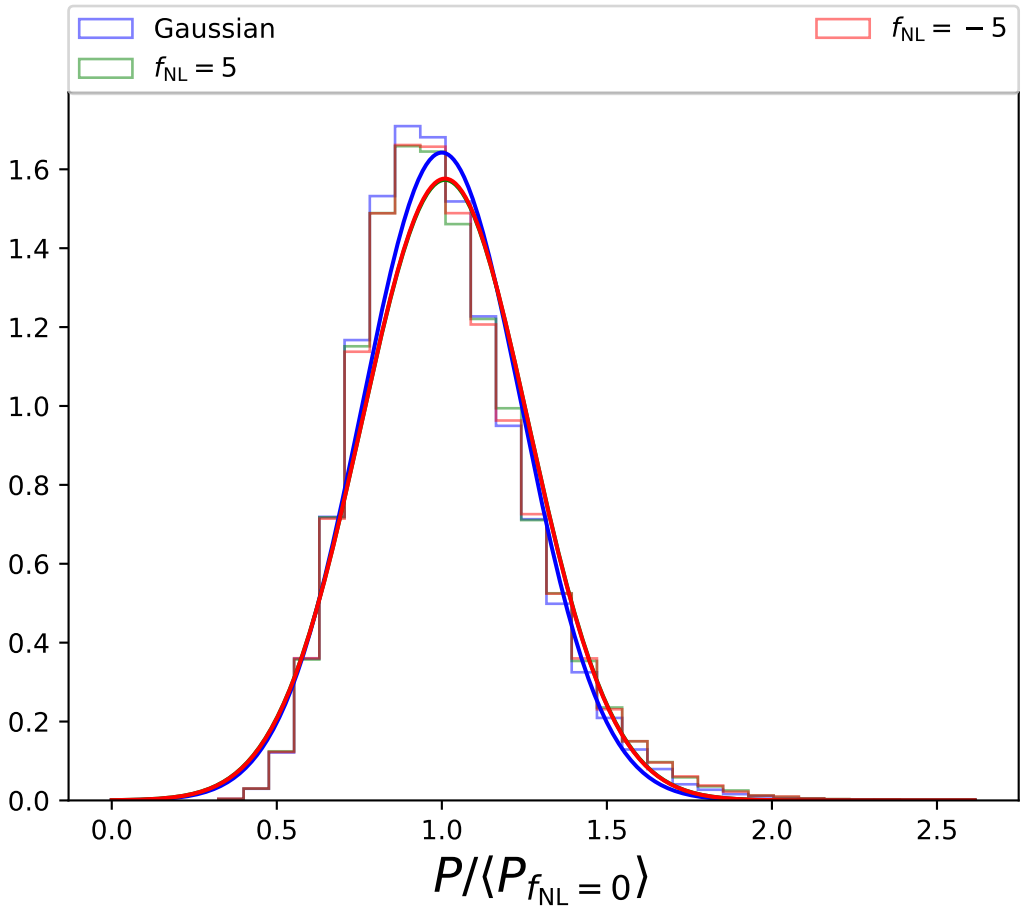

Figure A.1: Histogram of the power spectra of 100,000 Gaussian random fields (blue) and random fields with $f_{\mathrm{NL}} = -5$ (red) and $f_{\mathrm{NL}} = 5$ (green). A Gaussian distribution is fitted to all histograms and plotted as solid lines in the same colour as their corresponding histograms.

than assuming a Gaussian distribution for the power spectrum. Therefore, the results of Chapter 3 can be applied even when testing non-Gaussianity in the density field.

Table A.1: Kullback-Leibler divergences of approximating the distribution of power of a field with given value of $f_{\rm NL}$ with either the power of a Gaussian random field or a Gaussian power spectrum.

$f_{\rm NL}$	Gaussian density	Gaussian power
-5	0.002	0.61
0	0	0.64
5	0.002	0.61

Appendix B

Understanding mode deprojection by means of analytic toy cases

In this appendix I present some simple analytic examples both to check the equations of Sec. 4.1 and to clarify their meaning.

B.1 Two bins containing two modes each

Consider a constant power spectrum $P(k) = 1\ \forall k$, four k-modes $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$ and $\mathbf{k}_4$ which are binned into two bins $\mathbb{k}_1 = \{\mathbf{k}_1, \mathbf{k}_2\}$ and $\mathbb{k}_2 = \{\mathbf{k}_3, \mathbf{k}_4\}$ with effective k-values k_1 and k_2, respectively. The matrices $\mathbf{E}$ are given by

$$\mathbf{E}(k_1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ \mathbf{E}(k_2) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{B.1}$$

The test cases for this setup are no contamination at all, one single mode is affected by systematics, and finally both modes in one bin are affected.

B.1.1 No contamination

If there is no contamination, i.e. $f(\mathbf{k}) = 0\ \forall \mathbf{k}$, one runs into a problem, because $R_P = 0$, by which one has to divide in Eq. (4.32). Going back to

Eq. (4.31), one sees that in this particular case, the updated covariance matrix $\tilde{\mathbf{C}}^{-1} = \mathbf{C}^{-1} - \lim_{\sigma\to\infty} \frac{\sigma \mathbf{C}^{-1}\mathbf{f}\mathbf{f}^\dagger \mathbf{C}^{-1}}{1+\sigma \mathbf{f}^\dagger \mathbf{C}^{-1}\mathbf{f}}$ converges towards $\mathbf{C}^{-1}$ as one would expect, because σ is only multiplied with 0.

B.1.2 One single mode is contaminated

One mode is contaminated in one bin, i.e. $f(\mathbf{k}_2)$, the others remain zero: The estimator matrix is updated according to Eq. (4.34):

$$\tilde{\mathbf{E}}_{\alpha\beta}(k_j) = \frac{\delta_{\alpha\beta}}{P_\alpha^2}\Theta_{\alpha j} - \frac{1}{R_P}\frac{f_\alpha f_\beta^*}{P_\alpha P_\beta}\left(\frac{\Theta_{\alpha j}}{P_\alpha} + \frac{\Theta_{\beta j}}{P_\beta} - \frac{t_j}{R_P}\right), \tag{B.2}$$

where

$$R_P = 1,\ t_1 = 1,\ t_2 = 0. \tag{B.3}$$

Calling the prefactor of the update

$$\Phi \equiv \frac{1}{R_P}\frac{f_\alpha f_\beta^*}{P_\alpha P_\beta}, \tag{B.4}$$

we have

$$\Phi = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{B.5}$$

Hence the $(2,2)$-components of the $\mathbf{E}$-matrices have to be updated:

$$\begin{aligned} \tilde{\mathbf{E}}_{22}(k_1) &= \mathbf{E}_{22}(k_1) - \Phi_{22}(1+1-1) = 0 \\ \tilde{\mathbf{E}}_{22}(k_2) &= \mathbf{E}_{22}(k_2) - \Phi_{22}(0+0-0) = 0. \end{aligned} \tag{B.6}$$

After updating, the $\mathbf{k}_2$-mode is thus ignored:

$$\tilde{\mathbf{E}}(k_1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ \tilde{\mathbf{E}}(k_2) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{B.7}$$

$$\tilde{\mathbf{N}}_{ij} = \sum_{\mathbf{k}_\alpha \in \mathbb{k}_j} \tilde{\mathbf{E}}_{\alpha\alpha}(k_i) \qquad \tilde{\mathbf{N}} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \qquad \tilde{\mathbf{N}}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{B.8}$$

	$\widehat{P}(k_1) = \frac{\lvert F(\mathbf{k}_1)\rvert^2 + \lvert F(\mathbf{k}_2)\rvert^2}{2}$	$\widehat{P}(k_1) = \lvert F(\mathbf{k}_1)\rvert^2$
mean	1.4990	0.9961
standard deviation	1.0038	0.9956

Table B.1: Mean and standard deviation of power spectra calculated from 100000 sets of $F(\mathbf{k}_1)$ and $F(\mathbf{k}_2)$ realised as Gaussian distributed pseudo-random numbers with variance $\frac{1}{\sqrt{P(k_1)}}$ and mean 0 or 1 respectively.

The resulting estimate of the power is then similar to Eq. (4.28), but without $\mathbf{k}_2$:

$$\begin{aligned} \widehat{P}(k_1) &= |F(\mathbf{k}_1)|^2 \\ \widehat{P}(k_2) &= \frac{|F(\mathbf{k}_3)|^2 + |F(\mathbf{k}_4)|^2}{2} \end{aligned} \tag{B.9}$$

A simple numerical test of equation Eq. (B.9) is to generate 100000 sets of Gaussian distributed pseudo random numbers for $F(\mathbf{k}_1)$ and $F(\mathbf{k}_2)$, and calculate the power according to Eq. (4.28) and (B.9). The resulting mean and standard deviation are given in Tab. B.1. As expected, not accounting for the systematics, provides a too high mean power spectrum, whereas estimating the power using Eq. (B.9) yields a mean value very close to the input value, while almost retaining the same standard deviation.

B.1.3 Fully contaminated bin

Let us now see what happens when all modes in a bin are contaminated: $R_P = \sum_\mu \frac{|f(\mathbf{k}_\mu)|^2}{P(\mathbf{k}_\mu)} = 2$, thus

$$\Phi = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{B.10}$$

Hence, the whole upper left block of $\mathbf{E}$ has to be updated:

$$\begin{aligned}
\tilde{\mathbf{E}}_{11}(k_1) = \quad & \mathbf{E}_{11}(k_1) - \quad \Phi_{11}(1+1-1) \quad = 1 \quad -\frac{1}{2} = \quad \frac{1}{2} \\
\tilde{\mathbf{E}}_{12}(k_1) = \quad & \mathbf{E}_{12}(k_1) - \quad \Phi_{12}(1+1-1) \quad = 0 \quad -\frac{1}{2} = \quad -\frac{1}{2} \\
\tilde{\mathbf{E}}_{21}(k_1) = \quad & \mathbf{E}_{21}(k_1) - \quad \Phi_{21}(1+1-1) \quad = 0 \quad -\frac{1}{2} = \quad -\frac{1}{2} \\
\tilde{\mathbf{E}}_{22}(k_1) = \quad & \mathbf{E}_{22}(k_1) - \quad \Phi_{22}(1+1-1) \quad = 1 \quad -\frac{1}{2} = \quad \frac{1}{2} \\
\tilde{\mathbf{E}}_{11}(k_2) = \quad & \mathbf{E}_{11}(k_2) - \quad \Phi_{11}(0+0-0) \quad = 0 \quad -\frac{1}{2} = \quad 0 \\
\tilde{\mathbf{E}}_{12}(k_2) = \quad & \mathbf{E}_{12}(k_2) - \quad \Phi_{12}(0+0-0) \quad = 0 \quad -\frac{1}{2} = \quad 0 \\
\tilde{\mathbf{E}}_{21}(k_2) = \quad & \mathbf{E}_{21}(k_2) - \quad \Phi_{21}(0+0-0) \quad = 0 \quad -\frac{1}{2} = \quad 0 \\
\tilde{\mathbf{E}}_{22}(k_2) = \quad & \mathbf{E}_{22}(k_2) - \quad \Phi_{22}(0+0-0) \quad = 0 \quad -\frac{1}{2} = \quad 0.
\end{aligned} \tag{B.11}$$

The normalisation is the same as in the previous case.

$$\begin{aligned}
\widehat{P}(k_1) &= \frac{|F(\mathbf{k}_1)|^2 + |F(\mathbf{k}_2)|^2 - F^*(\mathbf{k}_1)F(\mathbf{k}_2) - F(\mathbf{k}_1)F^*(\mathbf{k}_2)}{2} \\
&= \frac{|F(\mathbf{k}_1) - F(\mathbf{k}_2)|^2}{2} \\
\widehat{P}(k_2) &= \frac{|F(\mathbf{k}_3)|^2 + |F(\mathbf{k}_4)|^2}{2}
\end{aligned} \tag{B.12}$$

This gives as some insight of how the method works: If there was no contamination, $F(\mathbf{k}_1)$ and $F(\mathbf{k}_2)$ would be independent and identically distributed (iid) random variables X_1 and X_2. Now with the contaminations, they are shifted, and according to our template by the same amount: $F(\mathbf{k}_1) = X_1 + f$ and $F(\mathbf{k}_2) = X_2 + f$. The mixed term just subtracts the shift caused by f from the sum of the squared absolute values:

$$\begin{aligned}
\frac{|F(\mathbf{k}_1) - F(\mathbf{k}_2)|^2}{2} &= \frac{|X_1 + f - X_2 - f|^2}{2} \\
&= \frac{|X_1|^2 + |X_2|^2}{2} - \mathrm{Re}[X_1^* X_2].
\end{aligned} \tag{B.13}$$

The average of $|X_1|^2$ and $|X_2|^2$ is the true power, and the average $\langle \mathrm{Re}[X_1^* X_2] \rangle = \mathrm{Re}[\langle X_1^* \rangle \langle X_2 \rangle] = 0$, because X_1 and X_2 are zero-centred iid. One therefore finds that the estimate is on average the true power:

$$\langle \widehat{P}(k_1) \rangle = P(k_1). \tag{B.14}$$

Thus, we can get an unbiased estimate of the power even when all modes are contaminated, which shows that this is a much better methodology than just discarding contaminated modes.

The third term in Eq. (B.13) becomes important in the calculation for the variance, adding the uncertainty due to the contamination:

$$\begin{aligned}
\mathrm{Var}(\widehat{P}) &= \left\langle \widehat{P}^2 \right\rangle - \left\langle \widehat{P} \right\rangle^2 \\
\widehat{P}^2 &= \left(\frac{|X_1|^2 + |X_2|^2}{2} - \mathrm{Re}[X_1^* X_2] \right)^2 \\
&= \frac{\left(|X_1|^2 + |X_2|^2\right)^2}{4} - \left(|X_1|^2 + |X_2|^2\right) \mathrm{Re}[X_1^* X_2] + \mathrm{Re}[X_1^* X_2]^2 \qquad \text{(B.15)}
\end{aligned}$$

The first term is the square of the uncontaminated power $\widehat{P}_{\mathrm{clean}}$, the second term is zero on average, as it contains only uneven numbers of the same iid random variable, such as $X_1^* X_1 X_1^* X_2$, which can be split using Isserlis' or Wick's theorem such that $\langle X_1^* X_1 X_1^* X_2 \rangle = \langle X_1^* X_1 \rangle \langle X_1^* X_2 \rangle + \langle X_1^* X_1^* \rangle \langle X_1 X_2 \rangle + \langle X_1^* X_2 \rangle \langle X_1 X_1^* \rangle$, which is zero because X_1 and X_2 are independent and zero on average. The additional variance coming from cleaning the power using BMP is then given by

$$\begin{aligned}
\triangle \mathrm{Var}(\widehat{P}) &\equiv \mathrm{Var}(\widehat{P}) - \mathrm{Var}(\widehat{P}_{\mathrm{clean}}) \\
&= \langle \mathrm{Re}[X_1^* X_2]^2 \rangle \\
&= \frac{1}{4} \langle [X_1^* X_2 + X_1 X_2^*]^2 \rangle \\
&= \frac{\langle X_1^* X_2 X_1^* X_2 + 2 X_1^* X_2 X_1 X_2^* + X_1 X_2^* X_1 X_2^* \rangle}{4} \qquad \text{(B.16)}
\end{aligned}$$

Again applying Isserlis' or Wick's theorem and ignoring zero valued mixing terms of X_1 and X_2, one finds

$$\triangle \mathrm{Var}(\widehat{P}) = \frac{\langle X_1^* X_1^* \rangle \langle X_2 X_2 \rangle + 2 \langle |X_1|^2 \rangle \langle |X_2|^2 \rangle + \langle X_1 X_1 \rangle \langle X_2^* X_2^* \rangle}{4} \qquad \text{(B.17)}$$

The central term can be expressed by the true power $\frac{2\langle |X_1|^2 \rangle \langle |X_2|^2 \rangle}{4} = \frac{P^2}{2}$. For the other terms, one has to evaluate the autocorrelation of the complex random variables. To do so, consider that the complex random variables are composed by iid distributed real and imaginary parts R and I, such that $X = R + iI$ and $\langle XX \rangle = \langle (R + iI)(R + iI) \rangle = \langle RR \rangle + i \langle RI \rangle - \langle II \rangle$. The term in the middle is zero because R and I are independent, and the other two terms cancel each

	$\widehat{P}_1 = \frac{\lvert F_1\rvert^2+\lvert F_2\rvert^2}{2}$	$\widehat{P}_1 = \frac{\lvert F_1-F_2\rvert^2}{2}$	$\widehat{P}_2 = \frac{\lvert F_3\rvert^2+\lvert F_4\rvert^2}{2}$
mean	2.00835	1.00978	0.994913
standard deviation	1.22939	0.986795	0.713499
variance	1.51139	0.973764	0.509081

Table B.2: Mean and standard deviation of power spectra calculated from 10000 sets of Gaussian distributed pseudo-random numbers. A constant real number has been added to the first two numbers F_1 and F_2, whereas the last two F_3 and F_4 are zero centred.

other, because R and I are identically distributed. Hence, overall the variance increases by

$$\triangle \mathrm{Var}(\widehat{P}) = \frac{P^2}{2}. \tag{B.18}$$

The analytic calculations in this paragraph can be again checked with a simple numerical test as in the paragraph before: I generate again a set of 10000 random numbers corresponding to $F(\mathbf{k}_1)$ and $F(\mathbf{k}_2)$, and measure their power using Eq. (4.28). Again, not accounting for the systematics yields a too high mean power spectrum, and using basic mode projection provides a power spectrum close to the input value (cf. Tab. B.2). Basic mode projection changes the distribution of the power estimates, as can be seen in Fig. B.1. Tab. B.2 also confirms the result of Eq. (B.18): $\widehat{P}_2$ is a clean power spectrum and $\widehat{P}_1$ is affected by systematics and has been cleaned using mode deprojection. Its variance is about a half larger than $\widehat{P}_2$, which is what Eq. (B.18) predicts if the true power is equal to 1.

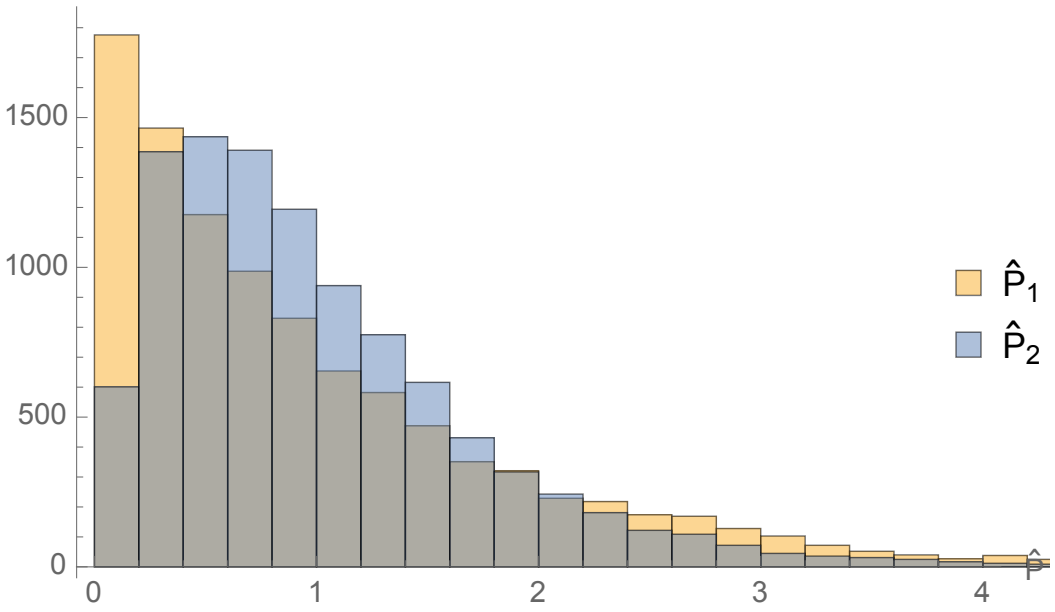

Figure B.1: Distribution of the power spectra estimated from 10000 realisations of 4 bivariate Gaussian random fields after BMP.

Bibliography

[1] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, "The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects," *Phys. Rev.* **D77** (2008) 123514, arXiv:0710.4560 [astro-ph].

[2] V. Desjacques, D. Jeong, and F. Schmidt, "Accurate Predictions for the Scale-Dependent Galaxy Bias from Primordial Non-Gaussianity," *Phys. Rev.* **D84** (2011) 061301, arXiv:1105.3476 [astro-ph.CO].

[3] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XVII. Constraints on primordial non-Gaussianity," *Astron. Astrophys.* **594** (2016) A17, arXiv:1502.01592 [astro-ph.CO].

[4] B. Kalus, W. J. Percival, and L. Samushia, "Cosmological parameter inference from galaxy clustering: The effect of the posterior distribution of the power spectrum," *Mon. Not. Roy. Astron. Soc.* **455** no. 3, (2016) 2573–2581, arXiv:1504.03979 [astro-ph.CO].

[5] B. Kalus, W. J. Percival, D. Bacon, and L. Samushia, "Unbiased contaminant removal for 3D galaxy power spectrum measurements," *Mon. Not. Roy. Astron. Soc.* **463** no. 1, (2016) 467–476, arXiv:1607.02417 [astro-ph.CO].

[6] A. J. Ross *et al.*, "The Clustering of Galaxies in SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Constraints on Primordial Non-Gaussianity," *Mon. Not. Roy. Astron. Soc.* **428** (2013) 1116–1127, arXiv:1208.1491 [astro-ph.CO].

[7] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.* **594** (2016) A13, arXiv:1502.01589 [astro-ph.CO].

[8] D. Wands, "Local non-Gaussianity from inflation," *Class. Quant. Grav.* **27** (2010) 124002, arXiv:1004.0818 [astro-ph.CO].

[9] A. Font-Ribera, P. McDonald, N. Mostek, B. A. Reid, H.-J. Seo, and A. Slosar, "DESI and other dark energy experiments in the era of neutrino mass measurements," *JCAP* **1405** (2014) 023, arXiv:1308.4164 [astro-ph.CO].

[10] M. Tellarini, A. J. Ross, G. Tasinato, and D. Wands, "Galaxy bispectrum, primordial non-Gaussianity and redshift space distortions," *JCAP* **1606** no. 06, (2016) 014, arXiv:1603.06814 [astro-ph.CO].

[11] M. Ata *et al.*, "The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: First measurement of Baryon Acoustic Oscillations between redshift 0.8 and 2.2," arXiv:1705.06373 [astro-ph.CO].

[12] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, "The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant," *Mon. Not. Roy. Astron. Soc.* **416** (2011) 3017–3032, arXiv:1106.3366 [astro-ph.CO].

[13] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, "The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at $z = 0.15$," *Mon. Not. Roy. Astron. Soc.* **449** no. 1, (2015) 835–847, arXiv:1409.3242 [astro-ph.CO].

[14] **BOSS** Collaboration, S. Alam *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample," *Mon. Not. Roy. Astron. Soc.* **470** no. 3, (2017) 2617–2652, arXiv:1607.03155 [astro-ph.CO].

[15] E. A. Kazin *et al.*, "The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature," *Mon. Not. Roy. Astron. Soc.* **441** no. 4, (2014) 3524–3542, arXiv:1401.0358 [astro-ph.CO].

[16] J. E. Bautista *et al.*, "Measurement of BAO correlations at $z = 2.3$ with SDSS DR12 Lyα-Forests," *Astron. Astrophys.* **603** (2017) A12, arXiv:1702.00176 [astro-ph.CO].

[17] C. Heymans *et al.*, "CFHTLenS: The Canada-France-Hawaii Telescope Lensing Survey," *Mon. Not. Roy. Astron. Soc.* **427** (2012) 146, arXiv:1210.0032 [astro-ph.CO].

[18] F. Köhlinger *et al.*, "KiDS-450: The tomographic weak lensing power spectrum and constraints on cosmological parameters," arXiv:1706.02892 [astro-ph.CO].

[19] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity," *Astron. Astrophys.* **571** (2014) A24, arXiv:1303.5084 [astro-ph.CO].

[20] G. Efstathiou, "H0 Revisited," *Mon. Not. Roy. Astron. Soc.* **440** no. 2, (2014) 1138–1152, arXiv:1311.3461 [astro-ph.CO].

[21] W. J. Percival, "Large Scale Structure Observations," in *Proceedings, International School of Physics 'Enrico Fermi': New Horizons for Observational Cosmology: Rome, Italy, June 30-July 6, 2013*, vol. 186, pp. 101–135. 2015. arXiv:1312.5490 [astro-ph.CO]. http://inspirehep.net/record/1272804/files/arXiv:1312.5490.pdf. [,317(2015)].

[22] A. Lewis, A. Challinor, and A. Lasenby, "Efficient computation of CMB anisotropies in closed FRW models," *Astrophys. J.* **538** (2000) 473–476, arXiv:astro-ph/9911177 [astro-ph].

[23] J. Lesgourgues and S. Pastor, "Massive neutrinos and cosmology," *Phys. Rept.* **429** (2006) 307–379, arXiv:astro-ph/0603494 [astro-ph].

[24] B. A. Reid, H.-J. Seo, A. Leauthaud, J. L. Tinker, and M. White, "A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies," *Mon. Not. Roy. Astron. Soc.* **444** no. 1, (2014) 476–502, arXiv:1404.3742 [astro-ph.CO].

[25] B. Reid *et al.*, "SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large scale structure catalogues," *Mon. Not. Roy. Astron. Soc.* **455** no. 2, (2016) 1553–1573, arXiv:1509.06529 [astro-ph.CO].

[26] **BOSS** Collaboration, A. J. Ross *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Analysis of potential systematics," *Mon. Not. Roy. Astron. Soc.* **424** (2012) 564, arXiv:1203.6499 [astro-ph.CO].

[27] M. R. Blanton *et al.*, "Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies and the Distant Universe," arXiv:1703.00052 [astro-ph.GA].

[28] **DES** Collaboration, T. Abbott *et al.*, "The Dark Energy Survey: more than dark energy – an overview," *Mon. Not. Roy. Astron. Soc.* **460** no. 2, (2016) 1270–1299, arXiv:1601.00329 [astro-ph.CO].

[29] K. Lemmens, "An investigation, implementation and comparison of 3 important particle simulation techniques: Pp : Particle-particle pm : Particle-mesh and tc : Tree-code," Master's thesis, University of Technology, Delft, 2, 1998.

[30] M. Vargas-Magana *et al.*, "An optimized correlation function estimator for galaxy surveys," *Astron. Astrophys.* **554** (2013) A131, arXiv:1211.6211 [astro-ph.CO].

[31] F. Leclercq, A. Pisani, and B. D. Wandelt, "Cosmology: from theory to data, from data to theory," *Proc. Int. Sch. Phys. Fermi* **186** (2014) 189–233, arXiv:1403.1260 [astro-ph.CO].

[32] P. Norberg, C. M. Baugh, E. Gaztanaga, and D. J. Croton, "Statistical Analysis of Galaxy Surveys - I. Robust error estimation for 2-point

clustering statistics," *Mon. Not. Roy. Astron. Soc.* **396** (2009) 19, arXiv:0810.1885 [astro-ph].

[33] A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik* **354** no. 7, (1916) 769–822. http://dx.doi.org/10.1002/andp.19163540702.

[34] S. Weinberg, *Gravitation and cosmology*. Wiley, New York, 1972.

[35] M. P. Hobson, G. Efstathiou, and A. N. Lasenby, *General relativity*. Cambridge Univ. Press, Cambridge, 2006.

[36] A. Friedmann, "Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes," *Zeitschrift fur Physik* **21** (Dec., 1924) 326–332.

[37] 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** (1927) 49–59.

[38] H. P. Robertson, "Kinematics and World-Structure," *Astrophys. J.* **82** (Nov., 1935) 284.

[39] **Supernova Search Team** Collaboration, A. G. Riess *et al.*, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.* **116** (1998) 1009–1038, arXiv:astro-ph/9805201 [astro-ph].

[40] **Supernova Cosmology Project** Collaboration, S. Perlmutter *et al.*, "Measurements of Omega and Lambda from 42 high redshift supernovae," *Astrophys. J.* **517** (1999) 565–586, arXiv:astro-ph/9812133 [astro-ph].

[41] R. Kippenhahn and A. Weigert, *Stellar structure and evolution*. Springer, Berlin, 1990.

[42] **SNLS** Collaboration, A. Conley *et al.*, "Supernova Constraints and Systematic Uncertainties from the First 3 Years of the Supernova

Legacy Survey," *Astrophys. J. Suppl.* **192** (2011) 1, arXiv:1104.1443 [astro-ph.CO].

[43] **SDSS** Collaboration, M. Betoule *et al.*, "Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples," *Astron. Astrophys.* **568** (2014) A22, arXiv:1401.4064 [astro-ph.CO].

[44] A. A. Penzias and R. W. Wilson, "A Measurement of excess antenna temperature at 4080-Mc/s," *Astrophys. J.* **142** (1965) 419–421.

[45] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, "Cosmic Black-Body Radiation," *Astrophys. J.* **142** (1965) 414–419.

[46] D. J. Fixsen, "The Temperature of the Cosmic Microwave Background," *Astrophys. J.* **707** (2009) 916–920, arXiv:0911.1955 [astro-ph.CO].

[47] T. Erben *et al.*, "CFHTLenS: The Canada-France-Hawaii Telescope Lensing Survey - Imaging Data and Catalogue Products," *Mon. Not. Roy. Astron. Soc.* **433** (2013) 2545, arXiv:1210.8156 [astro-ph.CO].

[48] **CFHTLenS** Collaboration, T. D. Kitching *et al.*, "3D Cosmic Shear: Cosmology from CFHTLenS," *Mon. Not. Roy. Astron. Soc.* **442** no. 2, (2014) 1326–1349, arXiv:1401.6842 [astro-ph.CO].

[49] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, and R. Chornock, "A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3," *Astrophys. J.* **730** (2011) 119, arXiv:1103.2976 [astro-ph.CO]. [Erratum: Astrophys. J.732,129(2011)].

[50] V. Marra, L. Amendola, I. Sawicki, and W. Valkenburg, "Cosmic variance and the measurement of the local Hubble parameter," *Phys. Rev. Lett.* **110** no. 24, (2013) 241305, arXiv:1303.3121 [astro-ph.CO].

[51] I. Ben-Dayan, R. Durrer, G. Marozzi, and D. J. Schwarz, "The value of H_0 in the inhomogeneous Universe," *Phys. Rev. Lett.* **112** (2014) 221301, arXiv:1401.7973 [astro-ph.CO].

[52] D. J. Heath, "The growth of density perturbations in zero pressure friedmann–lemaître universes," *Monthly Notices of the Royal Astronomical Society* **179** no. 3, (1977) 351–358, /oup/backfile/content_public/journal/mnras/179/3/10.1093_mnras_179.3.351/3/ +http://dx.doi.org/10.1093/mnras/179.3.351.

[53] O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees, "Dynamical effects of the cosmological constant," *Mon. Not. Roy. Astron. Soc.* **251** (1991) 128–136.

[54] Ya. B. Zeldovich, "Gravitational instability: An Approximate theory for large density perturbations," *Astron. Astrophys.* **5** (1970) 84–89.

[55] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz, "Perturbative Lagrangian approach to gravitational instability," *Astron. Astrophys.* **296** (1995) 575, arXiv:astro-ph/9406013 [astro-ph].

[56] J. A. Peacock, *Cosmological physics*. Cambridge Univ. Press, Cambridge [u.a.], reprint. ed., 2002.

[57] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika* **12** no. 1-2, (1918) 134. +http://dx.doi.org/10.1093/biomet/12.1-2.134.

[58] E. Komatsu and D. N. Spergel, "Acoustic signatures in the primary microwave background bispectrum," *Phys. Rev.* **D63** (2001) 063002, arXiv:astro-ph/0005036 [astro-ph].

[59] D. Babich, P. Creminelli, and M. Zaldarriaga, "The Shape of non-Gaussianities," *JCAP* **0408** (2004) 009, arXiv:astro-ph/0405356 [astro-ph].

[60] X. Chen, "Multi-throat brane inflation," *Phys. Rev.* **D71** (2005) 063506, arXiv:hep-th/0408084 [hep-th].

[61] L. Senatore, K. M. Smith, and M. Zaldarriaga, "Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data," *JCAP* **1001** (2010) 028, arXiv:0905.3746 [astro-ph.CO].

[62] B. S. Ryden, *Introduction to cosmology*. Pearson education. Addison Wesley, San Francisco, 2003.

[63] V. F. Mukhanov, *Physical foundations of cosmology*. Cambridge Univ. Press, Cambridge, 2005.

[64] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *JHEP* **05** (2003) 013, arXiv:astro-ph/0210603 [astro-ph].

[65] S. Mollerach, "Isocurvature Baryon Perturbations and Inflation," *Phys. Rev.* **D42** (1990) 313–325.

[66] A. D. Linde and V. F. Mukhanov, "Nongaussian isocurvature perturbations from inflation," *Phys. Rev.* **D56** (1997) R535–R539, arXiv:astro-ph/9610219 [astro-ph].

[67] K. Enqvist and M. S. Sloth, "Adiabatic CMB perturbations in pre - big bang string cosmology," *Nucl. Phys.* **B626** (2002) 395–409, arXiv:hep-ph/0109214 [hep-ph].

[68] D. H. Lyth and D. Wands, "Generating the curvature perturbation without an inflaton," *Phys. Lett.* **B524** (2002) 5–14, arXiv:hep-ph/0110002 [hep-ph].

[69] T. Moroi and T. Takahashi, "Effects of cosmological moduli fields on cosmic microwave background," *Phys. Lett.* **B522** (2001) 215–221, arXiv:hep-ph/0110096 [hep-ph]. [Erratum: Phys. Lett.B539,303(2002)].

[70] D. H. Lyth and Y. Rodriguez, "The Inflationary prediction for primordial non-Gaussianity," *Phys. Rev. Lett.* **95** (2005) 121302, arXiv:astro-ph/0504045 [astro-ph].

[71] M. Sasaki, J. Valiviita, and D. Wands, "Non-Gaussianity of the primordial perturbation in the curvaton model," *Phys. Rev.* **D74** (2006) 103003, arXiv:astro-ph/0607627 [astro-ph].

[72] N. Bartolo, S. Matarrese, and A. Riotto, "On nonGaussianity in the curvaton scenario," *Phys. Rev.* **D69** (2004) 043503, arXiv:hep-ph/0309033 [hep-ph].

[73] C. T. Byrnes and K.-Y. Choi, "Review of local non-Gaussianity from multi-field inflation," *Adv. Astron.* **2010** (2010) 724525, arXiv:1002.3110 [astro-ph.CO].

[74] K. Kohri, D. H. Lyth, and C. A. Valenzuela-Toledo, "Preheating and the non-gaussianity of the curvature perturbation," *JCAP* **1002** (2010) 023, arXiv:0904.0793 [hep-ph]. [Erratum: JCAP1009,E01(2011)].

[75] C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, "k - inflation," *Phys. Lett.* **B458** (1999) 209–218, arXiv:hep-th/9904075 [hep-th].

[76] X. Chen, M.-x. Huang, S. Kachru, and G. Shiu, "Observational signatures and non-Gaussianities of general single field inflation," *JCAP* **0701** (2007) 002, arXiv:hep-th/0605045 [hep-th].

[77] P. Creminelli and L. Senatore, "A Smooth bouncing cosmology with scale invariant spectrum," *JCAP* **0711** (2007) 010, arXiv:hep-th/0702165 [hep-th].

[78] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, "Non-Gaussianities from ekpyrotic collapse with multiple fields," *JCAP* **0711** (2007) 024, arXiv:0708.4321 [hep-th].

[79] A. Ricciardone, "Primordial Gravitational Waves with LISA," *J. Phys. Conf. Ser.* **840** no. 1, (2017) 012030, arXiv:1612.06799 [astro-ph.CO].

[80] **Virgo, LIGO Scientific** Collaboration, B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.* **116** no. 6, (2016) 061102, arXiv:1602.03837 [gr-qc].

[81] **Virgo, LIGO Scientific** Collaboration, B. P. Abbott *et al.*, "GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence," *Phys. Rev. Lett.* **116** no. 24, (2016) 241103, arXiv:1606.04855 [gr-qc].

[82] **VIRGO, LIGO Scientific** Collaboration, B. P. Abbott *et al.*, "GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2," *Phys. Rev. Lett.* **118** no. 22, (2017) 221101, arXiv:1706.01812 [gr-qc].

[83] **Virgo, LIGO Scientific** Collaboration, B. Abbott *et al.*, "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," *Phys. Rev. Lett.* **119** no. 16, (2017) 161101, arXiv:1710.05832 [gr-qc].

[84] **Planck** Collaboration, R. Adam *et al.*, "Planck intermediate results. XXX. The angular power spectrum of polarized dust emission at intermediate and high Galactic latitudes," *Astron. Astrophys.* **586** (2016) A133, arXiv:1409.5738 [astro-ph.CO].

[85] **BICEP2** Collaboration, P. A. R. Ade *et al.*, "Detection of B-Mode Polarization at Degree Angular Scales by BICEP2," *Phys. Rev. Lett.* **112** no. 24, (2014) 241101, arXiv:1403.3985 [astro-ph.CO].

[86] S. Dodelson, *Modern cosmology*. Acad. Press, Amsterdam, 2006.

[87] P. Meszaros, "The behaviour of point masses in an expanding cosmological substratum," *Astron. Astrophys.* **37** (1974) 225–228.

[88] D. J. Eisenstein and W. Hu, "Baryonic features in the matter transfer function," *Astrophys.J.* **496** (1998) 605, arXiv:astro-ph/9709112 [astro-ph].

[89] C. Howlett, A. Lewis, A. Hall, and A. Challinor, "CMB power spectrum parameter degeneracies in the era of precision cosmology," *JCAP* **1204** (2012) 027, arXiv:1201.3654 [astro-ph.CO].

[90] J. Lesgourgues, "The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview," arXiv:1104.2932 [astro-ph.IM].

[91] D. Blas, J. Lesgourgues, and T. Tram, "The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes," *JCAP* **1107** (2011) 034, arXiv:1104.2933 [astro-ph.CO].

[92] J. Lesgourgues, "The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparision with CAMB for LambdaCDM," arXiv:1104.2934 [astro-ph.CO].

[93] J. Lesgourgues and T. Tram, "The Cosmic Linear Anisotropy Solving System (CLASS) IV: efficient implementation of non-cold relics," *JCAP* **1109** (2011) 032, arXiv:1104.2935 [astro-ph.CO].

[94] W. H. Press and P. Schechter, "Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation," *Astrophys. J.* **187** (1974) 425–438.

[95] E. Sefusatti, C. Vale, K. Kadota, and J. Frieman, "Primordial non-Gaussianity and Dark Energy constraints from Cluster Surveys," *Astrophys. J.* **658** (2007) 669–679, arXiv:astro-ph/0609124 [astro-ph].

[96] S. Matarrese, L. Verde, and R. Jimenez, "The Abundance of high-redshift objects as a probe of non-Gaussian initial conditions," *Astrophys. J.* **541** (2000) 10, arXiv:astro-ph/0001366 [astro-ph].

[97] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, "The Statistics of Peaks of Gaussian Random Fields," *Astrophys. J.* **304** (1986) 15–61.

[98] S. Cole and N. Kaiser, "Biased clustering in the cold dark matter cosmogony," *Mon. Not. Roy. Astron. Soc.* **237** (1989) 1127–1146.

[99] V. Desjacques, D. Jeong, and F. Schmidt, "Large-Scale Galaxy Bias," arXiv:1611.09787 [astro-ph.CO].

[100] F. Schmidt and M. Kamionkowski, "Halo Clustering with Non-Local Non-Gaussianity," *Phys. Rev.* **D82** (2010) 103002, arXiv:1008.0638 [astro-ph.CO].

[101] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, "Constraints on local primordial non-Gaussianity from large scale structure," *JCAP* **0808** (2008) 031, arXiv:0805.3580 [astro-ph].

[102] T. Giannantonio and C. Porciani, "Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N-body simulations," *Phys. Rev.* **D81** (2010) 063530, arXiv:0911.0017 [astro-ph.CO].

[103] U. Seljak, "Extracting primordial non-gaussianity without cosmic variance," *Phys. Rev. Lett.* **102** (2009) 021302, arXiv:0807.1770 [astro-ph].

[104] K. M. Smith, S. Ferraro, and M. LoVerde, "Halo clustering and g_{NL}-type primordial non-Gaussianity," *JCAP* **1203** (2012) 032, arXiv:1106.0503 [astro-ph.CO].

[105] M. Bruni, R. Crittenden, K. Koyama, R. Maartens, C. Pitrou, and D. Wands, "Disentangling non-Gaussianity, bias and GR effects in the galaxy distribution," *Phys. Rev.* **D85** (2012) 041301, arXiv:1106.3999 [astro-ph.CO].

[106] **CMB-S4** Collaboration, K. N. Abazajian *et al.*, "CMB-S4 Science Book, First Edition," arXiv:1610.02743 [astro-ph.CO].

[107] **CORE** Collaboration, F. Finelli *et al.*, "Exploring Cosmic Origins with CORE: Inflation," arXiv:1612.08270 [astro-ph.CO].

[108] B. Leistedt, H. V. Peiris, and N. Roth, "Constraints on Primordial Non-Gaussianity from 800 000 Photometric Quasars," *Phys. Rev. Lett.* **113** no. 22, (2014) 221301, arXiv:1405.4315 [astro-ph.CO].

[109] H. A. Feldman, N. Kaiser, and J. A. Peacock, "Power spectrum analysis of three-dimensional redshift surveys," *Astrophys. J.* **426** (1994) 23–37, arXiv:astro-ph/9304022 [astro-ph].

[110] C. Alcock and B. Paczynski, "An evolution free test for non-zero cosmological constant," *Nature* **281** (1979) 358–359.

[111] **BOSS** Collaboration, F. Beutler *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space," *Mon. Not. Roy. Astron. Soc.* **464** no. 3, (2017) 3409–3430, arXiv:1607.03149 [astro-ph.CO].

[112] **BOSS** Collaboration, A. J. Ross *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function," *Mon. Not. Roy. Astron. Soc.* **464** no. 1, (2017) 1168–1191, arXiv:1607.03145 [astro-ph.CO].

[113] **BOSS** Collaboration, A. J. Ross *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function," *Mon. Not. Roy. Astron. Soc.* **464** no. 1, (2017) 1168–1191, arXiv:1607.03145 [astro-ph.CO].

[114] N. Kaiser, "Clustering in real space and in redshift space," *Mon.Not.Roy.Astron.Soc.* **227** (1987) 1–27.

[115] L. Samushia, W. J. Percival, L. Guzzo, Y. Wang, A. Cimatti, *et al.*, "Effects of cosmological model assumptions on galaxy redshift survey measurements," *Mon.Not.Roy.Astron.Soc.* **410** (2011) 1993–2002, arXiv:1006.0609 [astro-ph.CO].

[116] Y. Wang, D. N. Spergel, and E. L. Turner, "Implications of cosmic microwave background anisotropies for large scale variations in Hubble's constant," *Astrophys.J.* **498** (1998) 1, arXiv:astro-ph/9708014 [astro-ph].

[117] M. Maltoni, T. Schwetz, M. A. Tortola, and J. W. F. Valle, "Status of global fits to neutrino oscillations," *New J. Phys.* **6** (2004) 122, arXiv:hep-ph/0405172 [hep-ph].

[118] J. Lesgourgues and S. Pastor, "Neutrino cosmology and Planck," *New J. Phys.* **16** (2014) 065002, arXiv:1404.1740 [hep-ph].

[119] **SDSS** Collaboration, D. G. York *et al.*, "The Sloan Digital Sky Survey: Technical Summary," *Astron. J.* **120** (2000) 1579–1587, arXiv:astro-ph/0006396 [astro-ph].

[120] **SDSS** Collaboration, D. J. Eisenstein *et al.*, "SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems," *Astron. J.* **142** (2011) 72, arXiv:1101.1529 [astro-ph.IM].

[121] **SDSS** Collaboration, J. E. Gunn *et al.*, "The Sloan digital sky survey photometric camera," *Astron. J.* **116** (1998) 3040, arXiv:astro-ph/9809085 [astro-ph].

[122] **SDSS** Collaboration, D. J. Eisenstein *et al.*, "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies," *Astrophys. J.* **633** (2005) 560–574, arXiv:astro-ph/0501171 [astro-ph].

[123] **SDSS** Collaboration, H. Aihara *et al.*, "The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III," *Astrophys. J. Suppl.* **193** (2011) 29, arXiv:1101.1559 [astro-ph.IM]. [Erratum: Astrophys. J. Suppl.195,26(2011)].

[124] N. Padmanabhan *et al.*, "An Improved Photometric Calibration of the Sloan Digital Sky Survey Imaging Data," *Astrophys. J.* **674** (2008) 1217–1233, arXiv:astro-ph/0703454 [ASTRO-PH].

[125] A. J. Cuesta *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the correlation function of LOWZ and CMASS galaxies in Data Release 12," *Mon. Not. Roy. Astron. Soc.* **457** no. 2, (2016) 1770–1785, arXiv:1509.06371 [astro-ph.CO].

[126] **SDSS** Collaboration, C. Stoughton *et al.*, "The Sloan Digital Sky Survey: Early Data Release," *Astron. J.* **123** (2002) 485–548.

[127] **SDSS** Collaboration, K. Abazajian *et al.*, "The Second data release of the Sloan digital sky survey," *Astron. J.* **128** (2004) 502–512, arXiv:astro-ph/0403325 [astro-ph].

[128] D. J. Schlegel, D. P. Finkbeiner, and M. Davis, "Maps of dust IR emission for use in estimation of reddening and CMBR foregrounds," *Astrophys. J.* **500** (1998) 525, arXiv:astro-ph/9710327 [astro-ph].

[129] **BOSS** Collaboration, L. Anderson *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples," *Mon. Not. Roy. Astron. Soc.* **441** no. 1, (2014) 24–62, arXiv:1312.4877 [astro-ph.CO].

[130] D. Bianchi and W. J. Percival, "Unbiased clustering estimation in the presence of missing observations," arXiv:1703.02070 [astro-ph.CO].

[131] A. J. Ross *et al.*, "Ameliorating Systematic Uncertainties in the Angular Clustering of Galaxies: A Study using SDSS-III," *Mon. Not. Roy. Astron. Soc.* **417** (2011) 1350, arXiv:1105.2320 [astro-ph.CO].

[132] G.-B. Zhao *et al.*, "The extended Baryon Oscillation Spectroscopic Survey: a cosmological forecast," *Mon. Not. Roy. Astron. Soc.* **457** no. 3, (2016) 2377–2390, arXiv:1510.08216 [astro-ph.CO].

[133] D. Yamauchi, K. Takahashi, and M. Oguri, "Constraining primordial non-Gaussianity via a multitracer technique with surveys by Euclid and the Square Kilometre Array," *Phys. Rev.* **D90** no. 8, (2014) 083520, arXiv:1407.5453 [astro-ph.CO].

[134] D. W. Pearson, L. Samushia, and P. Gagrani, "Optimal weights for measuring redshift space distortions in multitracer galaxy catalogues," *Mon. Not. Roy. Astron. Soc.* **463** no. 3, (2016) 2708–2715, arXiv:1606.03435 [astro-ph.CO].

[135] G. J. Hill *et al.*, "The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results," *ASP Conf. Ser.* **399** (2008) 115–118, arXiv:0806.0183 [astro-ph].

[136] **BigBoss** Collaboration, D. Schlegel *et al.*, "The BigBOSS Experiment," arXiv:1106.1706 [astro-ph.IM].

[137] **EUCLID** Collaboration, R. Laureijs *et al.*, "Euclid Definition Study Report," arXiv:1110.3193 [astro-ph.CO].

[138] C. Fedeli, C. Carbone, L. Moscardini, and A. Cimatti, "The clustering of galaxies and galaxy clusters: constraints on primordial non-Gaussianity from future wide-field surveys," *Mon. Not. Roy. Astron. Soc.* **414** (2011) 1545–1559, arXiv:1012.2305 [astro-ph.CO].

[139] A. Cimatti, R. Laureijs, B. Leibundgut, S. Lilly, R. Nichol, A. Refregier, P. Rosati, M. Steinmetz, N. Thatte, and E. Valentijn, "Euclid Assessment Study Report for the ESA Cosmic Visions," arXiv:0912.0914 [astro-ph.CO].

[140] T. Giannantonio, C. Porciani, J. Carron, A. Amara, and A. Pillepich, "Constraining primordial non-Gaussianity with future galaxy surveys," *Mon. Not. Roy. Astron. Soc.* **422** (2012) 2854–2877, arXiv:1109.0958 [astro-ph.CO].

[141] J. Fonseca, R. Maartens, and M. G. Santos, "Probing the primordial Universe with MeerKAT and DES," *Mon. Not. Roy. Astron. Soc.* **466** no. 3, (2017) 2780–2786, arXiv:1611.01322 [astro-ph.CO].

[142] **DES** Collaboration, J. Elvin-Poole *et al.*, "Dark Energy Survey Year 1 Results: Galaxy clustering for combined probes," arXiv:1708.01536 [astro-ph.CO].

[143] J.-Q. Xia, M. Viel, C. Baccigalupi, G. De Zotti, S. Matarrese, and L. Verde, "Primordial Non-Gaussianity and the NRAO VLA Sky Survey," *Astrophys. J.* **717** (2010) L17, arXiv:1003.3451 [astro-ph.CO].

[144] S. Chen and D. J. Schwarz, "Angular two-point correlation of NVSS galaxies revisited," *Astron. Astrophys.* **591** (2016) A135, arXiv:1507.02160 [astro-ph.CO].

[145] S. Camera, M. G. Santos, P. G. Ferreira, and L. Ferramacho, "Cosmology on Ultra-Large Scales with HI Intensity Mapping: Limits on Primordial non-Gaussianity," *Phys. Rev. Lett.* **111** (2013) 171302, arXiv:1305.6928 [astro-ph.CO].

[146] A. Pourtsidou, D. Bacon, R. Crittenden, and R. B. Metcalf, "Prospects for clustering and lensing measurements with forthcoming intensity mapping and optical surveys," *Mon. Not. Roy. Astron. Soc.* **459** no. 1, (2016) 863–870, arXiv:1509.03286 [astro-ph.CO].

[147] S. Passmoor, C. Cress, A. Faltenbacher, R. Johnston, M. Smith, A. Ratsimbazafy, and B. Hoyle, "Probing the bias of radio sources at high redshift," *Mon. Not. Roy. Astron. Soc.* **429** (2013) 2183, arXiv:1211.5589 [astro-ph.CO].

[148] S. Camera, M. G. Santos, and R. Maartens, "Probing primordial non-Gaussianity with SKA galaxy redshift surveys: a fully relativistic analysis," *Mon. Not. Roy. Astron. Soc.* **448** no. 2, (2015) 1035–1043, arXiv:1409.8286 [astro-ph.CO].

[149] **BOSS** Collaboration, F. Beutler *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles," *Mon. Not. Roy. Astron. Soc.* **443** no. 2, (2014) 1065–1089, arXiv:1312.4611 [astro-ph.CO].

[150] K. Yamamoto, M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka, "A Measurement of the quadrupole power spectrum in the clustering of the 2dF QSO Survey," *Publ. Astron. Soc. Jap.* **58** (2006) 93–102, arXiv:astro-ph/0505115 [astro-ph].

[151] D. Bianchi, H. Gil-Marín, R. Ruggeri, and W. J. Percival, "Measuring line-of-sight dependent Fourier-space clustering using FFTs," *Mon. Not. Roy. Astron. Soc.* **453** no. 1, (2015) L11–L15, arXiv:1505.05341 [astro-ph.CO].

[152] S. D. Landy and A. S. Szalay, "Bias and variance of angular correlation functions," *Astrophys. J.* **412** (1993) 64.

[153] M. J. Wilson, J. A. Peacock, A. N. Taylor, and S. de la Torre, "Rapid modelling of the redshift-space power spectrum multipoles for a masked density field," *Mon. Not. Roy. Astron. Soc.* **464** no. 3, (2017) 3121–3130, arXiv:1511.07799 [astro-ph.CO].

[154] **BOSS** Collaboration, F. Beutler *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Anisotropic galaxy clustering in Fourier-space," *Mon. Not. Roy. Astron. Soc.* **466** no. 2, (2017) 2242–2260, arXiv:1607.03150 [astro-ph.CO].

[155] E. T. Jaynes, *Probability theory*. Cambridge Univ. Press, Cambridge, reprint. with corr., 4. print. ed., 2006.

[156] O. Cromwell and T. Carlyle, *Oliver Cromwell's Letters and Speeches: with Elucidations by Thomas Carlyle*. No. Bd. 4 in Collection of British authors. Tauchnitz, 1861. https://books.google.co.uk/books?id=RnLY9pXxfAoC.

[157] O. Häggström, *Finite Markov Chains and Algorithmic Applications*. London Mathematical Society Student Texts. Cambridge University Press, 2002. http://books.google.de/books?id=hpLxIJ9LwRgC.

[158] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika* **57** (1970) 97–109.

[159] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.* **21** (1953) 1087–1092.

[160] A. Lewis and S. Bridle, "Cosmological parameters from CMB and other data: A Monte Carlo approach," *Phys. Rev.* **D66** (2002) 103511, arXiv:astro-ph/0205436 [astro-ph].

[161] A. Lewis, "Efficient sampling of fast and slow cosmological parameters," *Phys. Rev.* **D87** (2013) 103529, arXiv:1304.4473 [astro-ph.CO].

[162] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Phys. Lett.* **B195** (1987) 216–222.

[163] M. H. Quenouille, "Problems in plane sampling," *Ann. Math. Statist.* **20** no. 3, (09, 1949) 355–375. http://dx.doi.org/10.1214/aoms/1177729989.

[164] J. W. Tukey, "Bias and confidence in not-quite large samples (abstract)," *Annals of Mathematical Statistics* **29** no. 2, (June, 1958) 614–614.

[165] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.* **7** no. 1, (01, 1979) 1–26. http://dx.doi.org/10.1214/aos/1176344552.

[166] **BOSS** Collaboration, K. S. Dawson *et al.*, "The Baryon Oscillation Spectroscopic Survey of SDSS-III," *Astron. J.* **145** (2013) 10, arXiv:1208.0022 [astro-ph.CO].

[167] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, "The Evolution of Large Scale Structure in a Universe Dominated by Cold Dark Matter," *Astrophys. J.* **292** (1985) 371–394.

[168] R. Scoccimarro and R. K. Sheth, "PTHalos: A Fast method for generating mock galaxy distributions," *Mon. Not. Roy. Astron. Soc.* **329** (2002) 629–640, arXiv:astro-ph/0106120 [astro-ph].

[169] M. Manera *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues," *Mon. Not. Roy. Astron. Soc.* **428** no. 2, (2012) 1036–1054, arXiv:1203.6609 [astro-ph.CO].

[170] T. R. Quinn, N. Katz, J. Stadel, and G. Lake, "Time stepping N body simulations," *Submitted to: Astrophys. J.* (1997) , arXiv:astro-ph/9710043 [astro-ph].

[171] S. Tassev, M. Zaldarriaga, and D. Eisenstein, "Solving Large Scale Structure in Ten Easy Steps with COLA," *JCAP* **1306** (2013) 036, arXiv:1301.0322 [astro-ph.CO].

[172] P. Monaco, T. Theuns, and G. Taffoni, "Pinocchio: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field," *Mon. Not. Roy. Astron. Soc.* **331** (2002) 587, arXiv:astro-ph/0109323 [astro-ph].

[173] C.-H. Chuang, F.-S. Kitaura, F. Prada, C. Zhao, and G. Yepes, "EZmocks: extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics," *Mon. Not. Roy. Astron. Soc.* **446** (2015) 2621–2628, arXiv:1409.1124 [astro-ph.CO].

[174] F.-S. Kitaura, H. Gil-Marín, C. Scoccola, C.-H. Chuang, V. Müller, G. Yepes, and F. Prada, "Constraining the halo bispectrum in real and redshift space from perturbation theory and non-linear stochastic bias," *Mon. Not. Roy. Astron. Soc.* **450** no. 2, (2015) 1836–1845, arXiv:1407.1236 [astro-ph.CO].

[175] A. Taylor, B. Joachimi, and T. Kitching, "Putting the Precision in Precision Cosmology: How accurate should your data covariance matrix be?," *Mon. Not. Roy. Astron. Soc.* **432** (2013) 1928, arXiv:1212.4359 [astro-ph.CO].

[176] W. J. Percival and M. L. Brown, "Likelihood methods for the combined analysis of CMB temperature and polarisation power spectra," *Mon. Not. Roy. Astron. Soc.* **372** (2006) 1104–1116, arXiv:astro-ph/0604547 [astro-ph].

[177] **WMAP** Collaboration, L. Verde *et al.*, "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Parameter estimation methodology," *Astrophys. J. Suppl.* **148** (2003) 195, arXiv:astro-ph/0302218 [astro-ph].

[178] A. F. Heavens and A. N. Taylor, "A Spherical Harmonic Analysis of Redshift Space," *Mon. Not. Roy. Astron. Soc.* **275** (1995) 483–497, arXiv:astro-ph/9409027 [astro-ph].

[179] **2dFGRS** Collaboration, W. J. Percival *et al.*, "The 2dF Galaxy Redshift Survey: Spherical harmonics analysis of fluctuations in the final catalogue," *Mon. Not. Roy. Astron. Soc.* **353** (2004) 1201, arXiv:astro-ph/0406513 [astro-ph].

[180] S. Hamimeche and A. Lewis, "Likelihood Analysis of CMB Temperature and Polarization Power Spectra," *Phys. Rev.* **D77** (2008) 103013, arXiv:0801.0554 [astro-ph].

[181] L. Blot, P. S. Corasaniti, J.-M. Alimi, V. Reverdy, and Y. Rasera, "Matter power spectrum covariance matrix from the DEUS-PUR ΛCDM simulations: mass resolution and non-Gaussian errors," *Mon. Not. Roy. Astron. Soc.* **446** (2015) 1756–1764, arXiv:1406.2713 [astro-ph.CO].

[182] M. Tegmark, "Measuring cosmological parameters with galaxy surveys," *Phys. Rev. Lett.* **79** (1997) 3806–3809, arXiv:astro-ph/9706198 [astro-ph].

[183] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.* **22** no. 1, (03, 1951) 79–86. http://dx.doi.org/10.1214/aoms/1177729694.

[184] J. R. Bond, A. H. Jaffe, and L. E. Knox, "Radical compression of cosmic microwave background data," *Astrophys. J.* **533** (2000) 19, arXiv:astro-ph/9808264 [astro-ph].

[185] S. Smith, A. Challinor, and G. Rocha, "What can be learned from the lensed cosmic microwave background b-mode polarization power spectrum?," *Phys. Rev.* **D73** (2006) 023517, arXiv:astro-ph/0511703 [astro-ph].

[186] S. Matarrese and L. Verde, "The effect of primordial non-Gaussianity on halo bias," *Astrophys. J.* **677** (2008) L77–L80, arXiv:0801.4826 [astro-ph].

[187] N. Afshordi and A. J. Tolley, "Primordial non-gaussianity, statistics of collapsed objects, and the Integrated Sachs-Wolfe effect," *Phys. Rev.* **D78** (2008) 123507, arXiv:0806.1046 [astro-ph].

[188] P. Valageas, "Mass function and bias of dark matter halos for non-Gaussian initial conditions," *Astron. Astrophys.* **514** (2010) A46, arXiv:0906.1042 [astro-ph.CO].

[189] M. Tellarini, A. J. Ross, G. Tasinato, and D. Wands, "Non-local bias in the halo bispectrum with primordial non-Gaussianity," *JCAP* **1507** no. 07, (2015) 004, arXiv:1504.00324 [astro-ph.CO].

[190] A. Orsi, C. M. Baugh, C. G. Lacey, A. Cimatti, Y. Wang, and G. Zamorani, "Probing dark energy with future redshift surveys: A comparison of emission line and broad band selection in the near infrared," *Mon. Not. Roy. Astron. Soc.* **405** (2010) 1006, arXiv:0911.0669 [astro-ph.CO].

[191] **Euclid Theory Working Group** Collaboration, L. Amendola *et al.*, "Cosmology and fundamental physics with the Euclid satellite," *Living Rev. Rel.* **16** (2013) 6, arXiv:1206.1225 [astro-ph.CO].

[192] L. Pozzetti, C. M. Hirata, J. E. Geach, A. Cimatti, C. Baugh, O. Cucciati, A. Merson, P. Norberg, and D. Shi, "Modelling the number density of Hα emitters for future spectroscopic near-IR space missions," *Astron. Astrophys.* **590** (2016) A3, arXiv:1603.01453 [astro-ph.GA].

[193] F. Elsner, B. Leistedt, and H. V. Peiris, "Unbiased methods for removing systematics from galaxy clustering measurements," *Mon. Not. Roy. Astron. Soc.* **456** no. 2, (2016) 2095–2104, arXiv:1509.08933 [astro-ph.CO].

[194] M. Tegmark, "How to measure CMB power spectra without losing information," *Phys. Rev.* **D55** (1997) 5895–5907, arXiv:astro-ph/9611174 [astro-ph].

[195] G. B. Rybicki and W. H. Press, "Interpolation, realization, and reconstruction of noisy, irregularly sampled data," *Astrophys. J.* **398** (1992) 169–176.

[196] B. Leistedt and H. V. Peiris, "Exploiting the full potential of photometric quasar surveys: Optimal power spectra through blind mitigation of systematics," *Mon. Not. Roy. Astron. Soc.* **444** no. 1, (2014) 2–14, arXiv:1404.6530 [astro-ph.CO].

[197] R. Raz, "On the complexity of matrix product," in *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing*, STOC '02, pp. 144–151. ACM, New York, NY, USA, 2002. http://doi.acm.org/10.1145/509907.509932.

[198] A. Slosar, U. Seljak, and A. Makarov, "Exact likelihood evaluations and foreground marginalization in low resolution WMAP data," *Phys. Rev.* **D69** (2004) 123003, arXiv:astro-ph/0403073 [astro-ph].

[199] S. Ho *et al.*, "Clustering of Sloan Digital Sky Survey III Photometric Luminous Galaxies: The Measurement, Systematics and Cosmological Implications," *Astrophys. J.* **761** (2012) 14, arXiv:1201.2137 [astro-ph.CO].

[200] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.* **21** no. 1, (03, 1950) 124–127. http://dx.doi.org/10.1214/aoms/1177729893.

[201] **WMAP** Collaboration, C. Bennett *et al.*, "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Foreground emission," *Astrophys. J. Suppl.* **148** (2003) 97, arXiv:astro-ph/0302208 [astro-ph].

[202] R. Scoccimarro, "Fast Estimators for Redshift-Space Clustering," *Phys. Rev.* **D92** no. 8, (2015) 083532, arXiv:1506.02729 [astro-ph.CO].

[203] A. R. Pullen and C. M. Hirata, "Systematic effects in large-scale angular power spectra of photometric quasars and implications for constraining primordial nongaussianity," *Publ. Astron. Soc. Pac.* **125** (2013) 705–718, arXiv:1212.4500 [astro-ph.CO].

[204] E. F. Schlafly and D. P. Finkbeiner, "Measuring Reddening with SDSS Stellar Spectra and Recalibrating SFD," *Astrophys. J.* **737** (2011) 103, arXiv:1012.4804 [astro-ph.GA].

[205] J. Yoo, "General Relativistic Description of the Observed Galaxy Power Spectrum: Do We Understand What We Measure?," *Phys. Rev.* **D82** (2010) 083508, arXiv:1009.3021 [astro-ph.CO].

[206] E.-M. Mueller, W. J. Percival, and R. Ruggeri, "Optimising primordial non-Gaussianity measurements from galaxy surveys," arXiv:1702.05088 [astro-ph.CO].

[207] F. Zhu, N. Padmanabhan, and M. White, "Optimal Redshift Weighting For Baryon Acoustic Oscillations," *Mon. Not. Roy. Astron. Soc.* **451** no. 1, (2015) 236–243, arXiv:1411.1424 [astro-ph.CO].

[208] R. Ruggeri, W. Percival, H. Gil-Marín, F. Zhu, G.-b. Zhao, and Y. Wang, "Optimal redshift weighting for redshift-space distortions," *Mon. Not. Roy. Astron. Soc.* **464** no. 3, (2017) 2698–2707, arXiv:1602.05195 [astro-ph.CO].

FORM UPR16

Research Ethics Review Checklist

University of Portsmouth

Please include this completed form as an appendix to your thesis (see the Postgraduate Research Student Handbook for more information

Postgraduate Research Student (PGRS) Information				**Student ID:**	712556
PGRS Name:	Benedict Kalus				
Department:	ICG/Tech	**First Supervisor:**	Will Percival		
Start Date: (or progression date for Prof Doc students)		**01/10/2013**			
Study Mode and Route:	Part-time ☐ Full-time ☒	MPhil ☐ PhD ☒		MD ☐ Professional Doctorate ☐	

Title of Thesis:	The Distribution of Galaxies as a Test of Primordial Non-Gaussianity
Thesis Word Count: (excluding ancillary data)	53,165

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES ☒ NO ☐
b) Have all contributions to knowledge been acknowledged?	YES ☒ NO ☐
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES ☒ NO ☐
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES ☒ NO ☐
e) Does your research comply with all legal, ethical, and contractual requirements?	YES ☒ NO ☐

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	B087-24F2-830A-1402-A6F1-DB5E-8F4D-E909

If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of

questions a) to e), please explain below why this is so:		
[redacted]		
Signed *(PGRS)*:	[redacted]	**Date:** 13/09/2017