

## Durham E-Theses

---

### *Search for non-Gaussianity in Large Scale Structure surveys*

KARAGIANNIS, DIONYSIOS

#### How to cite:

---

KARAGIANNIS, DIONYSIOS (2013) *Search for non-Gaussianity in Large Scale Structure surveys* , Durham theses, Durham University. Available at Durham E-Theses Online:  
<http://etheses.dur.ac.uk/6994/>

#### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

# Search for non-Gaussianity in Large Scale Structure surveys

Dionysios Karagiannis

A thesis submitted to the University of Durham  
in accordance with the regulations for  
admittance to the Master by Research in Astrophysics.

January 2013

Extragalactic & Cosmology Group  
Department of Physics  
University of Durham

# Abstract

In this work we put constraints on the primordial non-Gaussianities by using recent Large Scale Structure (LSS) surveys. The importance of measuring the amplitude of the primordial non-Gaussianity lies in the fact that it is the most prominent observational probe of the very early Universe. The plethora of the inflationary scenarios describing the early Universe makes it urgent to decide between them and create a solid physical theory for this era. The different inflation models predict different amount of non-Gaussianity in the primordial density perturbations, which will seed the LSS we observe. Therefore here we use the clustering results of prominent LSS surveys in order to test if they have the statistical power to constrain the primordial non-Gaussianity.

We review the clustering of the radio sources from the NRAO VLA Sky Survey at  $z \sim 1$ . The non-Gaussianity measured is one of the best determinations coming from LSS in the literature,  $f_{NL} = 62 \pm 27$  (68% CL). We also use the full scale range clustering of the LRG's from the CMASS of SDSS BOSS DR8 at  $z = 0.55$ . By using the scale dependence of the bias, originating from the existence of primordial non-Gaussianity, we fit non-Gaussian models to the large scales of our sample in order to measure the  $f_{NL}^{loc}$ . The resulting fits show that there is room in this sample for non-Gaussianity. Although due to the large scale uncertainty errors the standard  $\Lambda$ CDM model cannot be excluded. Recently the measured non-Gaussianity from the SDSS BOSS CMASS sample,  $-92 < f_{NL} < 398$  at 95% CL, shows that the constraints are not tight. This was expected because of the large scale statistical uncertainties in the clustering of this sample. Our best-fit measured  $f_{NL}^{loc} = 71 \pm 11$  ( $1\sigma$ ) is consistent with their measurements. The H-a emitters from HiZELS at a narrow redshift selection  $z = 2.23$  are a promising survey for non-Gaussianity, but in order to gain any interesting constraints we have to wait for a larger sample.

Finally we analyze the clustering of the  $\sim 30,000$  quasar sample of SDSS BOSS DR9 at an effective redshift of  $z_{eff} = 2.4$ . The results show an amplitude excess in the clustering of the sample at the large scales. By fitting non-Gaussian models to the correlation function we measure,  $f_{NL}^{loc} = 135 \pm 9$  at  $1\sigma$  *CL*.  $\Lambda$ CDM fits the clustering results until  $40 h^{-1}\text{Mpc}$ . However we cannot exclude the standard cosmological model since at the large scales that constrain  $f_{NL}$ , our results remain sensitive to the effects of systematic errors. We check the quasar sample for any potential systematics and particularly for the systematic effects of galactic extinction, seeing, sky brightness and foreground stars. Similar to previous studies the largest systematic comes from the presence of foreground stars. When we correct for such systematics we find,  $f_{NL}^{loc} = 63 \pm 16$  ( $1\sigma$ ). The measured amount of non-Gaussianity after correcting for the systematic effects is consistent with the results coming from the NVSS radio sources sample. Since the large scale amplitude of the clustering results is directly affected by systematics, we need to apply a more sophisticated method for correcting such effects. In any case, our original results show that the quasar sample shows excellent potential for determining the amplitude of primordial non-Gaussianity.

## **Statement of Copyright**

“The copyright of this thesis rests with the author. No quotation from it should be published without the author’s prior written consent and information derived from it should be acknowledged.”

## Acknowledgements

First of all I would like to thank my supervisor Prof. Tom Shanks for his tremendous help, guidance and support. I would also like to thank the postgrads and colleges with whom I shared the same office, Nikolaos Nikoloudakis, Dr. Utane Sawangwit and Joe Whitbourn for their help, guidance and time.

I would like to thank all the other postgrads and people of the Physics Department for the interesting conversations and useful information. I also thank Dr. Nicolas Ross for trusting me with his codes for the quasar selection of the SDSS BOSS DR9. Finally I would like to thank my family for their patience, financial support and help. My special thanks for her proofreading, support and love to Ms. Pallavi Vengsarkar.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theoretical Review</b>	<b>5</b>
2.1	The Spectrum of Cosmological Perturbations . . . . .	5
2.2	Primordial fluctuations from inflation . . . . .	11
2.3	Primordial non-Gaussianities . . . . .	19
<b>3</b>	<b>Primordial non-Gaussianity in Large Scale Structure</b>	<b>29</b>
3.1	Structure formation and bias . . . . .	30
3.2	Press-Schechter Theory . . . . .	31
3.3	Non-Gaussianities in the mass function . . . . .	34
3.4	Non-Gaussianities in the galaxy bias . . . . .	36
<b>4</b>	<b>Search for non-Gaussianity</b>	<b>40</b>
4.1	NRAO VLA Sky Survey from Xia et. al. 2010 . . . . .	41
4.2	The SDSS DR9 BOSS-CMASS sample from Sanchez et al. 2012 . . . . .	43
4.3	The HiZELS sample from Geach et al. 2012 . . . . .	51
<b>5</b>	<b>Search for non-Gaussianity in the quasars of SDSS BOSS DR9</b>	<b>55</b>
5.1	Data . . . . .	56
5.1.1	Quasar selection in BOSS . . . . .	57
5.1.2	Subsample and angular completeness . . . . .	58

5.2	Clustering analysis . . . . .	61
5.2.1	Random catalogue . . . . .	62
5.2.2	Error estimators . . . . .	63
5.2.3	Clustering results . . . . .	65
5.3	Test for non-Gaussianity . . . . .	69
5.4	Check for systematic errors . . . . .	73
5.5	Conclusions and Summary . . . . .	81
<b>6</b>	<b>Looking into the future: The 2QDES survey</b>	<b>83</b>
<b>7</b>	<b>Conclusions</b>	<b>85</b>
	<b>Bibliography</b>	<b>89</b>

# Chapter 1

## *Introduction*

One of the biggest challenges of cosmology is to explain the origin of the observable *Large Scale Structures* (LSS), namely galaxies, galaxy clusters and superclusters. Modern theories are trying to give an adequate solution to this fundamental physical problem via the evolution of perturbations of the primordial density field. These density fluctuations, due to their self-gravity, will not only grow over time, but the over-dense regions will attract more matter increasing their density and gravitational potential as well. The effect of gravitational instability will evolve these primordial fluctuations to the structures we observe today in the Universe. Primordial inhomogeneities are not just a theoretical idea physicists used to provide a logical explanation for the structures in the universe. Solid observational evidence can be derived from the surface of the last scattering. At redshift  $z \sim 1000$ , electrons recombine into atoms lowering the rate of photon scattering. At some point scattering rate is so low that photons can propagate freely through the universe. The spectrum of this scattered light, known as the *Cosmic Microwave Background* (CMB), carries information about the initial conditions for the formation of cosmic structures. These primordial density perturbations are related to the temperature anisotropies of the CMB, which are of order  $\Delta T/T \sim 10^{-5}$ .

The theoretical framework of standard cosmology is the Hot Big Bang model (HBB). It uses the *Friedmann–Lemaître–Robertson–Walker* (FLRW) metric, which is an exact solution of Einstein’s field equations of general relativity, to describe the accelerating expansion of a homogeneous and isotropic Universe. This model successfully explains the evolution of the Universe from a hot, dense, radiation dominated initial state, to a cool, low-density, non-relativistic dominated present state. The two more distinct features of our universe, its homogeneity and isotropy, are both just intelligent assumptions known as

the '*Cosmological Principle*'. But how can such an assumption be right when astronomers observe inhomogeneities (galaxy clusters, superclusters, voids, etc.) on small scales? Is there any particular scale beyond which universe appears homogeneous and isotropic? Current redshift surveys of galaxies have shown that the universe has these features for scales larger than 100 Mpc. Moreover the size of the observable patch of the universe is equal to the Hubble radius,  $R_H = c/H_0 \approx 3000\text{Mpc}$ , where  $H_0$  is the present Hubble constant. Thus it is obvious that the Cosmological Principle is valid only for a short range of scales. Besides its success, the standard Hot Big Bang model is inadequate to describe the very early universe, or even to explain the origins of the primordial inhomogeneities.

The inflation paradigm came as a supplement to the HBB model and elegantly solved its major problems. The first model of cosmological inflation was introduced by Guth in 1981 [1] as a solution to the horizon and flatness problems, which arose from the standard cosmological model. Moreover an improved inflationary model, called '*new*' *inflation*, was introduced soon after [2, 3]. Inflation is an era in the early history of the universe that provides a mechanism for driving an exponentially accelerated expansion. During the inflation epoch, the universe is dominated by a scalar field  $\phi$ , called the inflaton, and its self-interaction potential  $V(\phi)$  which is related to the vacuum energy density. For an explicit review on this subject, the reader should address [4]. The universe undergoes a phase transition while  $\phi$  slowly rolls down  $V(\phi)$  from an unstable pseudo-vacuum state, with high energy density, towards a stable vacuum state, represented by a local minimum at  $V(\phi_0)$ . The slope of the potential must be quite flat so inflation can last enough e-folds to solve the HBB model problems without spoiling the successful predictions of the Friedmann model. So the inflation era must smoothly reach to an end (graceful exit) and into a Friedmann expansion stage, otherwise the homogeneity of our observable patch of the universe will be destroyed.

Inflation after 30 years is still the most popular paradigm that describes the early universe; not only does it solve major cosmological problems, but it also explains the

---

production of the primordial density fluctuations that seed the LSS and the temperature anisotropies we observe in the CMB spectrum. One of the reasons that inflation was introduced in the first place, was to delude the topological defects (monopoles, cosmic strings, domain walls, etc.) that are created in the early universe after a phase transition, due to spontaneous symmetry breaking, in *Grand Unification Theories* (GUT). During the exponential expansion, inhomogeneities, including the topological defects, are stretched away and the universe ends up having the desired feature of homogeneity. Thus inflation can produce an observable homogeneous universe from a small domain, regardless of the fact that the universe may be quite inhomogeneous outside that domain. The existence of an event horizon prevents us from knowing the characteristics of the whole universe. Inflation predicts that, despite the homogeneity of our observable patch, universe is completely inhomogeneous on scales larger than the horizon. Therefore these inhomogeneous regions are not causally connected to our observable universe, and in turn cannot affect the homogeneity of the observable patch. Even though inflation is capable of creating such homogeneous patches, it can also give birth to the primordial density inhomogeneities. In the inflationary scenario, primordial perturbations can be created by quantum fluctuations of the scalar fields that drive the expansion. These fluctuations are stretched during inflation from Planckian size to galactic scales, maintaining their initial amplitudes nearly unchanged. They are created out of the horizon during the early stages, and they re-enter at a later time seeding the LSS and shaping the CMB spectrum anisotropies we observe now [5, 6, 7, 8, 9]. The tiny primordial fluctuations, which can be studied within linear perturbation theory, can be approximately treated as Gaussian. The density fluctuations, which characterize the primordial inhomogeneities, can be described as *random fields*, with the common notation  $\delta(x)$ . In other words, the phase of each of the Fourier modes of  $\delta(x)$  are random and uncorrelated; thus we can use Gaussian statistics for the density field. In most simple slow-roll inflation models quantum fluctuations of the inflaton have random phases and hence they can produce a Gaussian random density field.

---

More information considering the primordial inhomogeneities and the mechanism behind their generation can come from the deviation of the density field from Gaussian statistics. Non-Gaussianities reveal interactions beyond the linear theory. Even though current experiments, like WMAP, verify that the primordial density field follows a Gaussian distribution to a good approximation, there are inflationary models that predict primordial non-Gaussianities. Each model leaves a precise imprint of non-Gaussianities in the primordial curvature perturbations; some of them are constrained to produce non-Gaussianities much smaller than the predominant Gaussian component. Hence in order to gain information for the physics of the very early universe and the field interactions during inflation, we have to determine the statistical properties of the primordial inhomogeneities created from inflation. To do that we have to go beyond the power spectrum and measure the amount of deviation from the Gaussian initial conditions by calculating higher order correlation functions from the different observational probes, e.g. CMB, LSS clustering, Integrated Sachs-Wolfe effect [10, 11, 12] and gravitational lensing.

# Chapter 2

## *Theoretical Review*

In the following chapter, we will present a quick review of the theory describing the cosmological perturbations and their spectrum, and how we can generate such primordial perturbations in the density field from the simplest inflationary models. An extended description of the linear perturbation theory and a complete introduction to inflation can be found in many textbooks [12, 13, 14, 15]. A comprehensive review will follow on non-Gaussianity and on the different inflationary models that can produce deviation from the Gaussian conditions in the primordial perturbations.

### 2.1 The Spectrum of Cosmological Perturbations

To characterize the statistical properties of the cosmic density field we divide the Universe into large volumes considering a configuration  $\delta(x) = [\rho(\mathbf{x}) - \langle \rho \rangle] / \langle \rho \rangle$  within each of these regions as a realization of a random process. Density at each point  $\mathbf{x}$  is  $\rho(\mathbf{x})$ , and the mean density in each volume  $V$  is  $\langle \rho \rangle$ . Averaging over all these volumes is equivalent to the average over the whole volume of the universe. We describe the perturbation field by using Fourier methods, where the Fourier expansion of the function  $\delta(\mathbf{x})$  in a given volume  $V$  is

$$\delta(\mathbf{x}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \quad (1)$$

The Fourier coefficients  $\delta_{\mathbf{k}}$  are complex, but due to the reality of  $\delta(\mathbf{x})$ , we have  $\delta_{\mathbf{k}}^* = \delta_{-\mathbf{k}}$ . One can observe that the phases of the Fourier modes vary from one volume to another, as well as between each other in the same cell. In the case of random phases the different Fourier modes evolve independently and the consequent superposition ( $\delta(\mathbf{x})$ ) can approach

a Gaussian distribution as it is dictated by the central limit theorem. Therefore the primordial density perturbation field can be well appropriated by a Gaussian random field. In the limit  $V \rightarrow \infty$  the sum in Eq. 1 is replaced by the integral

$$\delta(\mathbf{x}) = \int_V \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \frac{d^3k}{(2\pi)^3} \quad (2)$$

By applying the Gaussian statistics to the density field, where the average of  $\delta(\mathbf{x})$  over volumes is zero, we can define the standard deviation  $\sigma^2$  as

$$\sigma^2 = \langle \delta^2 \rangle = \frac{1}{V} \sum_{\mathbf{k}} \langle |\delta_{\mathbf{k}}|^2 \rangle \quad (3)$$

For a homogeneous and isotropic density field we have for the large number of volumes

$$\sigma^2 = \frac{1}{2\pi^2} \int_0^\infty P(k) k^2 dk \quad (4)$$

where  $P(k) = \langle |\delta_k^2| \rangle$  for the limit  $V \rightarrow \infty$ . The quantity  $P(k)$  is called *power spectrum*. Roughly speaking,  $P(k)$  describes the amplitude of the primordial perturbations and thus the level of structures. Power spectrum is one of the most important quantities in determining the primordial fluctuation field. Its importance lies in the fact that structure formation depends on it to a large extent. Furthermore, through it we can determine whether or not non-Gaussianities exist in the primordial perturbations through the calculation of higher order correlation functions, which are directly connected to the power spectrum. The most common form to assume for the primordial power spectrum is the power-law form  $P(k) = Ak^n$ , where  $n$  is the spectral index. For  $n = 1$  we have the so called *Harrison-Zeldovich spectrum*.

Another way to describe the primordial density field is through the spacial *two-point correlation function*. The prerequisite is that the phases of the Fourier modes of  $\delta(x)$  are random and therefore we can apply Gaussian statistics if the volume number is large. We can define the two-point correlation function as

$$\langle \delta(\mathbf{x})\delta(\mathbf{y}) \rangle = \xi(|\mathbf{x} - \mathbf{y}|) = \xi(r) \quad (5)$$

where  $\mathbf{y} = \mathbf{x} + \mathbf{r}$ . Here we have used the simplifying assumptions of a homogeneous and isotropic overdensity random field. Correlation function provides a complete statistical characterization of the density field, as far as Gaussian statistics can be applied. It gives us information about the size of the fluctuations on different scales, and clearly depends only on the spacial difference between two points; this is valid if we accept the assumption that the Universe is statistically homogeneous and isotropic. Two-point correlation function can be used to relate primordial fluctuations with galaxy-clustering data. From substituting equation (2) into (5) we can define, for the limit  $V \rightarrow \infty$ , a relation for  $\xi(\mathbf{r})$

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int P(k) e^{i\mathbf{k}\cdot\mathbf{r}} dk \quad (6)$$

where  $P(k)$  is the power spectrum of the fluctuation and is defined as

$$\langle \delta_{\mathbf{k}}\delta_{\mathbf{k}'} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{k}') P(\mathbf{k}) \quad (7)$$

where  $\delta^{(3)}(\mathbf{k} + \mathbf{k}')$  is the Dirac function. Power spectrum is  $P(k) = |\delta_k|^2 = \delta_k \delta_k^*$ , therefore we can write

$$P(k) = \frac{1}{V} \int \delta(\mathbf{x})\delta(\mathbf{y}) e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} dV_x dV_y = \int \xi(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r} \quad (8)$$

To derive this result we used,  $\xi(\mathbf{r}) = \langle \delta(\mathbf{x} + \mathbf{r})\delta(\mathbf{x}) \rangle = \frac{1}{V} \int \delta(\mathbf{x} + \mathbf{r})\delta(\mathbf{x}) dV_x$ .

So the power spectrum is the Fourier transformation of the correlation function. If we integrate over all angles, by taking into consideration the isotropy of the universe, the above equation becomes

$$P(k) = 4\pi \int_0^\infty \xi(r) \frac{\sin kr}{kr} r^2 dk \quad (9)$$

and its inverse is going to be

$$\xi(r) = \frac{1}{2\pi^2} \int_0^\infty P(k) \frac{\sin kr}{kr} k^2 dk \quad (10)$$

As a rule of thumb  $\xi(r)$  can be described by a power law of the form  $\xi(r) = (r/r_0)^{-\gamma}$ , where  $\gamma = 1.8$  and  $r_0$  is approximately  $5 h^{-1}\text{Mpc}$ .

The two point correlation function can be also defined through the discrete distribution of masses. Galaxies are not randomly distributed all over universe, but rather form clusters and superclusters through the attractive force of gravity. Therefore it is probable to find a galaxy near another one. The probability of finding one galaxy in a volume  $dV$  is  $P_1 = n_V dV$ , where  $n_V = m^{-1} \rho(\mathbf{x}) dV$  is the mean number of galaxies in that volume. The probability of finding one galaxy at a point inside the volume  $dV_1$  and an other one in volume  $dV_2$ , with  $r$  being their separation distance, is

$$\begin{aligned} \delta P &= \frac{\langle \rho(\mathbf{x}) \rho(\mathbf{y}) \rangle}{m^2} dV_1 dV_2 \\ &= n_V^2 \frac{\langle \rho(\mathbf{x}) \rho(\mathbf{y}) \rangle}{\langle \rho \rangle^2} dV_1 dV_2 \\ &= n_V^2 (1 + \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle) dV_1 dV_2 \\ &= n_V^2 (1 + \xi(r)) dV_1 dV_2 \end{aligned} \quad (11)$$

So two-point correlation function is connected with the probability of finding two galaxies separated by a distance  $r$  in a three dimensional space.

The most popular way of measuring  $\xi(r)$  in a galaxy redshift survey is by counting pairs of galaxies with separation  $r$  and divide them by the number of pairs counted in a randomly distributed unclustered sample. The random catalogue of galaxies must have the same sky coverage as the data, including smooth redshift distribution, and also be large enough to reduce the Poisson errors. A very famous formula calculating  $\xi(r)$  is [16]

$$\xi = \frac{DD \cdot RR}{(DR)^2} - 1 \quad (12)$$

where  $DD$  is the number of galaxy pairs, in a particular bin of separation  $r$ , in the data catalogue.  $DR$  is the count of galaxy pairs between the random catalogue and the real data, and finally  $RR$  is the number of pairs in the random catalogue. The most common expression calculating the two-point correlation function is the Landy & Szalay estimator [17]

$$\xi = \frac{1}{RR} \left( DD \left( \frac{n_R}{n_D} \right)^2 - 2DR \left( \frac{n_R}{n_D} \right) + RR \right) \quad (13)$$

where  $n_R$  and  $n_D$  are the mean number density of galaxies in the random and data catalogues respectively.

In order to calculate the two-point correlation function we need to know the position of galaxies in the 3-D space, i.e. we need to know the right ascension, declination and redshift of each galaxy in the sample. It's not always easy to derive information about the redshift of galaxies. The difficulty lies on the fact that to calculate the redshift of galaxies with minimum uncertainty we need their spectra, which is not possible some times, therefore a spectroscopic survey must provide the data. However such surveys are time consuming and they need a high amount of resources. Another way to derive redshift information for galaxies is through photometric surveys using the photometry of galaxies in various bands. Despite the large sky area covering and depth, they lose in accuracy, providing redshift data with significant uncertainty. The angular two-point correlation function  $w(\theta)$  is the 2-D projection of  $\xi(r)$  on the sky plane, therefore we don't need redshift data to derive results for the galaxy clustering. However measuring  $w(\theta)$  we lose information existing in the 3-D clustering, that's obvious as  $w(\theta)$  contains information for the 2-D clustering. Generalizing Eq. 11 we gain the probability of two galaxies to have angular separation  $\theta$

$$\delta P = N^2(1 + w(\theta))d\Omega_1 d\Omega_2 \quad (14)$$

where  $N$  is the mean number density of galaxies per steradian  $d\Omega$ . All the above formulas

calculating  $\xi(r)$  can be generalized to the angular correlation function. Observationally it can also be described by a power law similar to  $\xi(r)$  but with different slope  $w(\theta) = (\theta/\theta_0)^{1-\gamma}$  [18], where  $\theta_0$  is the angular correlation scale. The angular correlation function can be related to the two-point correlation function from Limber's equation [19, 20]

$$w(\theta) = \frac{\int_0^\infty dz_1 r_1^2 \left(\frac{dr_1}{dz_1}\right) \phi(z_1) n(z_1) \int_0^\infty dz_2 r_2^2 \left(\frac{dr_2}{dz_2}\right) \phi(z_2) n(z_2) \xi(r_{12})}{\left[\int_0^\infty r^2 \left(\frac{dr}{dz}\right) \phi(z) n(z) dz\right]^2} \quad (15)$$

where  $\phi(z)$  is the redshift selection function of the sample,  $n(z)$  is the comoving number density and  $r_{12}$  is the comoving separation between the objects.

The above ideas for the two point correlation function can be generalized for more than two points. For  $N=3$  we have the three-point correlation function defined as

$$\zeta(r, s, t) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \delta(\mathbf{x} + \mathbf{s}) \rangle \quad (16)$$

The mean is taken over all points that define a triangle with sides  $r, s, t = |\mathbf{r} - \mathbf{s}|$ . The analogous of equation (7) will be now for the Fourier modes

$$\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_\delta(k_1, k_2, k_3) \quad (17)$$

The function  $B_\delta(k_1, k_2, k_3)$  is the bispectrum of the density field. As we will see further on, this plays a crucial role in constraining non-Gaussianities, which might be present in the primordial fluctuations. In fact, bispectrum represents the lower order statistics that will allow us to parametrize the level of non-Gaussianities. In a similar way we can generalize the probability, given by (11), of finding three galaxies in three different volumes  $dV$  separated by distances  $r, s, t$  respectively.

We can generalize the whole process to define a correlation function for  $N$  points through the mean  $\langle \delta_1 \delta_2 \cdots \delta_N \rangle$  (where  $\delta_1 = \delta(\mathbf{x}_1)$ , etc.). The  $N$ -point correlation function will now have contributions from all lower than  $N$  order correlation functions.

## 2.2 Primordial fluctuations from inflation

Inflation was introduced by Guth to solve the problems of the standard Hot Big Bang model. It is a period of exponential acceleration of the universe before the radiation dominated era. The dominant perception for inflation indicates that it is driven by the energy of a scalar field  $\phi$  with negative pressure, called inflaton, whose potential energy represents the energy density of the vacuum. During inflation,  $\phi$  undergoes a phase transition from an unstable false-vacuum state to a stable state of true vacuum. The inflaton simply slow-rolls down its self-potential towards the steady state. The potential is sufficiently flat so that inflation will last for an adequate amount of time, solving the problems of the standard model. When it reaches the minimum of the potential, it oscillates around it; this releases the energy difference, heating the universe (*reheating*) and creating particles. The number of inflationary models is large enough to give inflation the character of a paradigm. A proven modern theory, that can physically derive an inflaton scalar field with a self-potential of the desired form, does not exist until now.

Every perturbation in a system can be written as a combination of two types of perturbations, of the *adiabatic* and the *entropy* types. The first ones, also known as *curvature* perturbations, are inhomogeneities of the actual spacial curvature. The later ones are also called *isocurvature* perturbations, induce fluctuations in the number density of the different components of the system, i.e. different particle types, while the total energy density remains homogeneous.

The most important aspect of inflation, that made it the dominant way to describe the very early universe, is that it can explain the production of the adiabatic type perturbations that seed the observable structures. Initially one of the problems that inflation tried to solve is that of the unwanted relics of GUT models. According to the modern theories of Grand Unification, the Universe underwent many phase transitions, during which spontaneous symmetry breaking occurred. As a result, unwanted relics can be created. Nevertheless the existence of an accelerated expansion period can dilute every topological

inhomogeneity. The only way that inflation can produce perturbations in the primordial density field, is through the quantum fluctuations of the scalar field itself. Quantum fluctuations cannot be wiped away because new ones will be generated via Heisenberg's uncertainty relation. The cosmological horizon at the time of inflation was very small, so the wavelength of the quantum fluctuations of the scalar field will exceed soon, due to the inflationary expansion, the Hubble radius. The horizon is the scale beyond which causal connected processes cannot operate. Thus the events inside it will not affect the evolution of the fluctuations with wavelength greater than that scale. On super-horizon scales the amplitude of the fluctuations remains unchanged, "*frozen*", while its wavelength grows in an exponential rate. Quantum fluctuations will grow due to gravitational effects outside the horizon. After the end of inflation, the scale-factor will not grow faster than the Hubble radius. Thus, as time passes, perturbations on even larger scales will reenter the horizon. Due to gravitational instabilities, these fluctuations will give rise to galaxies and finally to the large structures we observe today in the Universe. It is easy to understand that the formation of structures depends on the form of the primordial fluctuation spectrum.

The action for a scalar field coupled to gravity through the metric  $g_{\mu\nu}$  is

$$S = \int d^4x \sqrt{-g} \left( \frac{M_P^2}{2} R - \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - V(\phi) \right). \quad (18)$$

The background metric of an expanding Universe is the FLRW metric  $ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) d\mathbf{x}^2$ , where  $\mathbf{x}$  is the spacial comoving components of the metric,  $d\mathbf{x}^2 = dr^2 / (1 - kr^2) + r^2 d\Omega^2$ , with  $k$  being the curvature of space. The scalar field in homogeneous and isotropic universe is,  $\phi(t, \mathbf{x}) = \phi(t)$ . Therefore to be able to describe the fluctuations of the field we split it into a homogeneous part  $\phi_0$  and a part that describes the quantum fluctuations of the field around  $\phi_0$

$$\phi(t, \mathbf{x}) = \phi_0(t) + \delta\phi(t, \mathbf{x})$$

where  $\phi_0$  is the homogeneous part of the scalar field and  $\delta\phi$  denotes the quantum fluctuations around  $\phi_0$ . For  $\phi(\mathbf{x}, t)$  we can derive the equations of motion, for a FLRW metric, after varying the action of the scalar field

$$\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2\phi}{a^2} + \frac{dV}{d\phi} = 0. \quad (19)$$

From the zero and *ith* components of the scalar's energy-momentum tensor,  $T^{\mu\nu} = -2(\partial\mathcal{L}/\partial g^{\mu\nu}) - g^{\mu\nu}\mathcal{L}$ , we can derive the equations for the pressure and density of the homogeneous part of the field

$$\rho = \frac{\dot{\phi}_0^2}{2} + V(\phi) \quad (20)$$

$$p = \frac{\dot{\phi}_0^2}{2} - V(\phi) \quad (21)$$

To derive the above equations we used the energy-momentum tensor for a perfect fluid,  $T_{\mu\nu} = (\rho + p)u_\mu u_\nu + pg_{\mu\nu}$ , where  $u_\mu$  is the four-velocity and  $u_\mu u^\mu = 1$  for  $\mu = \nu$ . The standard steps for deriving the above equations can be found in most cosmological textbooks and reviews such as [12, 14, 21]. Substituting the above equations to the fluid equation,  $\dot{\rho} = -3H(\rho + p)$ , and Friedmann equations we get the equations of motion (EOM) for the homogeneous part of the field  $\phi_0$

$$\ddot{\phi}_0 + 3H\dot{\phi}_0 = -\frac{dV}{d\phi} \quad (22)$$

$$H^2 = \frac{1}{3M_P^2} \left( \frac{\dot{\phi}_0^2}{2} + V(\phi) \right) \quad (23)$$

where the Hubble constant is  $H = \dot{a}/a$ . The EOM for the homogeneous part has the expected form as  $\phi_0$  depending only on time  $t$ , hence  $\nabla^2\phi = 0$  in the equation (15). The slow-roll approximation is introduced through 'new' inflation, so that it lasts for a sufficient amount of  $e$ -folds to solve the cosmological problems. This implies two conditions

$$\epsilon = \frac{M_P^2}{2} \left( \frac{V'}{V} \right)^2 \ll 1 \quad (24)$$

$$\eta = M_P^2 \frac{V''}{V} \ll 1 \quad (25)$$

which respectively they mean  $\dot{\phi}^2 \ll V(\phi)$  and  $\ddot{\phi}_0 \ll 3H\dot{\phi}_0$ .

During inflation, the energy density of the inflaton is the dominant one, driving the accelerating expansion of the universe. Therefore, a fluctuation  $\delta\phi$  in the scalar field implies a perturbation in the energy density field. Moreover, through Einstein's field equations  $G_{\mu\nu} = 8\pi GT_{\mu\nu}$ , inflaton fluctuations are coupled to metric perturbations. Thus we can have curvature perturbations  $\mathcal{R}(\mathbf{x}, t)$  and hence a gravitational potential, which imply fluctuations in the density field. During inflation the inflaton and the metric perturbations will be stretched to cosmological scales. Density fluctuations will grow outside the horizon and re-enter at a later time, after the end of inflation to seed the structure formation. To study the primordial perturbations produced during inflation, we have to derive the equations of motion for the scalar field fluctuations  $\delta\phi$ . By analogy with (15) the EOM for the spacial fluctuation modes will be

$$\ddot{\delta\phi}(t, \mathbf{k}) + 3H\dot{\delta\phi}(t, \mathbf{k}) + \frac{k^2}{a^2}\delta\phi(t, \mathbf{k}) = 0 \quad (26)$$

where the Fourier modes of the field fluctuation are

$$\delta\phi_k = \delta\phi(t, \mathbf{k}) = \int \delta\phi(t, \mathbf{x}) e^{i\mathbf{k}\mathbf{x}} d^3x. \quad (27)$$

To derive an explicit form for the power spectrum through inflation, we need to quantize the perturbations  $\delta\phi$ . Here we will refer to some steps of the procedure; for an explicit overview, the reader should refer to [22, 23]. We follow the standard way of quantization of a scalar field, and we promote  $\delta\phi$  to an operator by decomposing it to annihilation operators

$$\delta\phi(t, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \left( u_k(t) a_k e^{i\mathbf{k}\mathbf{x}} + u_k^*(t) a_k^\dagger e^{-i\mathbf{k}\mathbf{x}} \right) \quad (28)$$

where  $u_k(\tau)$  is the mode function of the scalar fluctuations and is defined as  $u_k(\tau) = a\delta\phi_k$ , with  $\tau = \int a^{-1} dt$  being the conformal time. The annihilation and creation operators follow the commutator relations

$$[a_k, a_{k'}] = [a_k^\dagger, a_{k'}^\dagger] = 0, \quad [a_k, a_{k'}^\dagger] = \delta^{(3)}(\mathbf{k} - \mathbf{k}') \quad (29)$$

The equations of motion for the modes  $u_k(\tau)$  can be found by using the equation (15)

$$u_k'' + \left( k^2 - \frac{a''}{a} + m_\phi^2 a \right) u_k = 0, \quad (30)$$

where  $m_\phi^2 = \partial^2 V / \partial \phi^2$  is the effective mass of the field and the prime is the partial derivative in respect to  $\tau$ . The size of the horizon is proportional to the time  $t$ ; thus, for  $t \rightarrow 0$  the size will decrease very fast. Eventually the modes, which depend on the scale factor, will have superhorizon size. Therefore we can divide the solutions of (26) into two cases. The horizon crossing is  $k = \alpha H$ , which is simply the modes that have horizon scales at a particular time. For subhorizon scales  $k^2 \gg a''/a$  the mass of the field are negligible; thus the fluctuations are described by ordinary plane waves. For superhorizon scales  $k^2 \ll a''/a$  we have for the quantum fluctuations of a massless scalar field

$$|\delta\phi_k| = \frac{|u_k|}{a} = \frac{H}{\sqrt{2k^3}} \quad (31)$$

which is a constant. The above result is derived from the requirement that the solutions for the two cases must be equal at the horizon crossing ( $k = aH$ ) for a fluctuation with mode number  $k$ . For subhorizon scales, the amplitude of the fluctuations is oscillatory, while for superhorizon scales, the amplitude is constant ('frozen'). On the other hand, for a non-vanishing mass term in (26), the fluctuations of the scalar will not be constant outside the horizon

$$|\delta\phi_k| = \frac{H}{\sqrt{2k^3}} \left( \frac{k}{aH} \right)^{3/2-\nu}, \quad (32)$$

where we used the approximation for a light scalar field ( $m_\phi^2/3H^2 \ll 1$ ). The index  $\nu$  is defined as

$$\nu^2 = \frac{9}{4} - \frac{m_\phi^2}{H^2}.$$

Knowing the amplitude of the fluctuations of the scalar field on superhorizon scales, we can define a power spectrum for these perturbations

$$\langle \delta\phi_k \delta\phi_{k'}^* \rangle = \frac{|u_k|^2}{a^2} \delta^{(3)}(\mathbf{k} - \mathbf{k}'). \quad (33)$$

The real and the imaginary part of the quantum scalar field follow Gaussian statistics and hence they can be treated as random fields. Hence we can acquire a power spectrum similar to the one in Eq. 4. The general form of the power spectrum for a scalar field is

$$\mathcal{P}_\phi(k) = \frac{k^3}{2\pi^2} |\delta\phi_k|^2 \quad (34)$$

where  $\mathcal{P}_\phi(k) = P_\phi(k)(k^3/2\pi^2)$  and  $P_\phi(k)$  is defined by the corresponding equation (7) for the scalar field  $\phi$ .

For a massless scalar field ( $m_\phi^2 = 0$ ) the power spectrum of the fluctuations outside the horizon are frozen, which means that the growth of fluctuations is scale invariant. By substituting (27) into (30) we get

$$\mathcal{P}_\phi(k) = \left( \frac{H}{2\pi} \right)^2 \quad (35)$$

which is indeed a 'frozen' power spectrum. If the scalar field has a non-vanishing effective mass then we get a power law shape spectrum for the fluctuations on superhorizon scales

$$\mathcal{P}_\phi(k) = \left( \frac{H}{2\pi} \right)^2 \left( \frac{k}{aH} \right)^{3-2\nu} \quad (36)$$

The power spectrum of the scalar field fluctuations has the usual form of a power law on these large scales. Therefore, inflation provides a mechanism through which inhomogeneities re-enter the horizon; having the desired spectrum will seed eventually the large scale structure of the observable Universe.

As we mentioned before the primordial scalar field fluctuations can lead to perturbations in the metric and therefore in the spacial curvature itself. We can also describe the metric fluctuations by using General Relativity and the so called Newtonian gauge. We can write the perturbed metric as a sum of an unperturbed FLRW background part and a perturbed one. By using the Newtonian gauge as infinitesimal transformations we keep the background part as it is while the perturbed part is subject to changes. In the Newtonian gauge the observer will detect a velocity field while particles fall into the gravitational field of matter and will measure a gravitational potential, more details on this can be found in [24].

The curvature perturbation  $\mathcal{R}(\mathbf{x}, t)$  is a more useful quantity than the scalar fluctuations, because it remains well defined even after the scalar field decays at the end of inflation. Also  $\mathcal{R}_k(t)$ , which are the Fourier coefficients of the curvature perturbations, is constant outside the horizon while  $\delta\phi_k$  is not. The relation between the scalar and curvature fluctuation at an initial time  $t_0$ , which is few Hubble times after the horizon crossing, is

$$\mathcal{R}_k = - \left[ \frac{H}{\dot{\phi}} \delta\phi_k \right]_{t=t_0} . \quad (37)$$

From equation (35) and (37) we can derive the power spectrum of the curvature perturbations at an initial time

$$\mathcal{P}_{\mathcal{R}}(k) = \left( \frac{H}{\dot{\phi}} \right)^2 \mathcal{P}_{\phi}(k) \Big|_{t=t_0} . \quad (38)$$

The primordial curvature power spectrum depends on the scalar field power spectrum and therefore we can easily derive the relations for  $\mathcal{P}_{\mathcal{R}}$  in the subhorizon and superhorizon

cases. The curvature perturbations will give birth to a primordial gravitational potential which will create inhomogeneities in the density field and due to gravitational instabilities these fluctuations will seed the first galaxies in the way we explained before. By multiplying  $\mathcal{R}_k$  defined at the initial time with the transfer function we can relate it to the density field power spectrum at later times

$$\mathcal{P}_\delta(t) = T^2(t)\mathcal{P}_\mathcal{R}. \quad (39)$$

The power spectrum is the only statistical tool we need to define the primordial fluctuations and inflation. Depending on which inflationary model we use, it can provide us with a variety of different power spectra forms. In this simple slow-roll inflation model, we obtain a power law spectrum from a light scalar field when the superhorizon stretched fluctuations re-enter the horizon. In more complicated inflationary scenarios, where more than one scalar fields exists during the acceleration period of the universe, fluctuations are produced from other scalar fields instead of the one that drives inflation. A review on the fluctuations and power spectra generated by other inflationary models, as far as an analytical description of the models themselves, can be found in [25, 21, 22]. Despite the difference in the way that primordial inhomogeneities are produced through inflationary scenarios and the form of the fluctuation power spectrum, most of them agree on the Gaussian nature of the perturbations. This is because usually in an inflationary model one assumes that the origin of these inhomogeneities is the quantum fluctuations of a scalar field during the inflationary expansion. Quantum perturbations of a field are vacuum state fluctuations, like the ground state harmonic oscillator in quantum mechanics and is therefore Gaussian. Gaussian distribution characterizes the primordial inhomogeneities in the simplest models, making inflation the dominant candidate for describing the evolution of the early Universe and the production of the primordial perturbations.

### 2.3 Primordial non-Gaussianities

The different inflationary models predict different amount of deviation from the Gaussian conditions in the primordial density field. Non-Gaussianities produced by the most standard single field slow-roll inflation with canonical kinetic terms are not enough to be detectable [26, 27], producing nearly a Gaussian primordial perturbation density field. More complicated models can generate large amplitude of non-Gaussianities, which can be detected by future experiments, if they violate at least one of the conditions of the standard inflationary models [28]. For the *single field* inflation, namely if one scalar field drives inflation and creates the primordial perturbations, the non-Gaussian signal is too weak to be observed. Hence multi-field models can produce large detectable non-Gaussianities. Nevertheless there are single field inflation models that have non-canonical kinetic terms or violate temporarily the slow-roll condition, producing significant non-Gaussianities [29].

The *slow-roll* condition refers to the steepness of the potential on which the inflaton rolls. To have a slow-roll we need the potential to be flat enough to make the scalar field roll slowly down to the minimum, making inflation last for the appropriate amount of *e-folds*. The violation of this condition is not a necessity to generate non-Gaussianity. Slow-roll multi-field inflation models can generate observably large non-Gaussianity if certain conditions are satisfied [30, 31]. On the other hand, certain classes of multi-field models can produce primordial non-Gaussianities without focusing on the slow-roll regime, due to a strong break down in the slow-roll condition shortly before the end of inflation [32, 33].

The simplest inflationary models have canonical *kinetic term* in their action,  $T = -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$ . There is the possibility of having higher derivative kinetic terms dominating the dynamics. Non-canonical kinetic term is used to go beyond the standard inflationary models and generate large non-Gaussianities [34]. In the standard inflation models, as long as field theory applies, we have to specify the initial quantum state of the scalar fields, which will also be the initial state of their perturbations. We assume

that this state is the *Bunch-Davies vacuum*. Annihilated by  $\alpha_i(\mathbf{k})$ , as defined in (29), yields  $\alpha_i(\mathbf{k})|0\rangle = 0$ , where  $i$  is the index for the different fields and  $|0\rangle$  is the vacuum state of the free (non-interacting) quantum field theory in curved spaces, called Bunch-Davies vacuum. If the inflaton modes are quantized in terms of time dependent creation and annihilation operators, the Bunch-Davies vacuum can be defined as the vacuum which is annihilated by  $\alpha_{\mathbf{k}}(t)$  as  $t \rightarrow -\infty$ . Besides the use of the adiabatic Bunch-Davies ground state as initial conditions, other excitations can exist due to boundary conditions or low scales of new physics. A non-Bunch-Davies vacuum is not difficult to occur. In some Dirac-Born-Infeld inflation models the speed limit of inflation and the scale of the new physics are directly related to the warp factor of the extra dimensions, which appears in the metric of warped spaces concealing the extra dimensions. Such spaces are naturally present in most cosmological scenarios including extra dimensions due to the fluxes used for stabilizing the compactification of strings. Such models have the possibility of more vacuum choices [35, 36].

A plethora of inflationary models can produce large non-Gaussianities, some of them are the multi-field inflation [37, 38, 33, 39], the curvaton scenario [40, 41, 42], the ekpyrotic inflationary scenario [43, 44] and vector field populated inflation [45, 46, 47]. A nice review on the production of non-Gaussianities from inflationary models can be found in [48, 49].

All the above different types of inflation models violate one of the previous conditions and generate large non-Gaussianities. Therefore in order to distinguish between all these different mechanisms we have to gain additional information, besides the ones provided by the power spectrum, from the non-Gaussian part of the primordial perturbations. Power spectrum describes Gaussian random fields, hence any information on non-Gaussianity must be extracted from the higher order correlation functions, where by higher order we mean higher than the two-point correlation function. The first non-Gaussian correlator is the three-point correlation function which correlates density or temperature fluctuations at three different points in space. Its presence guarantees the departure from Gaussianity.

The Fourier transformation of the three-point correlation function, the *bispectrum*, is defined similarly to power spectrum in equation (7)

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \quad (40)$$

where  $\zeta_{\mathbf{k}}$  are the Fourier modes of the primordial curvature perturbations in real space  $\zeta(\mathbf{x})$  and the relation between them is

$$\zeta_{\mathbf{k}} = \int \zeta(\mathbf{x}) e^{-i\mathbf{k}\mathbf{x}} d^3\mathbf{x} \quad (41)$$

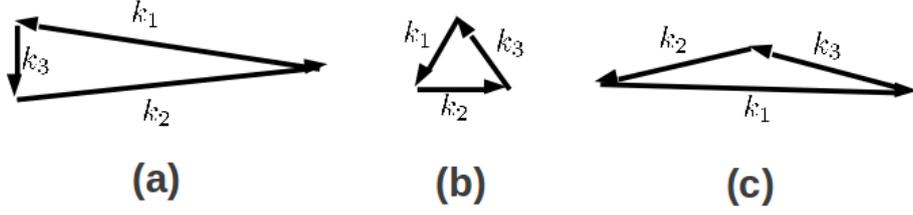
Here we use another symbol for the curvature perturbations used commonly in the literature, therefore we can write  $\zeta_{\mathbf{k}} \equiv \mathcal{R}_{\mathbf{k}}$ , where the last is related to the scalar fluctuations in (37).

Bispectrum correlates fluctuations at three points in Fourier space forming a triangle with the three wavevector due to momentum conservation, coming from the translation invariance. It is clear that the amount of information the bispectrum holds is far greater than that of the power spectrum, which correlates only two points. The number of shapes for the forming triangle is large and the different inflation models predict different shapes. The shape of the bispectrum refers to the dependence of the three-point correlation function on the ratios  $k_2/k_1$  and  $k_3/k_1$ , while we keep the overall momentum  $K = k_1 + k_2 + k_3$  fixed and restricted to zero [50, 51, 52]. The shape function contains the information about the momentum dependence and is defined [51]

$$S(k_1, k_2, k_3) = N^{-1} (k_1 k_2 k_3)^2 B_\zeta(k_1, k_2, k_3) \quad (42)$$

where in general  $1/N \sim f_{NL}$ .

There are generally three classes of bispectrum shapes in the momentum space (Fig. 1) which characterize the local, equilateral and folded type of primordial non-Gaussianity. The magnitude of each class is related to a dimensionless non-linear parameter  $f_{NL}$  [50]. This parameter defines the deviation from the Gaussian initial conditions and hence the



**Figure 1:** The shapes of Bispectrum, created by the three wave vectors. (a) The squeezed configuration with  $k_3 \ll k_2 \approx k_1$ . (b) The equilateral configuration, where  $k_1 = k_2 = k_3$ . (c) The folded configuration, where  $k_1 \approx 2k_2 \approx 2k_3$ .

amplitude of non-Gaussianity in each inflationary model, producing the relevant bispectrum shape. For the non-Gaussianity of the local type,  $S$  peaks at the squeezed triangle having the form  $S \sim k_1/k_3$ , hence the bispectrum is dominated by the squeezed configuration. For the equilateral type non-Gaussianity,  $S$  peaks at the equilateral triangle limit and bispectrum is governed by the equilateral configuration. The most studied type of non-Gaussianity in the literature is that of the local type. This type is generated by models that develop non-linearities outside of the horizon. Models where there is a contribution to the primordial fluctuations other than that of the inflaton have this characteristic. Non-linearities come from the evolution of these extra fields outside the horizon, which eventually pass into the density perturbations. Such model types are the multi-field and curvaton scenarios. Ekpyrotic inflation can also generate local non-Gaussianities [43]. Non-Gaussianities of the equilateral type can be produced by DBI inflation [53] and in models with higher derivative terms [34, 34]. Finally folded shape non-Gaussianities can occur if we consider a non-Bunch-Davies vacuum for the initial conditions. Hence different inflation models will produce different signature in the shape of bispectrum.

For the local configuration we can write the primordial curvature perturbations  $\zeta(\mathbf{x})$  as

$$\zeta(\mathbf{x}) = \zeta_G(\mathbf{x}) + \frac{3}{5} f_{NL}^{loc} [\zeta_G^2(\mathbf{x}) - \langle \zeta_G^2(\mathbf{x}) \rangle] \quad (43)$$

where  $\zeta_G(\mathbf{x})$  is a Gaussian random field. By adding a non-linear term to the Gaussian one, we made  $\zeta(\mathbf{x})$  non-Gaussian. The amplitude of non-Gaussianity in this template is parametrized by  $f_{NL}^{loc}$  as we explain before. The 3/5 is there for historical reasons and mainly because of the different conventions used in the literature to define  $f_{NL}$ . Here we use the convention used in [54], where to describe non-Gaussianity they expand in Taylor series the primordial gravitational potential perturbation, instead of the curvature ones, around the Gaussian part [54, 55, 56]

$$\Phi(\mathbf{x}) = \Phi_G(\mathbf{x}) + f_{NL} [\Phi_G^2(\mathbf{x}) - \langle \Phi_G^2(\mathbf{x}) \rangle] \quad (44)$$

where  $\Phi(\mathbf{x})$  is the Bardeen gauge-invariant potential and  $\Phi_G(\mathbf{x})$  is the Gaussian part of the potential and hence a random field. By using primordial here we mean the gravitational potential before the action of the transfer function. The Bardeen potential can be reduced to the usual Newtonian gravitational potential in the conformal Newtonian gauge for sub-Hubble scales, up to a minus sign. This type, given by equation (44), is predicted by the inflation models which generate non-Gaussianities outside of the horizon. In equation (43) we have achieved in introducing non-Gaussianity in the perturbations of the primordial gravitational potential and hence, since all the fields couple to gravity, to the density field. Adopting the large scale structure convention, we linearly extrapolate  $\Phi$  to  $z=0$  by  $\Phi(\mathbf{k}, z) = D(z)(1+z)\Phi(\mathbf{k})$ , where  $D(z)$  is the linear growth factor. In the CMB convention instead of extrapolating  $\Phi$  to  $z=0$ , we use the primordial one. The relation between the two  $f_{NL}$  parameters from the two conventions is:  $f_{NL}^{LSS} = g(z = \infty)/g(0)f_{NL}^{CMB} \sim 1.3f_{NL}^{CMB}$ , where  $g(z) = D(z)(1+z)$  is the linear growth suppression factor [57, 58].

The primordial curvature perturbation  $\zeta$  is related with the Bardeen potential for an equation of state  $p = w\rho$ , describing an perfect fluid, by

$$\zeta = \frac{5 + 3w}{3 + 3w}\Phi + \frac{2}{3(1+w)H}\dot{\Phi} \quad (45)$$

where for large scales and a period of constant  $w$ ,  $\dot{\Phi} = 0$ . In the case of matter domination

era, where  $w = 0$ , we get

$$\zeta = \frac{5 + 3w}{3 + 3w} \Phi = \frac{3}{5} \Phi \quad (46)$$

The primordial bispectrum for the local non-Gaussianity, described by (44), can be defined in momentum space as [54]

$$B_{\Phi}^{loc}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2f_{NL}^{loc} (P_{\phi}(k_1)P_{\phi}(k_2) + P_{\phi}(k_2)P_{\phi}(k_3) + P_{\phi}(k_1)P_{\phi}(k_3)) \quad (47)$$

where  $P_{\phi}(k)$  is the primordial gravitational potential power spectrum as in equation (7). Here and for the rest of the section  $\phi = \Phi_G$ . The above result is calculated by combining equation (44) and the version of equation (40) for the gravitational potential. It is obvious that we can calculate a similar bispectrum  $B_{\zeta}$  for the curvature perturbations since the two quantities are directly related. For the equilateral type non-Gaussianity, we can write a similar form to Eq. 47 for the equilateral bispectrum as the factorized form proposed in [59]

$$\begin{aligned} B_{\Phi}^{equil}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = & 6f_{NL}^{equil} \left( - (P_{\phi}(k_1)P_{\phi}(k_2) + cyc.) \right. \\ & - 2 (P_{\phi}(k_1)P_{\phi}(k_2)P_{\phi}(k_3))^{2/3} \\ & \left. + (P_{\phi}^{1/3}(k_1)P_{\phi}^{2/3}(k_2)P_{\phi}(k_3) + perms.) \right) \end{aligned} \quad (48)$$

where *cyc.* is all the cyclic permutations between  $k_1, k_2, k_3$ . It is easy to see that the signal of  $B_{\Phi}^{equil}$  is maximum for the equilateral template, where  $k_1 \approx k_2 \approx k_3$ . For the third non-Gaussian type, in which the bispectrum is dominated by the folded configuration, we have from [60]

$$\begin{aligned} B_{\Phi}^{fol}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = & 6f_{NL}^{fol} \left( (P_{\phi}(k_1)P_{\phi}(k_2) + cyc.) \right. \\ & \left. + 3 (P_{\phi}(k_1)P_{\phi}(k_2)P_{\phi}(k_3))^{2/3} \right. \\ & \left. - (P_{\phi}^{1/3}(k_1)P_{\phi}^{2/3}(k_2)P_{\phi}(k_3) + perms.) \right) \end{aligned} \quad (49)$$

where again the signal of this type is maximized by the folded configuration,  $k_1 \approx 2k_2 \approx 2k_3$ .

In order to use non-Gaussianities as a probe to the early universe and specifically to the aspects of inflation, we need to measure the shape of the primordial bispectrum and the magnitude of its signal by constraining the  $f_{NL}$  parameter. Mainly there are two ways to get information about the perturbations in the early universe, from the CMB anisotropies and from the abundance and clustering of the large-scale structures. The tighter constraints on  $f_{NL}$  and non-Gaussianity comes from the WMAP, they are summarized by: WMAP3 data [61],  $-36 < f_{NL}^{loc} < 100$  and  $-256 < f_{NL}^{equil} < 332$  at 95% CL, from WMAP5 [62, 63],  $-4 < f_{NL}^{loc} < 80$  and  $-151 < f_{NL}^{equil} < 253$  at 95% CL and finally from WMAP7 [64, 65],  $f_{NL}^{loc} = 32 \pm 21$  at 68% CL and  $-382 \leq f_{NL}^{equil} \leq 202$  at 95% CL.

A non-Gaussian primordial gravitational field can produce non-Gaussianity in the primordial density field. During the matter dominated era in the usual Newtonian gauge, the Bardeen potential satisfies the Poisson equation. Therefore we can relate the density fluctuations with the primordial curvature perturbations, which are directly related to the fluctuations of the primordial gravitational potential (Eq. 46). The cosmological Poisson equation is

$$k^2\Phi(\mathbf{x}) = 4\pi G\alpha^2\delta(\mathbf{x}) \quad (50)$$

where  $\delta(\mathbf{x})$  is the density fluctuation (see equation (2)). By linearly extrapolate  $\Phi(\mathbf{k}) \rightarrow \Phi(\mathbf{k}, z)$ , as noted before, and by using equation (46) and equation (50) in Fourier space we get the relation between  $\delta$ ,  $\zeta$  and  $\Phi$

$$\delta(\mathbf{k}, z) = \frac{2k^2c^2T(k)}{5H_0^2\Omega_{m,0}}D(z)\zeta(\mathbf{k}) = \frac{2k^2c^2T(k)D(z)}{3H_0^2\Omega_{m,0}}\Phi(\mathbf{k}) \quad (51)$$

where  $D(z)$  is normalized to  $1+z$  and  $z$  is the redshift,  $T(k)$  is the matter transfer function normalized to unity at  $k \rightarrow \infty$  and  $\Omega_{m,0} = \rho_m(z=0)/\rho_{crit}$  is the density parameter for the

matter in the present time, where  $\rho_{crit} = 3H_0^2/8\pi G$  is the critical density of the universe in order to have a flat geometry. The relation parameters in (51) are usually denoted in literature as

$$\mathcal{M}(\mathbf{k}, z) = \frac{2}{3} \frac{k^2 c^2 T(k) D(z)}{H_0^2 \Omega_m}. \quad (52)$$

When we compare observations with theoretical predictions, we have to apply some kind of smoothing. The usual linear matter power spectrum can be written now as,  $\mathcal{P}_m(k, z) = \mathcal{M}(\mathbf{k}, z)^2 \mathcal{P}_\phi(k)$ . In the case of LSS we need to smooth out non-linear dynamics in the collapsed regions, which later on will form galaxies, by assuming a spherical collapse. We choose the Fourier coefficient of the spherical top-hat filter,  $W(kR) = 3 \left( \frac{\sin(kR)}{(kR)^3} + \frac{\cos(kR)}{(kR)^2} \right)$ , with a characteristic radius  $R$ , in order to smooth out non-linearities in the small scales. Applying this window function to equation (51) we get the smoothed density field

$$\delta_R(\mathbf{k}, z) = \mathcal{M}(\mathbf{k}, z) W(kR) \Phi(\mathbf{k}) = \mathcal{M}_R(\mathbf{k}, z) \Phi(\mathbf{k}) \quad (53)$$

Using the above relations we can define the bispectrum for the density field

$$B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z) = \mathcal{M}_R(\mathbf{k}_1, z) \mathcal{M}_R(\mathbf{k}_2, z) \mathcal{M}_R(\mathbf{k}_3, z) B_\Phi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \quad (54)$$

Finally we can define from the density field bispectrum the three-point correlation function for density perturbations similar to equation (6)

$$\langle \delta_R^3 \rangle = \int \frac{dk_1}{2\pi^3} \int \frac{dk_2}{2\pi^3} \int \frac{dk_3}{2\pi^3} B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z) \quad (55)$$

To complete this short review on non-Gaussianities we'll refer to an other way, used in the literature, to describe the primordial curvature perturbations, the  $\delta N$  formalism [66, 67]

$$\zeta = \delta N \approx \sum_I N_{,I} \delta\phi_*^I + \frac{1}{2} \sum_{IJ} N_{,IJ} \delta\phi_*^I \delta\phi_*^J \quad (56)$$

where  $N(\mathbf{x}, t) = \ln \alpha$  is the number of e-folds calculated from the time that the modes exit the horizon, to a final uniform energy density hypersphere, which usually is the radiation dominated era. By substituting equation (56) to (40) and (47) we get

$$\begin{aligned} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle &= \sum_{IJK} N_{,I} N_{,J} N_{,K} \langle \delta\phi_{\mathbf{k}_1}^I \delta\phi_{\mathbf{k}_2}^J \delta\phi_{\mathbf{k}_3}^K \rangle + \\ &\frac{1}{2} \sum_{IJKL} N_{,I} N_{,J} N_{,KL} \langle \delta\phi_{\mathbf{k}_1}^I \delta\phi_{\mathbf{k}_2}^J (\delta\phi^K \delta\phi^L)_{\mathbf{k}_2} \rangle + perms. \end{aligned} \quad (57)$$

where  $\delta\phi_{\mathbf{k}}$  denotes the fluctuations of the scalar field driving the inflation. The  $f_{NL}$  parameter is defined now by

$$f_{NL} = \frac{6 N_{,I} N_{,J} N_{,IJ}}{5 (N_{,K} N_{,K})^2} \quad (58)$$

The above analysis for the three-point correlation function can be generalized and derive an even higher order correlation function, the four-point correlation function. We define the *trispectrum* in the same way as bispectrum

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_\zeta(k_1, k_2, k_3, k_4) \quad (59)$$

Now the trispectrum is the shape function of a quadrilateral. Since it correlates fluctuations in four points, it can be expressed using the four wave vectors and two diagonals. By Taylor expanding the curvature perturbation  $\zeta$  we will get equation (43) with an additional term coming from a higher order of the expansion

$$\zeta(\mathbf{x}) = \zeta_G(\mathbf{x}) + \frac{3}{5} f_{NL} [\zeta_G^2(\mathbf{x}) - \langle \zeta_G^2(\mathbf{x}) \rangle] + \frac{9}{25} g_{NL} \zeta_G^3(\mathbf{x}) \quad (60)$$

where again the number in front of the additional term is there for historical reasons. The parameter  $g_{NL}$  measures again the amplitude of non-Gaussianity in the specific regime originating from higher order than  $\langle \delta^3 \rangle$  correlations. We can define the trispectrum for the local regime through a similar relation to equation (47) for the bispectrum

$$T_{\zeta}(k_1, k_2, k_3, k_4) = \tau_{NL} (P_{\zeta}(k_1)P_{\zeta}(k_3)P_{\zeta}(k_4) + 11 \text{ perms}) + \frac{56}{25}g_{NL} (P_{\zeta}(k_2)P_{\zeta}(k_3)P_{\zeta}(k_4) + 3 \text{ perms}) \quad (61)$$

The parameter  $\tau_{NL}$  is directly related to  $f_{NL}$  for a multi field inflation through the relation  $\tau_{NL} = (36/25)f_{NL}^2$  [68]. In this higher order analysis of the primordial fluctuations to determine the amplitude of non-Gaussianities, we need to constrain both  $f_{NL}$  and  $g_{NL}$ .

The importance of determining and measuring non-Gaussianities is tremendous. They are apparently the only way to retrieve information about the very early Universe. Inflation is the dominant theory that describes this period of time but the variety of models is quite large. The detection of non-Gaussianities can provide us with a way to distinguish between different classes of inflation models and eliminate the ones that don't predict such deviations from exact Gaussian distribution. Each inflationary model leaves a unique imprint determining the shape of the bispectrum and trispectrum in Fourier space. The detection of non-Gaussian signals through the CMB anisotropy, LSS clustering, gravitational lensing, the abundance of galaxies and the Lyman-a forest, can give us the information we need in order to understand the physics of the early universe and the growth of the density inhomogeneities that take place during this early stage of the evolution and eventually seed the observable structures.

# Chapter 3

## *Primordial non-Gaussianity in Large Scale Structure*

As discussed in the previous section, in order to get information about the physics of the very early Universe and constrain the plethora of inflation models describing this era, we need to measure the amplitude of primordial non-Gaussianity. The most important observables to constrain the  $f_{NL}$  parameter and determine the shape of the bispectrum, is the CMB anisotropies [69] and the clustering signal of the LSS [70, 71]. The CMB data can give information about the cosmological fluctuations when they are very close to their original primordial form. By measuring the bispectrum of the CMB anisotropy, which is directly related to the bispectrum of the primordial gravitational potential  $\Phi$ , we should be able to detect primordial non-Gaussianity [54] at this level. Some constraints on  $f_{NL}$  have been presented in the previous section. The other prominent way of measuring non-Gaussianity in the initial conditions, is by measuring the bispectrum of the galaxy distribution. The statistics of the galaxy clustering can probe the fluctuations at a time close to the present, by measuring the probability distribution function (PDF) of the matter perturbations. Therefore, it is possible to measure primordial non-Gaussianity from LSS [72]. By using the abundance of clusters and the statistics of LSS, we have better chances to detect a deviation from the Gaussianity since the signal is stronger than the CMB, as well as LSS can probe matter fluctuations on smaller scales.

However, non-linearities in LSS can also come from the gravitational instability and the biasing of galaxies relative to the underlying matter. Hence a non-zero bispectrum

can be generated, even for a Gaussian primordial field, similar to the one produced by the non-Gaussian perturbations. Adding to this, the instrumental systematic errors that one comes across in LSS surveys, make the measurement of a non-Gaussian signal in the primordial perturbations a tough deal. Before we analyze the effects of non-Gaussian primordial fluctuations in the LSS structures, we'll make a small introduction on the formation of structures. A review on the extended Press-Schechter theory and the excursion sets can be found in [73].

### 3.1 Structure formation and bias

In the standard CDM cosmologies, structure formation begins from the gravitational collapse of dark matter overdense regions in the primordial density field, into bound, roughly spherical, virialized objects called *halos*. In the fiducial cosmology, dark matter density overweights by far the density of baryons. Hence baryons trapped in the gravitational well of the dark matter halos will cool and concentrate to create galaxies. As a result of this formation process is the existence of bias between the galaxy and the underlying dark matter distribution [74]. Different galaxy types at different redshifts trace halos with different characteristics.

Halos collapse on the peaks of the initial matter density distribution which are above a particular threshold  $\delta_c$ , therefore larger halos will be created on higher- $\sigma$  peaks. As a consequence, for a Gaussian random field, the halo bias will increase with the rarity of the peak. High-*sigma* peaks in a Gaussian density distribution are more rare and hence massive halos will have higher bias. The threshold, above which the dark matter density peaks collapse, depends on the background density. At higher redshift when the universe is more dense high mass halos will be less rare and strongly biased<sup>1</sup>, while low mass halos

---

<sup>1</sup>By changing the background density, as the redshift increases, by  $\delta$  is like changing the threshold by,  $\delta_c - \delta$ . Hence more peaks will be over the threshold hosting collapsed halos. The high-mass peaks, where massive halos will emerge, will be more biased and clustered.

following the hierarchical structure formation, will be removed from the halo population through merging (anti-biasing).

The bias between the galaxy and dark matter distributions can be defined [75]

$$\delta_g = \sum \frac{1}{k!} b_k \delta_m^k \quad (62)$$

where  $b$  denotes the bias. In the LSS studies, a linear and scale independent bias is assumed,  $\delta_g = b\delta_m$ , where  $b = b_1$  in equation (62). Such a bias will only rescale the power spectrum of the field. The galaxy density field on later time will be,  $\delta_g(\alpha) = \delta_L + \delta_m(\alpha)$ , where the Lagrangian overdensity field  $\delta_L$  corresponds to the clustering of the peaks in the initial density field. Therefore it is time independent, and  $\delta_m(\alpha)$  is the matter overdensity field at a later time  $\alpha$ . Therefore by making the standard assumption that halos move coherently with the underlying dark matter and that halo merging does not take place

$$b = b_E = \frac{\delta_g}{\delta_m} = \frac{\delta_L + \delta_m}{\delta_m} = 1 + b_L \quad (63)$$

where  $b_L$  is the Lagrangian bias and  $b_E$  is the Eulerian bias.

## 3.2 Press-Schechter Theory

In order to understand the clustering of galaxies we have to understand the formation and clustering of the dark matter halos. The formalism proposed by Press and Schechter [76] and its extensions based on the excursion set [77, 78], accounting to solve the “*cloud-in-cloud*” problem of PS theory, is the usual way to understand the clustering of halos. In this framework the mass function, which is the number density of halos,  $n(M, z)$ , with mass  $M$  at a redshift  $z$  is given by

$$n(M, z) = \frac{dn}{dM}(M, z) = \frac{\bar{\rho}}{M^2} \nu f(\nu) \frac{d \ln \nu}{d \ln M} \quad (64)$$

where  $\bar{\rho} = \Omega_m \rho_{crit}$  is the average comoving background matter density and  $\nu(M, z) = \delta_c(z)/\sigma_M(R, z)$  is the height of the peak, where the dark matter halo collapse takes place. Here  $\sigma_M(R, z)$  is the variance of the initial density fluctuation smoothed on a particular scale  $R$  (or mass scale  $M = 4\pi/3R^3\bar{\rho}$ ) by the top-hat window function, defined in the previous section, and linearly extrapolated to present epoch through the growth function

$$\sigma_M^2(R, z) = D(z)^2 \sigma_M^2(R) = D(z)^2 \int \frac{k^2}{2\pi^2} P_m(k) W^2(kR) dk \quad (65)$$

$\delta_c(z)$  is the overdensity threshold above which a peak in the initial overdensity field can collapse to form a halo at redshift  $z$  and  $P_m(k, z) = \mathcal{M}(\mathbf{k}, z)^2 P_\phi(k)$  originates from the primordial potential perturbation power spectrum as explained in the previous section. For the spherical approximation  $\delta_c = \delta_{sc}(z) = 1.687 \cdot D(0)/D(z)$ , where the value of the constant appearing in this expression is for  $\Omega_m = 1$  cosmologies. We can also calculate the critical overdensity for the spherical collapse from [79],  $\delta_c = \delta_c^0(\Omega_m(z))/D(z)$ , where for the  $\Lambda$ CDM cosmology  $\delta_c^0 = 0.15(12\pi)^{2/3} \Omega_m^{0.0055}(z)$  and  $\Omega_m(z)$  is the matter density for redshift  $z$ . Finally  $\nu f(\nu)$  is the probability that an overdensity with mass  $M$  will have value bigger than the threshold  $\delta_c(z)$ . In other words,  $f(\nu)d\nu$  gives the probability distribution of the first crossing of the barrier  $B(\nu)$ , where in the spherical collapse approximation with a constant barrier,  $B(\nu) = \delta_{sc}(z)$ . In the standard excursion set theory,  $\nu f(\nu)$  depends on the barrier shape which led [78] to derive a simple formula for it in using the spherical collapse approximation and assuming Gaussian initial conditions

$$\nu f(\nu) = \sqrt{\frac{2}{\pi}} \nu \exp\left(\frac{-\nu^2}{2}\right) \quad (66)$$

where we require the integral of  $f(\nu)$  over all  $d\nu$  to be unity. For the mass function in the standard Press-Schechter theory and the spherical collapse approximation, a simple relation for the Eulerian bias of the dark matter halos has been calculated in [80, 81]

$$\delta_g = b_E \delta_m = \left(1 + \frac{\nu^2 - 1}{\delta_c}\right) \delta_m \quad (67)$$

A better fit to simulations of the halo clustering bias and the  $\nu f(\nu)$  comes from [82, 83]. Besides its phenomenological characteristic such fit can be motivated within excursion theory [83, 84] by modifying the barrier shape. In [83] they assume an ellipsoidal collapse of the dark matter halos with a moving barrier. The overdensity threshold will be in this case

$$\delta_{ec}(\sigma_M, z) = \delta_c(z) \left( 1 + \beta \left( \frac{\sigma_M^2}{\delta_{sc}^2(z)} \right)^\gamma \right) \quad (68)$$

where  $\beta = 0.47$ ,  $\gamma = 0.615$  and  $\delta_c$  is the critical density of the spherical collapse. The mass function for case of the ellipsoidal collapse is

$$\nu f(\nu) = A \sqrt{\frac{2}{\pi}} \sqrt{q} \nu \left[ 1 + \frac{1}{(q\nu^2)^p} \right] \exp\left(\frac{-q\nu^2}{2}\right) \quad (69)$$

where  $q = 0.707$ ,  $p = 0.3$  and  $A = \sqrt{\pi}/(1 + (2^{-p})\Gamma(0.5 - p)) \approx 0.322184$ , which is the normalization constant requesting  $\int f(\nu)d\nu = 1$ , here  $\Gamma$  is the gamma function. To calculate the Eulerian halo bias for this framework, we use the expression for the ellipsoidal overdensity threshold  $\delta_{ec}$  and the bias calculated in equation (67)

$$b_E(\nu) = 1 + \frac{1}{\sqrt{q}\delta_c} \left[ \sqrt{q}(q\nu^2) + \sqrt{q}b(q\nu^2)^{1-c} - \frac{(q\nu^2)^c}{(q\nu^2)^c + b(1-c)(1-c/2)} \right] \quad (70)$$

Information on the amplitude of non-Gaussianity in the primordial fluctuation field, by using LSS observations, can come from the abundance of rare collapsed objects (e.g. galaxy clusters), i.e. events corresponding to the high peaks of the underlying dark matter density field and by measuring higher-order statistics in the clustering of galaxies and galaxy clusters. The measurement of the the existence of primordial non-Gaussianity in the LSS bispectrum is sensitive at high redshifts [56, 72]. On the other hand, even a small amplitude of non-Gaussianity in the primordial density fluctuations can produce significant changes in the tail of the halo distribution. The large-scale structures tracing the dark matter halos will also be affected by the existence of primordial non-Gaussianity due to the bias effect between the galaxy and the traced underlying matter distribution.

At higher redshifts a dense background can generate more high peaks, corresponding to rare events in the matter distribution, which can host high biased systems. The statistics of such high-redshift ( $z \geq 1$ ) systems can contain detectable deviations from the Gaussian conditions, making them prominent LSS probes of primordial non-Gaussianity. Further on we'll refer to the most prominent LSS probes to constrain primordial non-Gaussianity, the mass function and the galaxy bias.

### 3.3 Non-Gaussianities in the mass function

The mass function of the matter distribution, in the extended Press-Schechter theory, can be used to measure non-Gaussianity in the primordial density field from LSS observables. The existence of non-Gaussianities in the primordial perturbations will directly affect the peak distribution, where the halos collapse and hence the mass function. In order to introduce non-Gaussianity, we need to derive an expression for the PDF of the matter distribution having this characteristic. There are mainly two approaches in the literature to address this issue, the MVJ [85] and the LMSV [52]. In both cases they calculate the PDF,  $\mathcal{P}(\delta_R)$  of the smoothed initial matter overdensity field  $\delta_R$  (defined in equation (53)) accounting for the non-Gaussian initial conditions. In MVJ case, they use the saddle point approximation to calculate the PDF, while in LMSV they do that by following the Edgeworth expansion, where analytical details about it and its validity can be found in the actual paper. The non-Gaussian effects will pass to the mass function from the Press-Schechter theory, where the probability distribution of the initial overdensity being above the threshold  $\delta_c$  is defined

$$\mathcal{P}( > \delta_c, M, z) = \int_{\delta_c(z)}^{\infty} \mathcal{P}(\delta_R) d\delta_R \quad (71)$$

From that they finally calculate the mass function, since  $n(M, z)dM \propto |d\mathcal{P}( > \delta_c, M, z)/dM|$ .

The non-Gaussian version of the above equation can be derived in the Press-Schechter

theory, while this is not true in the excursion set approach of [78]. The latest assumes that the Fourier modes of the density fluctuation field are evolving independently, which is not the case for an overdensity field with a distribution that deviates from Gaussian statistics. The non-Gaussian mass function (in the Press-Schechter approach) can be written also as the product of the Gaussian mass function with a correction factor  $\mathcal{R}_{NG}(\sigma_M, f_{NL}) = f(\sigma_M, f_{NL})/f(\sigma_M, 0)$

$$n_{NG}(\sigma_M, z) = n_G(M, z)\mathcal{R}_{NG}(M, z, f_{NL}) \quad (72)$$

where the  $z$  and  $M$  dependence comes from the linear extrapolation of  $\sigma_M$  with the growth factor and the fact that the variance is a function of  $M$ . In [86] they follow the above philosophy to write the results of [85] and [52] for the non-Gaussian multiplication factor in the mass function. In the case of MVJ [85], we have for the multiplication factor of non-Gaussian mass function

$$\begin{aligned} \mathcal{R}_{NG}(M, z) &= \exp\left(\frac{\delta_c^3 S_3(\sigma_M)}{6\sigma_M^2}\right) \\ &\times \left| \frac{1}{6} \frac{\delta_c}{\sqrt{1 - \frac{\delta_c S_3(\sigma_M)}{3}}} \frac{dS_3(\sigma_M)}{d \ln \sigma_M} + \sqrt{1 - \frac{\delta_c S_3(\sigma_M)}{3}} \right| \end{aligned} \quad (73)$$

In the case of LMSV [52] we have

$$\begin{aligned} \mathcal{R}_{NG}(M, z) &= 1 + \frac{1}{6} \frac{\sigma_M^2}{\delta_c} \\ &\times \left[ S_3(\sigma_M) \left( \frac{\delta_c^4}{\sigma_M^4} - 2 \frac{\delta_c^2}{\sigma_M^2} - 1 \right) + \frac{dS_3(\sigma_M)}{d \ln \sigma_M} \left( \frac{\delta_c^2}{\sigma_M^2} - 1 \right) \right] \end{aligned} \quad (74)$$

where  $S_3(\sigma_M) = \langle \delta_R^3 \rangle / \langle \delta^2 \rangle^2$  is the normalized smoothed skewness of the density field, with  $\langle \delta^2 \rangle = \sigma^2(R)$  being the usual smoothed variance of the density fluctuations from equation (65) and  $\langle \delta_R^3 \rangle$  is the smoothed three-point correlation function as defined in (55). Both approximations in the limit of small non-Gaussianity limit ( $\sigma/\delta_c \ll 1$ ) become

$$\mathcal{R}_{NG}(M, z) = 1 + \frac{\delta_c^3 S_3(\sigma_M)}{6\sigma_M^2} \quad (75)$$

In order to fit N-body simulation data, we substitute in both approximations for the non-Gaussian correction factor  $\delta_c \rightarrow \delta_{ec}$  [87], where  $\delta_{ec} \sim \sqrt{q}\delta_c$  for high peaks and  $q = 0.75$  coming from the data fitting. A physical motivation for this substitution comes from [86, 88], by altering the nature of the barrier in the ellipsoidal collapse to a diffusing one they introduce naturally a similar constant  $q = 1/(1 + D_B)$ , where  $D_B$  is the diffusion coefficient of the barrier with value  $D_B \sim 0.25$  coming from N-body simulations in [88].

### 3.4 Non-Gaussianities in the galaxy bias

The effect of bias in the large scale structures introduces a relation between the galaxy distribution and that of the matter density field. Thus for Gaussian initial conditions the two-point correlation function of galaxies will be also related to the correlation function of the underlying dark matter

$$\xi(r) = b_E^2 \xi_m(r) \quad (76)$$

A derivation from Gaussianity in the initial matter density field will affect directly the halo distribution and therefore the distribution of the LSS tracers. Different galaxy types trace the dark matter halos distribution differently, i.e. quasars trace unusually massive dark matter halos, acquiring different biases. Therefore we can gain observational information from LSS for the amplitude of non-Gaussianity in the primordial density perturbations by measuring the correlation function of high biased systems, like high redshift galaxies and clusters. To go beyond the linear bias, applied in the Gaussian case, we have to add more terms besides the linear  $b_1$  in the bias relation (62). The non-linear bias between the dark matter overdensity field and the galaxies will be

$$\delta_g(x) = b_1\delta_m(x) + b_2\delta_m^2(x) \quad (77)$$

$b_1$  and  $b_2$  can be assumed scale independent. Recently in [89, 90] they showed that this assumption can break down since non-Gaussianity for the primordial density field can introduce a large-scale dependent dark matter halo bias. The results for the non-Gaussian bias in the local type of primordial non-Gaussianity has been derived with many different ways [91, 92, 93] acquiring generally the same results. Here we'll refer shortly to the result of [90], where they derive a bias formula from the bispectrum and hence it can be generalized to non-local types of primordial non-Gaussianity. In this approximation the correction in the halo correlation function, originating from the non-zero three-point correlator, is given by

$$\begin{aligned} \Delta\xi_h &= \frac{\nu_R^3}{2\sigma_R^3} \left[ \xi_R^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2) + \xi_R^{(3)}(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_2) \right] \\ &= \frac{\nu_R^3}{2\sigma_R^3} \xi_R^{(3)}(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_2) \end{aligned} \quad (78)$$

where  $\nu_R$  and  $\sigma_R$  are the smoothed to a scale  $R$  quantities mentioned previously, while  $\xi_R^{(3)}(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_2) = \langle \delta_R^3 \rangle$  is calculated as in equation (55). The fraction  $\nu_R^3/\sigma_R^3$  in the above relation, comes from the Lagrangian bias of the Press-Schechter theory in the spherical collapse case (equation (67)), where the extra  $1/\delta_c$  has been dropped out due to the high peak approximation. The Fourier transformation of equation (78) together with equation (55) are used to derive the non-Gaussian contribution to the halo power spectrum. From that the non-Gaussian correction of the Lagrangian bias will be,  $b_{NG}^h = b_h(1 + \Delta b_h/b_h)$ . From the relation between the bias and the power spectrum,  $\Delta b_h^L/b_h^L = \Delta P_h/(2P_h)$ , we have

$$\begin{aligned} \Delta b_h(k, f_{NL}) &= \frac{b_G^L \delta_c(z)}{8\pi^2 \sigma_R^2 \mathcal{M}_R(k)} \int_0^\infty dk_1 k_1^2 \mathcal{M}_R(k_1) \\ &\quad \times \int_{-1}^1 d\mu \mathcal{M}_R(\sqrt{\alpha}) \frac{B_\phi(k_1, \sqrt{\alpha}, k)}{P_\phi(k)} \end{aligned} \quad (79)$$

where  $\alpha = k_1^2 + k^2 + 2k_1 k \mu$ ,  $\mathcal{M}(k) = \mathcal{M}(k, 0)$  comes from the equation (53) for  $z = 0$  and  $b_L$  is the linear Lagrangian bias for the Gaussian case given by equation (70). For a detailed derivation of the above results, the reader can refer to [90]. All the above methods for calculating the non-Gaussian contribution to the halo bias give consistent, between them, results for the Lagrangian halo bias between the smoothed galaxy and matter distribution is generalized in the form

$$\frac{\Delta b(k, f_{NL}^{loc})}{b_G^L} = 2f_{NL}^{loc} \delta_c(z) \mathcal{M}_R^{-1}(k) \quad (80)$$

This result is for primordial non-Gaussianity of the local type and it can be generalized for the non-local cases as shown in [94]. The full non-Gaussian Eulerian halo bias is given by

$$b_{NG}^E(k, z, f_{NL}) = b_G^E(z) + 3f_{NL}(b_G^E(z) - 1) \frac{H_0^2 \Omega_m \delta_c(0)}{c^2 D(z) T(k) k^2} \quad (81)$$

The linear growth factor  $D(z)$  comes from the linearly extrapolated  $\mathcal{M}_R(k, z)$  of the previous relation. We can get the Gaussian bias from [82] for the spherical collapse, or in the case we substitute  $\delta_c \rightarrow \delta_{ec} \sim \sqrt{q} \delta_c$  for the ellipsoidal collapse from equation (70), extrapolating also to  $z = 0$ .

Non-Gaussianities have induced a  $\sim 1/k^2$  scale dependence in the halo bias. This means that we'll observe a deviation in the correlation function of galaxies at large scales, since  $\xi_g(r) = b_{NG}^2 \xi_m(r)$ . Such deviations are negligible for  $k > 0.1 h\text{Mpc}^{-1}$  in the case of local and folded types of non-Gaussianity [95], although they increase rapidly for smaller  $k$ .

We can take advantage of this characteristic of the non-Gaussian bias in order to measure deviations from the Gaussian conditions in the initial matter density field. Bias is redshift dependent (besides the obvious dependence introduced by the growth factor  $D(z)$ ), since at higher redshifts where the background is denser, high-sigma peaks are less rare and more clustered producing highly biased systems.

We can gain information on the amplitude of non-Gaussianity from such high-redshift systems (i.e. galaxy clusters) and objects (i.e. quasars), due to the fact that high biased tracers will be affected more from the existence of primordial non-Gaussianities through their bias relation, and hence it will be easier to detect in their clustering statistics. In fact tight constraints on the amplitude of primordial non-Gaussianity come from existing data using non-Gaussian bias as a probe. Recently for the local type  $f_{NL}$  we have constraints from [96],  $25 < f_{NL}^{loc} < 117$  at 95% Cl, where they measured the correlation function of extragalactic radio sources at redshift  $z \approx 1$ , and from [91] with,  $-29 < f_{NL}^{loc} < 70$  at 95% CL. In the next section we'll present prominent LSS data in the literature where non-Gaussianity is not ruled out. The reader can find a general review about the effects of the primordial non-Gaussianities in LSS, briefly reviewed in this section, and the different probes that can be used to constrain such effects from the LSS in [70, 95].

# Chapter 4

## *Search for non-Gaussianity*

Large-scale structures can be used successfully, as we analyzed in the previous section, to probe the primordial matter distribution for any deviation from the Gaussian conditions. The existence of such deviations would affect the initial density field of the underlying dark matter and as a result, the biased LSS tracers.

Tight constraints on the amount of primordial non-Gaussianity could be achieved by measuring the abundance and clustering of structures formed on the high peaks of the initial matter distribution. High redshift systems are formed in higher background densities, where the bias effect between the distribution of such systems and the underlying dark matter is high. Non-Gaussian effects induce correction factors in such basic quantities of the Press-Schechter theory as the mass-function and the bias.

The usual way to put constraints on the amount of such effects, using LSS as a probe, is the bias. Non-Gaussian bias has a  $1/k^2$  scale dependence and therefore large scales will mostly be affected by the existence of a non-zero  $f_{NL}$  parameter. Tight constraints can come from the correlation function of high redshift objects, where deviation of the two-point correlation function from the  $\Lambda$ CDM model at large-scales can be fitted by models accounting for the non-Gaussian corrections in bias. In this section we'll present some of the existing LSS data in the literature with promising features in their correlation functions that leave room for non-Gaussian models without excluding them completely.

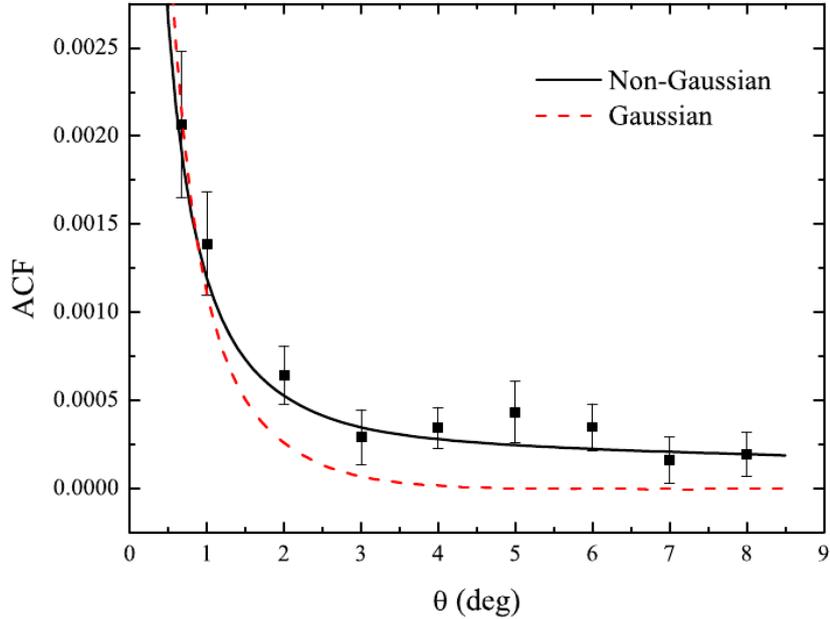
## 4.1 NRAO VLA Sky Survey from Xia et. al. 2010

In [96] they use data from the NRAO VLA Sky Survey [97]. NVSS has scanned the whole celestial sphere with a coverage of 82% at 1.4 GHz north from  $\delta = -40^\circ$ . The extragalactic radio sources of NVSS are promising candidates for clustering measurements, since they can be found at substantially high redshifts. Hence, because of the higher bias, they can be used to check the correlation function for any large scales deviations originating from the scale dependent bias contribution of a non-Gaussian density field. As an addition, radio surveys are not affected from the extinction of the galactic dust. In Xia et al. 2010, they measure the angular correlation function for radio sources having a median redshift of  $z \sim 1$ . Their main results of the angular correlation function are shown in Fig.2, together with the Gaussian and the non-Gaussian model.

The errors were estimated using a jackknife re-sampling, where they divide the data into, almost equally sized, subfields  $N_{sub} = 30$  measuring the ACF of the sample  $N_{sub}$  times, omitting each time a different subfield. In the plot we can see clearly that the angular correlation function for the Gaussian case fails to fit to the observations for the large scales. On the contrary, the case with the non-Gaussian bias has successfully modelled the large scale excess in the correlation function. The scale dependence of bias  $1/k^2$ , induced from the presence of non-Gaussian initial conditions, boosts the large scales of the galaxy power spectrum, for  $k < 0.03 h^{-1}\text{Mpc}$ . In order to incorporate the non-Gaussianities into the correlation function, they use a non-Gaussian bias as in equation (81),  $b_{NG}(M, z, k) = b_G(M, z) + 2(b_G(M, z) - 1)\delta_c(z)\alpha_M(k)$ . The parameter  $\alpha_M \sim 1/k^2$  is the factor that includes the scale dependent into the bias and it is affected directly from  $f_{NL}$ . The Gaussian bias is calculated from [83] in equation (70). Then an effective weighted bias is used to calculate the galaxy non-Gaussian angular correlation function

$$b_{NG}^{eff}(M_{\min}, z, k, f_{NL}) = \frac{\int_{M_{\min}}^{\infty} b_{NG} \frac{dn_{NG}}{dM} dM}{\int_{M_{\min}}^{\infty} \frac{dn_{NG}}{dM} dM} \quad (82)$$

where  $M_{\min}$  is the halo's minimum mass in order to collapse and create objects like the ones



**Figure 2:** The angular correlation function (ACF) of the NVSS dataset from [96]. The black solid line is the model accounting for the non-Gaussian corrections in the bias. The red-dashed line is the model assuming Gaussian initial conditions. The errors are from a jackknife re-sampling.

studied in this paper and  $dn_{NG}/dM$  is the mass function with the correction multiplying factor for the primordial non-Gaussianity, as in equation (72). The minimum mass and  $f_{NL}$  are both free parameters of the model, although the Gaussian bias affects more the small scales since the non-Gaussian corrections are negligible. Therefore, constraints for  $M_{\min}$  come from the small-scales, while for the  $f_{NL}$  come from the large ones where the scale dependent correction factor of the bias rules. In [96] they indicate that the non-Gaussian corrections appear  $r > 12$  Mpc. Strong constraints come from the NVSS radio sources for the amount of the local type of primordial non-Gaussianity after fixing  $M_{\min}$  and the cosmological parameters,  $f_{NL} = 62 \pm 27$  (68% *CL*), which agree with the other  $f_{NL}$  limits.

As we can see, the standard cosmological models fail to fit the data for large-scales.

However, this does not mean that the  $\Lambda$ CDM is ruled out. However, we can use the correlation function of large scale structure high redshift datasets in order to fit models with a non-Gaussian scale dependent correction in their bias relation and put constraints on  $f_{NL}$  parameter, competing even the constraints from WMAP. A complete analysis on the amount of primordial non-Gaussianity from the NVSS radio sources and the SDSS quasars can be found in [71], where the main result coming from the NVSS sample gives  $f_{NL} = 58 \pm 24$ .

## 4.2 The SDSS DR9 BOSS-CMASS sample from Sanchez et al. 2012

Here we'll discuss the data from the CMASS (DR9) sample of the Baryon Oscillation Spectroscopic Survey (BOSS) from [98]. In this study, they use the data in order to put constraints on the cosmological parameters by using the correlation function of the BOSS-CMASS galaxy sample combined with recent measurement from CMB and type Ia supernovae. The constraints on the cosmological parameters comes from the usage of the baryon acoustic oscillation peak position of the galaxies correlation function as a standard ruler. Standard rulers are objects with a well known comoving size as a function of redshift and hence they can be used to measure with a better accuracy angular diameter distances, together with the  $H(z)$ . Therefore, we can put constraints on the equation of state and the  $\Omega_k$  of the universe, providing us with better knowledge of the cosmological parameters.

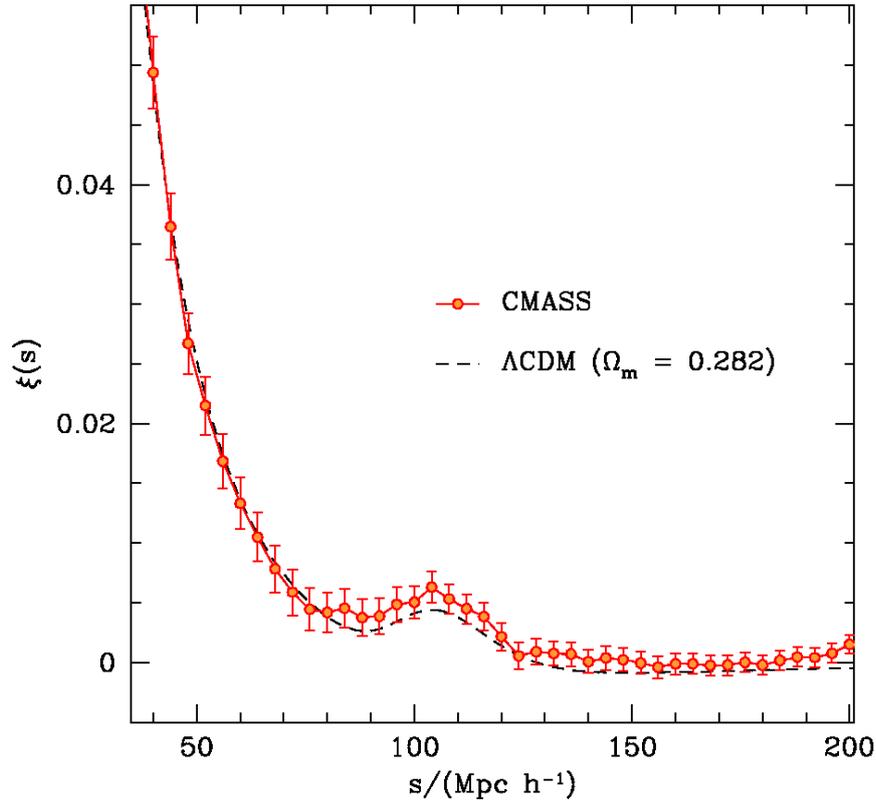
The BOSS survey is part of the SDSS-DR9 release and it consists of two spectroscopic surveys [99]. The first survey will measure the redshift of  $1.5 \times 10^6$  colour-magnitude selected massive galaxies up to  $z = 0.7$  with a magnitude limit at  $i < 19.9$ . It consists of two subsamples with different redshift ranges. The first is the LOWZ sample with range  $0.2 < z < 0.4$  and the second is the CMASS sample with redshifts above  $z = 0.4$ . In

this thesis the CMASS galaxy sample is used, consisting of 206,104 galaxies in the range discussed above. In this section we'll deal with the BOSS CMASS sample only.

The CMASS galaxies have a mean redshift at  $z = 0.55$ , for the ranges stated above. We will check in this section if the sample has any non-Gaussian signal in its correlation function. Such a non-Gaussian effect is induced from the halo bias, which imparts a scale-dependent correction. Besides the obvious large-scale dependence, bias depends on redshift. In small redshifts the bias is smaller and therefore we expect that the non-Gaussian signal and the  $f_{NL}$  constraints will be weaker. Hence we'll try to fit for a non-Gaussian model to see if there is room at these redshifts for non-Gaussianity.

We start by presenting their clustering results. To calculate the two-point correlation function  $\xi(s, \mu)$  they use the estimator of Landy & Szalay, as in equation (13). Here  $\mu = s_{\parallel}/|\vec{s}|$  and  $s_{\parallel}$  is the projected separation  $\vec{s}$  to the line of sight. The redshift space correlation function is calculated by integrating over all angles  $\mu$ ,  $2\xi(s) = \int_{-1}^1 \xi(s, \mu) d\mu$ . To acquire the randoms they used the method described in [100], where the randoms are generated using the selection function and redshift distribution of the actual sample. The errors are estimated from the diagonal elements of the covariance matrix constructed from 600 mock catalogues of the CMASS sample [101]. In order to take into account systematic errors, they apply weights to each object in the data set and in the randoms, following [100, 102]. Firstly they apply a radial weight  $w_r = 1/(1 + P_w \bar{n}(z))$ , where  $\bar{n}(z)$  is the mean number density of the particular set and  $P_w = 2 \times 1064 \text{ Mpc}^3 h^{-3}$ . This weight is used to combine data from regions with different mean number densities, which occur usually due to a redshift dependent selection, minimizing the variance of the estimator. Hence they have nothing to do with any systematic error correction [102].

To correct the data from the missing redshifts in some galaxies, originating from the fibre collisions of the spectrograph, they add a weight  $w_{mr} = w_{rf} + w_{fc} - 1$ , where the starting value of  $w_{rf}$  and  $w_{fc}$  is unity. For every missing redshift, they raise the value of  $w_{rf}$  of the closest galaxy by one and for every fibre collision they raise the value of  $w_{fc}$  for



**Figure 3:** The redshift-space two point correlation function from [98] together with the  $\Lambda$ CDM model.

galaxies closer than  $62''$ , by one. This way they manage to correct for the fact that the spectrograph cannot allocate on the plate fibres closer to each other than  $62''$  and hence under-sample the number of galaxies in clusters with smaller separations. Finally for the other systematics that affect the clustering signal, besides the ones mentioned above, they use a weight  $w_{sys}$ . Therefore the final weight that they use to the data and the randoms to measure the correlation function is,  $w_{tot} = w_r w_{mr} w_{sys}$ . Analytical details of these systematics and their effects on CMASS sample can be found in [102]. The clustering measurements of [98] together with the  $\Lambda$ CDM model with Gaussian initial conditions

are shown in Fig.3

The best fitting model in Fig. 3 is for  $\Omega_m = 0.282$  and assumes Gaussian initial conditions. As we can see, the model is in excellent agreement with the shape of the correlation function and the position of the BAO peak at  $105 \text{ h}^{-1}\text{Mpc}$ . They follow [103] by using the renormalized perturbation theory [104] to calculate the matter power spectrum and correlation function. We combine the results of  $\xi(s)$  in [105], where they have calculated the correlation function in redshift space for smaller scales, with the results shown in Fig.3. In this way we build the two-point correlation function of the CMASS sample for a bigger range of scales. In both papers they use the same fiducial cosmology ( $\Omega_m = 0.274$ ,  $h = 0.7$ ) and techniques to calculate the correlation function  $\xi(s)$ . The errors shown in Fig. 3 are estimated from the covariance matrix of the sample by using 600 individual mock catalogues. In order to calculate the non-Gaussian corrections of the correlation function, we need to measure the Gaussian bias and the mass function of the sample. In this way the only free parameter to constrain will be the  $f_{NL}$ . The Gaussian bias can be calculated from equation (70) for the ellipsoidal collapse. Together with the corrected mass function for the non-Gaussian effects, we can define the effective non-Gaussian halo bias from

$$b_{NG}^{eff}(M, k, z) = \frac{\int dM n_{NG} b_{NG}(M, k, z) \langle N(M) \rangle}{\int dM n_{NG} \langle N(M) \rangle} \quad (83)$$

where  $n_{NG} = dn_{NG}/dM$  is the non-Gaussian mass function from equation (72) and  $\langle N(M) \rangle$  is the mean number of galaxies in a halo of mass  $M$ .

As we described in the previous section, galaxies form inside halos of dark matter. The mass of halos needs to exceed a threshold  $M \geq M_{\min}$ , in order for enough regular matter to fall into halo's potential well, collapse under the gravitational attraction and form galaxies. Therefore halos with mass greater than the threshold  $M_{\min}$  will host one galaxy in their centre. According with the hierarchical formation of structures subhalos merge with other subhalos or halos to create higher mass bound objects. These subhalos

can host lower mass galaxies (satellites), where they orbit around the host halo. At a particular point these satellites will succumb under the gravitational force of the central galaxy, which eventually will consume them. The result are mergers of central galaxy with the satellites, which is a frequent event in highly populated galaxy clusters and superclusters.

Halo Occupation Distribution (HOD) models describe the way galaxies occupy halos as a function of the halo mass, giving theoretical prediction for the mass function as well as the bias of dark matter halos. In the HOD formalism in order to calculate the number of galaxies for a halo of mass  $M$ , we have to account also for the mean satellite number. According to the NFW profile [79] the satellites are distributed around the central galaxy, which resides in the halo. Hence we can assume that only halos with a central galaxy can host satellites. The usual assumption is that satellites follow a Poisson distribution [106], where here we take their mean number to be [107] for  $M \geq M_{\min}$

$$\langle N_s(M) \rangle = \left( \frac{M}{M_1} \right)^\alpha e^{-M_{\min}/M} \quad (84)$$

where  $M_1$ ,  $M_{\min}$  and  $\alpha$  are the free parameters of the HOD model, with  $M_{\min}$  being the minimum mass of a halo to have one central galaxy and  $M_1$  is the mass of the halo hosting one satellite. If we also take into account that not all halos contain a central galaxy due to the cut off in mass, we can define the number of the central galaxy,  $\langle N_c(M) \rangle = \exp(-M_{\min}/M)$ , being zero or one for  $M \geq M_{\min}$ . Hence we can calculate the mean number of galaxies for a halo of mass  $M$

$$\langle N(M) \rangle = \langle N_c(M) \rangle + \langle N_s(M) \rangle = e^{-M_{\min}/M} \left( 1 + \left( \frac{M}{M_1} \right)^\alpha \right) \quad (85)$$

The free parameters of the HOD model are calculated after the best fitting to clustering data. There are HOD models with more free parameters that can have a better fit to the clustering data [108]. For a review on HOD models the reader is referred to [109].

In order for us to fit a non-Gaussian model on the CMASS data, we'll take the values

of the best fit free parameters of the HOD model calculated in [105], where  $\log(M_{\min}) = 13.07 \pm 0.40$ ,  $\log(M_1) = 14.25 \pm 0.17$  and  $\alpha = 0.94 \pm 0.42$ . In [110] they find a 10% satellite fraction in the galaxy sample of CMASS from the fitting with the correlation function.

In the case of a non-Gaussian model our interest lies in the large scales of the correlation function, which is mostly affected from the scale dependent correcting factor of the non-Gaussian bias. Hence, we will calculate only the 2-halo term of the two-point correlation function, which is the part of the correlation function originating from pairs of galaxies located in different halos. The 2-halo term of the correlation function is the dominant one at large scales. On the other hand the 1-halo term, which comes from the pairs of galaxy-satellite and satellite-satellite of the same halo, dominates the small-scale regime. To calculate the linear matter power spectrum we use the same fiducial cosmology in [98]. We define the initial matter perturbation powerspectrum as

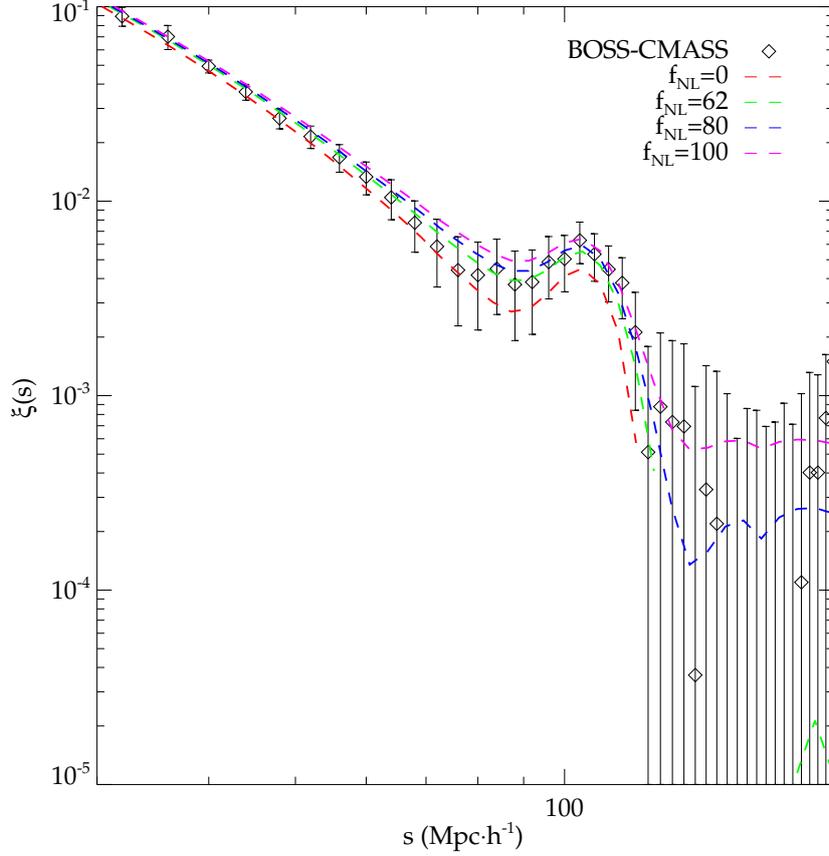
$$P_m(k) = Ak^{n_s}T^2(k) \quad (86)$$

where  $n_s = 0.96$  is the spectral index and A is the normalization constant normalizing the power spectrum at  $z = 0$  to give  $\sigma_8 = \sigma(R = 8 h^{-1}\text{Mpc}) = 0.8$ , with  $\sigma(R)$  being the smoothed variance of the initial density field at scale R

$$A = \frac{1}{2\pi^2} \frac{\sigma_8}{\int_0^\infty P(k)k^2W(k \cdot 8 h^{-1}\text{Mpc})} \quad (87)$$

$W(kR)$  is the Fourier coefficient of the window function defined in the previous section. For the transfer function we use the one proposed in [111], which accounts for the baryonic wiggles. Finally, we use the fitted formula from [112] to calculate the skewness in the multiplication correcting term of the non-Gaussian mass function. The correlation function is calculated from equation (10) after multiplying  $P(k)$  with the square of the non-Gaussian bias,  $P_{NG}^g(k) = (b_{NG}^{eff})^2 P_m(k)$ .

In order to calculate the effective non-Gaussian bias we have to calculate the non-Gaussian mass function together with the correction term in the bias. We use the three



**Figure 4:** The redshift-space two point correlation function of the BOSS-CMASS sample. We plot the  $\Lambda$ CDM model ( $f_{NL} = 0$ ) together with the non-Gaussian models of the local type for different values of the  $f_{NL}$  parameter.

free values of [105] in the HOD model together with equation (81) and (84). To test our measurements and see if we are on the right track we compare our finding for the bias of the Gaussian case, coming from the HOD, with the one measured in [105]. We find that the Gaussian bias is  $b = 1.959$  which is consistent with the measurement of [105],  $b = 1.98 \pm 0.05$ , and of [98],  $b = 1.96 \pm 0.09$ . Finally we calculate the correlation function in redshift space by using the Kaiser formula [113]. The results for the  $\Lambda$ CDM and three different  $f_{NL}$  values are plotted in Fig.4.

The  $\Lambda$ CDM model in the plot is for  $f_{NL} = 0$  and in order to be consistent with the best fit model of Fig. 3 we used  $\Omega_m = 0.282$  to calculate the correlation function for both Gaussian and non-Gaussian cases. The  $\Lambda$ CDM model fits well to the CMASS correlation function and it is consistent with the one calculated in [98]. However, it fails to fit the correlation function at large scales. After the  $120 h^{-1}\text{Mpc}$ , the Gaussian  $\Lambda$ CDM model declines while the measured clustering signal does not.

We have also calculated and plotted non-Gaussian models for 3 different values of the  $f_{NL}$  parameter. These non-Gaussian models due to the scale-dependent bias can have higher clustering amplitude at large scales, being able to fit well the correlation function of the CMASS sample, which has an increased clustering at large scale. The non-Gaussian models fit well to the shape of the correlation function till the  $120 h^{-1}\text{Mpc}$  like the  $\Lambda$ CDM model did. Further on at larger scales, due to their scale dependent bias, the non-Gaussian models instead of dropping down have a flattening which fits well to the CMASS sample at these scales. However the errors of the correlation function coming from the mock catalogues are big enough to rule out the Gaussian case. Uncertainty at these scales make the constraints on non-Gaussianities weak. All three different values of non-Gaussianity fit well at scales beyond  $120 h^{-1}\text{Mpc}$ .

Due to the large uncertainties of the large scales, we cannot exclude  $\Lambda$ CDM since it fits well for most of the scale range. Non-Gaussian models can neither be excluded since with a scale dependent bias it can fit well to the shape and BAO position of the correlation function for the whole range of scales. However in order to constrain the  $f_{NL}$  parameter we have to correct the sample for any systematic errors that may exist and will directly affect the data and hence the measured  $\xi(s)$ . Possible systematics in the BOSS survey are discussed in [102].

Recently in [114] they tried to put a constraint on the  $f_{NL}$  for the local type non-Gaussianity from the BOSS-CMASS sample, by measuring the power spectrum of the galaxies in the sample. Taking into consideration potential systematic errors contributing

to the power spectrum, they measured the amount of primordial non-Gaussianity with the method analyzed above, resulting in  $-92 < f_{NL}^{loc} < 398$  at 95% CL. The best-fit  $f_{NL}$  parameter measured from the minimization of the  $\chi^2$  goodness-of-fit test gave for the BOSS-CMASS sample,  $f_{NL}^{loc} = 71 \pm 11$  with  $\chi_{red}^2 = 0.2$ , which is consistent with the value measured in [114, 91, 96]. However, the constraints coming from [114] are weak compared to other measurement from LSS surveys like in [91, 96]. This was expected as we explained in the above analysis and easily saw in Fig. 4. Non-Gaussian models cannot be excluded from this LSS sample, since they can explain easily large scale excess in the clustering signal such as that observed in the CMASS sample.

### 4.3 The HiZELS sample from Geach et al. 2012

In [115] they present the clustering of 370 H $\alpha$  emitters (HAE) at  $z = 2.223$ . The HAE are selected in the Hi-Z Emission Line Survey (HiZELS), which is a survey targeting H $\alpha$  emitting galaxies in very narrow redshift range at  $z = 0.84$ ,  $z = 1.47$  and  $z = 2.23$  [116]. HiZELS has so far searched a  $1.2 \text{ deg}^2$  in the United Kingdom Infrared Deep Sky Survey (UKIDSS) and  $0.75 \text{ deg}^2$  in the Ultra Deep Survey (UDS) detecting 230 and 140 HAE respectively. The galaxies are selected in a narrow band from their emitting lines in a particular filter. Therefore they will have a very narrow redshift distribution where the selected population does not evolve. Galaxies at different redshifts than the selected band can introduce contamination in the sample since redshifted emitting lines can enter the narrow selection window of the survey. At high redshift such contaminations can be removed easily, since they are usually low redshift objects and a broad band colour selection can discard them from the survey. Multiple detection of the same emission line source from different narrow band filters, limits the contamination to the minimum. The contamination in the sample is expected to be less than 10%. Details for the survey and the selection can be found in [117].

They calculate the two-point correlation function of the HAE from equation (13),

where the final pair counts from each UDS and UKIDSS are unified to get the combined results for the angular correlation function. The errors are calculated from the jackknife re-sampling method calculating the full covariance matrix.

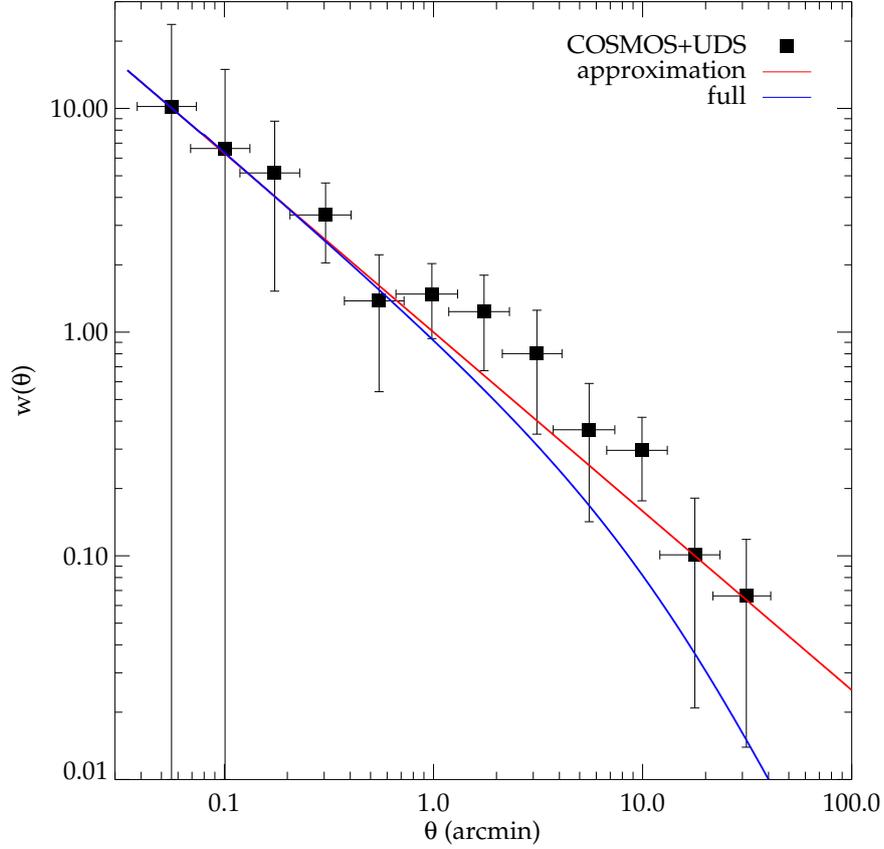
The angular correlation function can be fitted from a simple power law,  $w(\theta) = A\theta^{1-\gamma}$ . The slope  $\gamma \simeq 1.8$  is the usual power law for the two point correlation function discussed in a previous section. The power law fit of the angular correlation function is derived from the Limber's equation as the projection of the two point correlation, which as a rule of thumb can be fitted by a single power law of the form,  $\xi(r) = (r/r_0)^{-\gamma}$ . Despite the good fit to the angular correlation, the simple power law deviates at large  $\theta$  due to the divergence of the Limber's approximation for samples with a very narrow redshift distribution [118]. The approximated version of Limbers formula, which is used in [115] to calculate the amplitude of the power law, is given by [19, 119, 120]

$$A = r_0^\gamma \frac{\Gamma((\gamma - 1)/2)\Gamma(\gamma/2)}{\Gamma(1/2)} \int_0^\infty dz n(z)^2 \left(\frac{dr}{dz}\right)^{-1} r(z)^{1-\gamma} \quad (88)$$

where  $A$  is the amplitude of the angular correlation power law fit,  $n(z)$  is the redshift distribution,  $\Gamma$  is the gamma function and  $r(z)$  is the co-moving distance at redshift  $z$ .

The main reason for this deviation is that the full formula (equation 15) takes into account that the two object that form a pair are located in two different redshift bins, while the approximation assumes that pairs with a small separation ( $\theta < 1'$ ) have the same redshift. In the case where the distribution is very narrow the approximation fails and calculations with the full Limber formula is needed [20]. In Fig.5 we calculate the simple power law model from the approximate and the full Limber's formula.

In order to derive these results, we used the redshift distribution proposed in the paper [115], a Gaussian with  $z = 2.233$  being the centre and  $\sigma = 0.0126$  being the width of the band. We adopt, in our measurements, the best fitted value  $r_0 = 3.7 \pm 0.3 h^{-1}\text{Mpc}$ , for the power law of the spatial correlation function calculated in [115] after using the narrow band redshift distribution.



**Figure 5:** The angular correlation function of HAE from the COSMOS and USD fields. We fit a single power law calculated from the approximated and the full Limber’s formula. The break-down of the approximating formula for  $\theta > 1'$  is evident.

A single power law is not a sufficient model for the full angular correlation function. We have to account for the 1-halo and 2-halo terms of the correlation function, where complicated HOD models for HAE are required. However, we can use this model to fit the data and take it as a simple case for the Gaussian initial density perturbation field.

The calculations of the angular correlation power law from the full Limber’s formula shows the break-down point of the approximation at around  $\theta > 1'$ , for this narrow band selection. By using the full Limber’s formula outcome as a model for  $w(\theta)$ , we can see

that there is a clear deviation at large angles ( $\theta > 1'$ ) which cannot be fitted by this single power law.

The deviation of the power law, calculated from the full Limber's formula, at all angles is within the error limits of the angular correlation function and therefore the single power law representing the Gaussian model is not excluded.

A non-Gaussian model could fit the correlation function at these large angles, due to the scale dependence of the non-Gaussian bias. The HAE of HiZELS could be good candidates for the search of non-Gaussianity, by using their correlation function, since they are high redshift and high biased tracers. Although measurement of the angular correlation function at larger angles is needed in order to decide if its deviation from the single power law extends or it is just a local trend. The non-Gaussian signal, if any, will appear at these large scales. The constraints on the HAE HOD model are poor since the size of the sample consists of only 370 objects. Tighter constraints are needed in order to measure with higher accuracy the effective non-Gaussian bias.

Finally, the errors in the plot coming from the jackknife re-sampling are too big for us to exclude any of the models. If we consider the single power law as the Gaussian model, we can see that the deviation observed at large scales can leave some room for non-Gaussian models. It is obvious that more objects in the HiZELS sample are needed in order to be able to put tight constrain in the  $f_{NL}$  parameter.

## Chapter 5

# *Search for non-Gaussianity in the quasars of SDSS BOSS DR9*

The importance of measuring primordial non-Gaussianity lies in the fact that it is the most promising way to distinguish between the plethora of inflationary models, shedding light on the very early Universe. The most important probes to measure the amount of the primordial non-Gaussianity are the CMB and the LSS.

In order to measure non-Gaussianities from a LSS surveys, we have to take advantage of some characteristics of structure formation. Non-Gaussian initial conditions can affect the primordial gravitational potential field, which directly affects the matter density field. The mass function of the dark matter halos is affected directly, as we showed in a previous chapter. Ordinary matter trace the underlying dark matter and forms finally galaxies inside halos, hence the bias factor between their distributions will be also affected. It is easy to understand that non-Gaussianities in the primordial density field can affect the clustering of galaxies and hence we can use LSS surveys to measure any deviation from the Gaussian initial conditions.

The most important quantity measured in LSS survey is the two-point correlation function and shows the amount of clustering of the sample at different scales. Non-Gaussianities lead to higher order correlation function, which are zero in the case of purely Gaussian primordial density field. Non-Gaussian signals in the primordial density

field can be traced by measuring these higher order correlation function of galaxies and super/clusters of galaxies. The usual way to measure any deviation from Gaussianity is to use models that incorporate the corrected for non-Gaussian effects mass function and bias. Therefore we can fit a non-Gaussian model to the two-point correlation function of a LSS sample and constrain the  $f_{NL}$  parameter by measuring its best-fit value.

In this work, as we did before for the other samples, we will use the non-Gaussian bias as a probe for measuring non-Gaussianities in the sample. The non-Gaussian bias is scale dependent, therefore in order to put tighter constraints on the amplitude of non-Gaussianity from a LSS sample, we need to choose high biased tracers at high redshifts. In this way we will have better chances on finding any signal of non-Gaussianity at the large scales of the correlation function of the sample.

In this section we will use the second of the two spectroscopic samples of the SDSS BOSS. It consists of 150,000 quasars in the redshift range of  $2.2 < z < 3.5$ , with a mid redshift of  $z \sim 2.5$ . Quasars are high biased objects and are located at high redshifts. Quasar clustering can shed light on critical matters of the galaxies formation and evolution, as well as black hole growth, wind and feedback models. On the other hand, they can make excellent candidates to constrain primordial non-Gaussianity due to their high redshift and bias.

## 5.1 Data

The Sloan Digital Sky Survey [121] scanned almost a quarter of the sky using the Sloan foundation 2.5-metre telescope [122]. SDSS is now in its third phase (SDSS-III), with Baryon Oscillation Spectroscopic Survey (BOSS) being one of the most important surveys. The main purpose of BOSS survey is to measure with great precision the cosmic distance scale and the expansion rate, by using the BAO peak as a standard ruler. In order to achieve that the BOSS survey will measure the spectroscopic redshift of  $1.5 \times 10^6$  red luminous galaxies and the Ly $\alpha$  forest of high redshift quasars. The whole BOSS

survey will cover an area of  $10,000 \text{ deg}^2$ . We referred to the details of the first sample and the constraints on the amplitude of non-Gaussianity in a previous section.

Quasars are the most luminous objects in the Universe hosting super-massive black holes in their centres. They are point sources and hence it is very difficult to trace them, due to star contamination, especially the faint high redshift quasars. The rarity of these objects means that we need a large survey in order to trace a decent number of them and therefore retrieve important information from their clustering. The BOSS survey, that we will use in this section, is such a survey. Clustering studies from other smaller samples of quasars can be found in [123, 124]

The second BOSS sample consists of  $1.5 \times 10^5$  quasars in the redshift range of  $2.2 < z < 3.5$  selected from  $4 \times 10^5$  objects. The specified redshifts come from the measurements of their Ly $\alpha$  forest, which is one of the main goals of BOSS; to measure the BAO feature in the Ly $\alpha$  forest [125, 126]. The magnitude limits of the BOSS quasar survey are  $g < 22.5$  or  $r < 21.85$ .

### 5.1.1 Quasar selection in BOSS

The difficulty of such a survey is the quasar selection, which due to the rarity of these objects and the presence of stars is a tough job. To measure the spectra of quasars in a survey one needs first to target the most promising objects for being quasars. After the selection, one can take their spectra and determine their redshift and which of the targeted objects are indeed quasars. In SDSS, and therefore in BOSS, the selection is done by using the colour-colour diagrams (i.e. u-g,g-r) of point sources in the sample. Objects that lie away from the star locus in the colour-colour diagrams are targeted for spectroscopy as promising quasar candidates [127].

The redshift range of the quasar selection in the BOSS survey was selected to be  $2.2 < z < 3.5$ , since at these redshifts the BOSS spectrograph is sensitive for measuring the quasar's Ly $\alpha$  forest [128]. At redshift 2-3, where the BOSS targets also lie, the number

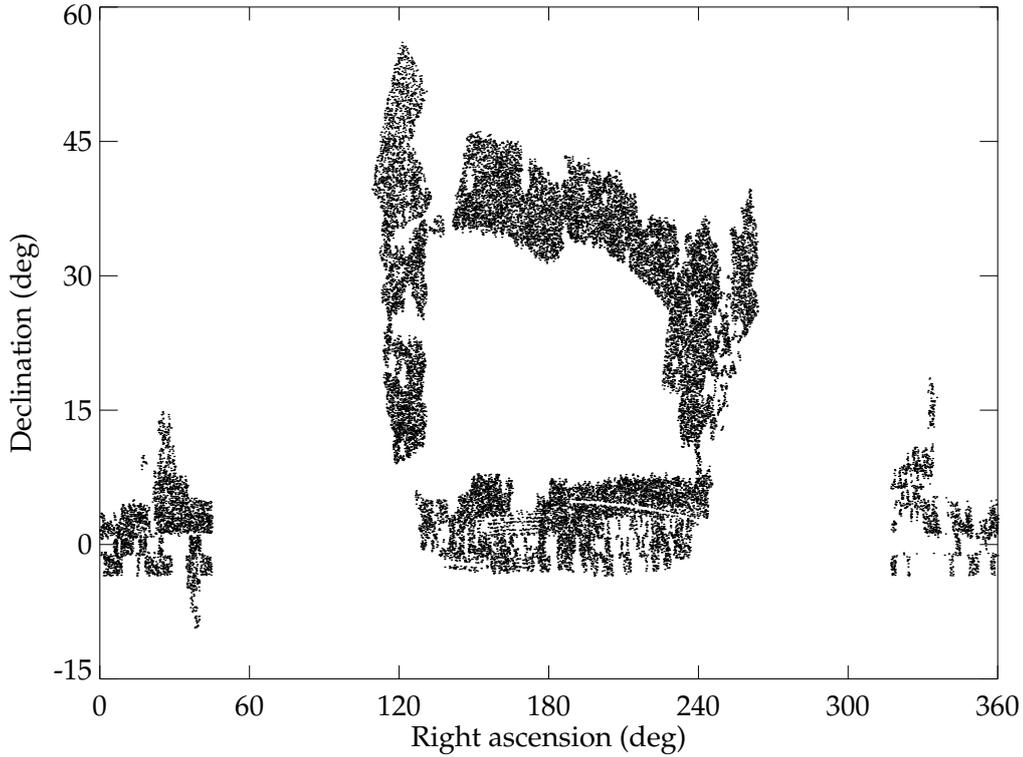
density of luminous quasars peak [123, 129]. However, the target selection is complicated at these redshift range. At redshift  $z = 2.7$  the quasar colours are similar to the colours of metal-poor A and F star populations [130] and hence the separation between these two makes the targeting of BOSS even more difficult. In addition to that, quasars at redshift  $z \sim 2.5$  are contaminated from lower redshift ( $z \sim 0.8$ ) less luminous quasars that have similar colour and flux with them [131, 132].

In order to use the SDSS BOSS quasars for statistical analysis (i.e. clustering studies), we need to produce a uniformly selected sample called the CORE. In addition to this, a BONUS sample is constructed by using as many additional data and techniques needed to reach the desired quasar density. The algorithm for the BOSS quasar selection is based on the *extreme deconvolution* algorithm (XD) of [133, 134]. After applying the XD method, every point source of SDSS BOSS is assigned with a XDQSO probability of being a quasar, by modeling the flux distribution of quasars and stars at different redshifts. In this way, a separation between targeted quasars and stars is achieved. More precisely, all the point sources of BOSS with XDQSO probability above 0.424 are targeted for spectroscopy in order to apply the CORE method [128]. More details on the CORE+BONUS method and the XDQSO technique used for the BOSS quasar selection, together with the details of the pipeline used are listed in [128].

### 5.1.2 Subsample and angular completeness

The CORE quasar sample produced with the methods described above will be used here, in order to put constraints on the amplitude of primordial non-Gaussianity. The same sample has been used before to analyze the clustering of quasars with  $z > 2.2$  in [135]. We will follow similar steps with [135] to construct and mask the CORE quasar subsample from BOSS.

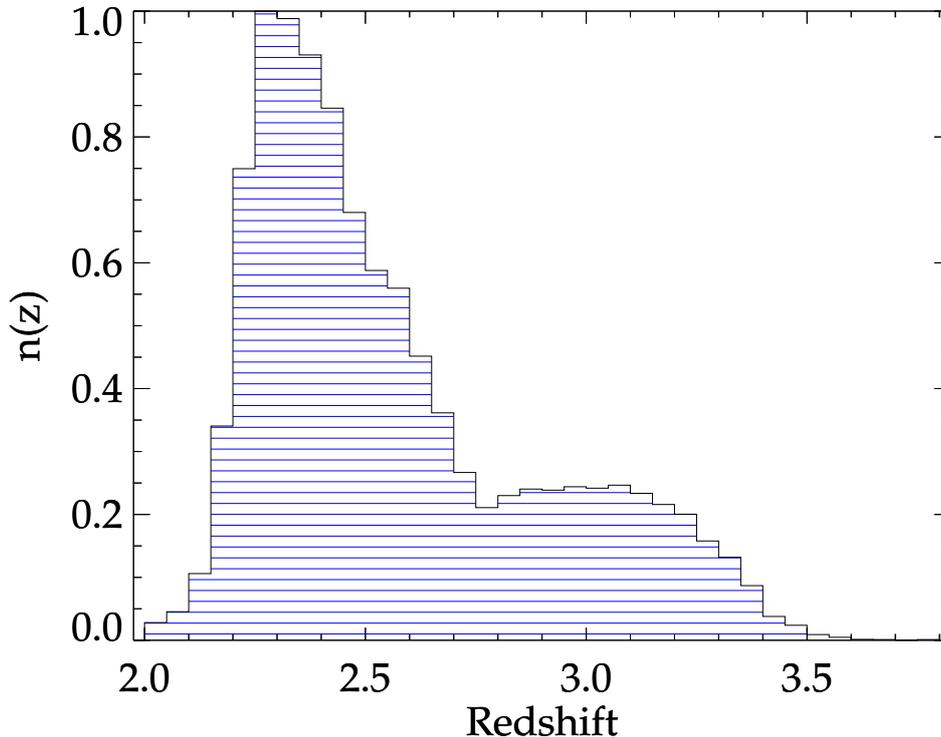
The data from SDSS Data Release 9 BOSS are used in order to apply the XDQSO technique. The *spAll - v5\_4\_45* from DR9 is used, which contains the spectroscopic



**Figure 6:** Our quasar sample, after applying the method of section 5.1.2, in J2000 equatorial coordinates. We discarded from the sample regions containing targeted quasars with less than 75% of them having received a fibre for spectroscopy.

classification and redshift of the objects from the the Spectro-1D pipeline together with their matched photometric details. The xdcare file from the SDSS DR9 is also used, which contains the quasar probabilities, used in the XDQSO selection method, for all the point sources of the SDSS DR9. These BOSS data are merged with the targets in the XDQSO CORE creating a sample that contains spectra and photometric details of the point sources matched, as well as their XDQSO probabilities.

The MANGLE software [136] is used to apply the angular mask of the BOSS DR9 survey. In order to achieve that, we use a set of polygons called 'bosspoly' covering a sector of the BOSS survey sky area. These MANGLE polygons contain the matched



**Figure 7:** The normalized redshift distribution,  $n(z)$ , of the BOSS quasar sample.

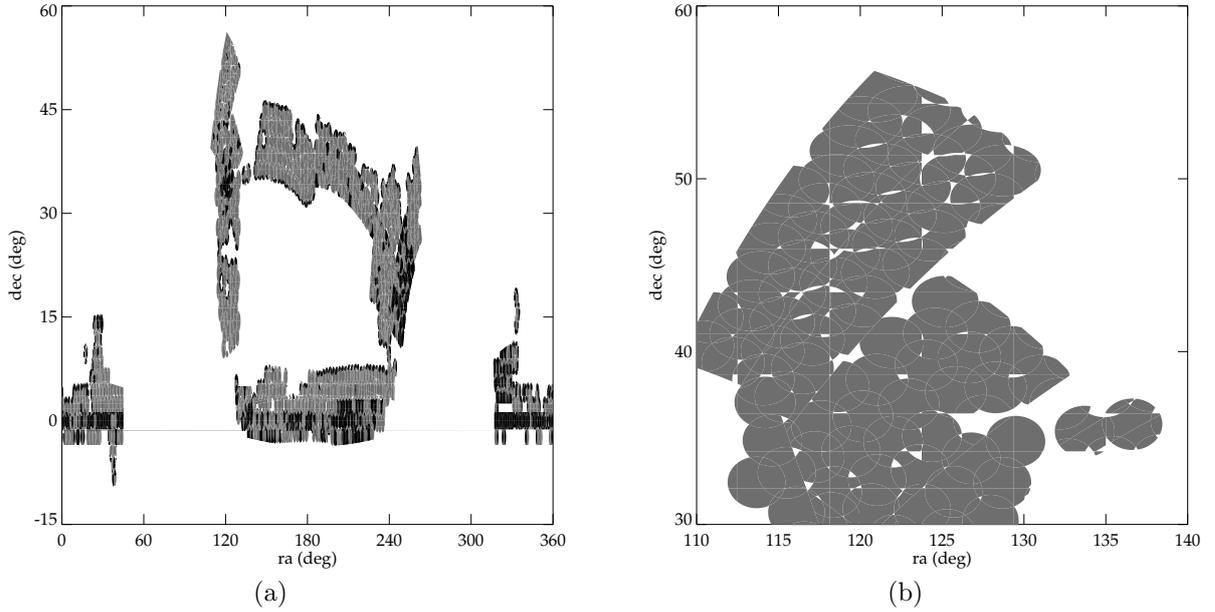
objects from BOSS and XDQSO CORE samples. The angular completeness is determined by the percentage of the targeting quasars, in a sector, that get a BOSS fibre. First we remove regions with bright stars, since no quasars can be observed there, as well as regions with bad u-band data. Further on we set a threshold of 75 per cent in the completeness. This simply means that we will keep the regions, and hence the objects inside them, in which 75% and more of the targeted XDQSO CORE quasars in them have been assigned a fibre obtaining their spectrum. More details on the above method, as well as on the redshift assignation of quasars from their spectrum and redshift errors can be found in [135]. Finally we keep in the sample only objects with  $zWARNING = 0$  indicating quasars with no known problem in their spectra. If the  $zWARNING$  flag [137], that is determined from the spectroscopic pipeline, is equal to zero then the redshift is accurate

at 99.7% level. We end up with a sample of 29,687 quasars (Fig. 6). The redshift distribution of the quasars utilized in this work is shown in Fig. 7.

In the redshift distribution of Fig. 7, the redshift range of the quasars in our sample is  $2 < z < 3.8$ . However, most of the objects are concentrated in the redshift range  $2.2 < z < 2.9$ , with the peak being at  $z \sim 2.3$ . Therefore, we will make a redshift cut and use only the quasars in redshift range  $2.2 < z < 2.9$  for our analysis, leaving 22,483 objects in the sample.

## 5.2 Clustering analysis

To measure the clustering of the quasar sample we follow the Landy and Szalay estimator, as in equation 13, and calculate the two-point correlation function  $\xi(s)$  in redshift space. To do that we count the pairs of objects that have separation  $s$ . We count the data-data  $DD(s)$ , the data-randoms  $DR(s)$  and randoms-randoms  $RR(s)$  pairs of the sample. The first are quasar-quasar pairs while the other two are pairs of quasars with randoms from the random catalogues and pairs of randoms between them. The DR and RR in the Landy & Szalay estimator are normalized by the ratio of the total number of random point  $N_{rad}$  and the total number of the quasars  $N_{dat}$ . We create the random catalogue to be  $\sim 20$  times bigger than the data and hence the ratio  $N_{rad}/N_{dat} \sim 20$ . We create also a catalogue  $\sim 50$  times bigger than the data in order to check the two results. The two correlation functions are almost the same and hence to save time in the calculation process we will use the  $\sim 20$  times bigger random catalogue. We also calculate the Hamilton estimator (equation 12) which has no normalization factor, but the difference between the two different estimators is negligible. The fiducial cosmology assumed in this work is  $\Omega_m = 0.27$ ,  $\Omega_\Lambda = 0.73$ ,  $\Omega_b = 0.045$ ,  $\sigma_8 = 0.8$ ,  $n_s = 0.96$  and  $h = 0.7$  (with  $H_0 = 100h \text{ km/s}$ ). Besides the density parameters, which define the cosmological framework of the analysis that follows, the spectral index  $n_s$  and the variance  $\sigma_8 = \sigma(R = 8 h^{-1} \text{ Mpc})$  also play a significant role. They are used in the construction of



**Figure 8:** (a) The MANGLE mask after removing the veto mask for bright stars and bad u-band fields. The polygons observed contain the XDQSO CORE objects (mask is one). The grey polygons are the ones that meet the 75% completeness threshold applied here, where the rest (black) do not and the objects inside are removed from the sample. (b) A zoom in a small region of the MANGLE mask in order to see the details of the polygons and the complexity of the applied mask. Here we do not separate with different color the polygons that meet the completeness threshold.

the linear powerspectrum that will give us the theoretical model needed to fit in the data set.

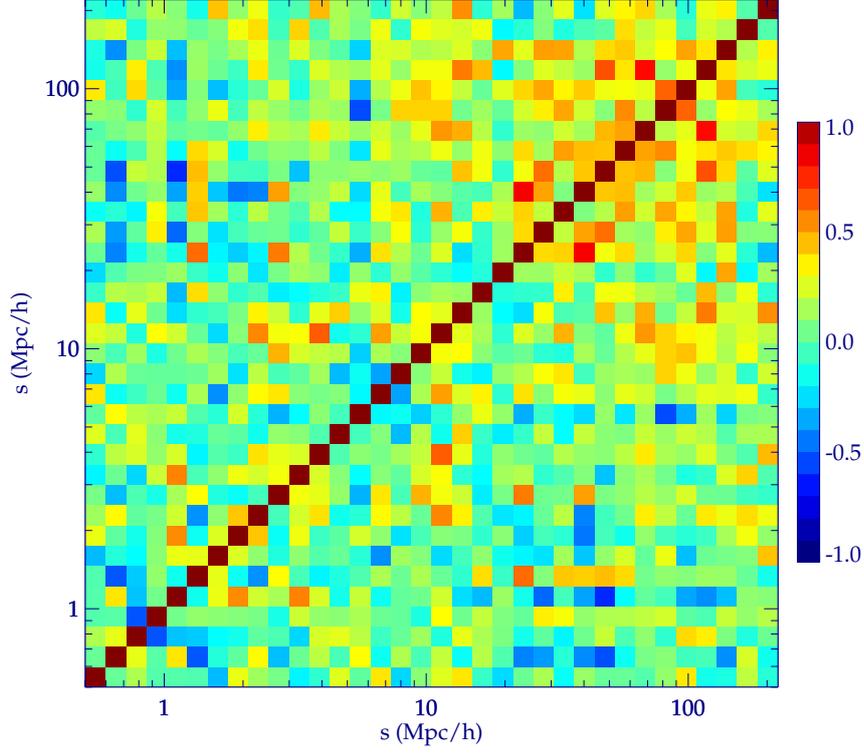
### 5.2.1 Random catalogue

In order to measure the correlation function of our sample from the Landy & Szalay estimator we need to construct a random catalogue in the angular mask of the quasars. Therefore, we have to create random points in the regions where the completeness is above

75%, in which the quasars of our sample are located. By using the *ransack* program of MANGLE we obtain the angular completeness of the survey, randomly generating angular coordinates of points inside the regions of the mask. An overview of the MANGLE mask used after removing the 'bad' survey regions is plotted in Fig. 8. Inside the mask lie the matched objects from the CORE targets and the BOSS data, where outside lie the rest objects of the BOSS survey. Each polygon is assigned with a completeness weight, which is according to the percentage of BOSS fibre assignment to the objects inside. In Fig. 8 the ones that meet the 75% completeness threshold are plotted in grey and the rest are plotted in black. The objects inside the first are used to construct the sample analysed here. We generate roughly 20 times more random points in the same regions where our quasars lie. To assign a redshift to each one of the random points we randomly take redshift from the range of the quasars following their redshift distribution. As it is also referred in [135], this method can produce artificial structures in redshift distribution of the random points, since it follows the distribution of the data. However, with the large angular size of the BOSS survey this method gives correct results.

### 5.2.2 Error estimators

In order to determine the statistical uncertainty of the measured quasar correlation function, we will use the jackknife re-sampling method. The sample is splitted in  $N_{sub} = 23$  angular regions (subfields) of equal size, where each subfield is roughly  $\sim 15 \text{ deg}^2$ . The jackknife is an internal method of error estimators. We reconstruct copies of the data by omitting in turn one subfield at a time, hence we construct  $N_{sub}$  different realizations of the original sample. The main idea of the jackknife re-sampling is to measure the correlation function of each realization and compare it with the mean correlation function of all the realizations, which in fact is the correlation function of the original data set. The jackknife error estimator is given by



**Figure 9:** The correlation coefficient,  $r_{ij}$ , which shows the level of correlation between each bin of separation  $s$ .

$$\sigma_{jk}^2 = \frac{N_{sub} - 1}{N_{sub}} \sum_{i=1}^{N_{sub}} [\xi_i(s) - \xi(s)]^2 \quad (89)$$

where the factor,  $(N_{sub} - 1)/N_{sub} = 22/23$ , corrects for the fact that the different sample realizations are not independent [138, 139]. The sum is over the square of the difference between the sample's correlation function measured without the  $i$ th subsample ( $i$ th realization) and the correlation function measured from the whole quasar sample. The jackknife error technique has been used before in many clustering analysis studies, such as in [140, 141, 142, 143]. A detailed analysis on the error estimators for two-point correlation functions can be found in [144].

The main purpose of this work is to fit non-Gaussian models to the correlation function

of the quasar sample, therefore we will have to calculate the full covariance matrix from

$$\mathbf{C}_{ij}^{jk} = \frac{N_{sub} - 1}{N_{sub}} \sum_{k=1}^{N_{sub}} [\xi_i^k - \bar{\xi}_i] [\xi_j^k - \bar{\xi}_j] \quad (90)$$

where  $\bar{\xi}(s)$  is the mean correlation function of all the realization,  $\xi_i^k(s)$  is the correlation function of the sample without the  $k$ th subsample and the subscript is the bin number. It is easy to understand that the jackknife error estimator of equation (89) is just the diagonal elements of the covariance matrix,  $\sigma_i^{jk} = \sqrt{\mathbf{C}_i^{jk}}$ . We can now compute the *correlation coefficient*,  $\mathbf{r}_{ij}$ , defined as

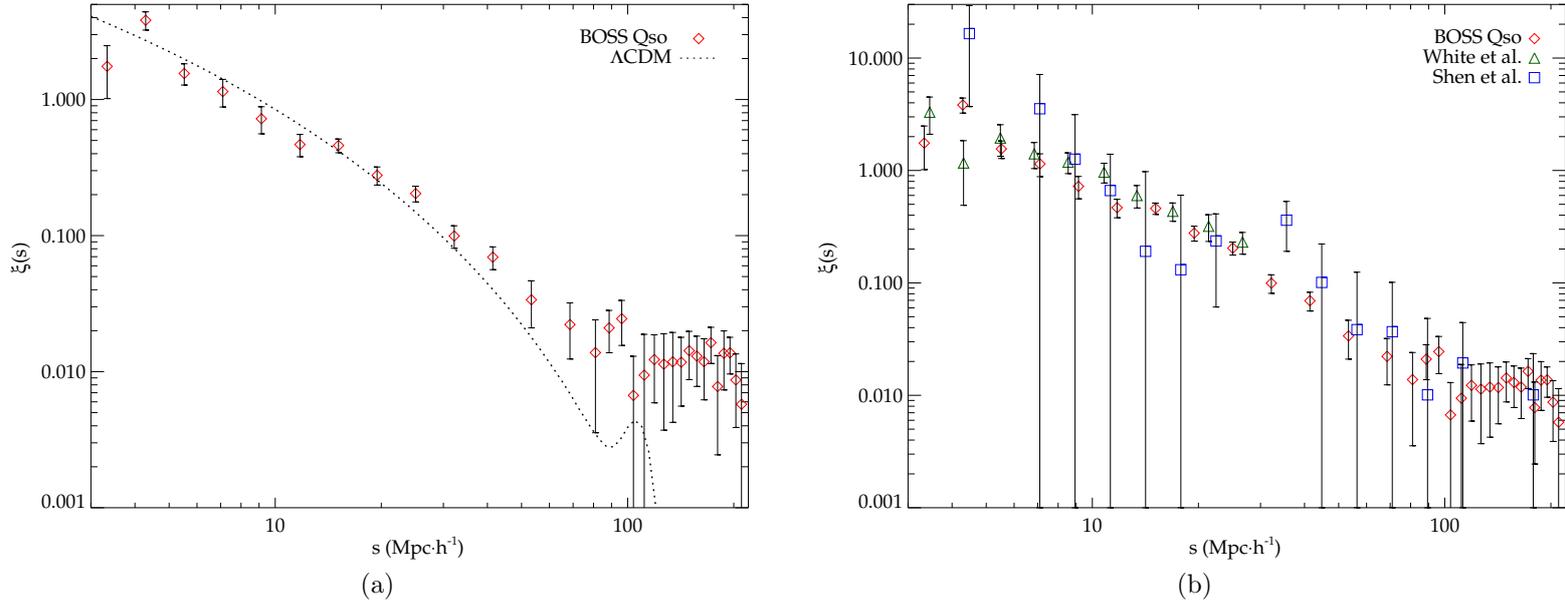
$$\mathbf{r}_{ij} = \frac{\mathbf{C}_{ij}}{\sqrt{\mathbf{C}_{ii} \cdot \mathbf{C}_{jj}}} \quad (91)$$

which is plotted in Fig 8. As we can see, the correlation of the different separation bins is negligible at small scales, while at large scales the correlation is higher but still small. Our main interest is the fitting of non-Gaussian models to the large scales of the sample, hence the covariance matrix from the jackknife re-sampling is used since  $\mathbf{r}_{ij}$  in Fig. 9 has a small correlation at large scales.

### 5.2.3 Clustering results

The two-point correlation function of the quasar BOSS sample is measured in redshift space, by using the estimators described above. For the pair counting of the Landy & Szalay formula we use the kd-tree code of [145]. The 3-D correlation results of the BOSS quasar sample is plotted in Fig. 10, together with the clustering results of the same quasar sample from [135] with redshift range  $2.2 < z < 2.8$  and the results of clustering from the quasars of Shen et al. 2007 [146], in redshift range  $2.9 < z < 5.4$ .

The sample of White et al. is the same CORE BOSS sample we use, with the same selection techniques and redshift distribution. Although they apply different redshift and magnitude cuts (for a detailed analysis see [135]). The quasars of Shen et al. 2007 [146]



**Figure 10:** (a) The measured redshift space two-point correlation function,  $\xi(s)$ , for the BOSS quasar sample. The errors are the square root of the diagonal elements of the jackknife covariance matrix. The dotted line is the best-fit  $\Lambda$ CDM model, as defined in the text below. (b) The clustering results of the quasar sample together with the clustering results of White et al. 2012, which uses the same quasar sample with different redshift cuts ( $2.2 < z < 2.8$ ), and with the redshift-space correlation function of the SDSS DR5 quasars from Shen et al. 2007, with redshift range  $2.9 < z < 5.4$ .

consists of 4,426 luminous optical quasars from SDSS DR5 at redshift range  $2.9 < z < 5.4$ . The error bars in these two samples are from the jackknife re-sampling, which is the same error estimator we applied in our sample.

The redshift space correlation function measured from the sample of White et al. is measured only for the small scales,  $3 h^{-1}\text{Mpc} < s < 30 h^{-1}\text{Mpc}$ . As we can see in Fig. 10(b) the  $\xi(s)$  measured in [135] is in very good agreement within the uncertainty limits, implied by the covariance matrix of each sample, with the correlation function measured in this work coming from the same sample as in White et al. Although we would expect the

two clustering results to differ by less than the uncertainty limits, since the two samples are almost the same and hence the error dominated by the sample cosmic variance are irrelevant. The Shen et al. quasars are one of the highest redshift clustering studies, with all the quasars having  $z > 2.9$ . We compare in Fig. 10(b) their finding with ours. Both redshift space correlation functions are consistent with each other within the uncertainty boundaries of the jackknife errors. Although most of the Shen et al. points reside near the points measured from the BOSS quasar sample, the huge error bars of the first doesn't allow us to make a fair comparison between the two. The redshift distribution of the two samples do not overlap, our sample is  $2.2 < z < 2.9$  and the Shen et al. is  $2.9 < z < 5.4$ , however they are close enough and the difference between the two correlation functions should not be large, which can be verified from Fig. 10(b). Therefore, the measured correlation function from the BOSS quasar sample is consistent with the results of White et al. and Shen et al., where they analyze the clustering of quasars in the redshift range similar or close to the one we are working on.

In Fig. 10(a), we have plotted only the correlation function of our sample together with a best fitting  $\Lambda$ CDM model. Our main goal is to try to constrain non-Gaussianities from the correlation function of the quasars by using the large scale excess of the non-Gaussian bias (equation (81)) described in the previous section. In order to achieve that, we need to measure the Gaussian bias between the measured correlation function and the correlation function generated by our model.

To generate the  $\Lambda$ CDM model we use the fiducial cosmology stated in the beginning of this section. The initial matter power spectrum is defined from,  $P_m(k) = Ak^{n_s}T^2(k)$ , which is the usual power-law with  $n_s$  being the spectral index and T being the transfer function from [111]. The normalization parameter is defined from equation (87), where we normalize the powerspectrum to give  $\sigma(R = 8 h^{-1}\text{Mpc}) = 0.8$ , with  $\sigma(R, z)$  being given by equation (65). After multiplying the power spectrum with the square of the growth factor in order to linearly extrapolate it to redshift  $z$ , we use the Fourier transformation from

equation (10) to calculate the two point correlation function of the matter distribution.

As we analyzed in previous sections there is a bias effect, originating from the way galaxies form and trace dark matter halos, between the galaxy power spectrum and the underlying matter. Since the correlation function is just the Fourier transformation of the power spectrum, the same bias effect exists between the correlation function measured from our quasar sample and the model,  $\xi_{QSO}(s) = b^2 \xi_{DM}(s)$ . Therefore, by taking into consideration the relation between the real space and the redshift space clustering [113], we defined the relationship between the linear generated correlation function and its redshift space analogue,  $\xi(s) = (b^2 + 2/3bf + f^2/5)\xi_{lin}$ , with  $f = \Omega_m(z)^{0.56}$  being the gravitational growth factor. The best-fit bias value will be measured by using the chi-squared goodness-of-fits

$$\chi^2 = \sum_{i,j=0}^N (\xi_i - \xi_i^m) \mathbf{C}_{ij}^{-1} (\xi_j - \xi_j^m) \quad (92)$$

where the sum is over the different bins  $i$  and  $j$ ,  $\mathbf{C}^{-1}$  is the inverse of the covariance matrix defined by the jackknife re-sampling method,  $\xi_i^m$  and  $\xi_i$  is the value of the model and measured correlation function respectively, at the  $i$ th bin. To calculate the model correlation function and finally measure the best fit bias, we need to define a redshift average  $\xi(s)$ . According to [135] if we cut our sample in redshift bins, big enough for bias to change from the one bin to the other, and calculate the correlation function in each bin we define a redshift average  $\xi(s)$ . This correlation function is equivalent with  $\xi(s)$  calculated at an effective redshift,  $z_{eff}$ , defined as

$$z_{eff} = \frac{\int dz n^2(z) (H(z)/d_A^2) z}{\int dz n^2(z) (H(z)/d_A^2)} \quad (93)$$

where  $n(z)$  is the redshift distribution, as shown in Fig. 7,  $d_A$  is the comoving angular diameter distance and  $H(z)$  is the Hubble parameter at redshift  $z$ . The effective redshift of our sample is  $z_{eff} = 2.4$  and is the redshift we will use to calculate all the models in this work. The  $z_{eff}$  measured here is consistent with the redshift measured in [135],

$z_{eff} = 2.39$ .

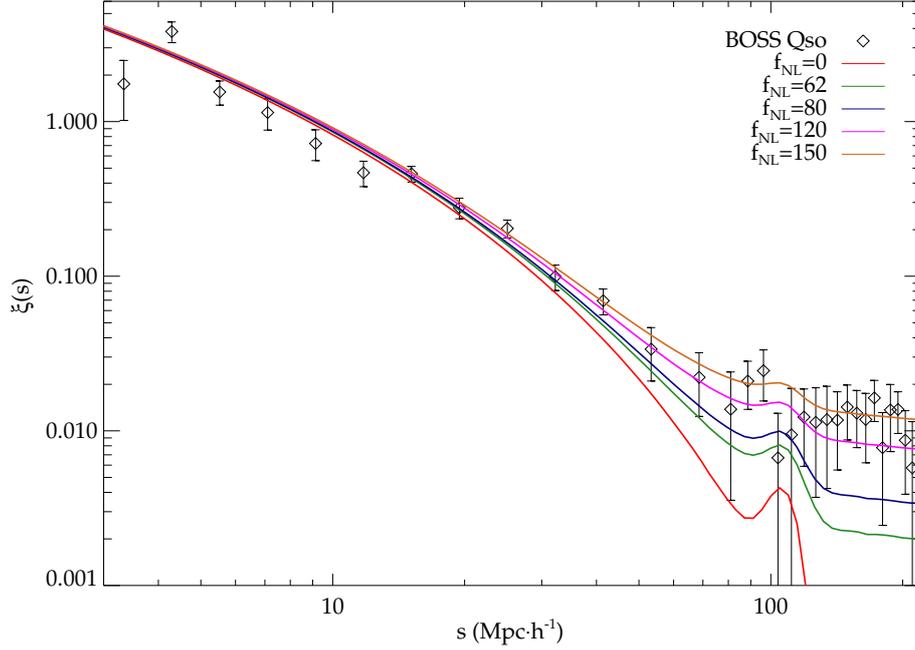
After fitting the model to the measured correlation function for scales  $3 < s < 130 h^{-1}\text{Mpc}$  and taking into account for the Kaiser effect, we measure the best-fit bias to be,  $b = 3.7 \pm 0.1$  with  $\chi_{red}^2 = 2.57$ . For the whole scale range,  $3 < s < 210 h^{-1}\text{Mpc}$ , the measured bias is  $b = 3.65 \pm 0.1$  with  $\chi_{red}^2 = 3.84$ . The difference between the two measured best fit biases is negligible. The full scale measured bias has a larger  $\chi_{red}^2$  value. As we can see from Fig. 10(a), the  $\Lambda\text{CDM}$  model fits well to the data till  $50 h^{-1}\text{Mpc}$  and not to the whole scale range. In addition to this the statistical uncertainties of the data at scales larger than  $100 h^{-1}\text{Mpc}$  lead to a bigger  $\chi_{red}^2$  value on the best-fit measured bias. On the other hand the value of  $\chi_{red}^2$  in the case where the bias comes from the best-fit of  $\Lambda\text{CDM}$  model on the data, at range  $s < 100 h^{-1}\text{Mpc}$ , is smaller since the standard model fits well on these small scales.

The measured best-fit bias for the Gaussian  $\Lambda\text{CDM}$  model is in good agreement with the one measured by [135],  $b = 3.8 \pm 0.3$ . Since we calculated the Gaussian bias for the quasar sample, we can now calculate non-Gaussian models that can fit the large scale excess of our sample, putting in this way constraints on the  $f_{NL}$  parameter.

An analytical description of the way quasars occupy the dark matter halos is mentioned in [135, 123]. In order to calculate an effective bias, we need to define a HOD model as well as to measure its best fit parameters. However, the difference between the bias and the effective bias of equation (82) is very small and hence we will not use a HOD model to measure the bias of the quasar sample. Instead, we will use the best fit bias measured above.

### 5.3 Test for non-Gaussianity

The main purpose of this work is to use scale dependent non-Gaussian models to fit to the BOSS quasar sample putting constraints on the amplitude of primordial non-Gaussianity. As we analyzed in a previous section, the existence of primordial non-



**Figure 11:** The clustering results of the BOSS quasar sample. In red colour is the  $\Lambda$ CDM best-fit model. We have also plotted 4 non-Gaussian model with different values of the  $f_{NL}$  parameter. It is obvious that these models can fit well for the large scale excess of the quasar correlation function.

Gaussianity leads to a non-Gaussian mass function of dark matter halos (equation (72)). If the amplitude of non-Gaussianity is non-zero the changed mass function will lead to higher mass halos and therefore to higher biased objects tracing them. As a result the non-Gaussian bias has an extra term,  $\Delta b_{\text{ng}} = (b - 1)f_{NL}^{\text{loc}}A(k)$ , where  $b$  is the Gaussian bias and  $A(k)$  is a scale dependent parameter given by equation (81).

The extra term in the bias is scale dependent therefore it will produce a large scales excess in clustering of our model. Hence, we will fit a non-Gaussian model to our data and especially to the large scales, where  $f_{NL}$  constraints can come from. We will use the correlation function originating from the simple  $k^n$  spectrum as defined in the previous section and we will use now the non-Gaussian bias,  $b_{\text{ng}} = b + \Delta b_{\text{ng}}$ . Hence if we also

take into consideration the Kaiser effect, since we have assumed linear theory to create our model correlation function, the non-Gaussian model will be  $\xi_{\text{ng}}(s) = (b_{\text{ng}}^2 + 2/3b_{\text{ng}}f + f^2/5)\xi_{\text{lin}}$ . We plot the original clustering results from the quasar sample together with the  $\Lambda$ CDM model and some non-Gaussian models with different values of  $f_{NL}$ .

The fitted  $\Lambda$ CDM model fits well to our data till the  $30 h^{-1}\text{Mpc}$ , as we can see in Fig. 11. At larger scales the standard model fails since it goes to zero too fast while a flattening is observed in the clustering of the data after the  $110 h^{-1}\text{Mpc}$ . The BAO peak can be seen at  $97 h^{-1}\text{Mpc}$ , where it is expected at  $\sim 105 h^{-1}\text{Mpc}$ . Although the peak is just a  $1\sigma$  detection, we can not omit the fact that the models' BAO peak is at the expected position while the peak of the data, if we accept that it is real, is not. Nevertheless the non-Gaussian models for the different  $f_{NL}$  parameters, due to the scale dependent bias can fit well to the large scale flattening observed in the sample, making them more consistent with the data. Therefore, a best-fit  $f_{NL}$  can be measured by fitting the non-Gaussian models to the observed large scale flattening of the the BOSS quasar sample correlation function. All the non-Gaussian models have been calculated at the effective redshift calculated in the previous section,  $z_{\text{eff}} = 2.4$ , and by using the value of the best-fit linear bias,  $b_{\text{lin}} = 3.7$ .

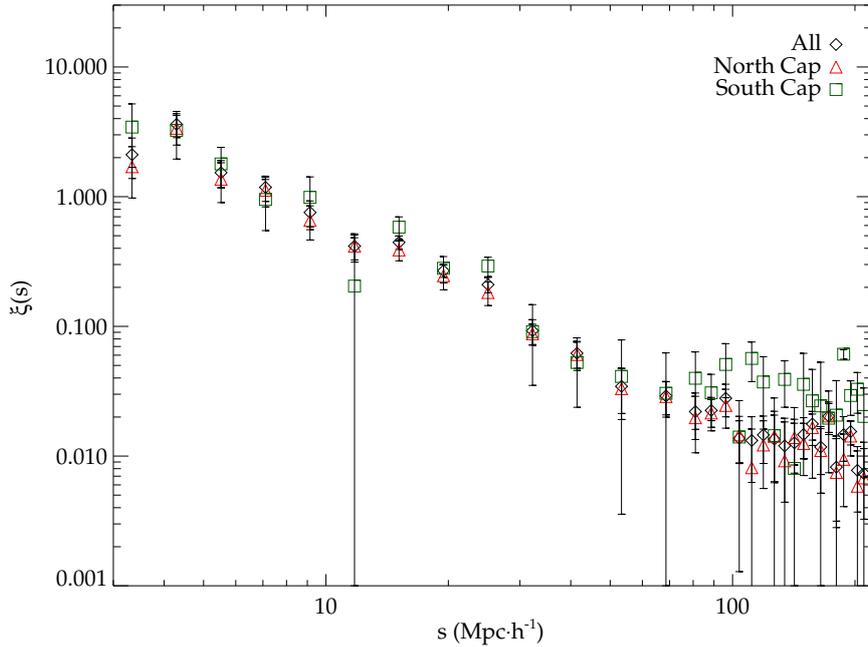
We use the full covariance matrix defined above together with the minimization of the  $\chi^2$  test defined from equation (92), in order to calculate the best-fit  $f_{NL}^{\text{loc}}$  parameter. The resulting value is,  $f_{NL}^{\text{loc}} = 135 \pm 9$  at  $1\sigma$  CL, with  $\chi_{\text{red}}^2 = 1.21$ . The value of the reduced  $\chi^2$  indicates that the model is a good fit to the data for the whole scale range and hence the constraints can be taken into consideration. It is evident from Fig 11 that the non-Gaussian models fit better than the  $\Lambda$ CDM standard model and especially at large scales, since the first has a scale dependent correction term in the bias. Although the uncertainty limits of the quasar clustering does not exclude the  $\Lambda$ CDM model since it can fit well for smaller scales. The constraints put on the amplitude of non-Gaussianity are tight in the sense that the values of  $f_{NL}$  lie in the range,  $117 < f_{NL}^{\text{loc}} < 153$  at 95%

confidence level.

The number of free parameters play a crucial role in the  $\chi^2$  goodness-of-fit test and hence in the best-fit parameter and its variation. If we allow the Gaussian bias as a free parameter, the  $\chi^2$  test will give different  $f_{NL}$  with larger errors, since the smaller the Gaussian bias the larger the  $f_{NL}$  must be to fit our data and vice versa. This leads to bigger uncertainty of the  $f_{NL}$  value measured from this sample and hence weaker constraints. The same would happen if we calculated the bias from a HOD model, where the plethora of different values of the fitting free HOD parameters would give a bigger number of best-fitting combinations of the bias and the  $f_{NL}$ , leading to higher uncertainty in the amplitude of non-Gaussianity, as measured from our sample. However, we calculated the Gaussian bias from the best fit of the  $\Lambda$ CDM model with the data and hence the constraints in the measured  $f_{NL}$  are tight, since all the free parameters are well defined.

Comparing our findings with the ones measured in [91],  $-29 < f_{NL}^{loc} < 70$ , and in [96],  $5 < f_{NL}^{loc} < 84$ , we find that the measured value of  $f_{NL}^{loc}$  from the BOSS quasar sample is not consistent with the values measured in the above studies. On the other hand our result is consistent with the less constrained values measured in [114, 147],  $-92 < f_{NL}^{loc} < 398$  and  $-268 < f_{NL}^{loc} < 164$  respectively and in [143],  $30 < f_{NL}^{loc} < 150$ . All the above  $f_{NL}$  values are at 95% significance level.

The above clustering studies are at lower redshift than our sample and use data from LRGs, less biased objects than quasars. These could explain to a point the difference in the  $f_{NL}$  values measured in all these studies. In addition to that, we would like to test our sample for systematic errors that affect the clustering signal and hence the amplitude of non-Gaussianity measured and constrained from it. As we referred in the beginning of the analysis, LSS surveys include systematic errors which are difficult to measure and correct. These errors affect the clustering signal and therefore the sensitive region of large scales, where the non-Gaussian constraints mainly comes from. Correcting for the systematic errors will directly change the value of, and constraints on,  $f_{NL}$ .



**Figure 12:** The two point correlation function from the hemisphere-split quasar sample, given that the southern sample is much smaller than the northern sample, is compared with the correlation function of the fiducial sample. The excess at large scales in the South Cap sample clustering can be easily distinguished.

## 5.4 Check for systematic errors

Systematic errors can affect the large scale clustering measurements and therefore the measured  $f_{NL}$  parameter. Hence it is important to search and reduce any potential sources of systematics that can affect our measurements, acquiring robust constraints on  $f_{NL}$ .

We begin checking the change in clustering by changing the amount of completeness as described in the previous section. Then jackknife tests are performed by applying different cuts in our sample. We apply cuts in the north and south galactic hemisphere, g-band extinction, seeing and sky brightness. In this way, we divide our sample into

two parts and after measuring the correlation function in each of them, we perform a jackknife re-sampling technique to define the covariance matrix in each part. The two resulting correlation functions for each different cut are compared to determine if any of the cuts applied can affect the large scale clustering. No statistically significant difference was detected in any of the above mentioned cuts, apart from the north and south galactic hemisphere cut. In Fig. 12, we can see that the south galactic hemisphere quasars have a stronger clustering signal than the ones in the north galactic cap. These findings agree with the findings of [135] and as they also explain why we need a larger number of quasars in our sample to determine whether this excess is real or the product of statistical fluctuations.

The different cuts do not affect significantly the clustering of our sample, even the South galactic excess due to small size of the South Cap sample cannot affect the clustering of the total sample. In Fig. 12, the correlation function of the whole sample do not differ significantly from the correlation function of the North hemisphere sample.

Now we will follow the systematic uncertainty check of [102, 148, 149] for the robustness of our 3D correlation results, which is a widely used method in the literature. In particular we will test if the clustering of the sample is affected by the galactic extinction, seeing, sky brightness and star density.

In order to do that, we will use HEALPix [150] maps to calculate the auto and cross-correlation of the quasar and the potential systematics. We use  $N_{side} = 64$ , which creates roughly pixels of  $\sim 0.84 \text{ deg}^2$  all over the sky. We split also our sample in redshift bins of  $\Delta z = 0.03$  and after we apply the angular mask of the quasar sample, we define in each created pixel  $i$  and redshift shell the overdensity

$$\delta_{i,z} = \frac{x_{i,z}}{\bar{x}_z} - 1 \quad (94)$$

where  $x_{i,z}$  is the value of the systematic in pixel  $i$  and redshift slice  $z$ , e.g. the mean value of extinction of quasars in pixel  $i$ . The  $\bar{x}_z$  is the mean value of the quantity in

question over all pixels at redshift bin  $z$ . We can define, by using the pixel overdensity, the correlation function of the each systematic from

$$\xi(s) = \frac{\sum_{i,j,z1,z2} \delta_{i,z1} \delta_{j,z2} \Theta_{i,j,z1,z2}(s) N_1(z1) N_2(z2)}{\sum_{i,j,z1,z2} \Theta_{i,j,z1,z2}(s) N_1(z1) N_2(z2)} \quad (95)$$

where the sum is over the different pixels  $i,j$  at redshift slices  $z1$  and  $z2$ , which are the redshift bins of the pixel  $i$  and  $j$  respectively,  $\Theta_{i,j,z1,z2}(s)$  is 1 if the separation between two pixels is within the bin  $s \pm \delta s$  and 0 otherwise,  $N_2(z2)$  is the number of quasars in the redshift slice  $z2$ . In order to calculate the 3D cross-correlation of quasars with the systematics, which follow an angular distribution, we just keep the overdensity field of the systematics and the number density  $N(z)$  of the angular map constant with redshift.

This method for calculating the correlation function requires smaller size of pixels (e.g.  $N_{side} = 256$ ). The reason is that this method measures the correlation function from the overdensity of pixels (or cubes in the 3D case) which comes from the objects inside. If the pixel size is big then the overdensity will average out information from the clustering of objects inside the pixel, since this method calculates the correlation function from the pixels and not from each object individually. The pixelization method leads to a ‘smoothed’ correlation function. In order to avoid this loss of information one has to choose very small pixels.

In our case, we chose almost a square degree pixel and a redshift slice of  $\Delta z = 0.03$ . The reason for the choice of these values is that the BOSS quasar sample is a low number density sample, in  $10,000 \text{ deg}^2$  we have  $\sim 22,000$  objects which means that we roughly have  $\sim 2 \text{ qso/deg}^2$ . Therefore, in order to have enough pixels with overdensity coming from more than one object, we had to use big enough pixels which can contain more objects.

In order to define the systematic effect caused to the clustering of the quasars from an observational parameter (e.g. galactic extinction, foreground stars, etc.), we will follow [148, 149], where they describe analytically the fluctuation that systematics may cause to

the large scales clustering. To the first order the overdensity field is given by

$$\delta_{obs} = \delta_t + \sum_i \epsilon_i \delta_i \quad (96)$$

where  $\delta_{obs}$  is the observed overdensity of the quasar sample,  $\delta_t$  is the true overdensity,  $\delta_i$  is the overdensity of the  $i$ th systematic and finally  $\epsilon_i$  is the effect of each potential systematic on the observed overdensity. From equation (95) and (96), we can define the true correlation function of our sample after correcting for the potential systematics

$$\xi_t(s) = \xi_{obs}(s) - \sum_i \epsilon_i^2 \xi_i(s) - \sum_{i,j>i} 2\epsilon_i \epsilon_j \xi_{i,j}(s) \quad (97)$$

where the  $\xi_i(s)$  is the auto-correlation of systematic  $i$  and  $\xi_{i,j}(s)$  is the auto/cross-correlation of the different potential systematics. In the Appendix of [148], they describe the case where they assume 3 different systematic errors. In our case we will calculate separately the effect of each systematic, which means that the cross-correlation terms between the different systematics will be zero. Hence, we just need to calculate  $\epsilon_i$  of equation (97) and subtract it from the observed correlation function separately for each systematic  $i$ . The correlation function corrected for each systematic at a time will be

$$\xi_t(s) = \xi_{obs}(s) - \frac{\xi_{q,i}^2(s)}{\xi_i(s)} \quad (98)$$

where  $\xi_{q,i}$  is the cross-correlation of the systematic  $i$  with the quasar sample. We will test our sample for 4 different potential systematics. The first three are the observable parameters of g-band extinction, seeing and sky brightness. The final systematic is the effect of the foreground stars.

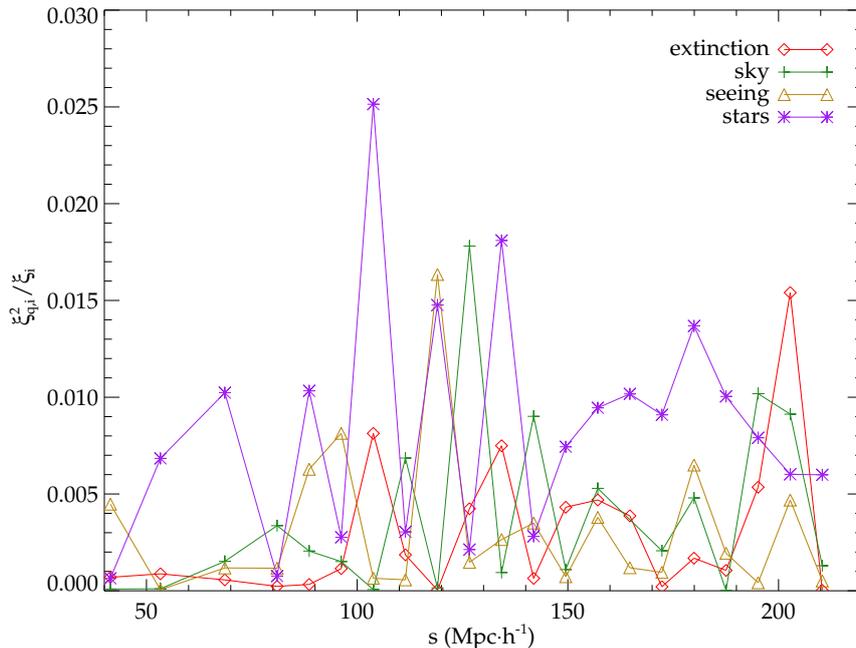
It was found in [148] that the number density of galaxies drops 10% from regions of high to regions of low stellar density. 3% is caused, as stated in [102, 148], from the fact that galaxies close to stars are not easily detectable. The other 7% may come from the change in the photometric pipeline from DR7 to DR8. This change leads to the difficulty

of the deblending code to separate more than 25 overlapping objects in regions of high stellar density. Since our data come from the SDSS BOSS DR9 they will also be affected by this effect and hence we will have to correct our sample for this systematic. In [102, 148] they both find that, the systematic effect from the foreground stars is the most important one, and to correct it they use a method where they apply weights to the sample. These weights correct the sample at a point where no other weights from other systematics need to be applied. Here we will only use the simple correlation function method to define the amount of correction from each of the systematic errors.

Besides these systematic effects, stellar contamination can also affect our sample. The XDQSO technique that we have applied here, uses models of quasar and star distributions in flux space to minimize the stellar contamination. However, this is a very difficult task since the stellar loci crosses the redshift  $z = 2.7$  quasars in the flux space. The quasar selection technique, as described in [134], performs well for mid-redshift quasars ( $2.2 < z < 3.5$ ) like the ones in our sample. Hence we will not try to correct for any stellar contamination in this work, although a study for measuring and correcting for these contaminants would be interesting. Detailed description on the correction of the star contamination in a quasar sample can be found in [124].

To check for the effect of the foreground stars we choose stars in the magnitude limit,  $17.5 < i_{mod} < 22.5$ , which are the selection limits of the XDQSO selection technique. We also apply the angular mask of BOSS quasar sample. The resulting correcting ratios,  $\xi_{q,i}^2(s)/\xi_i(s)$ , for the four systematics considered in this work are plotted in Fig. 13.

In Fig. 13 we can see that the ratio that corrects the correlation function for the different potential systematics is small for all the systematics except the foreground stars. Besides some peaks the effect of extinction, sky and seeing is smaller than that of the stars. This makes the effect from the presence of stars to be the larger systematic, with the highest correlation function correction, in agreement with [102, 148]. The ratio from the foreground stars is by far the highest on large scales, where mainly the non-Gaussianity

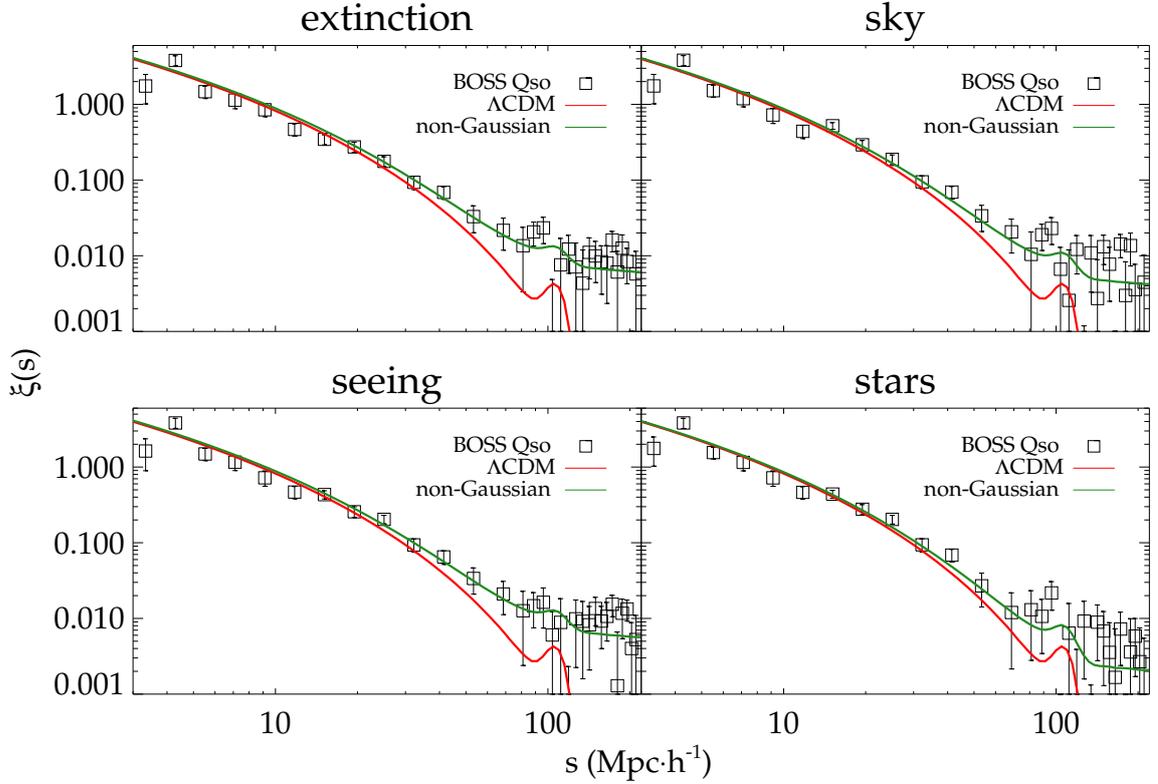


**Figure 13:** The ratio of the squared cross-correlation  $\xi_{q,i}(s)$  of the quasars and systematics with the auto-correlation  $\xi_i(s)$  of the systematic, for the galactic extinction, sky brightness, seeing and the effect of foreground stars.

constraints originate. Second comes the systematic from sky, while last is the effect of extinction and seeing. According to Fig. 13, we expect a maximum reduction of 0.005 – 0.01 of the measured correlation function by correcting for the systematics.

The main reason of the systematic error tests is to check the sample for potential effects that can affect the clustering of the large scales and hence the  $f_{NL}$  value measured by the best-fit non-Gaussian models. Since each systematic will affect differently the correlation function and therefore the  $f_{NL}$  value, we will apply the correction for each systematic separately. The resulting correlation function for each systematic is presented, together with the  $\Lambda$ CDM and the best-fit non-Gaussian model, in Fig. 14. We concentrate the best fitting value of  $f_{NL}$  and  $\chi^2$  results for all the four systematics in Table 1.

After the correction for the systematics, the clustering of the data at large scales



**Figure 14:** The correlation function of the BOSS quasars after we correct it for each potential systematic. We plot the  $\Lambda$ CDM standard model together with the best-fit  $f_{NL}^{loc}$  non-Gaussian model on each corrected correlation function. In all cases the non-Gaussian model is the best fit, especially at the large scales. Although the statistical uncertainty is such that it does not permit to us to exclude the standard cosmological model.

has been reduced. The effect from the foreground stars affects our sample the most. Correcting for this systematic will reduce the clustering more than any other systematic, as we expected from the results of Fig. 13. Since the large scale clustering has been reduced the best-fit value of  $f_{NL}$  will be also reduced. The maximum reduction of course

Systematics	$f_{NL}^{loc} (1\sigma)$	$\chi_{red}^2$
Extinction	$107 \pm 10$	1.66
Sky	$89 \pm 12$	3.23
Seeing	$103 \pm 11$	3.1
Stars	$63 \pm 16$	1.89

**Table 1:** The best-fit  $f_{NL}$  parameter after correcting the correlation function of the sample for each systematic error separately. The value of the reduced  $\chi^2$ -test is also presented.

happens when we correct for the stars, after that is the reduction caused by correcting for the sky brightness. The measured  $f_{NL}$  parameter after the correction of extinction and seeing are not significantly reduced compared to our fiducial results ( $f_{NL}^{loc} = 134 \pm 9 (1\sigma)$ ).

The reduction in the clustering from the stars is such that the best-fit  $f_{NL}$  measured in this case,  $31 < f_{NL}^{loc} < 95$  at 95% CL, is now consistent with the results from other LSS surveys of, [91, 96],  $-29 < f_{NL}^{loc} < 70$  and  $5 < f_{NL}^{loc} < 84$  respectively (both at  $2\sigma$ ), as well as from [143]  $f_{NL}^{loc} = 90 \pm 30$ . Our result after the stellar correction is consistent within the uncertainty limits with the results from WMAP7 [65],  $f_{NL}^{loc} = 32 \pm 21 (1\sigma)$ . Even after we corrected for the systematic errors, the large scale clustering did not reduce to a level where the non-Gaussian models are excluded. On the contrary the  $f_{NL}$  parameter was constrained and measured as being consistent with previous results of former studies. Nevertheless we cannot exclude the standard  $f_{NL} = 0$  model since as we can see the systematic errors significantly affect the large scales of the quasar clustering, which is the main source of constraining  $f_{NL}$ .

The precision of the measured value of non-Gaussianity depends on the validity of the tests for systematics. Hence the data must be examined more thoroughly and the systematics must be corrected without under estimating them, in order to obtain tight constraints and consistent values for  $f_{NL}$ . It would be interesting to apply the weight

method to correct the data from the major systematics and compare the results with the ones obtained in this work. But even if we do not fully accept the systematic reduced results, because of the simplicity of the method used here, the basic measurements show enough excess to leave room for non-Gaussian models.

## 5.5 Conclusions and Summary

We have measured the two-point correlation function  $\xi(s)$  in redshift space for  $\sim 22,000$  quasars from the SDSS BOSS DR9 in the range of  $2.2 < z < 2.9$ . The selection of the quasars is based on the XDQSO method, where the distribution of quasars and stars in flux space assigns a probability for all the point sources of SDSS BOSS for being a quasar or star. We measure the clustering of the quasar sample in order to put constraints on the amplitude of primordial non-Gaussianity. To calculate the correlation function of the sample we use the Landy & Szalay estimator and therefore create a random catalogue  $\sim 20$  times larger than our sample.

On the correlation function of the quasars we fit a  $\Lambda$ CDM model in order to measure the best-fit linear bias, which is  $b_{lin} = 3.7 \pm 0.1$ . The standard cosmological model fit well to our data till the  $40 h^{-1}\text{Mpc}$  and fails to explain the excess of the correlation function at the large scales. To put constraints on non-Gaussianity, we create a model with a non-Gaussian bias, which originates from the fact that primordial non-Gaussianity affects the mass function of the dark matter halos and therefore a scale dependent correction term is added to the linear bias. The effect of this non-Gaussian term in the bias is measured by the  $f_{NL}^{loc}$  parameter which at the same time measures the amplitude of primordial non-Gaussianity in the local regime. The non-Gaussian models can fit in the large scale excess of our data sample and therefore measure the best-fit  $f_{NL}^{loc}$  parameter.

The best-fitting value measured from the quasar sample is  $f_{NL}^{loc} = 134 \pm 9 (1\sigma)$ , which is consistent with previous measurement from other LSS clustering studies where the constraints on the parameter are loose. On the other hand, studies [91, 96] with more

constrained values of the  $f_{NL}$ , do not include our measured result. The big disadvantage of measuring cosmological parameters from LSS surveys is the fact that systematic errors can affect the large scale clustering of the data sample. Hence, we test our sample for a series of potential systematics in order to determine whether or not the excess at large scales, on which the measured best-fit value of  $f_{NL}$  mostly depends on, is real or the result of systematic uncertainties. We test and correct the sample for galactic extinction, sky brightness, seeing and the effect of foreground objects separately. The largest reduction to the large scales clustering signal comes from the presence of foreground stars, in total agreement with [102, 148]. The resulting best-fit value of the  $f_{NL}^{loc}$  after the correction for each systematic error are shown on Table 1. The biggest reduction on the fiducial value comes after the correction of the effect of the foreground stars,  $f_{NL}^{loc} = 63 \pm 16 (1\sigma)$ . This value is consistent with the results of the studies with tight constraints on the  $f_{NL}$  value.

The correction of systematic errors must be investigated further for this sample, by applying the weights method for correcting the effects of systematics. The method we followed here is simple and can overestimate the effect of each systematic in the large scales of the clustering results. While the raw data show some preference for non-Gaussianity, the  $\Lambda$ CDM model cannot currently be excluded, because of statistical and systematic uncertainties. Further analysis of the systematic errors in particular is needed to determine finally the value of  $f_{NL}$  implied by these data. Future quasar surveys will measure non-Gaussianity to even higher accuracy, by checking for excess power in the correlation function at large scales.

## Chapter 6

### *Looking into the future: The 2QDES survey*

The BOSS DR9 quasar sample gave some tight constraints on the  $f_{NL}$  value, leaving room for non-Gaussian effects in the clustering of high biased objects. Such objects at high redshifts can provide stringent constraints on  $f_{NL}$ , since the results of non-Gaussianity can be easier detected in their large scale clustering due to the scale dependent bias of the fitted non-Gaussian models. Therefore, tighter constraints can come from future high redshift surveys of strongly biased objects such as quasars. One very promising future quasar survey, which is going to obtain robust constraints on  $f_{NL}^{loc}$ , is the 2dF Quasar Dark Energy Survey (2QDES) of the VLT Survey Telescope (VST) ATLAS.

The VST is a 2.6 m wide field optical survey telescope in the southern hemisphere and in particular at ESO's platform in Cerro Paranal, Chile. The telescope is equipped with a  $16k \times 16k$  pixel CCD camera, the OmegaCAM. One of the three planned public surveys is the VST ATLAS survey, where its main objective is to reach a depth compatible with those of SDSS in the southern hemisphere. The VST ATLAS will scan  $4,500 \text{ deg}^2$  of the southern sky in the u, g, r, i, z bands of SDSS. The original aim was to measure the baryonic wiggles in the clustering of LRGs in order to constrain the dark energy equation of state.

One of the very interesting surveys is the Two-Degree Field Quasar Dark Energy Survey (2QDES), which will be a two degrees field quasar survey. The survey will use the 2df AAOmega spectrograph [151] of the 3.9m Anglo-Australian Telescope (AAT), together with the results of VST ATLAS to detect up to  $\sim 500,000$  quasars at  $z < 2.2$  in the southern sky. The fiber density of the AAOmega spectrograph, used to measure

the redshift of the quasars, is  $\sim 110 \text{ deg}^{-2}$ . Hence to achieve the detection of the desired number of quasars up to the magnitude limit of  $g = 22.5$  over  $4,500 \text{ deg}^2$ , a  $\approx 200$  nights survey is needed. The selection of quasars is achieved by the simple ugr and gri colour-colour diagram selection technique as well as with the XDQSO method of [134] analyzed in the previous section, which is the same selection process applied in this work and [135] to construct the BOSS quasar sample.

The main scientific goals of the 2QDES quasar survey is to measure the position of the BAO peak and the gravitational growth rate from the clustering of  $z \sim 1.6$  quasars. The BAO peak position and the gravitational growth rate can be used to define the equation of state of the dark energy and constrain crucial cosmological parameters. In addition to this a large quasars survey, like the 2QDES, can provide robust constraints on the amplitude of primordial non-Gaussianity from the large scale clustering of the quasars. Such a survey can provide high redshift,  $z < 2.2$ , quasars whose correlation function can give tight constraints on the  $f_{NL}$  as the BOSS quasar survey. In fact the 2QDES survey was designed to compete with the depth and numbers of the BOSS quasar survey and create a large quasar sample in the south hemisphere. A comparison of the 2QDES clustering results with the results of the BOSS quasars of this work would be interesting for the value and constraints of non-Gaussianity, since both sample are consisting of quasars and their redshift range is complementary. In addition to this, a large quasar sample in the southern hemisphere could provide us with vital information on the nature of the clustering excess observed in the correlation function of the southern hemisphere BOSS quasar over the quasars located at the northern hemisphere (Fig. 11). The combination of the 2QDES and the full BOSS DR9 quasar sample can provide us with stringent constraints on non-Gaussianity and other important cosmological parameters, like the gravitational growth factor and the parameters of the dark energy equation of state.

# Chapter 7

## *Conclusions*

The origin of observable structures is one of the most interesting subjects in cosmology. Fluctuations of the energy density field in the very early Universe can produce, from Einstein's field equations, perturbations in the primordial gravitational potential. These gravitational fluctuations will create over-densities of matter and more specifically dark matter since it constitutes roughly  $\sim 85\%$  of matter in the Universe. The self gravity of these concentrations will lead the ones located on high fluctuation peaks to collapse into bound virialized objects called dark matter halos. Due to the gravitational interaction of dark matter and ordinary (baryonic) matter, the halos will become the hosts for the creation of galaxies. The hierarchical formation of structure dictates that more massive halos will attract and eventually merge with less massive ones creating mergers, clusters and superclusters of galaxies.

The most dominant and acceptable theory for the early universe, which naturally predicts the origin of structures, is the inflationary paradigm. Inflation is an era of exponential acceleration expansion of the very early Universe. The expansion is driven by a scalar field, called the inflaton, slowly rolling down its own energy potential. The inflaton field is a quantum field and therefore quantum fluctuations, due to the uncertainty principle, can occur during the scalar field's slow roll phase. The fluctuation generated from the scalar field due to their quantum nature follow Gaussian statistics. The random fields created from the quantum perturbations of the inflaton are tiny and their life time short, but when they manifest during inflation they can grow to scales larger than the horizon as the expansion of the Universe accelerates, where they grow due to gravitational instabilities. When the energy fluctuation of the scalar field reenter the observable Universe, they will seed the structures we observe today.

The natural way inflationary scenarios explain the origin of structures is the most important aspect that has made inflation the dominant theory describing the early Universe. However, the plethora of inflationary theories makes it difficult to construct a robust theory for the beginning of the Universe. One of the most powerful observational probes that will help us decide upon the different inflation models is the observational measurement of the amplitude of non-Gaussianity in the primordial perturbation field. Although the fluctuations predicted by the simplest inflation scenarios follow Gaussian statistics, there are models that allow a deviation from Gaussian initial conditions. Different inflation models predict different amounts of primordial non-Gaussianity and therefore it is crucial to measure its value.

The two most important observational probes for non-Gaussianity is the anisotropy of CMB and the clustering of LSS. In this work we focused on constraining  $f_{NL}$ , the parameter measuring the amplitude of primordial non-Gaussianity, from LSS surveys. The deviation from Gaussian statistics in the primordial fluctuation field affects directly the mass function of the dark matter halos, since halos are created in the high sigma peaks of the energy perturbation field. The non-Gaussian mass function of halos will lead to a non-Gaussian bias between the dark matter distribution and the distribution of galaxies. This corrected for non-Gaussianity bias is scale dependent and therefore we can check the large scale clustering results of high biased objects, where such effects will be more prevalent, in order to measure and constrain  $f_{NL}$ .

We test the clustering results of promising high redshift surveys in order to find if non-Gaussianity can be measured from LSS surveys. We present the results of [96], where they measure  $f_{NL}^{loc} = 62 \pm 27$  ( $1\sigma$ ) from the clustering of extragalactic radio sources of the NVSS sample at  $z \sim 1$ . This results in one of the tightest constraints on  $f_{NL}$  coming from LSS surveys. Then we present the correlation function of the SDSS BOSS CMASS sample, where the sample consists of LRGs at  $z \sim 0.55$ . We calculate non-Gaussian models for different  $f_{NL}$  parameter values in order to fit for the large scales of the

correlation function. Due to the uncertainties in the large scale clustering of the sample, we can neither exclude the standard cosmological model nor the models incorporating non-Gaussianity. In [114] they measure and weakly constrain the amplitude of primordial non-Gaussianity,  $-92 < f_{NL} < 389$  ( $2\sigma$ ). We present also the angular correlation function of H-*alpha* emitters from the HiZELS sample. We use a power law as the standard model to fit the clustering results, where because of the Limber's formula approximation breaks down due to the narrow redshift selection of the sample, we can observe a deviation at large scales. A non-Gaussian model may be able to fit for such deviations, measuring the best-fit  $f_{NL}$  parameter.

Finally we measure the correlation function of the SDSS BOSS DR9 quasar sample, which consists of  $\sim 22,000$  quasars in  $2.2 < z < 2.9$ . High redshift and high biased objects, like quasars, are very promising candidates for measuring non-Gaussianity since any primordial non-Gaussian effects will be more evident in their clustering results. The measured best-fit  $f_{NL}^{loc} = 134 \pm 9$  ( $1\sigma$ ) is consistent with the weakly constrained results coming from other LSS clustering studies, but it is higher than the accepted amount coming from surveys that provided robust and stringent constraints on  $f_{NL}^{loc}$ . The  $\Lambda$ CDM model fits the data up to  $30 h^{-1}\text{Mpc}$ , but we cannot exclude it due to the large scale uncertainties, although it is obvious that the non-Gaussian models can fit the large scale flattening observed in the clustering of the quasar sample.

The big disadvantage of constraining non-Gaussianity from LSS surveys is the systematic errors that affect the large scale clustering, where the best-fit  $f_{NL}$  measurements mainly come from. Hence, we need to check for any potential systematic errors that can affect our clustering results and therefore the measured value of  $f_{NL}$ . After correcting for four potential systematics, we observe that the biggest reduction comes from the effect of foreground stars, in agreement with previous studies on the SDSS BOSS CMASS sample. The new best-fit measured  $f_{NL} = 63 \pm 16$  is in agreement with the results of previous studies providing the tightest non-Gaussianity constraints. However, the sim-

---

plicity of the method used to correct the systematic uncertainties in our sample does not make the corrected  $f_{NL}$  results robust. The more sophisticated weight method, presented in [148], must be used in order to check if the correlation function is over-corrected or not. New LSS surveys of high biased objects, like the proposed 2QDES quasar survey in the southern hemisphere, will provide us with vital information upon the amplitude of primordial non-Gaussianities. The precise and constrained measurement of the  $f_{NL}$  parameter will give us the observational resources we need in order to decide upon the plethora of inflationary models and build a robust model that will describe the very early Universe.

# Bibliography

- [1] A. H. Guth, “Inflationary universe: A possible solution to the horizon and flatness problems,” *Phys. Rev. D*, vol. 23, pp. 347–356, 1981.
- [2] A. Linde, “A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems,” *Physics Letters B*, vol. 108, no. 6, pp. 389 – 393, 1982.
- [3] A. Albrecht and P. J. Steinhardt, “Cosmology for grand unified theories with radiatively induced symmetry breaking,” *Phys. Rev. Lett.*, vol. 48, pp. 1220–1223, 1982.
- [4] P. J. E. Peebles and B. Ratra, “The cosmological constant and dark energy,” *Rev. Mod. Phys.*, vol. 75, pp. 559–606, 2003.
- [5] V. F. Mukhanov and G. V. Chibisov, “Quantum fluctuations and a nonsingular universe,” *ZhETF Pis ma Redaktsiiu*, vol. 33, pp. 549–553, 1981.
- [6] S. W. Hawking, “The development of irregularities in a single bubble inflationary universe,” *Phys. Lett. B*, vol. 115, pp. 295–297, 1982.
- [7] A. A. Starobinsky, “Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations,” *Phys. Lett. B*, vol. 117, pp. 175–178, 1982.

- 
- [8] A. H. Guth and S.-Y. Pi, “Fluctuations in the new inflationary universe,” *Phys. Rev. Lett.*, vol. 49, pp. 1110–1113, 1982.
- [9] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, “Spontaneous creation of almost scale-free density perturbations in an inflationary universe,” *Phys. Rev. D*, vol. 28, pp. 679–693, 1983.
- [10] R. K. Sachs and A. M. Wolfe, “Perturbations of a Cosmological Model and Angular Variations of the Microwave Background,” *ApJ*, vol. 147, p. 73, 1967.
- [11] W. Hu and N. Sugiyama, “Small scale integrated Sachs-Wolfe effect,” *Phys. Rev. D*, vol. 50, pp. 627–631, July 1994.
- [12] V. Mukhanov, *Physical Foundations of Cosmology*. Cambridge University Press, 2005.
- [13] P. Coles and F. Lucchin, *Cosmology: The Origin and Evolution of Cosmic Structure*. John Wiley and Sons, 2002.
- [14] A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure*. Cambridge University Press, 2000.
- [15] P. J. E. Peebles, *Principles of physical cosmology*. Princeton University Press, 1993.
- [16] A. J. S. Hamilton, “Linear Redshift Distortions: a Review,” *The Evolving Universe*, vol. 231, p. 185, 1998.
- [17] S. D. Landy and A. S. Szalay, “Bias and variance of angular correlation functions,” *ApJ*, vol. 412, pp. 64–71, July 1993.
- [18] R. A. C. Croft, G. B. Dalton, G. Efstathiou, W. J. Sutherland, and S. J. Maddox, “The richness dependence of galaxy cluster correlations: results from a redshift survey of rich APM clusters,” *MNRAS*, vol. 291, pp. 305–313, 1997.

- [19] D. N. Limber, “The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II.,” *ApJ*, vol. 119, p. 655, 1954.
- [20] S. Phillipps, R. Fong, R. S. E. S. M. Fall, and H. T. MacGillivray, “Correlation analysis deep galaxy samples - 1. Techniques with applications to a two-colour sample,” *MNRAS*, vol. 182, pp. 673–686, Mar. 1978.
- [21] D. Langlois, “Lectures on Inflation and Cosmological Perturbations,” vol. 800, pp. 1–57, 2010.
- [22] A. Linde, “Particle Physics and Inflationary Cosmology,” *ArXiv High Energy Physics - Theory e-prints*, 2005.
- [23] R. H. Brandenberger, “Lectures on the Theory of Cosmological Perturbations,” vol. 646, pp. 127–167, 2004.
- [24] L. Amendola and S. Tsujikawa, *Dark Energy: Theory and Observations*. Cambridge University Press, 2010.
- [25] K. A. Malik and D. Wands, “Cosmological perturbations,” *Physics Reports*, vol. 475, pp. 1–51, 2009.
- [26] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant second-order perturbations and non-Gaussianity from inflation,” *Nuclear Physics B*, vol. 667, pp. 119–148, 2003.
- [27] J. Maldacena, “Non-gaussian features of primordial fluctuations in single field inflationary models,” *Journal of High Energy Physics*, vol. 5, p. 13, 2003.
- [28] E. Komatsu, N. Afshordi, N. Bartolo, D. Baumann, J. R. Bond, E. I. Buchbinder, C. T. Byrnes, X. Chen, D. J. H. Chung, A. Cooray, P. Creminelli, N. Dalal, O. Dore, R. Easther, A. V. Frolov, J. Khoury, W. H. Kinney, L. Kofman, K. Koyama, L. Leblond, J.-L. Lehners, J. E. Lidsey, M. Liguori, E. A. Lim, A. Linde, D. H. Lyth,

- J. Maldacena, S. Matarrese, L. McAllister, P. McDonald, S. Mukohyama, B. Ovrut, H. V. Peiris, A. Riotto, Y. Rodrigues, M. Sasaki, R. Scoccimarro, D. Seery, A. Se-fusatti, K. M. Smith, A. A. Starobinsky, P. J. Steinhardt, F. Takahashi, M. Tegmark, A. J. Tolley, L. Verde, B. D. Wandelt, D. Wands, S. Weinberg, M. Wyman, A. P. S. Yadav, and M. Zaldarriaga, “Non-Gaussianity as a Probe of the Physics of the Pri-mordial Universe and the Astrophysics of the Low Redshift Universe,” *astro2010: The Astronomy and Astrophysics Decadal Survey*, vol. 2010, p. 158, 2009.
- [29] X. Chen, R. Easther, and E. A. Lim, “Large non-Gaussianities in single-field infla-tion,” *Journal of Cosmology and Astro-Particle Physics*, vol. 6, p. 23, June 2007.
- [30] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, “Conditions for large non-Gaussianity in two-field slow-roll inflation,” *Journal of Cosmology and Astro-Particle Physics*, vol. 10, p. 8, Oct. 2008.
- [31] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, “Large non-Gaussianity from two-component hybrid inflation,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2, p. 17, Feb. 2009.
- [32] C. T. Byrnes and G. Tasinato, “Non-Gaussianity beyond slow roll in multi-field inflation,” *Journal of Cosmology and Astro-Particle Physics*, vol. 8, p. 16, Aug. 2009.
- [33] C. T. Byrnes and K.-Y. Choi, “Review of Local Non-Gaussianity from Multifield Inflation,” *Advances in Astronomy*, vol. 2010, 2010.
- [34] X. Chen, M. x. Huang, S. Kachru, and G. Shiu, “Observational signatures and non-Gaussianities of general single-field inflation,” *Journal of Cosmology and Astro-Particle Physics*, vol. 1, p. 2, Jan. 2007.
- [35] X. Chen, “Inflation from warped space,” *Journal of High Energy Physics*, vol. 8, p. 45, Aug. 2005.

- [36] R. Bean, X. Chen, H. Peiris, and J. Xu, “Comparing infrared Dirac-Born-Infeld brane inflation to observations,” *Phys. Rev. D*, vol. 77, p. 023527, Jan. 2008.
- [37] V. F. Mukhanov and P. J. Steinhardt, “Density perturbations in multifield inflationary models,” *Phys. Lett. B*, vol. 422, pp. 52–60, Mar. 1998.
- [38] D. Polarski and A. A. Starobinsky, “Isocurvature perturbations in multiple inflationary models,” *Phys. Rev. D*, vol. 50, pp. 6123–6129, 1994.
- [39] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, “Large non-Gaussianity in multiple-field inflation,” *Phys. Rev. D*, vol. 73, p. 083522, Apr. 2006.
- [40] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity in the curvaton scenario,” *Phys. Rev. D*, vol. 69, no. 4, p. 043503, 2004.
- [41] M. Sasaki, J. Väliviita, and D. Wands, “Non-Gaussianity of the primordial perturbation in the curvaton model,” *Phys. Rev. D*, vol. 74, p. 103003, Nov. 2006.
- [42] D. H. Lyth and D. Wands, “Generating the curvature perturbation without an inflaton,” *Phys. Lett. B*, vol. 524, pp. 5–14, 2002.
- [43] J.-L. Lehners, “Ekpyrotic Nongaussianity: A Review,” *Advances in Astronomy*, vol. 2010, 2010.
- [44] J.-L. Lehners and P. J. Steinhardt, “Non-Gaussian density fluctuations from entropically generated curvature perturbations in ekpyrotic models,” *Phys. Rev. D*, vol. 77, p. 063533, Mar. 2008.
- [45] S. Yokoyama and J. Soda, “Primordial statistical anisotropy generated at the end of inflation,” *Journal of Cosmology and Astro-Particle Physics*, vol. 8, p. 5, 2008.
- [46] M. Karčiauskas, K. Dimopoulos, and D. H. Lyth, “Anisotropic non-Gaussianity from vector field perturbations,” *Phys. Rev. D*, vol. 80, no. 2, p. 023509, 2009.

- 
- [47] E. Dimastrogiovanni, N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity and Statistical Anisotropy from Vector Field Populated Inflationary Models,” *Advances in Astronomy*, vol. 2010, 2010.
- [48] X. Chen, “Primordial Non-Gaussianities from Inflation Models,” *Advances in Astronomy*, vol. 2010, 2010.
- [49] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation: theory and observations,” *Physics Reports*, vol. 402, pp. 103–266, 2004.
- [50] D. Babich, P. Creminelli, and M. Zaldarriaga, “The shape of non-Gaussianities,” *Journal of Cosmology and Astro-Particle Physics*, vol. 8, p. 9, 2004.
- [51] J. R. Fergusson and E. P. S. Shellard, “Shape of primordial non-Gaussianity and the CMB bispectrum,” *Phys. Rev. D*, vol. 80, p. 043510, Aug. 2009.
- [52] M. Lo Verde, A. Miller, S. Shandera, and L. Verde, “Effects of scale-dependent non-Gaussianity on cosmological structures,” *Journal of Cosmology and Astro-Particle Physics*, vol. 4, p. 14, Apr. 2008.
- [53] M. Alishahiha, E. Silverstein, and D. Tong, “DBI in the sky: Non-Gaussianity from inflation with a speed limit,” *Phys. Rev. D*, vol. 70, p. 123505, Dec. 2004.
- [54] E. Komatsu and D. N. Spergel, “Acoustic signatures in the primary microwave background bispectrum,” *Phys. Rev. D*, vol. 63, p. 063002, Mar. 2001.
- [55] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, “The three-point correlation function of the cosmic microwave background in inflationary models,” *ApJ*, vol. 430, pp. 447–457, Aug. 1994.
- [56] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, “Large-scale structure, the cosmic microwave background and primordial non-Gaussianity,” *MNRAS*, vol. 313, pp. 141–147, Mar. 2000.

- [57] O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees, “Dynamical effects of the cosmological constant,” *MNRAS*, vol. 251, pp. 128–136, 1991.
- [58] S. M. Carroll, W. H. Press, and E. L. Turner, “The cosmological constant,” *ARA&A*, vol. 30, pp. 499–542, 1992.
- [59] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, “Limits on non-Gaussianities from WMAP data,” *Journal of Cosmology and Astro-Particle Physics*, vol. 5, p. 4, May 2006.
- [60] P. D. Meerburg, J. P. van der Schaar, and P. Stefano Corasaniti, “Signatures of initial state modifications on bispectrum statistics,” *Journal of Cosmology and Astro-Particle Physics*, vol. 5, p. 18, May 2009.
- [61] P. Creminelli, L. Senatore, M. Zaldarriaga, and M. Tegmark, “Limits on  $f_{NL}$  parameters from Wilkinson Microwave Anisotropy Probe three-year data,” *Journal of Cosmology and Astro-Particle Physics*, vol. 3, p. 5, Mar. 2007.
- [62] E. Komatsu, J. Dunkley, M. R.olta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, “Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation,” *ApJS*, vol. 180, pp. 330–376, Feb. 2009.
- [63] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{NL}^{local}$  from WMAP 5-year data,” *Journal of Cosmology and Astro-Particle Physics*, vol. 9, p. 6, Sept. 2009.
- [64] A. Curto, E. Martínez-González, R. B. Barreiro, and M. P. Hobson, “Constraints on general primordial non-Gaussianity using wavelets for the Wilkinson Microwave Anisotropy Probe 7-year data,” *MNRAS*, vol. 417, pp. 488–494, Oct. 2011.

- [65] E. Komatsu, “Hunting for primordial non-Gaussianity in the cosmic microwave background,” *Classical and Quantum Gravity*, vol. 27, p. 124010, June 2010.
- [66] D. H. Lyth and Y. Rodríguez, “Inflationary Prediction for Primordial Non-Gaussianity,” *Phys. Rev. Lett.*, vol. 95, no. 12, p. 121302, 2005.
- [67] M. Sasaki and T. Tanaka, “Super-Horizon Scale Dynamics of Multi-Scalar Inflation,” *Progress of Theoretical Physics*, vol. 99, pp. 763–781, 1998.
- [68] C. T. Byrnes, M. Sasaki, and D. Wands, “Primordial trispectrum from inflation,” *Phys. Rev. D*, vol. 74, no. 12, p. 123519, 2006.
- [69] E. Komatsu and D. N. Spergel, “The Cosmic Microwave Background Bispectrum as a Test of the Physics of Inflation and Probe of the Astrophysics of the Low-Redshift Universe,” pp. 2009–2010, Dec. 2002.
- [70] L. Verde, “Non-Gaussianity from Large-Scale Structure Surveys,” *Advances in Astronomy*, vol. 2010, 2010.
- [71] J.-Q. Xia, A. Bonaldi, C. Baccigalupi, G. De Zotti, S. Matarrese, L. Verde, and M. Viel, “Constraining primordial non-Gaussianity with high-redshift probes,” *Journal of Cosmology and Astro-Particle Physics*, vol. 8, p. 13, Aug. 2010.
- [72] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, “Probing primordial non-Gaussianity with large-scale structure,” *Phys. Rev. D*, vol. 69, p. 103513, May 2004.
- [73] A. R. Zentner, “The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth,” *International Journal of Modern Physics D*, vol. 16, pp. 763–815, 2007.
- [74] N. Kaiser, “On the spatial correlations of Abell clusters,” *ApJ*, vol. 284, pp. L9–L12, Sept. 1984.

- [75] J. N. Fry, “The Evolution of Bias,” *ApJ*, vol. 461, p. L65, Apr. 1996.
- [76] W. H. Press and P. Schechter, “Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation,” *ApJ*, vol. 187, pp. 425–438, Feb. 1974.
- [77] J. A. Peacock and A. F. Heavens, “Alternatives to the Press-Schechter cosmological mass function,” *MNRAS*, vol. 243, pp. 133–143, Mar. 1990.
- [78] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, “Excursion set mass functions for hierarchical Gaussian fluctuations,” *ApJ*, vol. 379, pp. 440–460, Oct. 1991.
- [79] J. F. Navarro, C. S. Frenk, and S. D. M. White, “A Universal Density Profile from Hierarchical Clustering,” *ApJ*, vol. 490, p. 493, Dec. 1997.
- [80] S. Cole and N. Kaiser, “Biased clustering in the cold dark matter cosmogony,” *MNRAS*, vol. 237, pp. 1127–1146, Apr. 1989.
- [81] H. J. Mo, Y. P. Jing, and S. D. M. White, “The correlation function of clusters of galaxies and the amplitude of mass fluctuations in the Universe,” *MNRAS*, vol. 282, pp. 1096–1104, Oct. 1996.
- [82] R. K. Sheth and G. Tormen, “Large-scale bias and the peak background split,” *MNRAS*, vol. 308, pp. 119–126, Sept. 1999.
- [83] R. K. Sheth, H. J. Mo, and G. Tormen, “Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes,” *MNRAS*, vol. 323, pp. 1–12, May 2001.
- [84] R. K. Sheth and G. Tormen, “An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier,” *MNRAS*, vol. 329, pp. 61–75, Jan. 2002.

- 
- [85] S. Matarrese, L. Verde, and R. Jimenez, “The Abundance of High-Redshift Objects as a Probe of Non-Gaussian Initial Conditions,” *ApJ*, vol. 541, pp. 10–24, Sept. 2000.
- [86] M. Maggiore and A. Riotto, “The Halo Mass Function from Excursion Set Theory. III. Non-Gaussian Fluctuations,” *ApJ*, vol. 717, pp. 526–541, July 2010.
- [87] M. Grossi, L. Verde, C. Carbone, K. Dolag, E. Branchini, F. Iannuzzi, S. Matarrese, and L. Moscardini, “Large-scale non-Gaussian mass function and halo bias: tests on N-body simulations,” *MNRAS*, vol. 398, pp. 321–332, Sept. 2009.
- [88] M. Maggiore and A. Riotto, “The Halo mass function from Excursion Set Theory. II. The Diffusing Barrier,” *ApJ*, vol. 717, pp. 515–525, July 2010.
- [89] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, “Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects,” *Phys. Rev. D*, vol. 77, p. 123514, June 2008.
- [90] S. Matarrese and L. Verde, “The Effect of Primordial Non-Gaussianity on Halo Bias,” *ApJ*, vol. 677, pp. L77–L80, Apr. 2008.
- [91] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, “Constraints on local primordial non-Gaussianity from large scale structure,” *Journal of Cosmology and Astro-Particle Physics*, vol. 8, p. 31, Aug. 2008.
- [92] N. Afshordi and A. J. Tolley, “Primordial non-Gaussianity, statistics of collapsed objects, and the integrated Sachs-Wolfe effect,” *Phys. Rev. D*, vol. 78, p. 123507, Dec. 2008.
- [93] P. McDonald, “Primordial non-Gaussianity: Large-scale structure signature in the perturbative bias model,” *Phys. Rev. D*, vol. 78, p. 123519, Dec. 2008.

- [94] L. Verde and S. Matarrese, “Detectability of the Effect of Inflationary Non-Gaussianity on Halo Bias,” *ApJ*, vol. 706, pp. L91–L95, Nov. 2009.
- [95] V. Desjacques and U. Seljak, “Primordial Non-Gaussianity in the Large-Scale Structure of the Universe,” *Advances in Astronomy*, vol. 2010, 2010.
- [96] J.-Q. Xia, M. Viel, C. Baccigalupi, G. De Zotti, S. Matarrese, and L. Verde, “Primordial Non-Gaussianity and the NRAO VLA Sky Survey,” *ApJ*, vol. 717, pp. L17–L21, July 2010.
- [97] J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick, “The NRAO VLA Sky Survey,” *AJ*, vol. 115, pp. 1693–1716, May 1998.
- [98] A. G. Sánchez, C. G. Scóccola, A. J. Ross, W. Percival, M. Manera, F. Montesano, X. Mazzalay, A. J. Cuesta, D. J. Eisenstein, E. Kazin, C. K. McBride, K. Mehta, A. D. Montero-Dorta, N. Padmanabhan, F. Prada, J. A. Rubiño-Martín, R. Tojeiro, X. Xu, M. V. Magaña, E. Aubourg, N. A. Bahcall, S. Bailey, D. Bizyaev, A. S. Bolton, H. Brewington, J. Brinkmann, J. R. Brownstein, J. R. Gott, J. C. Hamilton, S. Ho, K. Honscheid, A. Labatie, E. Malanushenko, V. Malanushenko, C. Maraston, D. Muna, R. C. Nichol, D. Oravetz, K. Pan, N. P. Ross, N. A. Roe, B. A. Reid, D. J. Schlegel, A. Shelden, D. P. Schneider, A. Simmons, R. Skibba, S. Snedden, D. Thomas, J. Tinker, D. A. Wake, B. A. Weaver, D. H. Weinberg, M. White, I. Zehavi, and G. Zhao, “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the large-scale two-point correlation function,” *MNRAS*, vol. 425, pp. 415–437, Sept. 2012.
- [99] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot, and et al., “SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems,” *AJ*, vol. 142, p. 72, Sept. 2011.

- [100] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, A. J. Cuesta, L. N. A. da Costa, K. S. Dawson, R. de Putter, D. J. Eisenstein, J. E. Gunn, H. Guo, J.-C. Hamilton, P. Harding, S. Ho, K. Honscheid, E. Kazin, D. Kirkby, J.-P. Kneib, A. Labatie, C. Loomis, R. H. Lupton, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, C. K. McBride, K. T. Mehta, O. Mena, F. Montesano, D. Muna, R. C. Nichol, S. E. Nuza, M. D. Olmstead, D. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. Parejko, I. Paris, W. J. Percival, P. Petitjean, F. Prada, B. Reid, N. A. Roe, A. J. Ross, N. P. Ross, L. Samushia, A. G. Sanchez, D. J. S. D. P. Schneider, C. G. Scoccola, H.-J. Seo, E. S. Sheldon, A. Simmons, R. A. Skibba, M. A. Strauss, M. E. C. Swanson, D. Thomas, J. L. Tinker, R. Tojeiro, M. Vargas Magana, L. Verde, C. Wagner, D. A. Wake, B. A. Weaver, D. H. Weinberg, M. White, X. Xu, C. Yeche, I. Zehavi, and G.-B. Zhao, “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample,” *ArXiv e-prints*, Mar. 2012.
- [101] M. Manera, R. Scoccimarro, W. J. Percival, L. Samushia, C. K. McBride, A. J. Ross, R. K. Sheth, M. White, B. A. Reid, A. G. Sánchez, R. de Putter, X. Xu, A. A. Berlind, J. Brinkmann, B. Nichol, F. Montesano, N. Padmanabhan, R. A. Skibba, R. Tojeiro, and B. A. Weaver, “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues,” *ArXiv e-prints*, Mar. 2012.
- [102] A. J. Ross, W. J. Percival, A. G. Sánchez, L. Samushia, S. Ho, E. Kazin, M. Manera, B. Reid, M. White, R. Tojeiro, C. K. McBride, X. Xu, D. A. Wake, M. A. Strauss, F. Montesano, M. E. C. Swanson, S. Bailey, A. S. Bolton, A. M. Dorta, D. J. Eisenstein, H. Guo, J.-C. Hamilton, R. C. Nichol, N. Padmanabhan, F. Prada, D. J. Schlegel, M. V. Magaña, I. Zehavi, M. Blanton, D. Bizyaev, H. Brewington, A. J.

- Cuesta, E. Malanushenko, V. Malanushenko, D. Oravetz, J. Parejko, K. Pan, D. P. Schneider, A. Shelden, A. Simmons, S. Snedden, and G.-b. Zhao, “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: analysis of potential systematics,” *MNRAS*, vol. 424, pp. 564–590, July 2012.
- [103] M. Crocce and R. Scoccimarro, “Nonlinear evolution of baryon acoustic oscillations,” *Phys. Rev. D*, vol. 77, p. 023533, Jan. 2008.
- [104] M. Crocce and R. Scoccimarro, “Renormalized cosmological perturbation theory,” *Phys. Rev. D*, vol. 73, p. 063519, Mar. 2006.
- [105] S. E. Nuza, A. G. Sanchez, F. Prada, A. Klypin, D. J. Schlegel, S. Gottloeber, A. D. Montero-Dorta, M. Manera, C. K. McBride, A. J. Ross, R. Angulo, M. Blanton, A. Bolton, G. Favole, L. Samushia, F. Montesano, W. Percival, N. Padmanabhan, M. Steinmetz, J. Tinker, R. Skibba, D. Schneider, H. Guo, I. Zehavi, Z. Zheng, D. Bizyaev, O. Malanushenko, V. Malanushenko, A. E. Oravetz, D. J. Oravetz, and A. C. Shelden, “The clustering of galaxies at  $z \sim 0.5$  in the SDSS-III Data Release 9 BOSS-CMASS sample: a test for the  $\Lambda$ CDM cosmology,” *ArXiv e-prints*, Feb. 2012.
- [106] A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottlöber, B. Allgood, and J. R. Primack, “The Dark Side of the Halo Occupation Distribution,” *ApJ*, vol. 609, pp. 35–49, July 2004.
- [107] A. R. Wetzel and M. White, “What determines satellite galaxy disruption?,” *MNRAS*, vol. 403, pp. 1072–1088, Apr. 2010.
- [108] J. L. Tinker, D. H. Weinberg, Z. Zheng, and I. Zehavi, “On the Mass-to-Light Ratio of Large-Scale Structure,” *ApJ*, vol. 631, pp. 41–58, Sept. 2005.
- [109] A. Cooray and R. Sheth, “Halo models of large scale structure,” *Phys. Rep.*, vol. 372, pp. 1–129, Dec. 2002.

- [110] M. White, M. Blanton, A. Bolton, D. Schlegel, J. Tinker, A. Berlind, L. da Costa, E. Kazin, Y.-T. Lin, M. Maia, C. K. McBride, N. Padmanabhan, J. Parejko, W. Percival, F. Prada, B. Ramos, E. Sheldon, F. de Simoni, R. Skibba, D. Thomas, D. Wake, I. Zehavi, Z. Zheng, R. Nichol, D. P. Schneider, M. A. Strauss, B. A. Weaver, and D. H. Weinberg, “The Clustering of Massive Galaxies at  $z \sim 0.5$  from the First Semester of BOSS Data,” *ApJ*, vol. 728, p. 126, Feb. 2011.
- [111] D. J. Eisenstein and W. Hu, “Baryonic Features in the Matter Transfer Function,” *ApJ*, vol. 496, p. 605, Mar. 1998.
- [112] S. Chongchitnan and J. Silk, “A Study of High-order Non-Gaussianity with Applications to Massive Clusters and Large Voids,” *ApJ*, vol. 724, pp. 285–295, Nov. 2010.
- [113] N. Kaiser, “Clustering in real space and in redshift space,” *MNRAS*, vol. 227, pp. 1–21, July 1987.
- [114] A. J. Ross, W. J. Percival, A. Carnero, G.-b. Zhao, M. Manera, A. Raccanelli, E. Aubourg, D. Bizyaev, H. Brewington, J. Brinkmann, J. R. Brownstein, A. J. Cuesta, L. A. N. da Costa, D. J. Eisenstein, G. Ebelke, H. Guo, J.-C. Hamilton, M. Vargas Magana, E. Malanushenko, V. Malanushenko, C. Maraston, F. Montesano, R. C. Nichol, D. Oravetz, K. Pan, F. Prada, A. G. Sanchez, L. Samushia, D. J. Schlegel, D. P. Schneider, H.-J. Seo, A. Sheldon, A. Simmons, S. Snedden, M. E. C. Swanson, D. Thomas, J. L. Tinker, R. Tojeiro, and I. Zehavi, “The Clustering of Galaxies in SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Constraints on Primordial Non-Gaussianity,” *ArXiv e-prints*, Aug. 2012.
- [115] J. E. Geach, D. Sobral, R. C. Hickox, D. A. Wake, I. Smail, P. N. Best, C. M. Baugh, and J. P. Stott, “The clustering of H-alpha emitters at  $z=2.23$  from HiZELS,” *ArXiv e-prints*, June 2012.

- [116] J. E. Geach, I. Smail, P. N. Best, J. Kurk, M. Casali, R. J. Ivison, and K. Coppin, “HiZELS: a high-redshift survey of H-alpha emitters - I. The cosmic star formation rate and clustering at  $z = 2.23$ ,” *MNRAS*, vol. 388, pp. 1473–1486, Aug. 2008.
- [117] D. Sobral, I. Smail, P. N. Best, J. E. Geach, Y. Matsuda, J. P. Stott, M. Cirasuolo, and J. Kurk, “A large, multi-epoch H-alpha survey at  $z=2.23, 1.47, 0.84, 0.40$ : the 11 Gyr evolution of star-forming galaxies from HiZELS,” *ArXiv e-prints*, Feb. 2012.
- [118] P. Simon, “How accurate is Limber’s equation?,” *A&A*, vol. 473, pp. 711–714, Oct. 2007.
- [119] P. J. E. Peebles, *The large-scale structure of the universe*. 1980.
- [120] C. Blake, A. Collister, and O. Lahav, “Halo-model signatures from 380000 Sloan Digital Sky Survey luminous red galaxies with photometric redshifts,” *MNRAS*, vol. 385, pp. 1257–1269, Apr. 2008.
- [121] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery,

- P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, N. Yasuda, and SDSS Collaboration, “The Sloan Digital Sky Survey: Technical Summary,” *AJ*, vol. 120, pp. 1579–1587, Sept. 2000.
- [122] J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger, L. N. Carey, G. R. Knapp, D. G. York, W. N. Boroski, S. M. Kent, R. H. Lupton, C. M. Rockosi, M. L. Evans, P. Waddell, J. E. Anderson, J. Annis, J. C. Barentine, L. M. Bartoszek, S. Bastian, S. B. Bracker, H. J. Brewington, C. I. Briegel, J. Brinkmann, Y. J. Brown, M. A. Carr, P. C. Czarapata, C. C. Drennan, T. Dombeck, G. R. Federwitz, B. A. Gillespie, C. Gonzales, S. U. Hansen, M. Harvanek, J. Hayes, W. Jordan, E. Kinney, M. Klaene, S. J. Kleinman, R. G. Kron, J. Kresinski, G. Lee, S. Limmongkol, C. W. Lindenmeyer, D. C. Long, C. L. Loomis, P. M. McGehee, P. M. Mantsch, E. H. Neilsen, Jr., R. M. Neswold, P. R. Newman, A. Nitta, J. Peoples, Jr., J. R. Pier, P. S. Prieto, A. Prosapio, C. Rivetta, D. P. Schneider, S. Snedden, and S.-i. Wang, “The 2.5 m Telescope of the Sloan Digital Sky Survey,” *AJ*, vol. 131, pp. 2332–2359, Apr. 2006.
- [123] S. M. Croom, B. J. Boyle, T. Shanks, R. J. Smith, L. Miller, P. J. Outram, N. S. Loaring, F. Hoyle, and J. da Ângela, “The 2dF QSO Redshift Survey - XIV. Struc-

- ture and evolution from the two-point correlation function,” *MNRAS*, vol. 356, pp. 415–438, Jan. 2005.
- [124] A. D. Myers, R. J. Brunner, G. T. Richards, R. C. Nichol, D. P. Schneider, D. E. Vanden Berk, R. Scranton, A. G. Gray, and J. Brinkmann, “First Measurement of the Clustering Evolution of Photometrically Classified Quasars,” *ApJ*, vol. 638, pp. 622–634, Feb. 2006.
- [125] M. White, “The Ly- $\alpha$  forest,” in *The Davis Meeting On Cosmic Inflation*, Mar. 2003.
- [126] P. McDonald and D. J. Eisenstein, “Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman- $\alpha$  forest,” *Phys. Rev. D*, vol. 76, p. 063009, Sept. 2007.
- [127] J. W. Richards, D. L. Starr, N. R. Butler, J. S. Bloom, J. M. Brewer, A. Crellin-Quick, J. Higgins, R. Kennedy, and M. Rischard, “On Machine-learned Classification of Variable Stars with Sparse and Noisy Time-series Data,” *ApJ*, vol. 733, p. 10, May 2011.
- [128] N. P. Ross, A. D. Myers, E. S. Sheldon, C. Yèche, M. A. Strauss, J. Bovy, J. A. Kirkpatrick, G. T. Richards, É. Aubourg, M. R. Blanton, W. N. Brandt, W. C. Carithers, R. A. C. Croft, R. da Silva, K. Dawson, D. J. Eisenstein, J. F. Hennawi, S. Ho, D. W. Hogg, K.-G. Lee, B. Lundgren, R. G. McMahon, J. Miralda-Escudé, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, M. M. Pieri, J. Rich, N. A. Roe, D. Schiminovich, D. J. Schlegel, D. P. Schneider, A. Slosar, N. Suzuki, J. L. Tinker, D. H. Weinberg, A. Weyant, M. White, and W. M. Wood-Vasey, “The SDSS-III Baryon Oscillation Spectroscopic Survey: Quasar Target Selection for Data Release Nine,” *ApJS*, vol. 199, p. 3, Mar. 2012.

- [129] G. T. Richards, M. A. Strauss, X. Fan, P. B. Hall, S. Jester, D. P. Schneider, D. E. Vanden Berk, C. Stoughton, S. F. Anderson, R. J. Brunner, J. Gray, J. E. Gunn, Ž. Ivezić, M. K. Kirkland, G. R. Knapp, J. Loveday, A. Meiksin, A. Pope, A. S. Szalay, A. R. Thakar, B. Yanny, D. G. York, J. C. Barentine, H. J. Brewington, J. Brinkmann, M. Fukugita, M. Harvanek, S. M. Kent, S. J. Kleinman, J. Krzesiński, D. C. Long, R. H. Lupton, T. Nash, E. H. Neilsen, Jr., A. Nitta, D. J. Schlegel, and S. A. Snedden, “The Sloan Digital Sky Survey Quasar Survey: Quasar Luminosity Function from Data Release 3,” *AJ*, vol. 131, pp. 2766–2787, June 2006.
- [130] X. Fan, “Simulation of Stellar Objects in SDSS Color Space,” *AJ*, vol. 117, pp. 2528–2551, May 1999.
- [131] G. T. Richards, M. A. Weinstein, D. P. Schneider, X. Fan, M. A. Strauss, D. E. Vanden Berk, J. Annis, S. Burles, E. M. Laubacher, D. G. York, J. A. Frieman, D. Johnston, R. Scranton, J. E. Gunn, Ž. Ivezić, R. C. Nichol, T. Budavári, I. Csabai, A. S. Szalay, A. J. Connolly, G. P. Szokoly, N. A. Bahcall, N. Benítez, J. Brinkmann, R. Brunner, M. Fukugita, P. B. Hall, G. S. Hennessy, G. R. Knapp, P. Z. Kunszt, D. Q. Lamb, J. A. Munn, H. J. Newberg, and C. Stoughton, “Photometric Redshifts of Quasars,” *AJ*, vol. 122, pp. 1151–1162, Sept. 2001.
- [132] S. M. Croom, G. T. Richards, T. Shanks, B. J. Boyle, R. G. Sharp, J. Bland-Hawthorn, T. Bridges, R. J. Brunner, R. Cannon, D. Carson, K. Chiu, M. Colless, W. Couch, R. de Propris, M. J. Drinkwater, A. Edge, S. Fine, J. Loveday, L. Miller, A. D. Myers, R. C. Nichol, P. Outram, K. Pimbblet, I. Roseboom, N. Ross, D. P. Schneider, A. Smith, C. Stoughton, M. A. Strauss, and D. Wake, “The 2dF-SDSS LRG and QSO Survey: the spectroscopic QSO catalogue,” *MNRAS*, vol. 392, pp. 19–44, Jan. 2009.
- [133] Bovy Jo, D. W. Hogg, and S. T. Roweis, “Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations,”

- Annals of Applied Statistics*, vol. 5, pp. 1657–1677, June 2011.
- [134] J. Bovy, J. F. Hennawi, D. W. Hogg, A. D. Myers, J. A. Kirkpatrick, D. J. Schlegel, N. P. Ross, E. S. Sheldon, I. D. McGreer, D. P. Schneider, and B. A. Weaver, “Think Outside the Color Box: Probabilistic Target Selection and the SDSS-XDQSO Quasar Targeting Catalog,” *ApJ*, vol. 729, p. 141, Mar. 2011.
- [135] M. White, A. D. Myers, N. P. Ross, D. J. Schlegel, J. F. Hennawi, Y. Shen, I. McGreer, M. A. Strauss, A. S. Bolton, J. Bovy, X. Fan, J. Miralda-Escude, N. Palanque-Delabrouille, I. Paris, P. Petitjean, D. P. Schneider, M. Viel, D. H. Weinberg, C. Yèche, I. Zehavi, K. Pan, S. Snedden, D. Bizyaev, H. Brewington, J. Brinkmann, V. Malanushenko, E. Malanushenko, D. Oravetz, A. Simmons, A. Sheldon, and B. A. Weaver, “The clustering of intermediate-redshift quasars as measured by the Baryon Oscillation Spectroscopic Survey,” *MNRAS*, vol. 424, pp. 933–950, Aug. 2012.
- [136] M. E. C. Swanson, M. Tegmark, A. J. S. Hamilton, and J. C. Hill, “Methods for rapidly processing angular masks of next-generation galaxy surveys,” *MNRAS*, vol. 387, pp. 1391–1402, July 2008.
- [137] J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, K. S. J. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, C. A. L. Bailer-Jones, I. K. Baldry, J. C. Barentine, B. A. Bassett, A. C. Becker, T. C. Beers, E. F. Bell, A. A. Berlind, M. Bernardi, M. R. Blanton, J. J. Bochanski, W. N. Boroski, J. Brinchmann, J. Brinkmann, R. J. Brunner, T. Budavári, S. Carliles, M. A. Carr, F. J. Castander, D. Cinabro, R. J. Cool, K. R. Covey, I. Csabai, C. E. Cunha, J. R. A. Davenport, B. Dilday, M. Doi, D. J. Eisenstein, M. L. Evans, X. Fan, D. P. Finkbeiner, S. D. Friedman, J. A. Frieman, M. Fukugita, B. T. Gänsicke, E. Gates, B. Gillespie, K. Glazebrook, J. Gray, E. K. Grebel, J. E. Gunn, V. K. Gurbani, P. B. Hall, P. Harding, M. Harvanek, S. L. Hawley, J. Hayes, T. M. Heck-

- man, J. S. Hendry, R. B. Hindsley, C. M. Hirata, C. J. Hogan, D. W. Hogg, J. B. Hyde, S.-i. Ichikawa, Ž. Ivezić, S. Jester, J. A. Johnson, A. M. Jorgensen, M. Jurić, S. M. Kent, R. Kessler, S. J. Kleinman, G. R. Knapp, R. G. Kron, J. Krzesinski, N. Kuropatkin, D. Q. Lamb, H. Lampeitl, S. Lebedeva, Y. S. Lee, R. F. Leger, S. Lépine, M. Lima, H. Lin, D. C. Long, C. P. Loomis, J. Loveday, R. H. Lupton, O. Malanushenko, V. Malanushenko, R. Mandelbaum, B. Margon, J. P. Marriner, D. Martínez-Delgado, T. Matsubara, P. M. McGehee, T. A. McKay, A. Meiksin, H. L. Morrison, J. A. Munn, R. Nakajima, E. H. Nielsen, Jr., H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nieto-Santisteban, A. Nitta, S. Okamura, R. Owen, H. Oyaizu, N. Padmanabhan, K. Pan, C. Park, J. Peoples, Jr., J. R. Pier, A. C. Pope, N. Purger, M. J. Raddick, P. Re Fiorentin, G. T. Richards, M. W. Richmond, A. G. Riess, H.-W. Rix, C. M. Rockosi, M. Sako, D. J. Schlegel, D. P. Schneider, M. R. Schreiber, A. D. Schwobe, U. Seljak, B. Sesar, E. Sheldon, K. Shimasaku, T. Sivarani, J. A. Smith, S. A. Snedden, M. Steinmetz, M. A. Strauss, M. SubbaRao, Y. Suto, A. S. Szalay, I. Szapudi, P. Szkody, M. Tegmark, A. R. Thakar, C. A. Tremonti, D. L. Tucker, A. Uomoto, D. E. Vanden Berk, J. Vandenberg, S. Vidrih, M. S. Vogeley, W. Voges, N. P. Vogt, Y. Wadadekar, D. H. Weinberg, A. A. West, S. D. M. White, B. C. Wilhite, B. Yanny, D. R. Yocum, D. G. York, I. Zehavi, and D. B. Zucker, “The Sixth Data Release of the Sloan Digital Sky Survey,” *ApJS*, vol. 175, pp. 297–313, Apr. 2008.
- [138] A. J. Ross, W. J. Percival, and R. J. Brunner, “Evolution of the clustering of photometrically selected SDSS galaxies,” *MNRAS*, vol. 407, pp. 420–434, Sept. 2010.
- [139] M. Crocce, E. Gaztañaga, A. Cabré, A. Carnero, and E. Sánchez, “Clustering of photometric luminous red galaxies - I. Growth of structure and baryon acoustic feature,” *MNRAS*, vol. 417, pp. 2577–2591, Nov. 2011.
- [140] I. Zehavi, D. J. Eisenstein, R. C. Nichol, M. R. Blanton, D. W. Hogg, J. Brinkmann,

- J. Loveday, A. Meiksin, D. P. Schneider, and M. Tegmark, “The Intermediate-Scale Clustering of Luminous Red Galaxies,” *ApJ*, vol. 621, pp. 22–31, Mar. 2005.
- [141] N. P. Ross, J. da Ângela, T. Shanks, D. A. Wake, R. D. Cannon, A. C. Edge, R. C. Nichol, P. J. Outram, M. Colless, W. J. Couch, S. M. Croom, R. de Propris, M. J. Drinkwater, D. J. Eisenstein, J. Loveday, K. A. Pimblet, I. G. Roseboom, D. P. Schneider, R. G. Sharp, and P. M. Weilbacher, “The 2dF-SDSS LRG and QSO Survey: the LRG 2-point correlation function and redshift-space distortions,” *MNRAS*, vol. 381, pp. 573–588, Oct. 2007.
- [142] U. Sawangwit, T. Shanks, F. B. Abdalla, R. D. Cannon, S. M. Croom, A. C. Edge, N. P. Ross, and D. A. Wake, “Angular correlation function of 1.5 million luminous red galaxies: clustering evolution and a search for baryon acoustic oscillations,” *MNRAS*, vol. 416, pp. 3033–3056, Oct. 2011.
- [143] N. Nikoloudakis, T. Shanks, and U. Sawangwit, “Clustering analysis of high-redshift Luminous Red Galaxies in Stripe 82,” *ArXiv e-prints*, Apr. 2012.
- [144] P. Norberg, C. M. Baugh, E. Gaztañaga, and D. J. Croton, “Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics,” *MNRAS*, vol. 396, pp. 19–38, June 2009.
- [145] A. W. Moore, A. J. Connolly, C. Genovese, A. Gray, L. Grone, N. Kanidoris, II, R. C. Nichol, J. Schneider, A. S. Szalay, I. Szapudi, and L. Wasserman, “Fast Algorithms and Efficient Statistics: N-Point Correlation Functions,” in *Mining the Sky* (A. J. Banday, S. Zaroubi, and M. Bartelmann, eds.), p. 71, 2001.
- [146] Y. Shen, M. A. Strauss, M. Oguri, J. F. Hennawi, X. Fan, G. T. Richards, P. B. Hall, J. E. Gunn, D. P. Schneider, A. S. Szalay, A. R. Thakar, D. E. Vanden Berk, S. F. Anderson, N. A. Bahcall, A. J. Connolly, and G. R. Knapp, “Clustering of

- High-Redshift ( $z \approx 2.9$ ) Quasars from the Sloan Digital Sky Survey,” *AJ*, vol. 133, pp. 2222–2241, May 2007.
- [147] N. Padmanabhan, D. J. Schlegel, U. Seljak, A. Makarov, N. A. Bahcall, M. R. Blanton, J. Brinkmann, D. J. Eisenstein, D. P. Finkbeiner, J. E. Gunn, D. W. Hogg, Ž. Ivezić, G. R. Knapp, J. Loveday, R. H. Lupton, R. C. Nichol, D. P. Schneider, M. A. Strauss, M. Tegmark, and D. G. York, “The clustering of luminous red galaxies in the Sloan Digital Sky Survey imaging data,” *MNRAS*, vol. 378, pp. 852–872, July 2007.
- [148] A. J. Ross, S. Ho, A. J. Cuesta, R. Tojeiro, W. J. Percival, D. Wake, K. L. Masters, R. C. Nichol, A. D. Myers, F. de Simoni, H. J. Seo, C. Hernández-Monteagudo, R. Crittenden, M. Blanton, J. Brinkmann, L. A. N. da Costa, H. Guo, E. Kazin, M. A. G. Maia, C. Maraston, N. Padmanabhan, F. Prada, B. Ramos, A. Sanchez, E. F. Schlafly, D. J. Schlegel, D. P. Schneider, R. Skibba, D. Thomas, B. A. Weaver, M. White, and I. Zehavi, “Ameliorating systematic uncertainties in the angular clustering of galaxies: a study using the SDSS-III,” *MNRAS*, vol. 417, pp. 1350–1373, Oct. 2011.
- [149] S. Ho, A. Cuesta, H.-J. Seo, R. de Putter, A. J. Ross, M. White, N. Padmanabhan, S. Saito, D. J. Schlegel, E. Schlafly, U. Seljak, C. Hernández-Monteagudo, A. G. Sánchez, W. J. Percival, M. Blanton, R. Skibba, D. Schneider, B. Reid, O. Mena, M. Viel, D. J. Eisenstein, F. Prada, B. A. Weaver, N. Bahcall, D. Bizyaev, H. Brewinton, J. Brinkman, L. Nicolaci da Costa, J. R. Gott, E. Malanushenko, V. Malanushenko, B. Nichol, D. Oravetz, K. Pan, N. Palanque-Delabrouille, N. P. Ross, A. Simmons, F. de Simoni, S. Snedden, and C. Yèche, “Clustering of Sloan Digital Sky Survey III Photometric Luminous Galaxies: The Measurement, Systematics, and Cosmological Implications,” *ApJ*, vol. 761, p. 14, Dec. 2012.

- 
- [150] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere,” *ApJ*, vol. 622, pp. 759–771, Apr. 2005.
- [151] W. Saunders, T. Bridges, P. Gillingham, R. Haynes, G. A. Smith, J. D. Whittard, V. Churilov, A. Lankshear, S. Croom, D. Jones, and C. Boshuizen, “AAOmega: a scientific and optical overview,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (A. F. M. Moorwood and M. Iye, eds.), vol. 5492 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 389–400, Sept. 2004.