



UNIVERSIDAD DE VALPARAÍSO
FACULTAD DE CIENCIAS
INSTITUTO DE FÍSICA Y ASTRONOMÍA

The Effect of Non-Gaussian Primordial Perturbations on Large-Scale Structure

Tesis para Optar al Grado de Magíster en Astrofísica

Greco Alejandro Peña Pinto

Profesor Guía:
Graeme Candlish

Miembros de la Comisión:
Victor Cárdenas
Facundo Gómez

Valparaíso, Chile
Diciembre, 2021

© 2015, Vicente Villanueva Llanos.

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

Contents

Acknowledgements	8
Abstract	10
1 Introduction	11
1.1 Primordial Non-Gaussianity	13
1.2 Statistical tools for the analysis of Large-Scale Structures	16
1.2.1 2-point correlation function and power spectrum	17
1.2.2 3-point correlation function and bispectrum	21
2 Tomographic non-Gaussianity	24
2.1 Multifield inflation context	25
2.1.1 Two-field lagrangian	25
2.2 Tomographic non-Gaussianity mechanism	27
2.3 Reconstruction of the non-Gaussian Probability Distribution Function	29
2.3.1 Asymptotic reconstruction	33
2.3.2 Full reconstruction	34
3 Methodology	36
3.1 Generating the initial conditions	36
3.2 MUSIC	38
3.2.1 Initial condition for the particle field	38
3.3 RAMSES	40
3.4 L-PICOLA	41
3.5 The FFT Polyspectrum Estimator	42
3.6 Simulations	45
3.6.1 Initial conditions	45
3.6.2 Polyspectra analysis	50

4	Results	52
4.1	Comparing RAMSES and L-PICOLA	52
4.2	Variance among realizations	55
4.2.1	Power spectra	57
4.2.2	Bispectra	60
4.2.3	Trispectra	66
4.3	Fitting functions	69
4.4	CMB constraints	73
5	Summary and Conclusions	76

List of Figures

1.1	2dF Galaxy Redshift Survey. Carried out by the Anglo-Australian Observatory (AAO) between 1997 and 2002 (Colless et al., 2003). There are 221,414 Galaxies.	16
3.1	The AMR technique applied to a shock impacting a steep slope is shown here. Credits: Wikipedia.	40
3.2	The asymptotic form of the PDF, for various values of the amplitude (levels 1 to 4) and frequency (<i>left panel: f1; middle panel: f2; right panel: f3</i>). The levels and frequency values are defined in Table 3.1.	47
3.3	Illustration of accept-reject technique. Here we are representing only the NG4f3 model.	48
3.4	Non-Gaussian values versus Gaussian values before and after sorting (left and right respectively). For simplicity we have plotted only 5×10^4 points.	49
4.1	Projection density of RAMSES and L-PICOLA at $z=0$; the first and the second column respectively. <i>First row: Gaussian initial conditions (model Gr1); second row: the NG2f1r1 model; third row: the NG4f1r1 model</i> . We see a considerable similarity in the structure distribution, especially at large scales. At small scales, the structures are much more defined in RAMSES than in L-PICOLA.	53
4.2	<i>Left column: polyspectra of a single realization of a Gaussian simulation using RAMSES and L-PICOLA. Right column: The polyspectra of L-PICOLA (for one realization) normalized by those of RAMSES for models Gr1, NG2f1r1 and NG4f1r1. Top row: power spectrum, middle row: bispectrum, bottom row: trispectrum. Only equilateral triangle configurations are considered for the bispectrum and only square configurations for the trispectrum.</i>	54

4.3	Comparison of the normalized polyspectra for L-PICOLA and RAMSES. <i>Left column:</i> NG2f1r1; <i>Right column:</i> NG4f1r1. By normalising using the Gaussian polyspectra for both codes, we find that the results are extremely similar in both cases. We find an excellent agreement, less than 1% at all scales.	56
4.4	Non-Gaussian power spectra normalized with respect to the Gaussian power spectra, averaged across all realizations at $z = 0$. <i>Left column:</i> NG2 models, <i>right column:</i> NG4 models. <i>Top row:</i> frequency f1, <i>middle row:</i> frequency f2, <i>bottom row:</i> frequency f3. The red and blue lines indicate the average normalized power spectra (for the NG2 and NG4 models respectively) while the light grey strip shows the degree of variance around this average arising from the individual realizations.	58
4.5	Non-Gaussian power spectra normalized with respect to the Gaussian power spectra, as for Fig. 4.4, but at $z = 2.2$. The variance is reduced in all models. This is to be expected since the non-linear processes, are reduced as the redshift increases.	59
4.6	Non-Gaussian bispectra normalized with respect to the Gaussian bispectra, at $z = 0$. <i>Left column:</i> NG2 models, <i>right column:</i> NG4 models. <i>Top row:</i> frequency f1, <i>middle row:</i> frequency f2, <i>bottom row:</i> frequency f3. We can see, apparently a similar “dip” in the models NG2f2 and NG4f2 as in the power spectrum, although on a slightly wider range of scales: $0.3 \lesssim k \lesssim 0.8$ h/Mpc, and with a stronger suppression.	61
4.7	Non-Gaussian bispectra normalized with respect to the Gaussian bispectra, at $z = 2.2$. The panels are as for Fig. 4.4. Contrary to the case of the power spectrum, we see an unexpected increase in variance rather than a decrease, which is not seen in the power spectrum. . . .	62
4.8	NG4 normalized bispectra, at $z = 0$. <i>Top row:</i> frequency f1, <i>middle row:</i> frequency f2, <i>bottom row:</i> frequency f3. <i>Red line:</i> non-symmetric configurations; <i>blue line:</i> symmetric configurations. We have tested with the following configuration: $k_2 = 1.1k_1$, and $k_3 = 1.2k_1$	65

4.9	Non-Gaussian trispectra normalized with respect to the Gaussian trispectra, at $z = 0$. <i>Left column</i> : NG2 models, <i>right column</i> : NG4 models. <i>Top row</i> : frequency f1, <i>middle row</i> : frequency f2, <i>bottom row</i> : frequency f3. Again, we see, the “dip” in the models NG2f2 and NG4f2 as in the power spectrum and the bispectrum. Here, we have quadrilaterals in equilateral configurations, which are generally folded in 3-dimensional space. So, we take $k_1 = k_2 = k_3 = k_4$, where we have not set a restriction on the other two additional degrees of freedom.	67
4.10	Non-Gaussian trispectra normalized with respect to the Gaussian trispectra, at $z = 2.2$. The panels are as for Fig. 4.4. The variance is reduced in all models when we compare with the results for the trispectrum at $z = 0$. Again, this is to be expected since the non-linear processes, are reduced as the redshift increases.	68
4.11	Fit lines for the averaged normalized power spectra at $z = 0$ in the first row. A 4th order polynomial fit is used for all frequencies (f1, left panel; f2, middle panel; f3, right panel). In the second row, we show the residual histograms of the adjustments. The variance S^2 of the residual histograms is also given (multiplied by 10^3 to improve legibility of the plot).	69
4.12	Fit lines for the averaged normalized bispectra at $z = 0$ in the first row. A 4th order polynomial fit is used for all frequencies (f1, left panel; f2, middle panel; f3, right panel). In the second row, we show the residual histograms of the adjustments. The variance S^2 of the residual histograms is also given (multiplied by 10^3 to improve legibility of the plot).	70
4.13	Fit lines for the averaged normalized trispectra at $z = 0$ in the first row. A 4th order polynomial fit is used for all frequencies (f1, left panel; f2, middle panel; f3, right panel). In the second row, we show the residual histograms of the adjustments. The variance S^2 of the residual histograms is also given (multiplied by 10^3 to improve legibility of the plot).	71
4.14	The permitted region of parameter space for the full PDF, as determined from the trispectrum, is shown in orange. The excluded region is shown in white. The models we have considered are shown as blue dots.	75

List of Tables

3.1 Values of A and f used in the non-Gaussian correction term of the PDF for our models. 47

Acknowledgements

I don't know how to write this section so I'll just write what comes to me: *Agradezco a la vida*, because I always remember this song (*gracias a la Vida de Violeta Parra*) sung a lot of times by my mother until today. I thank loneliness because I learned to be with myself, I learned to listen to myself, I learned to feel comfortable with who I am, but always excited to keep learning, and this is practically who I am. I thank the circumstances, the people I met, the people I didn't know, the people with whom I generated links, people with whom I did not generate links. I am grateful for everything because it has all been part of my growth.

If I have to name names here they go: I thank my parents, Jimena Pinto and Renato Peña. In their way, they have helped me to grow as a professional. Both have given me values that today determine the person I am.

I thank my brothers Darco Peña and Franco Peña because they have always wanted the best for me. We have listened to each other, we have grown up together, we know our stories, we know how difficult it is to get ahead starting from poverty.

I thank a lot of my old and new friends, especially those who were in this stage of my life: Ana Lorenzo, Catalina Zamora, Rafael Fuentes, Daniela Palma, and Andrés Araya (*sucio jopes*).

I am grateful to my guidance professor, Graeme Candlish, who has been a great thesis tutor with many good virtues; he is patient, teaches very well and knows how to listen. And I am very grateful for the respect with which we have always treated each other.

Finally, I am grateful to every scientist, ethical, critical, committed, and pleas-

ant, who before anything else is a person, and who seeks to make the world in which we live a better world.

Abstract

The late-time effect of primordial non-Gaussianity offers a window into the physics of inflation and the very early Universe. In this thesis we study the consequences of a particular class of primordial non-Gaussianity that is fully characterized by initial density fluctuations drawn from a non-Gaussian probability density function, rather than by construction of a particular form for the primordial bispectrum. We numerically generate multiple realisations of cosmological structure and use the late-time matter power spectrum, bispectrum and trispectrum to determine the effect of these modified initial conditions. We show that the initial non-Gaussianity has only a small imprint on the first three polyspectra, when compared to a standard Gaussian cosmology. Furthermore, some of our models present an interesting scale-dependent deviation from the Gaussian case in the bispectrum and trispectrum, although the signal is at most at the percent level. The majority of our models are consistent with CMB constraints, while the others are only marginally excluded. Finally, we discuss further possible extensions of our study.

Chapter 1

Introduction

One of the principal problems in cosmology is to understand the physics of the early Universe, in particular the period referred to as inflation. This period is thought to be responsible for determining the initial conditions of our Universe and ultimately gives rise to the observed late-time large-scale structure.

During the exponential growth of the Universe in the inflationary epoch primordial quantum fluctuations grew to create the first gravitational wells where structure started to form. Thanks to the CMB, we have strong observational evidence for the existence of these quantum perturbations at earlier times, as they are imprinted upon the CMB temperature map. The extremely small size of the temperature anisotropies in the CMB accords with the concept of homogeneity of the Universe. In fact, the smoothness of the CMB temperature map is one of the “puzzles” in standard cosmology that is resolved by inflation (the so-called horizon problem). By studying the statistics of the fluctuations from the CMB and the large-scale structure of the Universe, it is possible to obtain constraints on the statistics of the primordial fluctuations in the early Universe, and therefore obtain constraints on the physics of the inflationary epoch.

The most simple inflationary models predict that the distribution of the primordial fluctuations has to be near Gaussian, and this is supported by the current observational evidence from the CMB that also shows a near Gaussian distribution of the temperature fluctuations (Planck Collaboration et al., 2020). This means that we can find all the information about the distribution of the fluctuations through only the two-point correlation function, or say its power spectrum. However, many inflationary models do not predict Gaussian statistics. For instance, inflationary

models such as multi-field inflation, through the interactions of several fields, can transfer their statistics from one to the other leading to deviations from Gaussian statistics in the primordial fluctuations.

Among the many probes of the statistics of the perturbations that we have available, the n -point correlation functions correspond to the moments of the probability density function (PDF) describing the statistical distribution of the perturbations. At present, the search for evidence of deviations from Gaussianity in the CMB has focused just on the lowest orders of the n -point correlation functions, for instance, by measuring the 3 and 4-point correlation functions (the Fourier transforms of these quantities are the bispectrum and trispectrum respectively). On the other hand, parameterising deviations from Gaussianity with the bispectrum or trispectrum is not a fully comprehensive way to find evidence of non-Gaussianity (see Bennett et al. (2003) and Buchert et al. (2017)). One could be missing additional information in correlation functions of order greater than 3 or 4, which could be related to physical processes in the early universe that have not yet been tested.

It is in this context that theories that can quantify these deviations through higher orders of correlation functions have begun to attract attention. Multi-field inflation can provide mechanisms that produce large amounts of non-Gaussianity through the interaction of several degrees of freedom. In this work, we will analyse the impact on the large-scale structure of a novel class of non-Gaussianity generated by the interaction of an isocurvature field and the primordial curvature field. This kind of non-Gaussianity can inherit the structure of the potential of the isocurvature field, so by studying the PDF associated with the primordial perturbations, we could theoretically find evidence of the structure of this primordial potential, in this sense, this kind of non-Gaussianity is called *tomographic non-Gaussianity*.

We will study the impact of these primordial non-Gaussianities on structure formation through the tool of N-body simulations as well as approximate methods referred to as mock catalogue generators. These tools allow us to model both linear and nonlinear structure formation, and will in principle allow us to contrast and compare with future observational studies of large-scale structures, such as LSST (Vera C. Rubin Observatory, 2021), EUCLID (2021), and SKA (Weltman et al., 2020).

The remainder of this thesis is organized as follows: In Sec. 1.1, we summarize attempts to search for primordial non-Gaussianities, and in Sec. 1.2, we describe some common tools for studying large-scale structure. In Chap. 2 we describe the mech-

anism that generates the primordial perturbations and subsequently the probability density function associated to these perturbations, which is used for the generation of the initial conditions of our simulations. In Chap. 3, we present the codes used in this work, as well as the method by which we generate our simulations, and the analysis of the results. In Chap. 4, we present and discuss our results. Finally, in Chap. 5, we summarize the work and present our conclusions.

1.1 Primordial Non-Gaussianity

The search for the initial conditions of our Universe has been discussed over time from an observational and theoretical perspective. From a historical context, observational evidence for large-scale structure (LSS) already indicated that the distribution of matter in the Universe was not Gaussian. Groth and Peebles (1977) measured the 3-point correlation function of galaxies from the Shane-Wirtanen catalog, which indicated that there was no preference for a linear distribution characteristic of galaxies. From this, the question arose as to whether the measured non-Gaussianity in the galaxy distribution was due only to the gravitational collapse of matter, or whether, in addition, there might be a hidden signal corresponding to primordial non-Gaussianities that could give rise to the initial conditions of the Universe.

The theory of cosmic inflation appeared as a solution to the problems of the standard model of cosmology (now referred to as Λ CDM, after detection of the dark energy component), which were the monopole problem, the horizon problem, the flatness problem, and the problem of the initial conditions leading to structure formation. Inflation was first developed by numerous authors, in particular Guth, 1981; Kazanas, 1980; Sato, 1981; Starobinsky, 1980. The simplest models of inflation postulate a scalar field ϕ (referred to as the inflaton) whose potential is such that the field evolves extremely slowly in a very flat region (the so-called *slow-roll* regime) in order to generate an exponential expansion of the Universe, after which the inflaton reaches a potential minimum. The inflaton quantum mechanically oscillates around this minimum, transferring the energy acquired when falling to the potential minimum to the standard model particles, a process known as reheating. In addition, inflation suggests a mechanism for the spontaneous generation of initial conditions: quantum fluctuations associated to this field generate the primordial curvature perturbations at the beginning of the Universe. These perturbations have been shown to have near Gaussian statistics in the simplest inflationary models. The usual way to quantify some deviation from Gaussianity has been through the f_{NL} parameter, which refers to the first parameter of local non-linearities in a power series around a Gaussian

contribution, i.e.

$$\mathcal{R} = \mathcal{R}_G + \frac{3}{5} f_{\text{NL}} \mathcal{R}_G^2, \quad (1.1)$$

where \mathcal{R} is the co-moving primordial curvature perturbation, a gauge-invariant quantity, and \mathcal{R}_G is the Gaussian component of this perturbation. This is related to the primordial gravitational potential Ψ (on a co-moving hypersurface) as

$$\mathcal{R} = \Psi + H \frac{\delta\phi}{\dot{\phi}}. \quad (1.2)$$

Here, H is the hubble parameter. So we can associate a bispectrum (the 3-point correlation function in Fourier space) to the f_{NL} parameter as follows

$$\begin{aligned} \langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) & \frac{6}{5} f_{\text{NL}} [P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_2) \\ & + P_{\mathcal{R}}(k_2) P_{\mathcal{R}}(k_3) + P_{\mathcal{R}}(k_3) P_{\mathcal{R}}(k_1)], \end{aligned} \quad (1.3)$$

where $P_{\mathcal{R}}$ is the primordial power spectrum. Since different inflation models produce different shapes and amplitudes for the bispectrum (see for example Chen (2010)) we can, in principle, discriminate different inflationary mechanisms that could have given rise to the primordial non-Gaussianities using this statistic. To first order in the slow-roll parameter f_{NL} is given by the Maldacena consistency relation:

$$f_{\text{NL}} = -\frac{5}{12} (n_s - 1), \quad (1.4)$$

where n_s is the spectral index related to the primordial power spectrum $P_{\mathcal{R}}(k)$ as

$$n_s - 1 \equiv \frac{d \ln(k^3 P_{\mathcal{R}}(k))}{d \ln k}. \quad (1.5)$$

Since inflation predicts a near scale-invariant primordial power spectrum we expect f_{NL} to be of order 10^{-2} . From Planck Collaboration et al. (2020) a value of $f_{\text{NL}} = -0.9 \pm 5.1$ has been reported. This is still far from tightly constraining inflationary models, and such constraints are not anticipated to improve substantially even for the next generation of cosmological studies Abazajian et al., 2016; EUCLID, 2021; Vera C. Rubin Observatory, 2021; Weltman et al., 2020.

Similar attempts have been made with the higher-order n-point functions, with similar results when considered from the point-of-view of single-field inflation. If

future observational constraints strongly indicated higher levels of non-Gaussianity than currently suggested with the weak constraints today, the simpler models of inflation could be discarded, and more complex inflationary mechanisms would have to be considered. These predict stronger signals in the n -point functions. Examples of more complex inflationary mechanisms are the dynamical interaction of the inflationary field with other degrees of freedom, self-interactions of primordial fields, non-canonical terms, etc. Many reviews provide a more complete overview of this topic: Abazajian et al., 2016; Bartolo et al., 2004; Celoria and Matarrese, 2018; Chen, 2010; Liguori et al., 2010. On the other hand, there have also been attempts to analytically explore non-Gaussianities through perturbative methods in the non-linear regime, using the effective field theory of large-scale structure (EFT of LSS). See Matarrese and Pietroni, 2007; Pietroni, 2008 and Angulo et al., 2015; Assassi et al., 2015; Baldauf et al., 2015.

In addition, analysis of three-dimensional large-scale structure data sources can provide other ways to constrain the primordial non-Gaussianities. Although one of the greatest difficulties in working with late-time observational data is the difficulty of distinguishing between the non-Gaussianity coming from the gravitational collapse and the primordial non-Gaussian signal. This difficulty is mainly because we only have one realization of the Universe. Therefore, N -body simulations and mock catalogs have been a widely used tool to generate different realizations of the Universe, to obtain statistically robust observables that can be used to compare with observations. For some examples see: Desjacques et al., 2009; Giannantonio and Porciani, 2010; Grossi et al., 2007, 2009; Hikage et al., 2008; Sefusatti et al., 2010; Wagner and Verde, 2012; Wagner et al., 2010.

The polyspectra (n -point functions in Fourier space) is one of the standard tools for LSS analysis, although other powerful tests of non-Gaussianity can be performed through the halo mass function, halo bias, etc (see Palma et al., 2020). Otherwise, they can also be used jointly to constrain late-time clustering observables (e.g. Gil-Marín et al., 2015 and Gil-Marín et al., 2017). In the following section we will describe some of the tools commonly used for large-scale structure analysis.

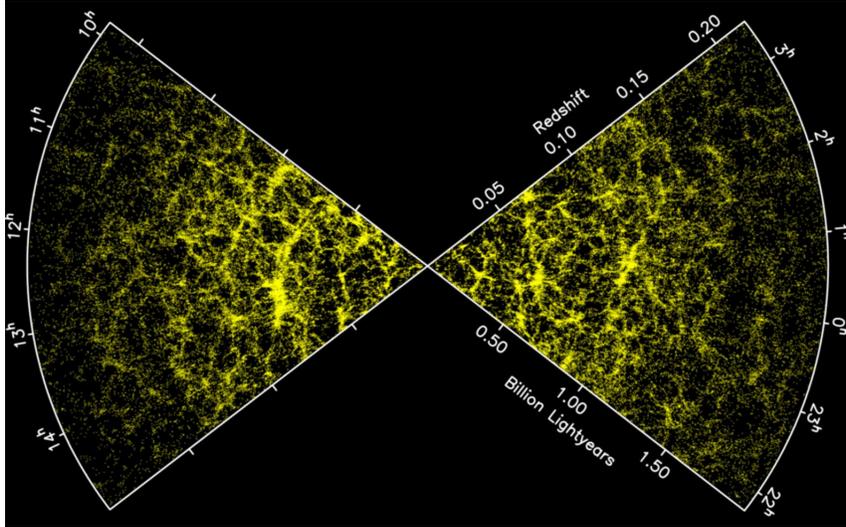


Figure 1.1: 2dF Galaxy Redshift Survey. Carried out by the Anglo-Australian Observatory (AAO) between 1997 and 2002 (Colless et al., 2003). There are 221,414 Galaxies.

1.2 Statistical tools for the analysis of Large-Scale Structures

As discussed earlier, one would expect that signals of primordial non-Gaussianity would propagate through the evolution of the Universe and ultimately be imprinted on late-time large-scale structure statistics, in addition to their effect on early-time observables such as the CMB. The challenge, however, is to extract the primordial signal from the non-Gaussianity induced due to gravitational collapse.

One way to study the large-scale structure of the Universe is by analyzing its statistical properties, through, for example, the n -point correlation functions and their analogues in Fourier space. To define these functions, we begin by defining clustering, which consists of counting objects separated by a certain distance. Suppose we have a map of galaxies as in Fig. 1.1, if we place on this map a grid, count the number of galaxies in each grid cell and divide by its volume, we can calculate the density of galaxies as

$$\rho(\mathbf{x}) = N_{cell}(\mathbf{x})/V_{cell}, \quad (1.6)$$

where $N_{cell}(\mathbf{x})$ is the number of objects within a grid cell and V_{cell} is the volume of

each cell.

An alternative way of representing the density field is by using the over-density field $\delta(\mathbf{x})$ (also known as density contrast), which is defined as the density at a point \mathbf{x} divided by the average cosmological density ρ_0 minus one:

$$\delta(\mathbf{x}) = \rho(\mathbf{x})/\rho_0 - 1, \quad (1.7)$$

where $\delta(\mathbf{x}) = -1$ in the case of a void; $\delta(\mathbf{x}) = 0$ when $\rho(\mathbf{x})$ is equal to the background density, and $\delta(\mathbf{x}) > 0$ when we have an overdense region, i.e. clustering.

1.2.1 2-point correlation function and power spectrum

In this section we will follow closely the description given in Amendola and Tsujikawa (2010). An important tool in the statistical analysis of random point distributions in cosmology are the correlation functions. The most commonly used is the two-point correlation function, and its counterpart in Fourier space, the power spectrum. These quantify the correlation that exists between the clustering of particles of two positions in the space.

Consider a box with volume V and number average density ρ_0 , the average number of particles n in an infinitesimal volume inside the box will be $\rho_0 dV$. Then, for two points in the space, separated by a distance $r_{ab} > 0$, the average number of pairs of particles $\langle n_a n_b \rangle$ in the volumes dV_a and dV_b , are related to the correlation function $\xi(r_{ab})$ by

$$dN_{ab} = \langle n_a n_b \rangle = \rho_0^2 dV_a dV_b [1 + \xi(r_{ab})], \quad (1.8)$$

where, we can see that if $\xi(r_{ab}) = 0$, then the particle distribution is uncorrelated (does not depend on position), and the number of pairs of particles is independently determined by $\rho_0^2 dV_a dV_b$. In case of correlation ($\xi(r_{ab}) \neq 0$) we have

$$\xi(r_{ab}) = \frac{dN_{ab}}{\rho_0^2 dV_a dV_b} - 1 = \langle \delta(r_a) \delta(r_b) \rangle, \quad (1.9)$$

where $\delta(r_a) = n_a/\rho_0 dV_a - 1$ is the contrast density.

if we average it over all the possible positions in a sample (i.e sample average), the correlation function is written as

$$\xi(\mathbf{r}) = \frac{1}{V} \int \delta(\mathbf{y} + \mathbf{r})\delta(\mathbf{y})dV_y = \langle \delta(\mathbf{y} + \mathbf{r})\delta(\mathbf{y}) \rangle. \quad (1.10)$$

Power spectrum

The power spectrum is a common descriptor of clustering at different scales in cosmology, and is defined as a real quadratic function of a perturbation variable in Fourier space, i.e.

$$P(\mathbf{k}) = V\delta(\mathbf{k})\delta(\mathbf{k})^* = V|\delta(\mathbf{k})|^2, \quad (1.11)$$

where V is the volume and $\delta(\mathbf{k})$ is the Fourier transform of the density contrast, given by

$$\delta(\mathbf{k}) = \frac{1}{V} \int \delta(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}dV. \quad (1.12)$$

We can insert the above equation in the expression for the power spectrum, equation (1.11), to find

$$P(\mathbf{k}) = \frac{1}{V} \int \delta(\mathbf{x})\delta(\mathbf{y})e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}dV_x dV_y. \quad (1.13)$$

Now, if we set $\mathbf{r} = \mathbf{x} - \mathbf{y}$, and using the expression (1.10), we can write the power spectrum as follows

$$P(\mathbf{k}) = \int \xi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}dV. \quad (1.14)$$

Thus, we see that the power spectrum is the Fourier transform of the 2-point correlation function (Wiener-Khinchin theorem). Likewise, the 2-point correlation function may be written as

$$\xi(\mathbf{r}) = (2\pi)^{-3} \int P(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}d^3k. \quad (1.15)$$

One can define the power spectrum in a more general way if we consider ensemble averages (averaging different realizations) for the power spectrum, i.e,

$$\begin{aligned}
V \langle \delta(\mathbf{k}) \delta(\mathbf{k}')^* \rangle &= \frac{1}{V} \int \langle \delta(\mathbf{y}) \delta(\mathbf{y} + \mathbf{r}) \rangle e^{-i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{y} - i\mathbf{k} \cdot \mathbf{r}} dV_r dV_y \\
&= \frac{1}{V} \int e^{-i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{y}} dV_y \int e^{-i\mathbf{k} \cdot \mathbf{r}} \xi(\mathbf{r}) dV_r \\
&= \frac{(2\pi)^3}{V} P(\mathbf{k}) \delta_D(\mathbf{k} - \mathbf{k}'),
\end{aligned} \tag{1.16}$$

where we have used Dirac's delta function in Fourier space

$$\delta_D(\mathbf{k}) = (2\pi)^{-3} \int e^{i\mathbf{k} \cdot \mathbf{x}} d^3x, \tag{1.17}$$

which is normalised to ensure that

$$\int \delta_D(\mathbf{k}) d^3k = 1. \tag{1.18}$$

The expression in (1.16) shows that the modes \mathbf{k} are uncorrelated when the correlation function $\xi(\mathbf{x})$ depends only on the separation \mathbf{r} and not on the position \mathbf{y} . This corresponds to (statistical) homogeneity. Moreover, since $\delta(x)$ is a real function then $\delta(\mathbf{k}) = \delta^*(\mathbf{k})$, and we can also write the power spectrum as

$$V \langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = \frac{(2\pi)^3}{V} P(\mathbf{k}) \delta_D(\mathbf{k} + \mathbf{k}'). \tag{1.19}$$

If we have an isotropic distribution, in addition to homogeneity, then the 2-point correlation function depends only on the distance between the two points, i.e. $r = |\mathbf{r}|$, and the power spectrum depends only on the magnitude of the wavevector $k = |\mathbf{k}|$.

Due to the discreteness of the matter distribution of the Universe, we do not obtain information about a continuous density field but instead we must use a sampling of this field. This sampling will introduce biases in the reconstruction of the density field. In what follows, we assume the simplest form of discrete sampling which is a Poissonian distribution of points in any location. For galaxies as tracers of the underlying dark matter field, we know that a Poissonian distribution is not the whole picture, and the concept of galaxy bias must be introduced. In this study we are interested in the dark matter distribution itself (using simulations) rather than the galaxy distribution, but we are still limited to a discrete sampling due to the limited particle resolution of our simulations.

To proceed, we introduce a window function $W(\mathbf{x})$ in order to quantify our particle selection procedure. The simplest approach is to select all particles within a volume and ignore all particles outside that volume. This corresponds to using the top-hat window function. This window function is normalised to a volume V as

$$\int W(\mathbf{x})dV = 1. \quad (1.20)$$

Thus the function is simply given by $W(\mathbf{x}) = 1/V$ inside the volume and zero everywhere else. The density field of our specific sample, δ_s , is then given by the full density field $\delta(\mathbf{x})$ multiplied by the window function $W(\mathbf{x})$ and the sample volume V , due to the normalisation of $W(\mathbf{x})$:

$$\delta_s(\mathbf{x}) = \delta(\mathbf{x})VW(\mathbf{x}). \quad (1.21)$$

The discrete density field can now be written as

$$\delta_s(\mathbf{x}) = \frac{V}{N} \sum w_i \delta_D(\mathbf{x} - \mathbf{x}_i) - VW(\mathbf{x}), \quad (1.22)$$

where we have used $\rho(\mathbf{x}) = \sum_i \delta_D(\mathbf{x} - \mathbf{x}_i)$, i.e. the discrete field is written as a sum of Dirac delta functions, and $w_i = VW(\mathbf{x}_i)$. In Fourier space the discretised field is

$$\delta_s(\mathbf{k}) = \frac{1}{N} \sum_i w_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} - W(\mathbf{k}), \quad (1.23)$$

where $W(\mathbf{k})$ is the k -space window function:

$$W(\mathbf{k}) = \int W(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}dV, \quad (1.24)$$

with normalisation $W(0) = 1$. Now, from Eq.(1.23), we have

$$\langle \Delta^2(\mathbf{k}) \rangle \equiv V \langle \delta_s(\mathbf{k})\delta_s(\mathbf{k})^* \rangle = P(\mathbf{k}) + P_n, \quad (1.25)$$

where we have used $\langle \delta_s(\mathbf{k}) \rangle = 0$. In the above expression, the power spectrum $P(\mathbf{k})$ and the pure noise spectrum P_n are given by:

$$\begin{aligned} P(\mathbf{k}) &= \frac{V}{N^2} \sum_{i \neq j} \langle w_i w_j \rangle e^{-i\mathbf{k}\cdot(\mathbf{x}_i - \mathbf{x}_j)} - VW(k)^2, \\ P_n &= \frac{V}{N^2} \sum_i w_i^2 = \frac{V}{N}, \end{aligned} \quad (1.26)$$

where, in the last equality, we have used the fact that w_i is either 0 (outside the volume) or 1 (inside the volume). The quantity P_n gives the power spectrum of a Poisson distribution of points, and is negligible when the number of points in the volume becomes extremely large $\rho_0 = N/V \rightarrow \infty$. Typically, however, we do not have a very large number of points available (in observations due to limitations of the survey, in simulations due to limitations in resolution) and thus the noise is usually not negligible and must be subtracted from the measured power spectrum in order to obtain a more accurate estimate:

$$\hat{P}(\mathbf{k}) = \Delta^2 \mathbf{k} - P_n. \quad (1.27)$$

1.2.2 3-point correlation function and bispectrum

It is easy to extend the correlation function to higher orders by simply adding another point in space to the product of density perturbations. Such higher order correlation functions are sensitive to non-Gaussianities (the two-point function fully characterises the distribution for Gaussian statistics). As we have discussed earlier, the 1-point correlation function may be taken simply as the ensemble or sample-averaged density perturbation, which would be zero. We have already shown above the 2-point function and its Fourier transform, the power spectrum. Now we consider the 3-point correlation function, defined as

$$\varsigma_{abc}(r_a, r_b, r_c) = \langle \delta(r_a) \delta(r_b) \delta(r_c) \rangle, \quad (1.28)$$

or

$$\begin{aligned} \varsigma_{abc}(r_a, r_b, r_c) &= \left\langle \left(\frac{n_a}{\rho_0 dV_a} - 1 \right) \left(\frac{n_b}{\rho_0 dV_b} - 1 \right) \left(\frac{n_c}{\rho_0 dV_c} - 1 \right) \right\rangle \\ &= \frac{\langle n_a n_b n_c \rangle}{\rho_0^3 dV_a dV_b dV_c} - \xi_{ab} - \xi_{bc} - \xi_{ac} - 1, \end{aligned} \quad (1.29)$$

with $\xi_{ij} \equiv \xi(r_{ij})$. Rewriting (1.29) we obtain the relation,

$$\langle n_a n_b n_c \rangle = \rho_0^3 dV_a dV_b dV_c (1 + \xi_{ab} + \xi_{bc} + \xi_{ac} + \varsigma_{abc}). \quad (1.30)$$

Here, the term ς_{abc} , is known as the “disconnected” part of the 3-point correlation function, and it provides information about the nonlinearities of a random field. If $\varsigma_{abc} = 0$, we are dealing with a Gaussian distribution, and in this case, the static properties of the field are completely described by the 2-point correlation function

(or the power spectrum in Fourier space). This notion can be carried to larger orders of the correlation function, and the same description applies. As we will see in the following subsection, the Fourier transform of the 3-point correlation function is the “bispectrum”.

Bispectrum

The bispectrum is the Fourier transform of the 3-point correlation function, and is defined as

$$\langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \delta(\mathbf{k}_3) \rangle = \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3). \quad (1.31)$$

In this case the Dirac delta function ensures that the bispectrum is computed for only closed triangular configurations, i.e., $\sum_i \mathbf{k}_i = 0$.

It is useful to define the reduced bispectrum $Q(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ as

$$Q(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \equiv \frac{B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{P(k_1)P(k_2) + P(k_1)P(k_3) + P(k_2)P(k_3)}. \quad (1.32)$$

This quantity is useful for observational studies as it is only weakly dependent on scale and cosmology (Peebles, 1980), which is beneficial as it breaks the degeneracy between cosmological parameters, and thus it can isolate the effects of gravity.

The bispectrum for any Gaussian distribution of points is zero, which means all information about physical processes is contained in the power spectrum. The bispectrum remains zero in a linear evolutionary process, i.e., as long as the k modes evolve independently. On the other hand, the bispectrum takes a relevant role when one wants to study non-linear processes; when an evolutionary process is no longer linear, the modes are coupled and the bispectrum is no longer zero. In this way, information about non-linear processes that influence the evolution of dark matter clusters can be extracted. In addition, the presence of non-Gaussianity will also lead to a non-zero bispectrum, which will mix with the contribution arising from non-linearities.

The analytical approach of the bispectrum is also useful when we want to contrast information with observational data and/or simulations. In Eulerian perturbation theory, at second order (tree-level), and in an Einstein de-Sitter Universe the bispectrum results in

$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2F_2^s(\mathbf{k}_1, \mathbf{k}_2) P_1^L P_2^L + \text{cyc. perm.}, \quad (1.33)$$

where $P_i^L = P^L(k_i)$ is the linear power spectrum, and F_2^s is the symmetrized twopoint kernel given by

$$F_2^s(\mathbf{k}_i, \mathbf{k}_j) = \frac{5}{7} + \frac{1}{2} \cos(\theta_{ij}) \left(\frac{k_i}{k_j} + \frac{k_j}{k_i} \right) + \frac{2}{7} \cos^2(\theta_{ij}), \quad (1.34)$$

where θ_{ij} is the angle between the vectors \mathbf{k}_i and \mathbf{k}_j . This is an accurate approximation for quasi-linear scales, but is inaccurate for moderately non-linear scales.

Chapter 2

Tomographic non-Gaussianity

We now introduce the specific model of primordial non-Gaussianity that we are interested in this work. We give a brief summary of the theory, as described in detail in Chen et al. (2018a). As is standard in studies of inflation, we present the statistical measures in terms of the primordial curvature perturbation ζ . This may then be straightforwardly related to the dark matter density perturbation δ through the gravitational potential. To begin with, the n -point correlation functions $\langle \zeta(\mathbf{x}_1) \cdots \zeta(\mathbf{x}_n) \rangle$ that we have been considering are statistical moments of an underlying multivariate probability density function $P(\zeta)$:

$$\langle \zeta(\mathbf{x}_1) \cdots \zeta(\mathbf{x}_n) \rangle = \int d\zeta P(\zeta) \zeta(\mathbf{x}_1) \cdots \zeta(\mathbf{x}_n), \quad (2.1)$$

where $\langle \cdot \rangle$ denotes the expectation value of a random variable. On the other hand, the non-Gaussianity considered in this work can be fully characterised by a reduced PDF that considers only the statistics of a single point of the ζ field:

$$\langle \zeta(\mathbf{x})^n \rangle = \int_{-\infty}^{\infty} d(\zeta(\mathbf{x})) \rho(\zeta(\mathbf{x})) (\zeta(\mathbf{x}))^n, \quad (2.2)$$

where $\rho(\zeta(\mathbf{x}))$ is the PDF which will be reconstructed in the Sec. 2.3.

In this chapter, we will describe the physical mechanism generating the primordial perturbations which give rise to the probability distribution function (PDF) of the primordial tomographic non-Gaussianity, and consequently, we will describe the statistics associated with this non-Gaussianity through the n -point functions. After describing the generating mechanism of the primordial fluctuations, in Sec. 2.3 we

will describe how to reconstruct the PDF through its relation to the n -point functions. We do not show the details of this calculation but instead refer the reader to the original study of Chen et al. (2018a).

2.1 Multifield inflation context

The CMB observational data constrain the bispectrum to be consistent with Gaussian statistic Planck Collaboration et al. (2020). On the other hand, parameterising the primordial statistics via the 3-point correlation function or higher orders may not be the best way to constrain the data (see Bennett et al. (2003) and (Buchert et al., 2017)). Instead, multi-field (with a light isocurvature field) or quasi-single-field (with a massive isocurvature field) inflation models can generate larger amounts of primordial non-Gaussianity by increasing the amplitude of the 3-point correlation function. See Enqvist and Väihkönen, 2004 and Achúcarro et al. (2017), for examples of these kinds of models. The isocurvature field ψ , during horizon-crossing and/or super-horizon scales, could affect the dynamics of the primordial perturbation field ζ due to a derivative coupling of the form $\mathcal{L}_{\text{int}} \propto \dot{\zeta}\psi$. This interaction can lead to deviations in the inflationary path of the inflaton through the landscape of the inflationary potential Gordon et al., 2000; Groot Nibbelink and van Tent, 2000, 2002, something that is clearly not possible in single-field inflation. These couplings produce different types of non-linearities capable of being transferred to the statistics of the primordial perturbation field ζ .

2.1.1 Two-field lagrangian

In a two-field inflationary theory, the interaction between the fields is given by a term at quadratic order in the fluctuations, and this interaction can turn the inflationary trajectory. The evolution of these fluctuations is then described by a Lagrangian which couples the comoving curvature perturbation ζ with a single isocurvature field ψ (see, for instance Gordon et al., 2000; Groot Nibbelink and van Tent, 2000, 2002). This is given by

$$\mathcal{L}(\zeta, \psi) = a^3 \left[\epsilon(\dot{\zeta} - \alpha\dot{\psi})^2 - \frac{\epsilon}{a^2}(\nabla\zeta)^2 + \frac{1}{2}\dot{\psi}^2 - \frac{1}{2a^2}(\nabla\psi)^2 - \frac{1}{2}\mu^2\psi^2 \right], \quad (2.3)$$

where $\epsilon \equiv -\dot{H}/H^2$ is the first slow-roll parameter, μ is called the entropic mass of ψ , a is the scale factor, and α (which depends on time) incorporates the coupling between ζ and ψ . It follows that ψ satisfies the following equation of motion in the long-wavelength limit:

$$\ddot{\psi} + 3H\dot{\psi} + \mu^2\psi = 0. \quad (2.4)$$

From this, we see that μ corresponds to the mass of ψ on super-horizon scales. On the other hand, if $\mu = 0$, then ψ becomes “ultralight” and the Lagrangian gains a translational symmetry

$$\psi \rightarrow \psi' = \psi + C, \quad (2.5)$$

$$\zeta \rightarrow \zeta' = \zeta + C \int^t dt\alpha, \quad (2.6)$$

where C is an arbitrary constant. Summarised from Achúcarro et al. (2017), this symmetry tell us two things: Firstly, from Eq. (2.5), the Lagrangian generates a constant solution for ψ (which can also be seen from Eq. (2.6)). This constant solution, ψ_* breaks the symmetry of the Lagrangian dominating on super-horizon scales. Secondly: from the symmetry of the lagrangian gained through Eq. (2.6), it follows that the solution ψ_* will be the source of the evolution of ζ on super-horizon scales. Considering that $\alpha \approx const.$, on super-horizon scales it can be shown that

$$\zeta \simeq \frac{\alpha}{H}\psi_*\Delta N. \quad (2.7)$$

If we consider that ψ_* has a value such that it corresponds to the horizon crossing, then ΔN will correspond to the number of e-folds at that time, so the statistical relationship between ζ and ψ , given by the n-point function is as follows

$$\langle \zeta^n \rangle \simeq \left(\frac{\alpha}{H}\Delta N \right)^n \langle \psi_*^n \rangle. \quad (2.8)$$

Then, the ratio of the power spectra ($n = 2$) for both fields is given by

$$P_\zeta \simeq \frac{\alpha^2\Delta N^2}{H^2}P_\psi. \quad (2.9)$$

Since the order of the interaction in the Lagrangian of Eq. (2.3) is quadratic, and there are no higher order terms for the self-interaction of ψ , then the statistics of ψ , which is then passed to ζ (given by Eqs. (2.8) and (2.9)), is expected to be fully Gaussian, and any deviation from the Gaussian case is expected to be suppressed by

the slow-roll parameters (Maldacena, 2003).

There are, however, other ways to transfer non-Gaussian statistics to the primordial curvature field. In the next section we will describe such a mechanism, in which a structure-rich potential is given by an axion-like particle, giving rise to a type of non-Gaussianity referred to as tomographic non-Gaussianity.

2.2 Tomographic non-Gaussianity mechanism

As described in Sec. 2.1, this type of non-Gaussianity arises from a two-field inflationary model, which includes an interaction between the isocurvature field ψ and the comoving curvature field ζ . If now, instead of the lagrangian of Eq. (2.3), we think of a more general case where the self-interaction of the isocurvature field ψ is described by the potential $\Delta V(\psi)$, then the Lagrangian would now be as follows

$$\mathcal{L}(\zeta, \psi) = a^3 \left[\varepsilon(\dot{\zeta} - \alpha\dot{\psi})^2 - \frac{\varepsilon}{a^2}(\nabla\zeta)^2 + \frac{1}{2}\dot{\psi}^2 - \frac{1}{2a^2}(\nabla\psi)^2 - \Delta V(\psi) \right], \quad (2.10)$$

where, in Chen et al. (2018a), the potential $\Delta V(\psi)$ is referred to as the *landscape* of the perturbation, since the goal is to explore situations where the fluctuations ψ are such that we cannot disregard the structure of $\Delta V(\psi)$.

Axions are particles widely studied in multifield inflation theories, and axion-like particles are well motivated by string theory (see Cicoli, 2013; Marsh, 2016, for instance). The characteristic potential associated with this particle is given by

$$\Delta V(\psi) = \Lambda^4 \left[1 - \cos\left(\frac{\psi}{f}\right) \right], \quad (2.11)$$

where Λ corresponds to the characteristic energy scale of the interaction, while f is the axion decay constant. The potential of Eq. (2.11) breaks the continuous translation symmetry previously held by the Lagrangian of Eq. (2.3) for $\mu = 0$, and replaces it with a discrete symmetry given by:

$$\psi \rightarrow \psi' = \psi + 2\pi n f, \quad (2.12)$$

$$\zeta \rightarrow \zeta' = \zeta + 2\pi n f \int^t dt \alpha. \quad (2.13)$$

The way ψ transfers its statistics to ζ is similar to that described using the symmetry arguments considering Eqs. (2.5) and (2.6), but this time the ψ field can acquire constant solutions that minimise the potential. These solutions dominate the behavior of ψ on superhorizon scales, managing to pass its information to the ζ field. The information that can be transferred from ψ to ζ will increase for solutions that minimise the potential, while this will decrease for solutions that maximise it. Therefore, the structure of ΔV will end up being inherited by the PDF of ζ , making it non-Gaussian.

The Lagrangian of Eq. (2.10) can be expected to come from perturbing a more fundamental system with a potential $V = V_0 + \Delta V$. If we stay in the regime $\Lambda^4/3H^2M_{\text{Pl}}^2 \ll 1$, then ΔV practically does not affect the background evolution given by the potential V_0 . In this regime, the ψ field will behave as an ultralight field, so before and during the horizon crossing it will be frozen, and during the horizon crossing the statistics of ψ will be transferred to ζ by Eq. (2.7). As time passes, the non-linear contributions from ΔV will increase and will be passed to ζ because ψ will no longer be frozen. Therefore these contributions will produce non-Gaussian statistics, with the relation between the fields given in Chen et al. (2018a) as

$$\langle \zeta^n \rangle_{\text{NG}} \simeq \frac{1}{2} \left(\frac{\alpha}{H} \Delta N \right)^n \langle \psi^n \rangle_{\text{NG}}, \quad (2.14)$$

where the n-point correlation functions for ψ were calculated in Palma and Riquelme (2017) using the *in-in formalism*, and are given by

$$\langle \psi(\mathbf{k}_1, \tau) \cdots \psi(\mathbf{k}_n, \tau) \rangle_c = (-1)^{n/2} (2\pi)^3 \delta^{(3)} \left(\sum_j \mathbf{k}_j \right) \frac{2}{3} \frac{\Lambda^4}{H^4} e^{-\frac{\sigma_0^2}{2f^2}} \left(\frac{H^2}{2f} \right)^n \frac{k_1^3 + \cdots + k_n^3}{k_1^3 \cdots k_n^3} \Delta N. \quad (2.15)$$

This is the statistics coming from the non-Gaussian contributions acquired by ψ due to its potential $\Delta V(\psi) = \Lambda^4[1 - \cos(\psi/f)]$. The correlation functions for ζ were obtained in Chen et al. (2018a), and are referred to as follows.

$$\tilde{G}_\zeta^{(n)}(\tau, \mathbf{k}_1, \cdots, \mathbf{k}_n) \equiv \langle \zeta(\mathbf{k}_1, \tau) \cdots \zeta(\mathbf{k}_n, \tau) \rangle. \quad (2.16)$$

After quantising the Lagrangian of Eq. (2.10) and applying the in-in formalism, in momentum space, the n-point functions are obtained as

$$\tilde{G}_\zeta^{(n)}(\tau, \mathbf{k}_1, \cdots, \mathbf{k}_n) = (-1)^{n/2} (2\pi)^3 \delta^{(3)} \left(\sum_i \mathbf{k}_i \right) \frac{\Lambda^4 e^{-\frac{\sigma_0^2}{2f^2}}}{3H^4} \left(\frac{\lambda H^2 \Delta N}{2f\sqrt{2\varepsilon}} \right)^n \frac{k_1^3 + \cdots + k_n^3}{k_1^3 \cdots k_n^3} \Delta N, \quad (2.17)$$

with n even and $\tilde{G}_\zeta^{(n)} = 0$ for n odd, and $\lambda = \frac{\sqrt{2}\varepsilon\alpha}{H}$ is a dimensionless coupling. σ_0^2 is the variance ψ , and defining the integration variable $q = k|\tau|$, it is given by

$$\sigma_0^2 = \frac{H^2}{4\pi^2} \int dq \left(q + \frac{1}{q} \right). \quad (2.18)$$

This expression contains divergences in the limits $q \rightarrow 0$ and $q \rightarrow +\infty$, and therefore it is necessary to introduce infrared and ultraviolet cutoff scales q_{IR} and q_{UV} . The UV cutoff refers to the deepest scales accessible within the horizon, according to the applied physics, while the IR cutoff derives from the largest scales of the observable Universe.

2.3 Reconstruction of the non-Gaussian Probability Distribution Function

Using the n -point correlation function given in Eq. (2.17), we can reconstruct the PDF of the non-Gaussian primordial perturbations. For this purpose, we must obtain the moments $\langle \zeta^n \rangle$ of this expression in coordinate space. These are calculated by evaluating a common value \mathbf{x} , in all coordinates of the inverse Fourier transform of Eq. (2.17):

$$\langle \zeta^n \rangle_c \equiv G_\zeta^{(n)}(\tau, \mathbf{x}, \dots, \mathbf{x}). \quad (2.19)$$

The subscript c indicates that the moments $\langle \zeta^n \rangle$ come from fully connected diagrams, indicating that this expression is proportional to a single Dirac delta function that encapsulates momentum conservation, so the moments $\langle \zeta^n \rangle$ are independent of \mathbf{x} .

On the other hand, the expression for $\tilde{G}_\zeta^{(n)}(\tau, \mathbf{k}_1, \dots, \mathbf{k}_n)$ has been calculated for long wavelength modes of ζ , in order to make inflation predictions for super-horizon perturbations, so a momentum cutoff k_L has been introduced, which is represented in terms of the physical momentum $q_{\text{phys}} \equiv k|\tau|$ (per unit of H). Thus this cutoff divides the curvature perturbations as follows:

$$\zeta = \zeta_S + \zeta_L. \quad (2.20)$$

In this case, ζ_L , includes modes of wavelength larger than a value given by $2\pi/q_L$. Consider that the horizon crossing happens at $q_{\text{phys}} = 1$, so that $q_L \leq 1$. Then q_L is given by

$$\zeta_L(\mathbf{x}, \tau) = \int_{k < k_L} \zeta(\mathbf{k}, \tau) e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (2.21)$$

with $k_L = q_L/|\tau|$. Therefore, the correlation function, $G_{\zeta,L}^{(n)}(\tau, \mathbf{x}_1, \dots, \mathbf{x}_n)$, is now integrated up to a value k_L resulting in

$$G_{\zeta,L}^{(n)}(\tau, \mathbf{x}_1, \dots, \mathbf{x}_n) = (-1)^{n/2} (2\pi)^3 \frac{\Lambda^4}{3H^4} \int_{k_1 < k_L} \dots \int_{k_n < k_L} \delta^{(3)}\left(\sum_j \mathbf{k}_j\right) e^{-i\sum_j \mathbf{k}_j \cdot \mathbf{x}_j} \\ \times e^{-\frac{\sigma_0^2}{2f^2}} \left(\frac{\lambda H^2}{2f\sqrt{2\varepsilon}} \Delta N\right)^n \frac{k_1^3 + \dots + k_n^3}{k_1^3 \dots k_n^3} \Delta N. \quad (2.22)$$

The cutoff introduced in Eq. (2.21) forces the quantity σ_0^2 to be divided into short wavelength contributions (σ_S^2) and large (σ_L^2), such that they receive larger and shorter contributions than k_L respectively. These quantities, according to Eq. (2.18), are as follows

$$\sigma_S^2 = \frac{H^2}{4\pi^2} \left(\frac{1}{2} (q_{UV}^2 - q_L^2) + \ln(q_{UV}/q_L) \right) \simeq \frac{H^2}{8\pi^2} q_{UV}^2, \quad (2.23)$$

$$\sigma_L^2 = \frac{H^2}{4\pi^2} \left(\frac{1}{2} (q_L^2 - q_{IR}^2) + \ln(q_L/q_{IR}) \right) \simeq \frac{H^2}{4\pi^2} \ln \xi, \quad (2.24)$$

where ξ corresponds to a range of accessible scales of long-wavelength modes between k_L and k_{IR} , which contribute to ζ_L . This is defined as

$$\xi = \frac{q_L}{q_{IR}} = \frac{k_L}{k_{IR}}. \quad (2.25)$$

Note that ξ parameterises the range of modes available per observation. For example, for the CMB case, $\ln \xi \sim 8$.

After applying the cutoff, if we proceed to evaluate a single coordinate value \mathbf{x} in Eq. (2.22) we get the expression

$$\langle \zeta_L^n \rangle_c = (-1)^{n/2} g_n A^2 e^{-\frac{\sigma_L^2}{2f^2}} \left[\frac{\lambda \sigma_L^2}{f\sqrt{2\varepsilon}} \Delta N \right]^n, \quad (2.26)$$

where $A^2 \equiv \frac{\Delta N}{6\sigma_L^2} \frac{\Lambda_{\text{ren}}^4}{H^2}$, and $\Lambda_{\text{ren}}^4 = e^{-\sigma_S^2/2f^2} \Lambda^4$ is the renormalised coupling resulting from a loop resummation of the calculation of the correlation function (see Chen et al. (2018a) for details), the coefficient g_n is

$$g_n \equiv \frac{(2\pi)^3}{(2\sigma_L^2/H^2)^{n-1}} I_n, \quad (2.27)$$

with $g_n = 0$ for odd n , and where I_n is the following integral

$$I_n \equiv \int_{k_1 < k_L} \dots \int_{k_n < k_L} \delta^{(3)} \left(\sum_j \mathbf{k}_j \right) \frac{k_1^3 + \dots + k_n^3}{k_1^3 \dots k_n^3}. \quad (2.28)$$

For convenience it is better to write Eq. (2.26) in terms of the variance σ_ζ^2 rather than σ_L^2 (which is associated with the ψ field). In relation to this, in Achúcarro et al. (2017), it was investigated that in the linear regime, and in the limit $\Lambda^4 \rightarrow 0$, the power spectrum of ζ and ψ are related as follows:

$$P_\zeta(k) = \frac{\lambda^2}{2\epsilon} \Delta N^2 P_\psi(k). \quad (2.29)$$

If we stay in the regime $\lambda^2 \Delta N^2 \gtrsim 1$ (to simplify the calculation) the above expression implies that

$$\sigma_\zeta^2 = \sigma_L^2 \frac{\lambda^2}{2\epsilon} \Delta N^2, \quad (2.30)$$

and defining $f_\zeta \equiv f \frac{\sigma_\zeta}{\sigma_L} = f \frac{\lambda}{\sqrt{2\epsilon}} \Delta N$, we find that

$$\langle \zeta_L^n \rangle_c = (-1)^{n/2} g_n A^2 e^{-\frac{\sigma_\zeta^2}{2f_\zeta^2}} \left[\frac{\sigma_\zeta^2}{f_\zeta} \right]^n. \quad (2.31)$$

This expression is the general form of the n -point correlation function, which requires the form of g_n (given in Eq. (2.27)), and hence the integral I_n (given in Eq. (2.28)), both to be used in the reconstruction of the PDF of ζ .

Using I_n we can introduce scale dependence into the n -point function. In appendix C of Chen et al. (2018a) it is shown that I_n can be written in terms of a single integration variable, as follows

$$I_n(\xi) = \frac{n}{(2\pi^2)^{n+1}} \int_0^\infty \frac{dx}{x} G(\xi, x) [F(\xi, x)]^{n-1}, \quad (2.32)$$

where the functions $G(\xi, x)$ and $F(\xi, x)$ are given by

$$\begin{aligned}
G(\xi, x) &= [\sin(x) - x \cos(x) - \sin(x/\xi) \\
&\quad + (x/\xi) \cos(x/\xi)], \\
F(\xi, x) &= \text{Ci}(x) - \frac{\sin(x)}{x} - \text{Ci}(x/\xi) + \frac{\sin(x/\xi)}{x/\xi},
\end{aligned} \tag{2.33}$$

and where $\text{Ci}(x)$ is the cosine integral function. Furthermore, it is shown that when we have access to an unbounded number of modes ($\xi \rightarrow \infty$), the integral I_n assumes its asymptotic form $I_n^0(\xi)$, given by

$$I_n^0(\xi) \equiv \frac{n\pi}{2(2\pi^2)^{n+1}} (\ln \xi)^{n-1}. \tag{2.34}$$

Note that the scale k_L of Eq. (2.25), corresponds to the horizon exit, and that the cutoff k_{IR} corresponds to the largest modes available from observations, so $\ln \xi \leq 60$. This tells us that, in principle, we cannot consider I_n as I_n^0 , as these functions lead to significantly different PDFs, as discussed in detail in Appendix C of Chen et al. (2018a).

However, in this work we have found that by applying a scaling to the A^2 parameter of the PDF derived from I_n^0 , we can obtain a close match with the PDF derived from the general integral I_n . This is relevant considering that, for practical reasons, and in order to obtain preliminary results, it is easier to work with the PDF reconstructed with I_n^0 than the one obtained with I_n (we will present the PDFs themselves in the following subsections). In Sec. 4.4, we discuss this issue in more detail, and we also show that most of our models are within the constraints given by the CMB.

The probability density function $\rho(\zeta)$ is related to its moments according to

$$\langle \zeta_L^n \rangle = \int d\zeta \rho(\zeta) \zeta^n. \tag{2.35}$$

To carry out the reconstruction of $\rho(\zeta)$ it is useful to understand the general structure of $\rho(\zeta)$. Let us start by considering the full n-point correlation function, which has both connected and disconnected diagrams, this quantity is given in Chen et al. (2018a) as

$$\langle \zeta_L^n \rangle = \sum_{m=0}^{n/2} \frac{n!}{m!(n-2m)!2^m} \sigma_\zeta^{2m} \langle \zeta_L^{n-2m} \rangle_c, \tag{2.36}$$

where we can note that the term $m = n/2$ in Eq. (2.36) corresponds to

$$\langle \zeta_L^n \rangle = \frac{n!}{(n/2)! 2^{n/2}} \sigma_\zeta^n. \quad (2.37)$$

This expression corresponds to the moments of a Gaussian distribution, which means that $\rho(\zeta)$ is given by a main Gaussian distribution with non-Gaussian corrections. A simple proposed ansatz is

$$\rho(\zeta) = \rho_G(\zeta) + \Delta\rho(\zeta), \quad (2.38)$$

where

$$\rho_G(\zeta) = \frac{\exp\left\{-\frac{\zeta^2}{2\sigma_\zeta^2}\right\}}{\sqrt{2\pi}\sigma_\zeta}, \quad (2.39)$$

is the Gaussian part arising from the n-point functions given by Eq. (2.37), and $\Delta\rho(\zeta)$ results from the nonlinear interactions proportional to A^2 .

2.3.1 Asymptotic reconstruction

As we mentioned before, in the formal limit $\xi \rightarrow \infty$ the integral I_n is defined by its asymptotic form $I_n^0(\xi)$. In this case, using Eq. (2.24), we find that $g_n = n$. Then for n even, the n-point functions are given by

$$\langle \zeta_L^n \rangle_c = (-1)^{n/2} n A^2 e^{-\frac{\sigma_\zeta^2}{2f_\zeta^2}} \left[\frac{\sigma_\zeta^2}{f_\zeta} \right]^n, \quad (2.40)$$

and for n odd $\langle \zeta_L^n \rangle_c = 0$. For the final step, in Chen et al. (2018a), the following ansatz is proposed

$$\Delta\rho(\zeta) = \frac{e^{-\frac{\zeta^2}{2\sigma_\zeta^2}}}{\sqrt{2\pi}\sigma_\zeta} \left[\sum_{m=0} B_m \zeta^{2m} \cos\left(\frac{\zeta}{f_\zeta}\right) + \sum_{m=0} C_m \zeta^{2m+1} \sin\left(\frac{\zeta}{f_\zeta}\right) \right], \quad (2.41)$$

which satisfies that it is even under the change $\zeta \rightarrow -\zeta$, so only the odd moments vanish. Using the relation of Eq. (2.35), we can find the B_0 and C_0 coefficients, which are the only ones that do not vanish. Finally $\rho(\zeta)$ is given by

$$\rho(\zeta) = \frac{e^{-\frac{\zeta^2}{2\sigma_\zeta^2}}}{\sqrt{2\pi}\sigma_\zeta} \left[1 + A^2 \frac{\sigma_\zeta^2}{f_\zeta^2} \cos\left(\frac{\zeta}{f_\zeta}\right) - A^2 \frac{\zeta}{f_\zeta} \sin\left(\frac{\zeta}{f_\zeta}\right) \right]. \quad (2.42)$$

It can be shown that this PDF complies with the normalisation $\int \rho(\zeta) d\zeta = 1$. Interestingly, the importance of this result lies in its simplicity and dependence on several parameters related to the shape of the landscape. If we note the second term, within square brackets in Eq. (2.42) we see that this increases the probability of finding a value of ζ sourced by values of ψ that minimise the cosine potential of Eq. (2.11). While the third term could be interpreted as the probability values that represent the propagation of ζ . It is also worth noting, however, that this third term could be absorbed into the second term by shifting f_ζ .

2.3.2 Full reconstruction

The full reconstruction of the PDF is performed in a similar way to its asymptotic form, but this time we must consider any value of $\ln \xi > 0$, so we have to deal with the integral of Eq. (2.32) which depends on the coordinate x . On the other hand, if we compare I_n and I_n^0 of Eqs. (2.32) and (2.34) respectively, we find the same dependence on n , but with the difference that in Eq. (2.32) we see that n is a factor for each value of x . Then, checking Eq. (2.35), by comparison we can identify a “decay constant” that now depends on x :

$$f_\zeta(x) \equiv f_\zeta \frac{\ln \xi}{F(\xi, x)} \geq f_\zeta, \quad (2.43)$$

which satisfies $f_\zeta(0) = f_\zeta$. Using again the ansatz presented in Eq. (2.41), and with some algebraic manipulations, we finally get that

$$\rho(\zeta) = \frac{e^{-\frac{\zeta^2}{2\sigma_\zeta^2}}}{\sqrt{2\pi}\sigma_\zeta} \left[1 + A^2 \int_0^\infty \frac{dx}{x} \mathcal{K}_\xi(x) \left(\frac{\sigma_\zeta^2}{f_\zeta(x)^2} \cos\left(\frac{\zeta}{f_\zeta(x)}\right) - \frac{\zeta}{f_\zeta(x)} \sin\left(\frac{\zeta}{f_\zeta(x)}\right) \right) \right], \quad (2.44)$$

where

$$\mathcal{K}_\xi(x) \equiv \frac{2G(\xi, x) \ln \xi}{\pi F(\xi, x)} \exp\left(-\frac{\sigma_\zeta^2 (f_\zeta^2(x) - f_\zeta^2)}{2f_\zeta^2 f_\zeta^2(x)}\right). \quad (2.45)$$

The oscillatory structure shown in Eq. (2.44) is very similar to that shown in Eq. (2.42). The main difference is the filtering produced by selecting a limited range

of modes, given by the ratio $\xi = k_L/k_{\text{IR}}$. As a result, the filter significantly suppresses the amplitude of the oscillations deviating its shape from the shape of the asymptotic reconstruction. Furthermore, small values of the parameter f_ζ decreases the amplitude of the oscillations, contrary to what happens with the asymptotic reconstruction, this is mainly due to the exponential in the kernel function $\mathcal{K}_\xi(x)$.

However, as discussed earlier and as also indicated in Chen et al. (2018a), by properly rescaling the A^2 parameter, the amplitude suppressed by this filter can be recovered, resulting in very similar forms for both PDFs, even if we construct PDFs with varying frequencies and amplitudes. In Sec. 4.4 we show that different PDFs generated from the asymptotic reconstruction with varying amplitudes and frequencies do not deviate significantly from those generated with the full reconstruction when we properly rescale A^2 ; our most extreme model, with very low frequency and very high amplitude, deviates up to $\sim 12\%$ from the full reconstruction.

It is important to note that in the limit $\xi \rightarrow \infty$, the PDF of Eq. (2.44) recovers the form given by the asymptotic reconstruction in Eq. (2.44). In the ideal case of having access to all modes, it would be possible to obtain full information about the shape of the landscape potential through the parameters Λ_{ren}^4 and f . Otherwise, with only a bounded range of modes, there will be a filter due to ξ in the function $\mathcal{K}_\xi(x)$, which will limit the amount of information we can obtain from the landscape potential. This is to be expected since ξ restricts the amount of available modes that can be used to reconstruct the shape of ΔV .

Chapter 3

Methodology

In the previous chapter, we described the mechanism and the way to reconstruct the PDF representing the non-Gaussianities used to generate the initial conditions of our simulations. In this chapter, we describe the codes used to generate these initial conditions and the corresponding cosmological simulations, as well as the method used to calculate the polyspectra. We begin by describing the process of generating initial conditions in the MUSIC code in Sec. 3.2. These initial conditions are later used in the full non-linear N-body code RAMSES as described in Sec. 3.3. We then go on to describe L-PICOLA in Sec.3.4, a code for generating mock catalogs quickly, which has an identical initial-condition-generating mechanism to MUSIC. Finally, we will describe the method for analysing our results in Sec. 3.5.

3.1 Generating the initial conditions

The standard way to generate initial conditions for cosmological simulations is to create a density field from a Gaussian noise sample. To set up the initial conditions required for our simulations, it is necessary to replace the Gaussian noise sample with a non-Gaussian sample. This process will eventually change the density field that will be used in Lagrangian perturbation theory to create the field of initial positions and velocities as discussed in the following section.

A density field $\delta(\mathbf{r})$ may be described by its power spectrum, as we have discussed earlier. The power spectrum of matter perturbations after recombination may be determined by using a transfer function $\mathcal{T}(k)$ applied to the primordial power spectrum of the fluctuations produced by inflation:

$$P(k) = \alpha k^{n_s} \mathcal{T}^2(k), \quad (3.1)$$

where α is a normalisation constant and n_s is the spectral index of the power spectrum after inflation.

To generate initial conditions for cosmological simulations we must produce a random density field with a power spectrum given by equation (3.1). This random density field may be constructed in two ways: the first option is working in Fourier space and calculating the product of a noise field $\tilde{\mu}(\mathbf{k})$ with the square root of the power spectrum:

$$\tilde{\delta}(\mathbf{k}) = \sqrt{P(\|\mathbf{k}\|)} \tilde{\mu}(\mathbf{k}) = \alpha \|\mathbf{k}\|^{n_s/2} \mathcal{T}(\|\mathbf{k}\|) \mu(\tilde{\mathbf{k}}), \quad (3.2)$$

so, by applying the inverse Fourier transformation one obtains the real-space overdensity field $\delta(\mathbf{r})$. This procedure is often called “ k -space sampling”. The second option is to work in real space and apply a convolution between a real-space transfer function $T(\|\mathbf{r}\|)$ and a real-space noise sample $\mu(\mathbf{r})$:

$$\delta(\mathbf{r}) = T(\|\mathbf{r}\|) * \mu(\mathbf{r}), \quad (3.3)$$

where $T(\mathbf{r})$ is the real-space counterpart of $\tilde{T}(\mathbf{k}) \equiv \alpha \mathbf{k}^{n_s/2} \mathcal{T}(\mathbf{k})$. The expression (3.3) is equivalent to Eq. (3.2), since a product in Fourier space is a convolution in the real space. From the physical point of view, the transfer function is responsible for encoding the information on the physical processes that occurred from the end of inflation until the recombination period, while the noise sample physically represents the primordial fluctuations during the inflationary period. In this context, the Eq. (3.3) has an intuitive physical interpretation, which is that the convolution operation imprints the perturbations of the density field in space, according to the physical processes that occurred after inflation and until recombination, as a result of the primordial fluctuations produced during inflation.

The noise sample $\mu(\mathbf{r})$ (or its Fourier space counterpart $\tilde{\mu}(\mathbf{k})$) is usually drawn from a Gaussian distribution with zero mean and unit standard deviation. In our case, we will generate $\mu(\mathbf{r})$ by drawing from a non-Gaussian PDF in order to include the non-Gaussianity in our N-body models. This requires us to Fourier transform the non-Gaussian noise sample $\mu(\mathbf{r})$ to Fourier space. Of course, for a Gaussian noise sample, the Fourier transform is also Gaussian, and so $\tilde{\mu}(\mathbf{k})$ may be obtained directly in that case.

In MUSIC (see Sec. 3.2) it is possible to obtain the density perturbations using either of the two forms given by Eqs. (3.3) and (3.2), both of which have certain advantages and disadvantages when implemented as discussed in Hahn and Abel (2011). In this thesis we have considered generating the initial conditions through the traditional form given by Eq. (3.2), after applying the Fourier transform to the non-Gaussian random numbers. These are subsequently used in the simulations of the N-body code RAMSES (see Sec. 3.3). The reason for this is that our results will be compared with the L-PICOLA code described in Sec. 3.4, in which the initial conditions are generated in the same way as described in this section, via 2LPT and Eq. (3.2).

3.2 MUSIC

MUSIC¹ is an algorithm to generate multi-scale initial conditions with multiple levels of refinements for cosmological simulations. This software uses the Lagrangian perturbation theory (LPT) at first (1LPT) or second (2LPT) order to generate a field of displacements and velocities as initial conditions that can be used by N-body codes. These displacements and velocities are generated by applying LPT to a homogeneous particle distribution in order to sample the appropriate underlying density field. This density field is constructed by convolving the matter power spectrum with a random noise field. Usually this random noise field would be constructed by drawing a random number from a Gaussian distribution at each spatial point. In our case we will replace this with our non-Gaussian PDF.

The aim of this section will be limited to describing the method used to generate the initial field of displacements and velocities. For a more detailed description of MUSIC see Hahn and Abel (2011).

3.2.1 Initial condition for the particle field

Lagrangian perturbation theory

The evolution of the density perturbations in the rest-frame of a fluid can be described using Lagrangian perturbation theory. In this frame, a fluid element with a position \mathbf{x} and velocity $\dot{\mathbf{x}}$, at a given time t , can be written as

¹The code is publicly available at <https://www-n.oca.eu/ohahn/MUSIC/>

$$\mathbf{x}(t) = \mathbf{q} + \mathbf{L}(\mathbf{q}, t), \quad \dot{\mathbf{x}}(t) = \frac{d}{dt}\mathbf{L}(\mathbf{q}, t), \quad (3.4)$$

where \mathbf{q} its the initial position of the fluid element, and $\mathbf{L}(\mathbf{q}, t)$ is called the *displacement field*, derived from perturbation theory. At first order in the perturbations, referred to as the Zel'dovich approximation (Zel'Dovich, 1970), the displacement field $\mathbf{L}(\mathbf{q}, t)$ is given by

$$\mathbf{L}(\mathbf{q}) = -\frac{2}{3H_0^2 a^2 D_+(t)} \nabla_q \phi(\mathbf{q}, t) \equiv D_+^{-1}(t) \nabla_q \Phi(\mathbf{q}, t), \quad (3.5)$$

where H_0 is the Hubble constant, a is the expansion factor at time t , $D_+(t)$ is the growth factor of linear density perturbations, and ϕ is the gravitational potential obeying Poisson's equation

$$\Delta_q \phi(\mathbf{q}, t) = \frac{3}{2} H_0^2 a^2 \delta(\mathbf{q}, t), \quad (3.6)$$

where $\delta(q, t)$ is the Gaussian density field after convolution with the matter power spectrum. It is important to note that δ is the source field of the displacements, not the density field created after displacing the fluid elements by the lagrange perturbation theory. In this approximation the velocities are irrotational, i.e. $\Delta \times \dot{x}(t) = 0$.

MUSIC can also use second-order Lagrangian perturbation theory to improve the precision in the calculation of the displacement field, since the first-order perturbation theory might underestimate higher-order moments from the density probability distribution of this displacement field (see e.g. Scoccimarro (1998)). The displacement field at second order contains a contribution of the second-order potential, in addition to the contribution of the first-order gravitational potential, so the displacement field for 2PT is

$$\mathbf{L}(\mathbf{q}, t) = D_+(t) \nabla_q \Phi(\mathbf{q}, t) + D_2(t) \nabla_q \Psi(\mathbf{q}, t), \quad (3.7)$$

where Ψ is the second-order potential obeying the Poisson equation $\nabla_q \Psi((q), t) = \tau((q), (t))$ and

$$\tau(\mathbf{q}, t) = -\frac{1}{2} \sum_{i,j} [(\partial_{q_i} \partial_{q_j} \Phi)^2 - (\partial_{q_i} \partial_i \Phi)(\partial_{q_j} \partial_{q_j} \Phi)]. \quad (3.8)$$

Here, $D_2(t) \simeq \frac{3}{7} D_+^2(t)$, where $D_+^2(t)$ is the growth factor of linear perturbations.

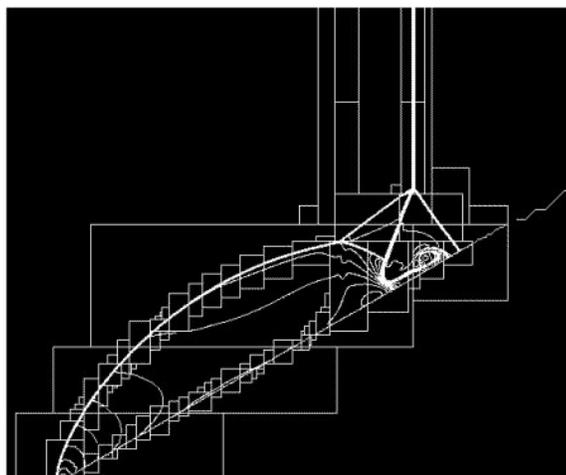


Figure 3.1: The AMR technique applied to a shock impacting a steep slope is shown here. Credits: Wikipedia.

3.3 RAMSES

RAMSES ² is an N-body and hydrodynamics code, which has been developed for the purpose of investigating structure formation with high spatial resolution (Teyssier, 2002). RAMSES is based on the Adaptive Refinement Tree (ART) (Kravtsov et al., 1997) code, but with minor differences in the implementation. Also, it implements Adaptive Mesh Refinement (AMR) with a tree-based data structure allowing recursive grid refinements on a cell-by-cell basis.

RAMSES calculates the gravitational potential on a computational grid (or mesh). By fixing the size of this grid, the numerical solution is determined at a uniform resolution throughout the entire space. However, in some cases a more accurate solution is required in specific areas, which may be obtained by increasing the grid resolution in those areas, a process referred to as refinement. AMR consists of dynamically adapting the accuracy of a solution in certain regions of the computational domain. It is dynamic in the sense that it may be modified every timestep in order to adapt to the solution as it evolves. Thus throughout the simulation there will be regions with higher levels of refinement, and therefore a more accurate solution, and these are nested within lower levels of refinement, where the accuracy of the solution is lower. The development of this technique is credited to Berger and Olinger (1984).

²The code is publicly available at <https://www.ics.uzh.ch/teyssier/ramses/RAMSES.html>

We have generated some of our simulations using RAMSES, and we consider those to be our reference simulations, due to the high accuracy of the solution in very over-dense regions, thanks to the AMR technique. However, due to the computational cost of running full N-body simulations (in particular, our simulations required about 8 hours per run), we generate the majority of our cosmological realizations using a mock catalog generator known as L-PICOLA, as discussed in the following section.

3.4 L-PICOLA

Due to the need to generate large numbers of simulations to obtain accurate measurements of statistical and systematic errors of large-scale structure surveys, different codes have been developed in recent years that quickly generate dark matter halo catalogs (or galaxy catalogs) known as *mocks* catalogs, which can be compared with the observed galaxy distribution of a survey (Howlett et al., 2015). L-PICOLA is a parallel code able to evolve a dark matter field from early times to the present day much faster than a full nonlinear N-body code. The precision of L-PICOLA is excellent at linear and quasi-linear scales and is poorer at non-linear scales (as usually seen in this type of mock catalog generator). As opposed to a code like RAMSES (see Sec. 3.3), L-PICOLA does not have adaptive mesh refinement (AMR) capability, it is based on a mesh with a fixed resolution.

L-PICOLA is based on the Comoving Lagrangian Acceleration method (COLA; Tassev et al. (2013)), which is more accurate than 2LPT. The COLA method solves the equation of motion of the particles by splitting the linear and quasi-linear contributions from the nonlinear ones. It then solves the larger scales using Lagrangian perturbation theory in a comoving reference frame in Lagrangian space (described in subsection 3.2.1), where the solution is exact. While the nonlinear component is solved using the particle mesh (PM) method (see Hockney and Eastwood (1988)), which is utilized in n-body codes. The EOM is discretized using the Kick-Drift-Kick method and solved iteratively. In each iteration, the positions and velocities are updated according to the gravitational potential felt by each particle. The positions, and velocities, are not calculated for the same point in time (except for the beginning and end), but they jump each other and the next iteration of the velocity depends on the position of the previous iteration.

In this work we use the full implementation of L-PICOLA, i.e. we use the COLA method to run our simulations. The method still needs a certain amount of timesteps

to reproduce CDM halo statistics. The amount recommended by the authors is 10 timesteps, which has been adopted for all our simulations with L-PICOLA. For this reason L-PICOLA is much faster than RAMSES, as the number of timesteps is fewer.

The L-PICOLA implementation automatically generates initial conditions through 2LPT which are used in the calculation of the density field at different redshifts. We have modified the L-PICOLA implementation to use our non-Gaussian random number distributions given by different models (see Fig. 3.2 from Sec. 3.6) generated from Eq. (2.42). A comparison between the density field results between RAMSES and L-PICOLA is given in Fig. 4.1 of Chap. 4. In the next section, we describe the implementation used to analyse the polyspectra.

3.5 The FFT Polyspectrum Estimator

Here, we present the Fast Fourier Transform Polyspectrum Estimator (Watkinson et al., 2017), the method that we used to calculate the polyspectra of our simulations. This is a general extension of the method described in the thesis of Jeong (2010) which in turn has been built from the thesis of Sefusatti (2005). To start we define the following FFT conventions:

$$\delta(\mathbf{x}) = \frac{1}{V} \sum \Delta(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (3.9)$$

$$\Delta(\mathbf{k}) = H \sum \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}. \quad (3.10)$$

The density field of the simulations can be determined on a regular lattice, thus it is useful to use dimensionless grid coordinates to write the polyspectrum estimator, i.e. $\mathbf{k} = k_{\text{F}} \mathbf{m}$, where \mathbf{m} is a dimensionless integer triplet (m_x, m_y, m_z) , $k_{\text{F}} = 2\pi/L$ is the fundamental frequency, and L is the large of the box.

Due to the discretisation of the data it is necessary to implement Δ_{FFT} (the discrete Fast Fourier Transform algorithm), which is implemented using the well-known FFTW³ library. This must be connected to the theoretical $\Delta(\mathbf{k})$ of Eq. (3.10) as follow

³Fastest Fourier Transform in the West (FFTW) is a software library for computing discrete Fourier transforms (DFTs). More information in: <https://www.fftw.org/>

$$\begin{aligned}
\Delta_{\text{FFT}}(\mathbf{m}) &= \sum_r \delta(\mathbf{x}) e^{-i\mathbf{x}\cdot\mathbf{k}} = \frac{\Delta(\mathbf{k})}{H}, \\
&= \sum_{\mathbf{n}} \delta(\mathbf{n}) e^{-i2\pi\mathbf{m}\cdot\mathbf{n}/N_{\text{side}}}.
\end{aligned} \tag{3.11}$$

Here, \mathbf{n} is an integer related to the co-ordinate of the grid point as $\mathbf{x} = \mathbf{n}L/N_{\text{side}}$, with N_{side} being the number of grid points of each dimension of the box. So, we can now write the polyspectrum, $\mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p)$, as a measure of a discretized data set as follow

$$\begin{aligned}
(2\pi)^3 \mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p) \delta^{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 \dots + \mathbf{k}_p) \\
&= \left\langle \prod_i^p \Delta(\mathbf{k}_i) \right\rangle, \\
(2\pi)^3 \mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p) \delta^{\text{K}}(\mathbf{k}_1 + \mathbf{k}_2 \dots + \mathbf{k}_p) \\
&\approx H^p \left\langle \prod_i^p \Delta_{\text{FFT}}(k_{\text{F}}\mathbf{m}_i) \right\rangle,
\end{aligned} \tag{3.12}$$

where we have converted the dirac delta function $\delta^{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 \dots + \mathbf{k}_p)$ to the Kronecker delta function $\delta^{\text{K}}(\mathbf{k}_1 + \mathbf{k}_2 \dots + \mathbf{k}_p)$. After some cancellations and enforcing the Kronecker delta function on the left, the polyspectrum is left as

$$\begin{aligned}
\mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p) &\approx H^p \frac{1}{V} \\
&\times \left\langle \prod_i^p \delta^{\text{K}}(\mathbf{m}_1 + \mathbf{m}_2 + \dots + \mathbf{m}_p) \Delta_{\text{FFT}}(\mathbf{m}_i) \right\rangle.
\end{aligned} \tag{3.13}$$

Because of the discretisation of the data, we must work with a bin s of width k_f (at least), then

$$\begin{aligned}
\mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p) &\approx H^p \frac{1}{V} \frac{1}{N_{\text{poly}}} \\
&\times \sum_{l_1 \pm s/2} \dots \sum_{l_p \pm s/2} \prod_i^p \delta^K(\mathbf{m}_1 + \mathbf{m}_2 \dots + \mathbf{m}_p) \Delta_{\text{FFT}}(\mathbf{m}_i) \\
&= H^p \frac{1}{V} \\
&\times \frac{\sum_{l_1 \pm s/2} \sum_{l_2 \pm s/2} \dots \sum_{l_p \pm s/2} \prod_i^p \delta^K(\mathbf{m}_1 + \mathbf{m}_2 \dots + \mathbf{m}_p) \Delta_{\text{FFT}}(\mathbf{m}_i)}{\sum_{l_1 \pm s/2} \dots \sum_{l_p \pm s/2} \delta^K(\mathbf{m}_1 + \mathbf{m}_2 \dots + \mathbf{m}_p)},
\end{aligned} \tag{3.14}$$

where $l_i = |(\mathbf{k}_i/k_F) - \mathbf{m}_i|$, and the summation is performed over all vectors \mathbf{m}_i within a bin width of \mathbf{k}_i/k_F , in other words, all k-space grids where $l_i \leq s/2$. Finally N_{poly} is the number of polygons given by $\mathbf{m}_1 + \mathbf{m}_2 \dots + \mathbf{m}_p = 0$. Note that, if the above requirements are satisfied, N_{poly} can be rewritten as a sum over the Kronecker delta function, as shown in the denominator of the second equality of Eq. (3.14). The Kronecker delta may be written as follows

$$\delta^K(\mathbf{m}_1 + \mathbf{m}_2 \dots + \mathbf{m}_p) = \frac{1}{N_{\text{pix}}} \sum_n^p \prod_i^p e^{i2\pi \mathbf{n} \cdot \mathbf{m}_i / N_{\text{side}}}, \tag{3.15}$$

where we used $\mathbf{x} = \mathbf{n}L/N_{\text{side}}$. Then, we can use Eq. (3.15) to rewrite Eq. (3.14):

$$\begin{aligned}
\mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p) &\approx H^p \frac{1}{V} \\
&\times \frac{\sum_{\mathbf{n}}^{N_{\text{pix}}} \left[\sum_{l_1 \pm s/2} \dots \sum_{l_p \pm s/2} \prod_i^p \Delta_{\text{FFT}}(\mathbf{m}_i) e^{i2\pi \mathbf{n} \cdot \mathbf{m}_i / N_{\text{side}}} \right]}{\sum_{\mathbf{n}}^{N_{\text{pix}}} \left[\sum_{l_1 \pm s/2} \dots \sum_{l_p \pm s/2} \prod_i^p e^{i2\pi \mathbf{n} \cdot \mathbf{m}_i / N_{\text{side}}} \right]}.
\end{aligned} \tag{3.16}$$

Finally, we can estimate the p^{th} -order polyspectrum rewriting $\mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p)$ as

$$\mathcal{P}(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_p) \approx H^p \frac{1}{V} \frac{\sum_{\mathbf{n}}^{N_{\text{grid}}} \prod_{i=1}^p \delta(\mathbf{n}, \mathbf{k}_i)}{\sum_{\mathbf{n}}^{N_{\text{grid}}} \prod_{i=1}^p I(\mathbf{n}, \mathbf{k}_i)}, \tag{3.17}$$

where the terms $\delta(\mathbf{n}, \mathbf{k}_i)$ and $I(\mathbf{n}, \mathbf{k}_i)$ are given by

$$\begin{aligned}
\delta(\mathbf{n}, \mathbf{k}_i) &= \sum_{l_i \pm s/2} \Delta_{\text{FFT}}(\mathbf{m}_i) e^{i2\pi \mathbf{n} \cdot \mathbf{m}_i / N_{\text{side}}}, \\
I(\mathbf{n}, \mathbf{k}_i) &= \sum_{l_i \pm s/2} e^{i2\pi \mathbf{n} \cdot \mathbf{m}_i / N_{\text{side}}}.
\end{aligned}
\tag{3.18}$$

To calculate these terms, we can create a new FFT box that gets the data $\Delta(\mathbf{k}_i)$, and where a grid vector satisfies that $\mathbf{k}_i/k_f \simeq \mathbf{m}_i$ and zero otherwise. Then, we can apply a Fourier transform to this box to move to real space to get $\delta(\mathbf{n}, \mathbf{k}_i)$. And equivalently, we can do the same to create $I(\mathbf{n}, \mathbf{k}_i)$.

For an example, we will write the bispectrum ($p = 3$) with this convention:

$$B(k_f m_1, k_f m_2, k_f m_3) \approx \frac{V^2}{N_{\text{grid}}^3} \frac{\sum_n^{N_{\text{grid}}} \delta(n, k_1) \delta(n, k_2) \delta(n, k_3)}{\sum_n^{N_{\text{grid}}} I(n, k_1) I(n, k_2) I(n, k_3)}.
\tag{3.19}$$

For our analysis, we have considered the lowest orders of the polyspectra, i.e., $p = 2$, $p = 3$, and $p = 4$. These correspond to the power spectrum, bispectrum, and trispectrum, respectively. We have used the Pylians code (Villaescusa-Navarro, 2018) where the polyspectra estimator of Eq. (3.17) is implemented. We have modified and added the trispectrum calculation to the code. Note that to apply the polyspectra estimator we must convert the particle distribution of the simulation to a density contrast field on a regular lattice. This is achieved using a standard Cloud-in-Cell assignment scheme within Pylians. In Sec. 4.1 we will show the results between the polyspectra from RAMSES and L-PICOLA, where we will see that although we have a large divergences at smaller scales, the normalized polyspectra shows a better agreement than 1%, at all scales considered in this work.

3.6 Simulations

We now give the technical details of the generation of the initial conditions, the manner in which we generate different realizations, and the details of how we implement the polyspectra analysis.

3.6.1 Initial conditions

Before explaining the generation of the initial conditions (ICs), and in order to provide clarity when indicating a model, due to the great variety that we will present, we have adopted the following naming:

- Initial condition type: G for Gaussian, NG for non-Gaussian.
- Level of non-Gaussianity: 1, 2, 3 or 4. This only applies for NG models. The meaning of these levels will be clarified shortly.
- Frequency of non-Gaussianity: f1, f2 or f3. This only applies for NG models. The meaning of these levels will be clarified shortly.
- Random realizations: r1, r2, r3, r4 or r5. This will only be necessary when we refer to one specific realization. For most results we average over all realizations.

To analyze the impact of this type of non-Gaussianity on the large-scale structure, we consider a range of amplitudes and frequencies for the oscillations of the probability density distribution of the primordial perturbations Eq. (2.42). As mentioned in Sec. 2.3, in this thesis we focus on the asymptotic approximation (i.e. $k \rightarrow \infty$) rather than the full reconstruction Eq. (2.44) for practical reasons. In Fig. 3.2 we show the form of Eq. (2.42) for different values of the amplitude A and different values of the frequency f . Note that all of our PDFs have a standard deviation of unity. We constructed 12 PDFs by choosing 4 amplitudes for 3 different frequencies (see table 3.1). The 3 panels correspond to frequencies f1, f2 and f3 respectively. Each panel shows orange, green, red and purple solid lines representing the non-Gaussian curves (hereafter non-Gaussianity levels). Each non-Gaussianity level has an overplotted Gaussian curve (dashed lines) that perfectly matches the central peaks of the oscillations of the 4 non-Gaussian levels. We can see how the non-Gaussian curves are constructed from a Gaussian PDF with an oscillatory correction. This ensures that the odd moments of the distribution are zero just as in the case of a Gaussian distribution (see Sec. 2.3 for more details). By comparison, a Gaussian distribution is also shown in each panel (blue dashed line). To construct the non-Gaussian PDFs, we vary the parameter f which modifies the frequency of the oscillations, as well as their amplitude (the f parameter appears in the coefficients that multiply the trigonometric functions in Eq. (2.42) and in their arguments), so we compensate this variation by modifying A , such that for each level of non-Gaussianity the central peak is at the same height for all frequencies. .

To generate our random numbers, which will be the seeds of the initial conditions (after convolution with the transfer function), we use the accept-reject technique. This technique consists of generating random values and accepting those which are equal to or below the PDF from which we wish to draw the sample. In more detail, two uniformly distributed random numbers are generated. The first,

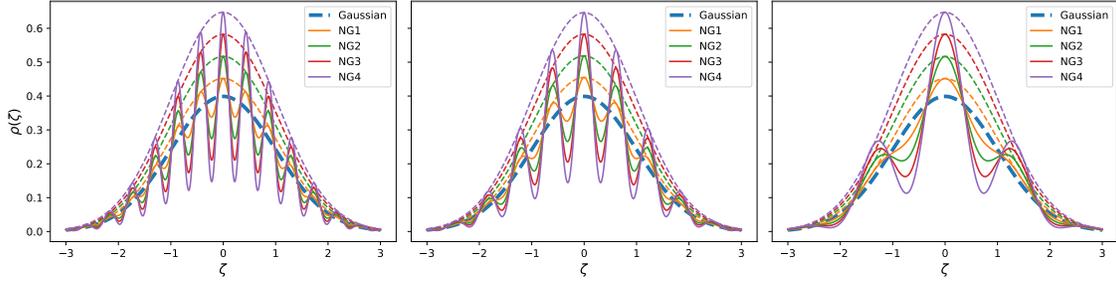


Figure 3.2: The asymptotic form of the PDF, for various values of the amplitude (levels 1 to 4) and frequency (*left panel*: f1; *middle panel*: f2; *right panel*: f3). The levels and frequency values are defined in Table 3.1.

Frequency	A , Level 1	A , Level 2	A , Level 3	A , Level 4
f1 ($f = 7 \times 10^{-2}$)	6.5×10^{-4}	1.45×10^{-3}	2.225×10^{-3}	3.05×10^{-3}
f2 ($f = 1 \times 10^{-1}$)	1.4×10^{-3}	3×10^{-3}	4.6×10^{-3}	6.2×10^{-3}
f3 ($f = 2.4 \times 10^{-1}$)	7.6×10^{-3}	1.7×10^{-2}	2.65×10^{-2}	3.585×10^{-2}

Table 3.1: Values of A and f used in the non-Gaussian correction term of the PDF for our models.

r_1 lies in the horizontal range of $\zeta_{min} = -4$ and $\zeta_{max} = 4$, while the other, r_2 , lies between zero and the maximum central peak height of the non-Gaussian PDF with the largest amplitude oscillations (level 4). Then we compare r_2 to $P(r_1)$ where P is one of the probability distribution functions that we are considering. If $r_2 \leq P(r_1)$ then the value r_1 is accepted, otherwise it is rejected. This process continues until each distribution has been sampled with N values, where N will ultimately be equal to the number of particles in our simulations and the number of grid points in the regular lattice used to calculate the polyspectra. Note that after a given PDF sample has reached N values we only continue to add points for samples that have not yet reached N values. For this work we use $N = 256^3$ for all our simulations.

The final step is to sort the non-Gaussian distributions according to the ordering of the Gaussian one (we use the same seed for the Gaussian model and the non-Gaussian models). To do this, we sort the values in the Gaussian distribution in ascending order, then we obtain the set of indices required to invert this sorting and restore the Gaussian distribution to its original ordering. These indices are then used to reorder the non-Gaussian distributions. By comparing the values of the samples one to one, we can see a small variability of the non-Gaussian values with respect

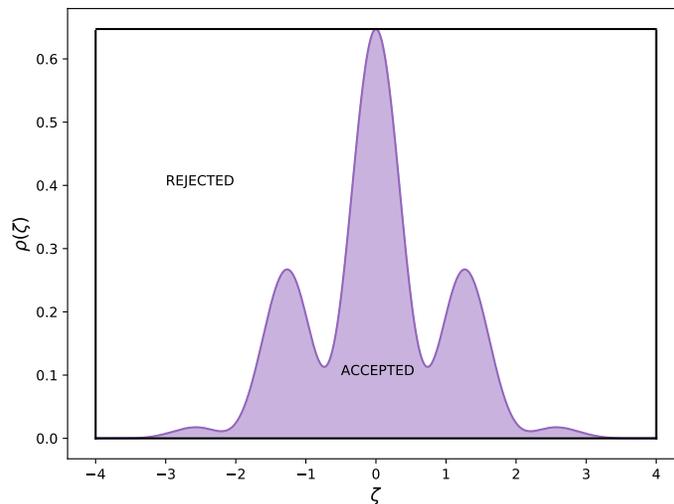


Figure 3.3: Illustration of accept-reject technique. Here we are representing only the NG4f3 model.

to the Gaussian ones, which translates to a deviation from the Gaussian case. In Fig. 3.4 we show a comparison between non-Gaussian and Gaussian values before and after the ordering. This ordering is important since it will eliminate the variance in the realizations produced by the difference in the random numbers. This will produce initial conditions that will evolve very similarly to form very similar structures, but with small differences due only to the effect of non-Gaussianity as will be shown in Chap. 4. This procedure was repeated for all levels of non-Gaussianity, all 3 frequencies and for 5 different values of the initial seed of the uniform random number generator in order to generate 5 different statistical realizations. These realizations will be averaged over to try to remove the contribution to any signal of non-Gaussianity at late times that comes from gravitational collapse, leaving only the contribution from the primordial statistics. In total, therefore, we produced initial conditions for 65 models, 5 of them being fiducial Gaussian models.

For our simulations we have used RAMSES and L-PICOLA (see sections 3.3 and 3.4). To generate the initial conditions for RAMSES we have passed the random numbers to MUSIC, where they are k-space sampled (transformed to Fourier space and multiplied with a transfer function generated by CAMB as seen in Eq. (3.2) and

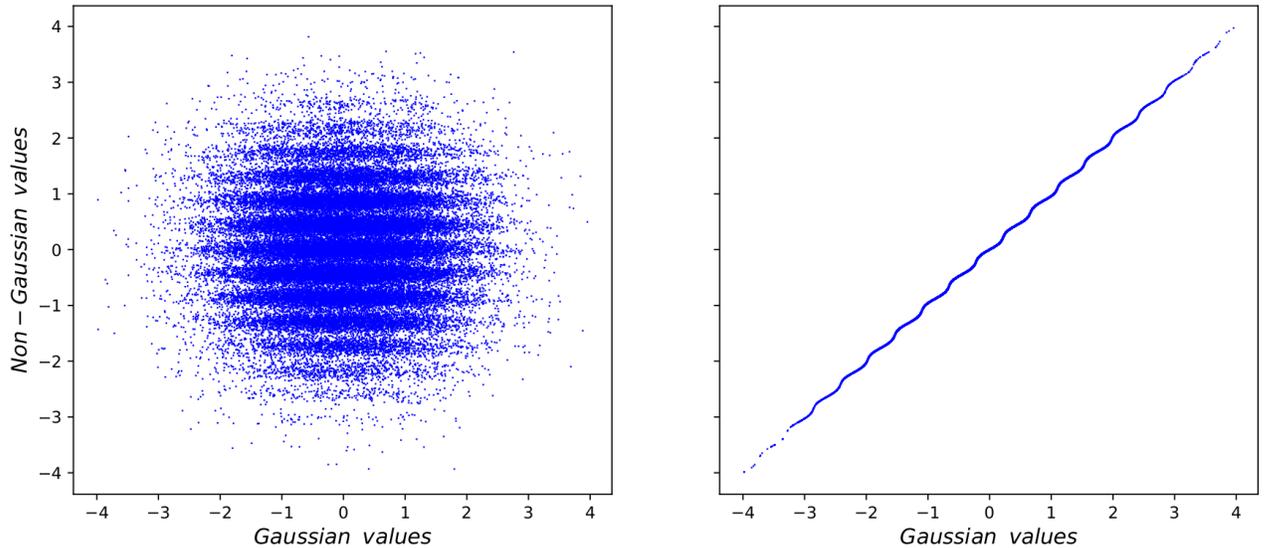


Figure 3.4: Non-Gaussian values versus Gaussian values before and after sorting (left and right respectively). For simplicity we have plotted only 5×10^4 points.

described in Sec. 3.2. Then the 2LPT method is applied to generate the initial conditions. In the case of L-PICOLA, we have modified the initial condition generation to read and use our random numbers, where the 2LPT method is also used. Therefore, the generation of the initial conditions of L-PICOLA is the same as MUSIC. Note that we use RAMSES for just one realization with frequency f1 (5 models). All other models are generated only using L-PICOLA.

For all our simulations, both in RAMSES and L-PICOLA, we have assumed a standard Λ CDM cosmology with the following parameters: $\Omega_m = 0.3$, $\Omega_b = 0.04$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.88$, and $n_s = 0.96$. In addition, for all realizations we have used a 500 Mpc box, and a particle number of 256^3 . In the case of RAMSES, which employs the AMR method, we have set a coarser grid resolution of 256^3 points (minimum level 8). The maximum refinement level (16) corresponds to a resolution of 7.6 kpc, sufficient to resolve massive halos. These conditions required approximately 8 hours of execution per simulation using 8 cores in RAMSES. In the case of L-PICOLA we have a fixed grid resolution of 256^3 points and 10 timesteps. These latter simulations required approximately 5 minutes of execution time using 1 core.

3.6.2 Polyspectra analysis

The analysis performed in this thesis will concentrate only on the lowest orders of the n -point correlation function in Fourier space. An important feature of the nature of the non-Gaussian PDF is that only the even n -point functions are non-zero, while the odd n -point functions are zero just like a Gaussian PDF. On the other hand, possible evidence of primordial non-Gaussianities is mixed with non-Gaussianities coming from the gravitational collapse of the large-scale structure at all orders of the n -point function making it difficult to decode its signature. Therefore, as preliminary work, we have decided to focus on the search for the possible signature of this primordial Gaussianity by considering the power spectrum, bispectrum, and trispectrum, which are the 2, 3, and 4-point correlation functions in Fourier space respectively.

Due to the assumption of isotropy, we can fully describe the power spectrum using only the modulus k of the spatial wavevector \mathbf{k} . This situation changes quite a bit when we consider correlation functions with $n \geq 2$. For example, for the bispectrum, we can form any triangular configuration with \mathbf{k}_1 , \mathbf{k}_2 and \mathbf{k}_3 , or any quadrilateral configuration with \mathbf{k}_1 , \mathbf{k}_2 , \mathbf{k}_3 and \mathbf{k}_4 in the case of the trispectrum. As a first step, we have focused our analyses on highly symmetric configurations. Specifically, for the bispectrum we choose $k_1 = k_2 = k_3$, and for the trispectrum we choose $k_1 = k_2 = k_3 = k_4$. It should be noted that a quadrilateral does not necessarily lie in a plane like a triangle, it is possible to have four non-coplanar points in a 3-dimensional space. If for example we take a square and fold it partially across one of its diagonals, what we would previously consider a square in 2-dimensional space would no longer be a square in 3-dimensions, i.e., a quadrilateral is not necessarily flat. This means that we have two additional degrees of freedom for the quadrilateral, given by the lengths of the diagonals $|\mathbf{k}_1 + \mathbf{k}_2|$ and $|\mathbf{k}_2 + \mathbf{k}_3|$. For our case, meanwhile, we have considered all (possibly folded) quadrilaterals with equal side lengths.

We have calculated the polyspectra using a public version of the Pylians code described in Villaescusa-Navarro (2018). We have modified the code and added the trispectrum calculation for our purposes. The calculation of all these polyspectra follows the method described in Watkinson et al. (2017) (see Sec. 3.5). We use 35 linearly-spaced bins in the range $2.2k_F \leq k \leq k_F N_k/3$, where the fundamental frequency is $k_F = 2\pi/L$, with $L = 500$ Mpc (box length), and N_k is the number of points used in our discretised Fourier space in each dimension, i.e. $N_k = 256$. We have avoided small values of k corresponding to scales close to the box size as they are highly affected by the sample variance due to the small number of k configurations.

At the other extreme we set the upper limit on k to $k_F N_k/3$ to avoid very high values of k where the estimator is expected to perform poorly (Sefusatti et al., 2016). This also ensures that we are below the fundamental limit set by the Nyquist frequency $k_{\text{Nyq}} = k_F N_k/2$.

Chapter 4

Results

4.1 Comparing RAMSES and L-PICOLA

In order to justify the use of L-PICOLA in this work, let us compare results with RAMSES. First, by checking the differences in the projected dark matter density distributions, and second, by comparing the low- n polyspectra. The projected density field at $z=0$ for RAMSES and L-PICOLA are shown in Fig. 4.1. The first row corresponds to results using Gaussian ICs (Gr1 model), while the second and third rows correspond to results using non-Gaussian ICs (NG2f1r1 and NG4f1r1 models respectively).

Comparing the simulation results between RAMSES and L-PICOLA we can see a considerable similarity in the structure distribution, especially at large scales. At small scales, the structures are much more defined in RAMSES than in L-PICOLA. The latter is expected due to the precision of the solutions in a full non-linear N-body code, which also features AMR to improve the spatial resolution of the overdensities. On the other hand, L-PICOLA has solutions based on approximations on a mesh with a fixed resolution. This mesh is equivalent to the coarse-level grid resolution that we used in RAMSES. When comparing the models (Gaussian and non-Gaussian) in both RAMSES and L-PICOLA, we see extremely similar results, making it much more difficult to see any difference. This is a product of the sorting applied to the seeds (random numbers) of the initial conditions. These differences will be more evident when we analyze the different levels of non-Gaussianity in the polyspectra. For now, we will continue to look at the differences between RAMSES and L-PICOLA.

The differences between RAMSES and L-PICOLA can be seen more easily in

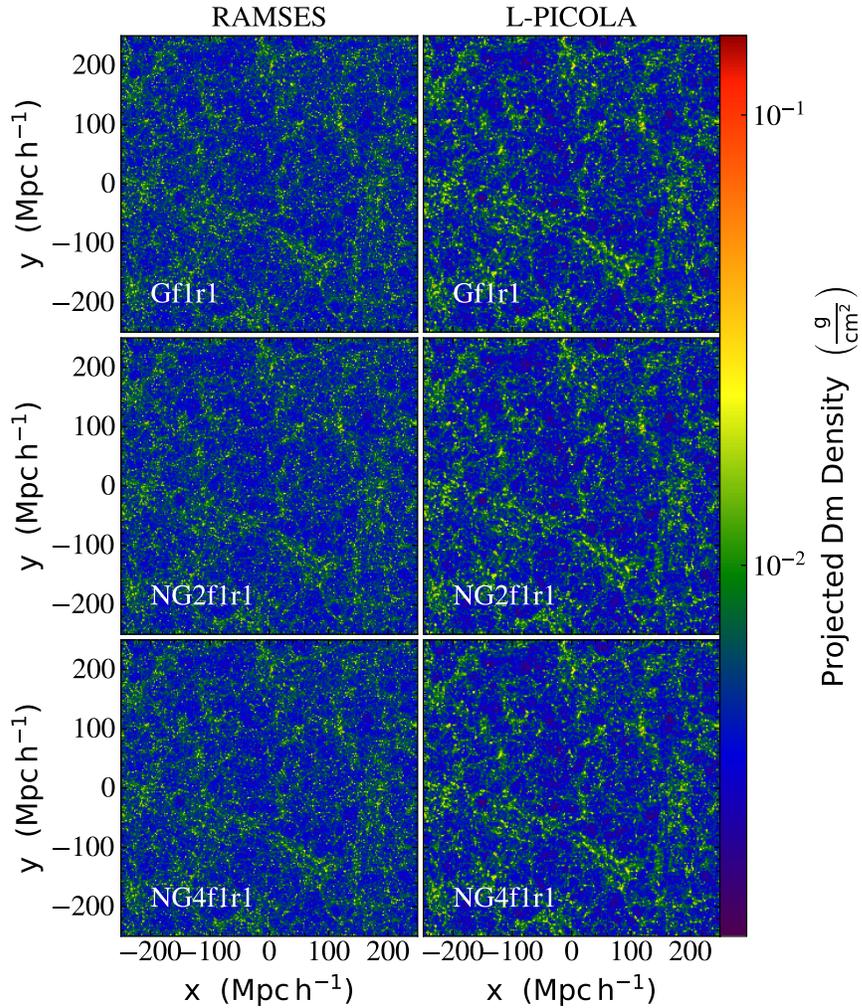


Figure 4.1: Projection density of RAMSES and L-PICOLA at $z=0$; the first and the second column respectively. *First row*: Gaussian initial conditions (model Gr1); *second row*: the NG2f1r1 model; *third row*: the NG4f1r1 model. We see a considerable similarity in the structure distribution, especially at large scales. At small scales, the structures are much more defined in RAMSES than in L-PICOLA.

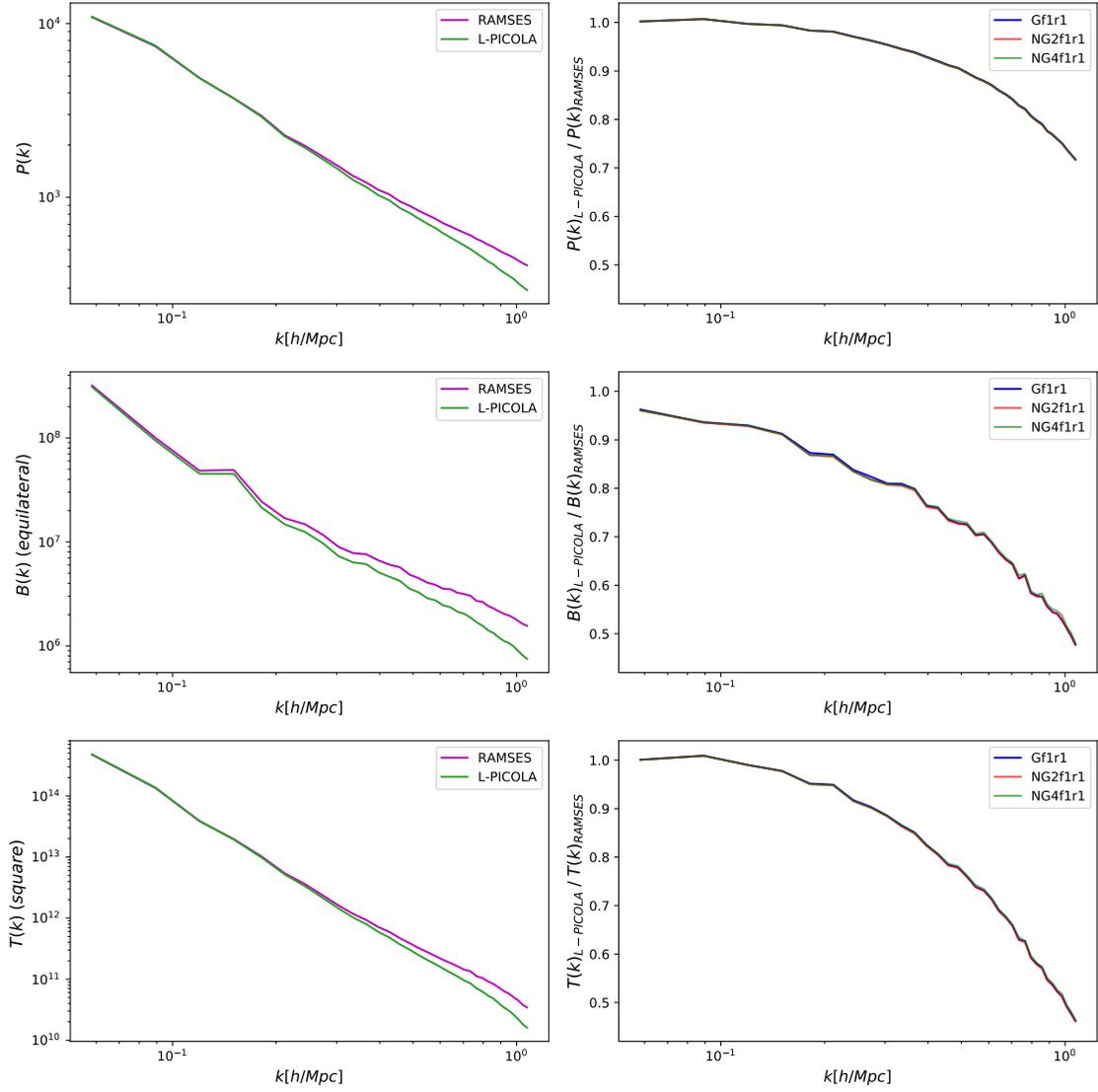


Figure 4.2: *Left column*: polyspectra of a single realization of a Gaussian simulation using RAMSES and L-PICOLA. *Right column*: The polyspectra of L-PICOLA (for one realization) normalized by those of RAMSES for models Gr1, NG2f1r1 and NG4f1r1. *Top row*: power spectrum, *middle row*: bispectrum, *bottom row*: trispectrum. Only equilateral triangle configurations are considered for the bispectrum and only square configurations for the trispectrum.

the polyspectra. In the left column of Fig. 4.2, we show the polyspectra: the power spectrum, the bispectrum, and the trispectrum (using equilateral configuration for the latter two) for RAMSES and L-PICOLA (purple and green lines, respectively) for one realization (r1). As we observed in the density field, we can see that there is good agreement at larger scales (k small), whereas, as the scale decreases (k large) the disagreement between the codes becomes more evident. In the right column, we have the L-PICOLA polyspectra normalized by the polyspectra from RAMSES, for the models Gf1r1, NG2f1r1, and NG4f1r1. The difference between L-PICOLA and RAMSES at scales larger than $k \leq 2$ h/Mpc is between 5% for the power spectrum and the trispectrum, while for the bispectrum this difference is just over 10%. For the smallest scales ($k \sim 1$ h/Mpc) the difference reaches about 25% for the power spectrum, while in the case of the bi- and trispectrum it is about 50%. We can observe two important points from the results in the right column. First, the similarity in the shapes of the curves of the Gaussian and non-Gaussian models, which is, again, a result of the sorting of the non-Gaussian random numbers. And second, there is no effective spread in the lines of the plots, indicating that the inclusion of non-Gaussianity does not affect the relationship between the results of RAMSES and L-PICOLA.

In Figure 4.3 we show the polyspectra for RAMSES and L-PICOLA for models NG2f1r1 (left column) and NG4f1r1 (right column), both normalized by the Gr1 polyspectra. Here we take advantage of what we noted earlier, that the non-Gaussianity does not affect the relationship between L-PICOLA and RAMSES. By normalising using the Gaussian polyspectra for both codes, we find that the results are extremely similar in both cases. Although the curves do not coincide exactly at all scales, in particular due to the lack of structure at small scales in L-PICOLA, we still find an excellent agreement of less than 1% at all scales. Therefore, we can have high confidence that the non-Gaussian polyspectra normalized by the Gaussian polyspectra are very well represented by L-PICOLA. Hereafter, we will refer to these quantities as *normalized polyspectra*.

4.2 Variance among realizations

The use of numerical simulations allows us to explore the statistical effect of having several realizations of the density field of the Universe. In this way, we can determine the amount of variability to be expected in measurements of n -point functions (or rather their associated polyspectra) as a product of the different structures

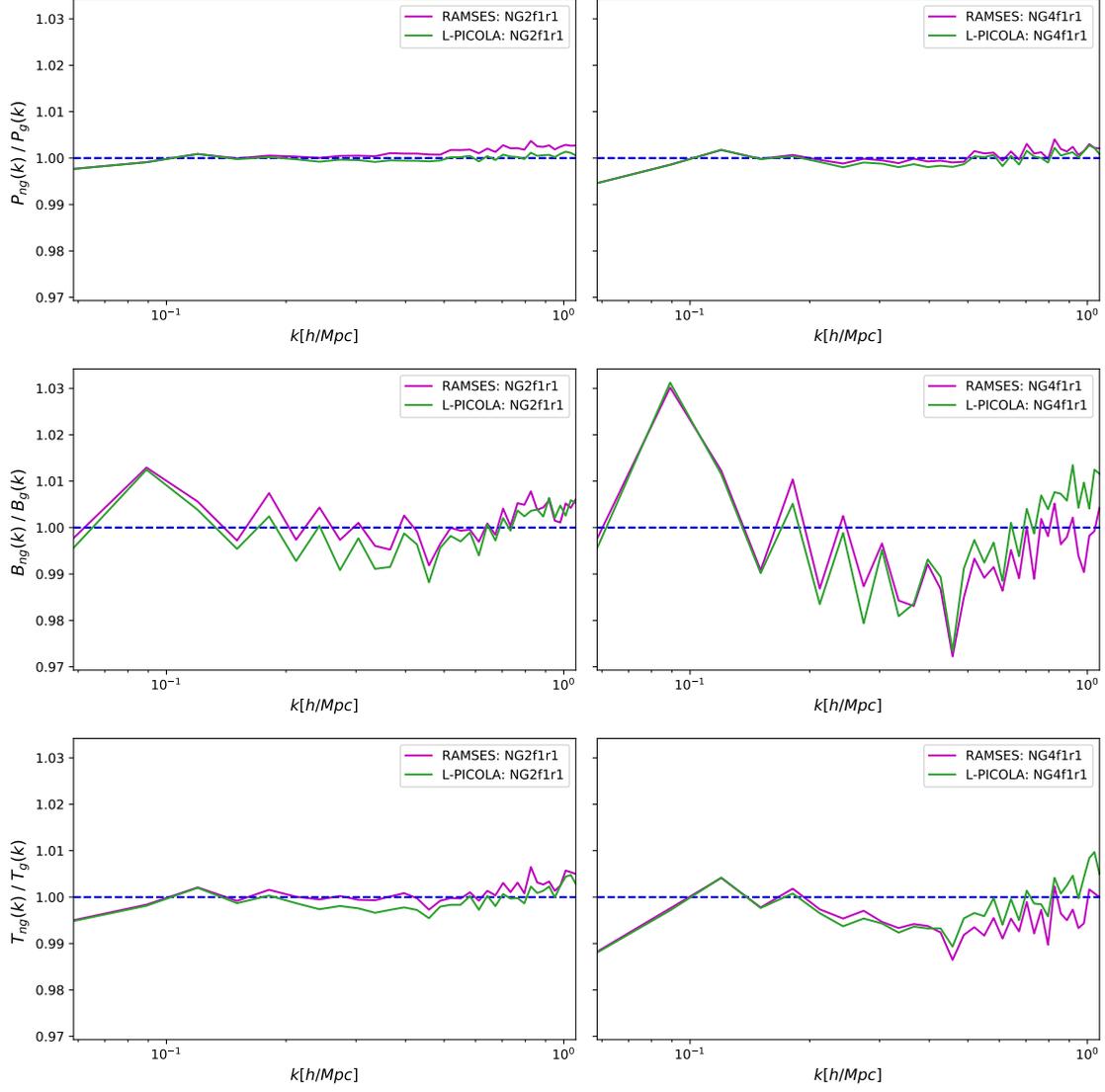


Figure 4.3: Comparison of the normalized polyspectra for L-PICOLA and RAMSES. *Left column*: NG2f1r1; *Right column*: NG4f1r1. By normalising using the Gaussian polyspectra for both codes, we find that the results are extremely similar in both cases. We find an excellent agreement, less than 1% at all scales.

arising from gravitational collapse. This variability limits the degree to which we can confidently assert that a deviation in a statistical measure is entirely due to primordial non-Gaussianity. To remove this variance, we average all the normalized polyspectra over the 5 realizations that we have for each model. We then define the quantity *averaged normalized polyspectrum*, $\Delta_{\mathcal{P}}(k)$, defined as

$$\Delta_{\mathcal{P}}(k) = \frac{1}{n} \sum_{i=1}^n \frac{\mathcal{P}_{ng}^{(i)}(k)}{\mathcal{P}_g^{(i)}(k)}, \quad (4.1)$$

where $\mathcal{P}_g(k)$ is a Gaussian polyspectrum, and $\mathcal{P}_{ng}(k)$ a non-Gaussian polyspectrum. Thus $\Delta_{\mathcal{P}}(k)$ will show deviations from the Gaussian case. The index i corresponds to the specific realization (generated by a specific initial seed), and n is the total number of realizations ($n = 5$, for our case). As mentioned before, we will only work with equilateral configurations of \mathbf{k} in this work, so we will show all polyspectra depending on the modulus k .

We will refer to the *variance* of our realizations as the range between the extreme values of the polyspectra, i.e., from the values of the polyspectra measured for each realization, we take those furthest away from the average (minimum and maximum value) and we consider them as the range in which the polyspectra vary.

In the following subsections (4.2.1, 4.2.2, and 4.2.3), but with the exception of the subsection *Deviation from the equilateral case* of 4.2.2, we will show six panels plots (two columns and three rows) with the results of the analysis of variance in the polyspectra, for models NG2 (left column) and NG4 (right column) for our frequencies f1, f2, and f3 (first, second, and third row, respectively) at $z = 0$ and $z = 2.2$. In all plots, the red solid line represents the NG2 model, and the blue solid line represents the NG4 model, while the light grey strip will show the range of variance around these lines.

4.2.1 Power spectra

In Fig. 4.4, we show the averaged normalized power spectra at $z = 0$. We can see that the variance in most of these models is always below 1%, except for the largest scales of some models. For example, for NG4f3, the variance reaches 3% at scales close to the box size. This increase in variance at larger scales is expected because of the small sample size for k -space configurations close to the size of the box.

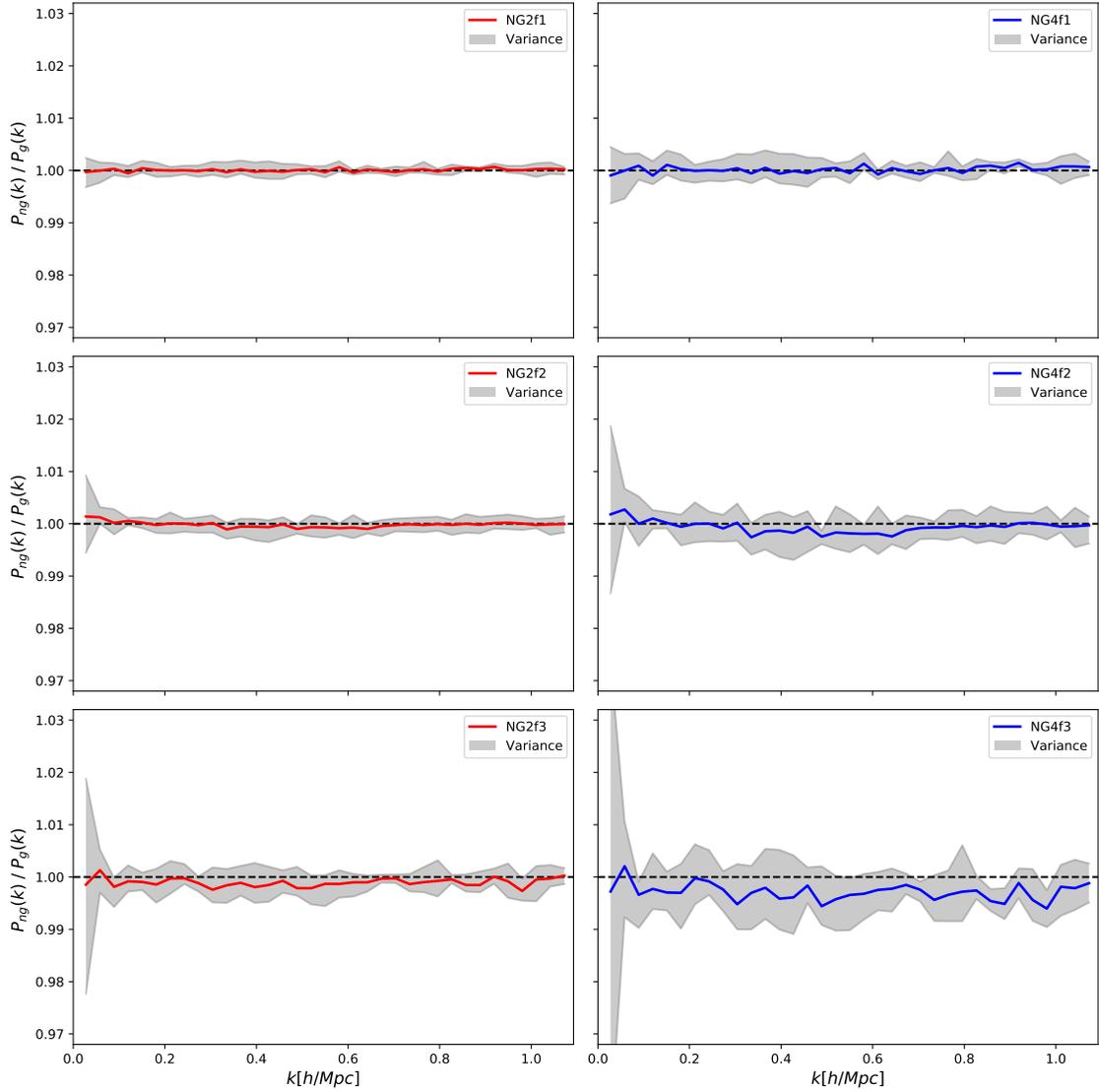


Figure 4.4: Non-Gaussian power spectra normalized with respect to the Gaussian power spectra, averaged across all realizations at $z = 0$. *Left column*: NG2 models, *right column*: NG4 models. *Top row*: frequency f1, *middle row*: frequency f2, *bottom row*: frequency f3. The red and blue lines indicate the average normalized power spectra (for the NG2 and NG4 models respectively) while the light grey strip shows the degree of variance around this average arising from the individual realizations.

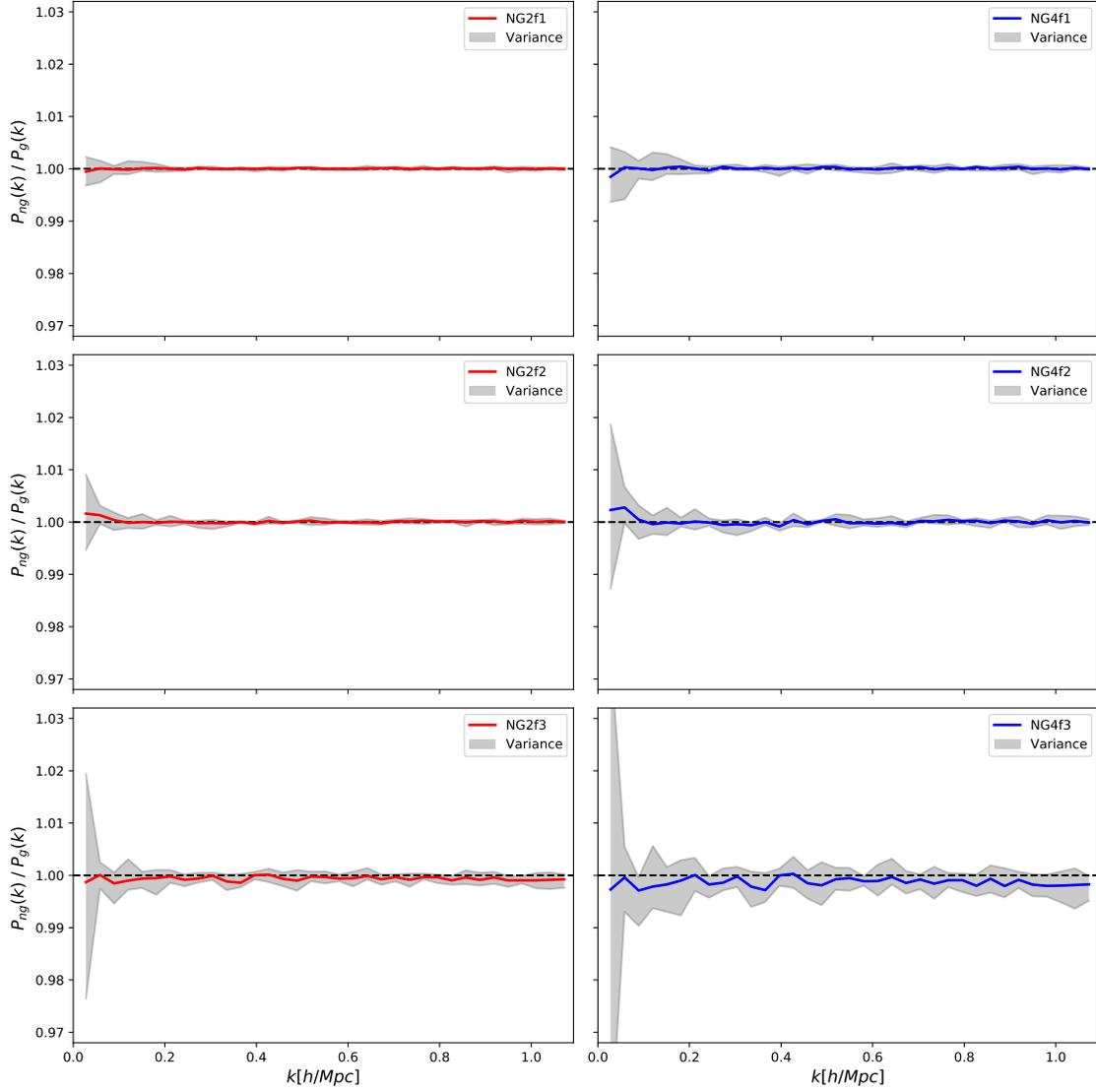


Figure 4.5: Non-Gaussian power spectra normalized with respect to the Gaussian power spectra, as for Fig. 4.4, but at $z = 2.2$. The variance is reduced in all models. This is to be expected since the non-linear processes, are reduced as the redshift increases.

If we now focus on analyzing the normalized power spectrum curve, at first look, there is no scale dependence, except for the NG4f2 model in the range between $k \sim 0.3$ h/Mpc and $k \sim 0.7$ h/Mpc, which is much more subtle for NG2f2. We will refer to this feature as a “dip”. In addition, we can see there is a rather subtle suppression of the power spectrum in the NG2f3 model, which is much more evident in NG4f3, at practically all scales. Taking into consideration the PDFs corresponding to frequencies f2 and f3 (see Fig. 3.2), we can see that a higher frequency (f2) leads to a weak suppression of the power spectrum of NG4f2, and in a limited range. While a lower frequency (f3) leads to suppression in all the scales that we show here. Note, however, that this suppression still lies within the variance band determined by considering all 5 realizations. Thus this signal is at the boundary of what we could consider to be statistically significant. Nevertheless, it is noteworthy that such a large deviation from Gaussian statistics in the initial conditions (as exemplified by the NG4f3 model) fails to give rise to a significant deviation in the late-time matter power spectrum.

In Fig. 4.5 we show the results of the normalized power spectrum averaged at $z = 2.2$. This result shows how the variance is reduced in all models. This is to be expected since it is known that the non-linear processes, which are the main cause of the formation of the structure, are reduced as the redshift increases. We also see a non-significant deviation of the power spectrum with respect to the Gaussian case, and only a suppression in the NG2f3 and NG4f3 models.

4.2.2 Bispectra

Equilateral configuration

Now, we will show the results for the bispectrum for an equilateral configuration, and three configurations deviating from the equilateral case:

In Figure 4.6 we show the results for the averaged normalized bispectra in the equilateral configuration, at $z = 0$. In this case, we see an increase of the variance in all the models at all scales, and more noticeably at large scales as expected. While the power spectrum has a variance always less than $\sim 1\%$, only the NG2f1 model has a similar variance to the rest of power spectra. The NG4f3 model shows a variance of $\sim 16\%$ (without considering the larger scales), and the rest of the models are below $\sim 10\%$. This shows us a much higher sensitivity of the bispectrum for different structures. On the other hand, here, we also have a small scale dependence, with a stronger signal than in the power spectrum (note that the vertical axis

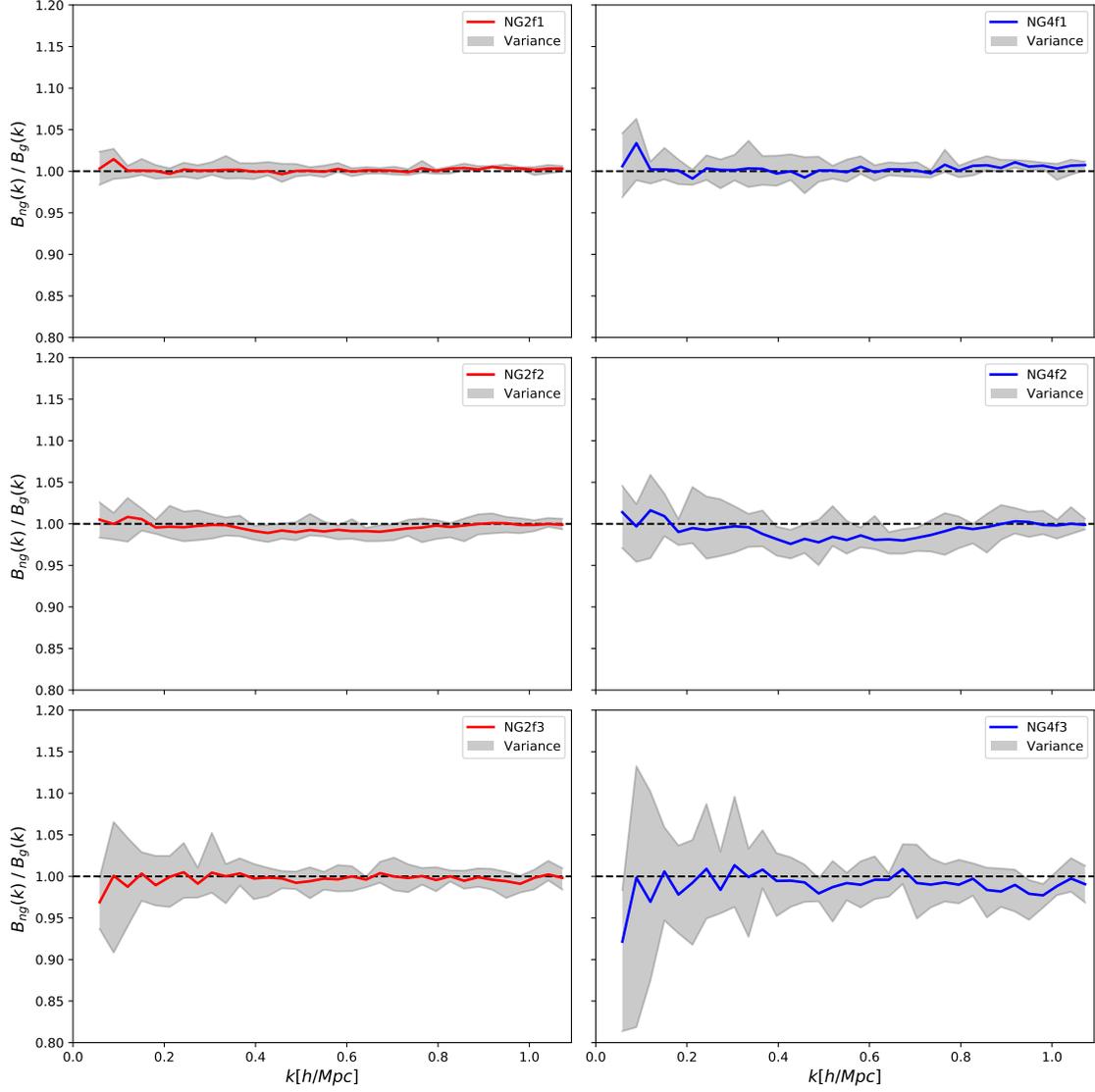


Figure 4.6: Non-Gaussian bispectra normalized with respect to the Gaussian bispectra, at $z = 0$. *Left column*: NG2 models, *right column*: NG4 models. *Top row*: frequency f1, *middle row*: frequency f2, *bottom row*: frequency f3. We can see, apparently a similar “dip” in the models NG2f2 and NG4f2 as in the power spectrum, although on a slightly wider range of scales: $0.3 \lesssim k \lesssim 0.8$ h/Mpc, and with a stronger suppression.

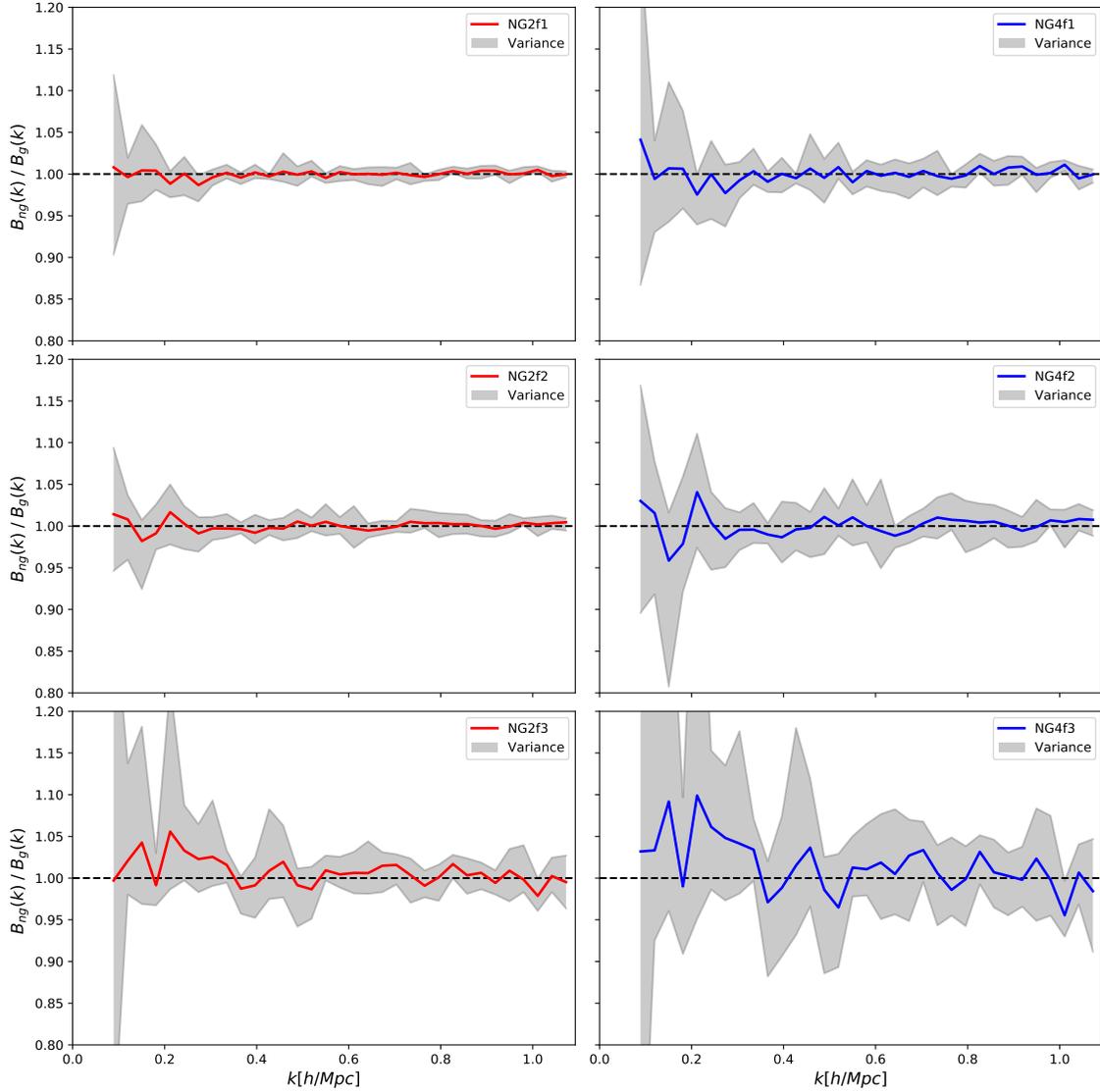


Figure 4.7: Non-Gaussian bispectra normalized with respect to the Gaussian bispectra, at $z = 2.2$. The panels are as for Fig. 4.4. Contrary to the case of the power spectrum, we see an unexpected increase in variance rather than a decrease, which is not seen in the power spectrum.

scale in Fig. 4.6 is five times larger than that of Fig. 4.4). For example, we see a slight elevation (within $\sim 1\%$) of the bispectrum in the NG2f1 and NG2f1 models at $0.8 \lesssim k \lesssim 1$ h/Mpc. Furthermore, we can see, apparently a similar “dip” in the models NG2f2 and NG4f2 as in the power spectrum, although on a slightly wider range of scales: $0.3 \lesssim k \lesssim 0.8$ h/Mpc, and with a stronger suppression due to the increased sensitivity of the bispectrum as discussed above. In addition, we do not see the same structure suppression for the NG4f3 model as seen in Fig. 4.4.

If we take into account the form of the non-Gaussianity with which we work, which is a symmetrical oscillatory correction of an underlying Gaussian PDF, we can expect that the bispectrum of this primordial non-Gaussianity is zero (as it is also for the rest of the odd n-point functions of a Gaussian distribution). On the other hand, it is known that the evolution in time of non-linearities in the structure, leads to such an asymmetry that the matter bispectrum is not zero. Now, when comparing the bispectrum resulting from the evolution of nonlinearities of the Gaussian case and the non-Gaussian case, we see that this type of primordial non-Gaussianity shows a deviation from the Gaussian one. That is, the primordial non-Gaussianity drives the nonlinearities to form structures (through gravitational collapse) that imprint a characteristic signal on the bispectrum. This is different from the signal we would expect from the nonlinearities coming from gravitational collapse driven by purely Gaussian initial conditions. This signal is marginally significant at the limit of the variance, at least with this choice of parameters.

In Fig. 4.7 we show the bispectra at $z = 2.2$. Recall that due to the large variance in the bispectrum at larger scales, resulting from the small number of k configurations used, we have decided to remove the first two values of the bispectrum for this result, so the range shown in this plot is $0.1 \lesssim k \lesssim 1$ h/Mpc. Contrary to the case of the power spectrum, we see an unexpected increase in variance rather than a decrease. This is probably indicative of the sensitivity of the bispectrum to the initial conditions, which is not seen in the power spectrum. As we have already noted, the variance tends to be much higher as the scale increases, and as the frequency is lower, with NG4f3 being the model with the highest variance, reaching a value even higher than 20%.

As a general comment, the noise of the average, minimum and maximum lines in all of the variance plots is likely due to the small number of realizations that we used. So, we could improve the smoothness of the lines with a larger number of realizations. Furthermore, the variance increases with lower frequencies due to the

larger deviation from the Gaussian PDF in these cases.

Deviation from the equilateral case

As a first test to find the signature of the non-Gaussianity in non-symmetric configurations in k -space, we have decided to test with a deviation from the equilateral case in the bispectrum. We have tested with the following configuration: $k_2 = 1.1k_1$, and $k_3 = 1.2k_1$. So in this case, we will show the results depending on one of the sides of the triangle, k_1 , as is shown in Figure 4.8. The first, second, and third panels correspond to the frequencies f1, f2, and f3, respectively. We have selected only the highest level of non-Gaussianity (NG4) to observe its effect. The red solid line represents the measurement for the non-symmetric configurations of the bispectrum. In contrast, we have plotted the symmetric configurations represented with a solid blue line.

For the first panel (frequency f1), we do not see a substantially significant deviation for the majority of scales, except for larger scales, where we see a deviation from Gaussianity greater than 15%, although the deviation is not statistically significant as it is within the limits of the variance. For the smaller scales ($k_1 > 0.3$ h/Mpc), the deviation from Gaussianity reaches $\sim 2\%$, which is not evident for the equilateral case. Between the range of scales $0.8 \lesssim k_1 \lesssim 1$ h/Mpc, we see a slight increase in the bispectrum for all k when compared to the Gaussian case, of order $\sim 1\%$. It is worth noting that the variance band is also somewhat shifted above the Gaussian line when compared to the equilateral case, thus this $\sim 1\%$ deviation is at the limit of detectability.

For frequencies f2 and f3, the differences are less obvious. We see no significant deviation from the symmetric configuration, although the variance tends to be slightly smaller in the non-symmetric configuration at larger scales, and this trend seems to be at all scales of frequency f3.

Despite these differences, there is no clear manifestation of the effect of our non-symmetric configuration, consequently, an exploration with a larger variety of non-symmetric configurations and at other orders of the n -point correlation function is needed. This, together with further exploration of amplitudes and frequencies (see the section 4.4 for constraints of the CMB), could lead us to more significant scale-dependent deviations arising from this type of non-Gaussianity.

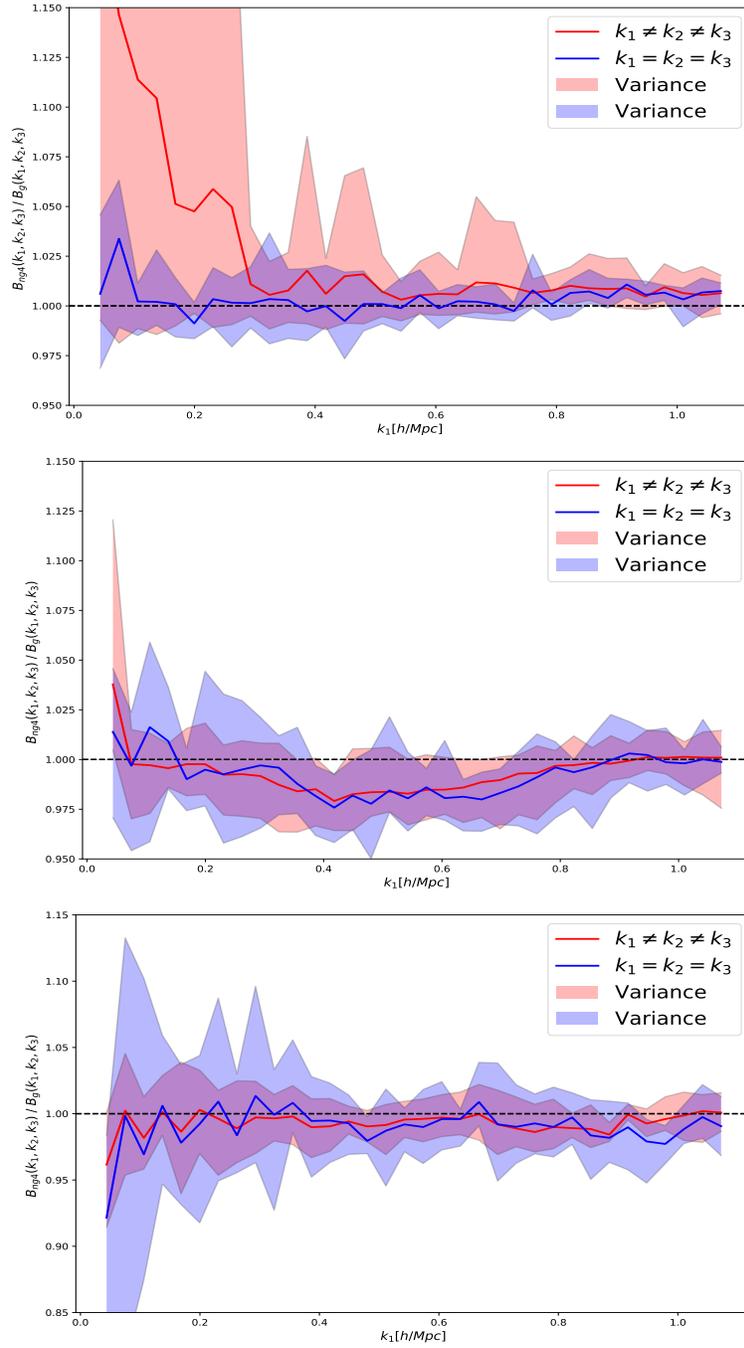


Figure 4.8: NG4 normalized bispectra, at $z = 0$. *Top row*: frequency f_1 , *middle row*: frequency f_2 , *bottom row*: frequency f_3 . *Red line*: non-symmetric configurations; *blue line*: symmetric configurations. We have tested with the following configuration: $k_2 = 1.1k_1$, and $k_3 = 1.2k_1$.

4.2.3 Trispectra

In the case of the trispectrum, we have considered working with quadrilaterals in equilateral configurations, which are generally folded in 3-dimensional space. Then we take $k_1 = k_2 = k_3 = k_4$, where we have not set a restriction on the other two additional degrees of freedom. On the other hand, it is worth remembering that the trispectrum is the analogue in Fourier space of the connected 4-point correlation function, where the disconnected parts are given by products of the power spectra, which correspond to the disconnected 2-point correlation functions (Verde and Heavens, 2001). Thus, including for the equilateral configuration considered here, the trispectrum is an independent statistical measure that goes beyond the power spectrum.

The results for the averaged normalized trispectra, at $z = 0$, are shown in Fig. 4.9. Here, we must take into consideration that we have the same range of values considered on the vertical axes of the power spectrum of Figs. 4.4 and 4.5, which shows a range of variation up to $\sim 6\%$, much smaller than the range of values shown on the vertical axes of the bispectrum of Figs. 4.6 and 4.7, with a range of variation up to 40%. With this in mind, we can more clearly analyze these results.

We see that there is an appreciably higher sensitivity in all models, in contrast to the power spectrum, but this is lower if we compare it to the bispectrum. The variance for the NG4f3 model is somewhat higher than 6% at larger scales, but averages $\sim 4\%$.

Concerning the normalized averaged curves of the models, we see an elevation in the scales $0.8 \lesssim k \lesssim 1$ h/Mpc for both models at frequency f1, similar to what occurs in the bispectrum in Fig. 4.6. For models of frequency f2, interestingly, we have again a “dip” in a similar range ($0.3 \lesssim k \lesssim 0.9$ h/Mpc) to that shown in the power spectrum and the bispectrum. In addition, the f3 frequency models show an indication of suppression, which is more evident at the smaller scales of the NG4f3 model, just as it is also the case for the power spectrum and the bispectrum. These similarities will be more carefully analyzed in section 4.3.

In the results at $z = 2.2$, for the averaged normalized trispectra of Fig. 4.10, we see a decrease in variance compared to the results at $z = 0$, and equivalent to what occurs in the power spectrum case (see Figs. 4.4 and 4.5), and different from what occurs in the bispectrum case (see Figs. 4.6 and 4.7). Additionally, we see an indication of trispectrum suppression for the f3 frequency models, with an increase

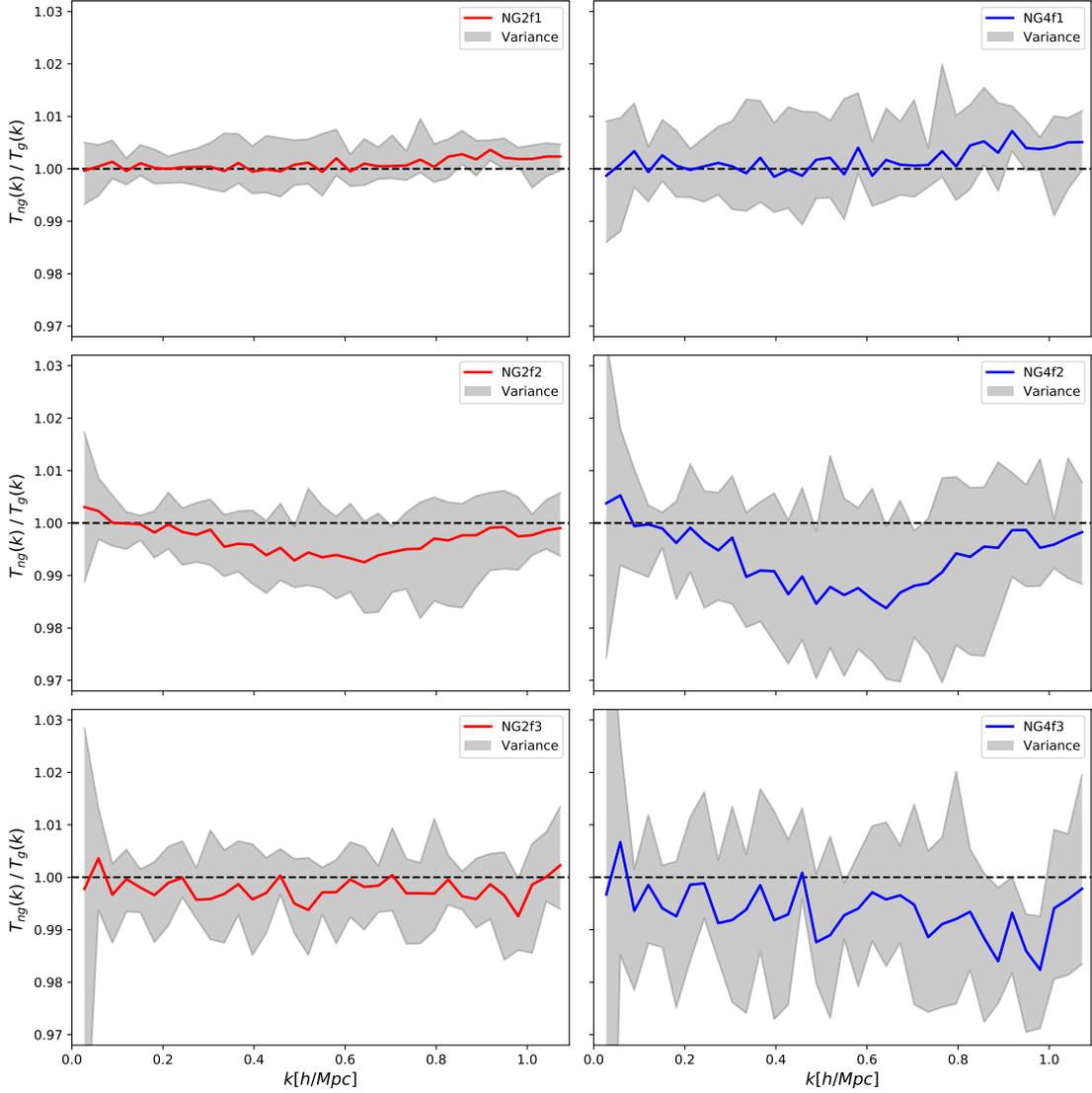


Figure 4.9: Non-Gaussian trispectra normalized with respect to the Gaussian trispectra, at $z = 0$. *Left column*: NG2 models, *right column*: NG4 models. *Top row*: frequency f1, *middle row*: frequency f2, *bottom row*: frequency f3. Again, we see, the “dip” in the models NG2f2 and NG4f2 as in the power spectrum and the bispectrum. Here, we have quadrilaterals in equilateral configurations, which are generally folded in 3-dimensional space. So, we take $k_1 = k_2 = k_3 = k_4$, where we have not set a restriction on the other two additional degrees of freedom.

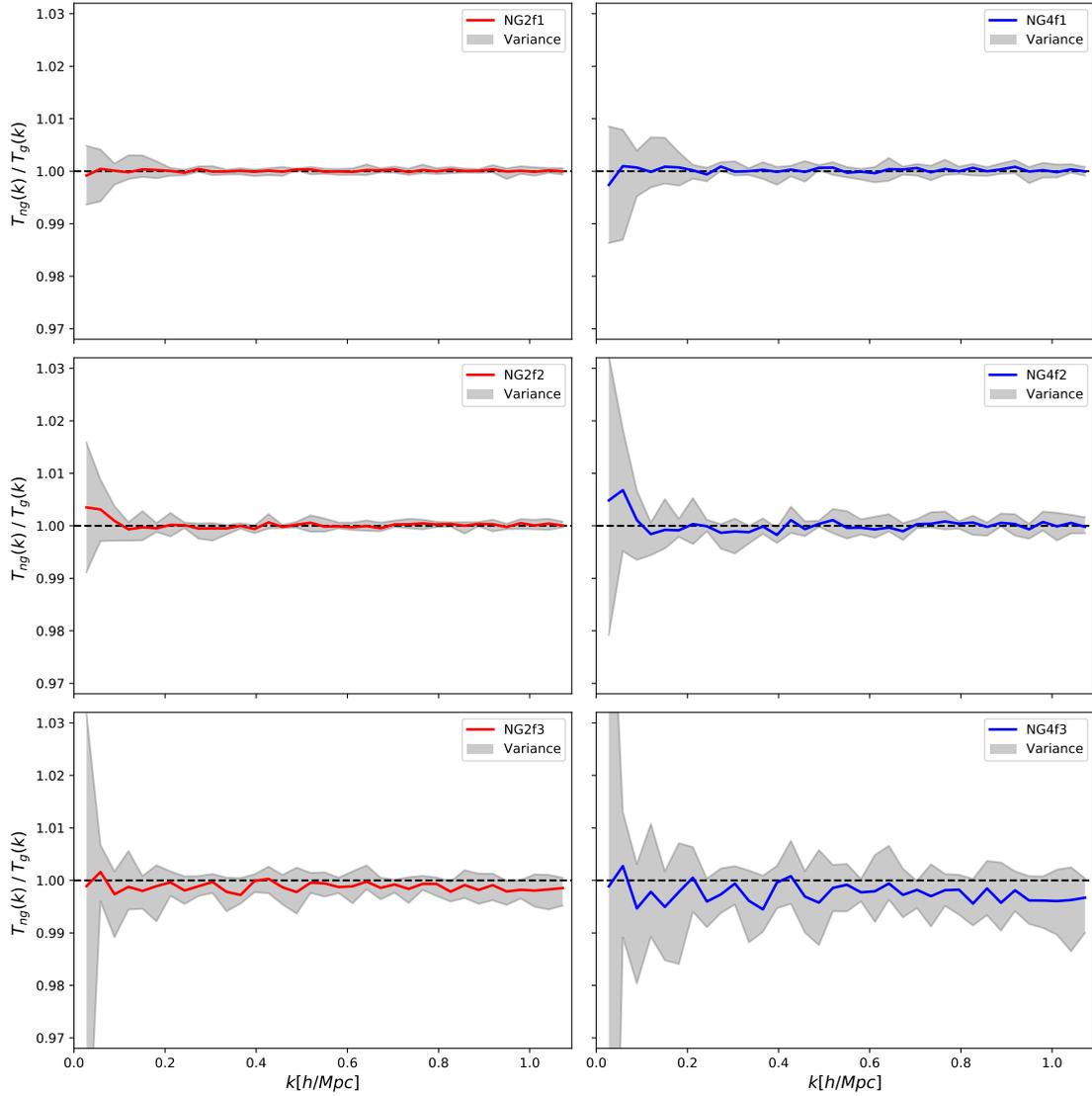


Figure 4.10: Non-Gaussian trispectra normalized with respect to the Gaussian trispectra, at $z = 2.2$. The panels are as for Fig. 4.4. The variance is reduced in all models when we compare with the results for the trispectrum at $z = 0$. Again, this is to be expected since the non-linear processes, are reduced as the redshift increases.

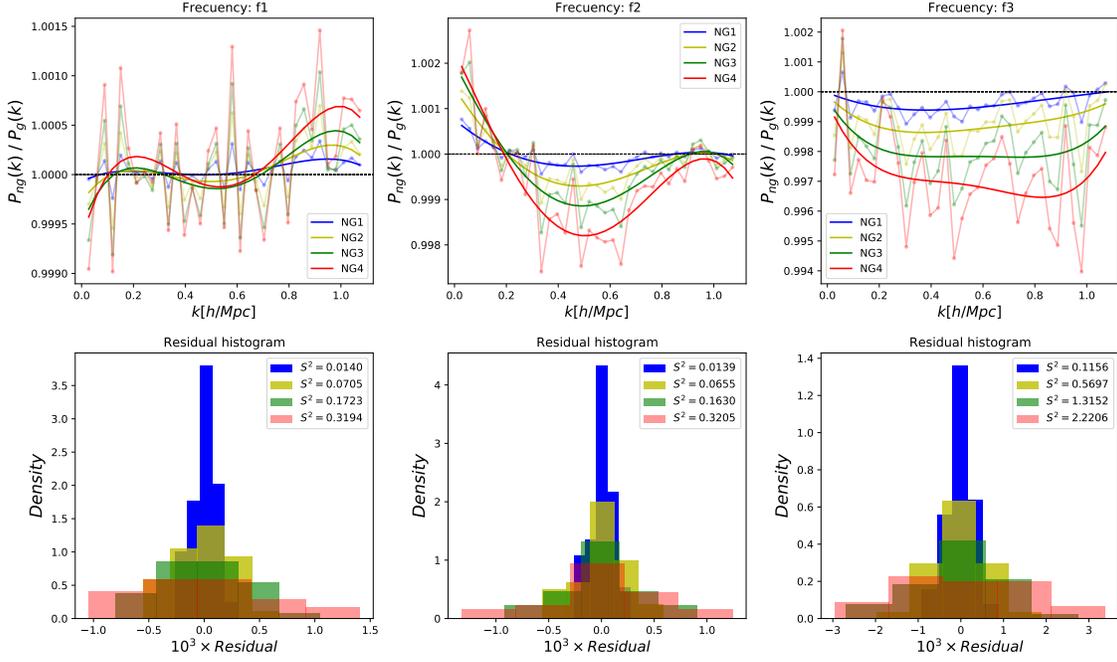


Figure 4.11: Fit lines for the averaged normalized power spectra at $z = 0$ in the first row. A 4th order polynomial fit is used for all frequencies (f1, left panel; f2, middle panel; f3, right panel). In the second row, we show the residual histograms of the adjustments. The variance S^2 of the residual histograms is also given (multiplied by 10^3 to improve legibility of the plot).

at smaller scales, just as happens for $z = 0$. The level of variance, however, implies that these results are not statistically significant.

4.3 Fitting functions

In this section, we analyze in more detail the different levels of non-Gaussianity in order to measure its degree of deviation from the Gaussian case. For this purpose, we will show results with the best-fit curve of the averaged normalized polyspectra $\Delta_{\mathcal{P}}(k)$, from which we hope to detect any systematic scale dependence and any intrinsic property of these measures in their dependence on k . We will present the results for all our models (the 4 levels of non-Gaussianity with the 3 different frequencies). In Figs. 4.11, 4.12, and 4.13, we show the results for the power spectra,

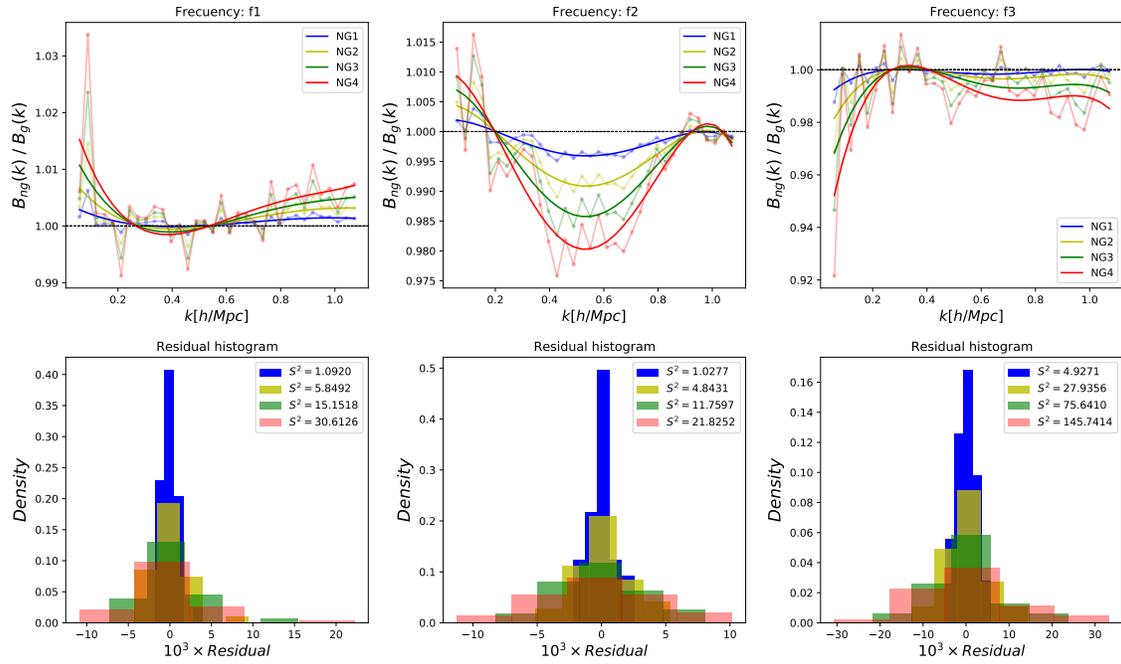


Figure 4.12: Fit lines for the averaged normalized bispectra at $z = 0$ in the first row. A 4th order polynomial fit is used for all frequencies (f1, left panel; f2, middle panel; f3, right panel). In the second row, we show the residual histograms of the adjustments. The variance S^2 of the residual histograms is also given (multiplied by 10^3 to improve legibility of the plot).

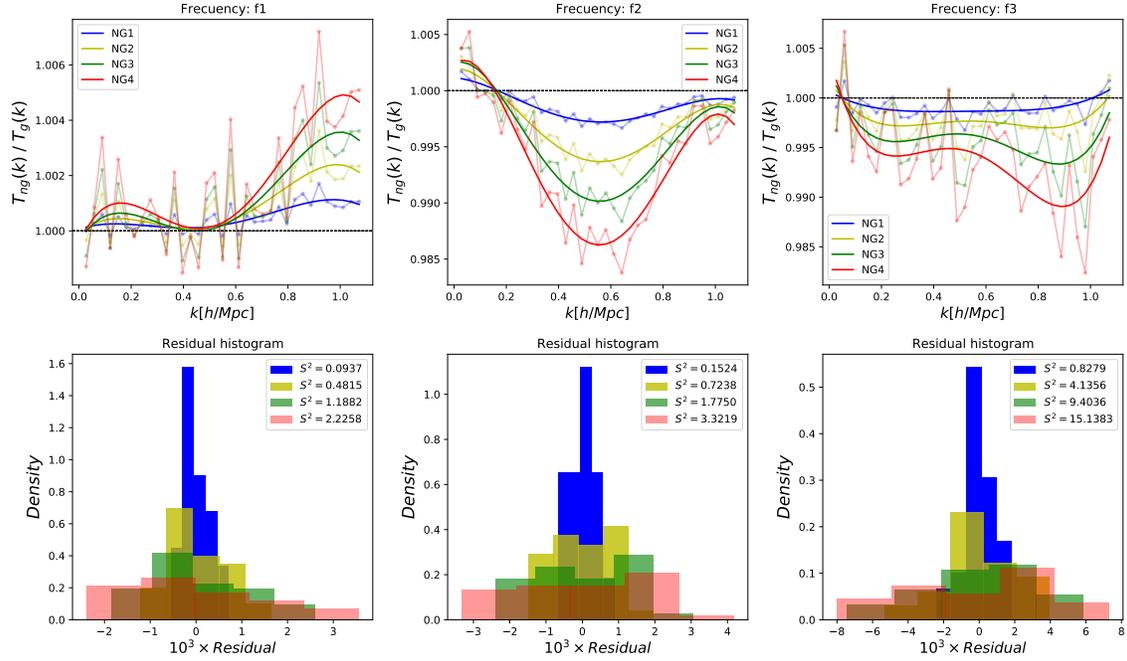


Figure 4.13: Fit lines for the averaged normalized trispectra at $z = 0$ in the first row. A 4th order polynomial fit is used for all frequencies (f1, left panel; f2, middle panel; f3, right panel). In the second row, we show the residual histograms of the adjustments. The variance S^2 of the residual histograms is also given (multiplied by 10^3 to improve legibility of the plot).

bispectra, and trispectra, respectively. The columns in all figures show the results for the frequencies f1, f2 and f3, where the frequency decreases from left to right. For all our results, we have used a polynomial of fourth degree, which represents reasonably well the trend of the data. The fitting of the different levels of non-Gaussianity (NG1, NG2, NG3, and NG4), are represented by the blue, yellow, green, and red solid lines, respectively. The goodness of fit can be seen represented in the residual histograms in each figure.

First of all, we will highlight the systematic effects that are clearly shown in the results of this section. Note that we have decreased the range of the vertical axis to have a clearer view and comparison between the fits. The clearest effect is a product of the amplitudes of the modulations of the 4 levels of non-Gaussianity. We see a systematic suppression or elevation in all polyspectra (Figs. 4.11, 4.12, and 4.13), and all frequencies. Furthermore, if we check the sample variance S^2 in the histograms, we see that it increases for each level of Gaussianity. On the other hand, comparing the sample variance of the fit across columns, we see that the frequencies f1 and f2 show very similar magnitudes (an expected result given that the frequencies f1 and f2 are comparable, as shown in Fig. 3.2). In some cases the sample variance for f1 is slightly larger than that for f2, for some levels of non-Gaussianity (see, for instance, Fig. 4.12), which is a product of the realization variance (defined in Eq.2.18) at larger scales. For the frequency f3 we see a clear effect where the residuals of the fit are much larger than for the other frequencies.

We will now take a closer look at each frequency individually. Comparing the results for frequency f1 across polyspectra (i.e. the left panels in Figs. 4.11, 4.12 and 4.13), we can see that the non-Gaussianity induces a systematic shift upwards at smaller scales in this case $k \gtrsim 0.7$ h/Mpc, $k \gtrsim 0.55$ h/Mpc, and $k \gtrsim 0.5$ h/Mpc, with maximum amplitudes of $\sim 0.06\%$, $\sim 0.7\%$, and $\sim 0.5\%$, for the power spectra, bispectra, and trispectra, respectively. While the scale dependencies in the polyspectra are very similar, the amplitude of the deviation produced in the power spectra is extremely small, while for the bispectra and trispectra they are very similar. It is worth noting that the residual histograms are unimodal and approximately centred at zero, with some mild skewness arising for the higher levels of non-Gaussianity.

Moving on to the frequency f2 (middle panel in the Figs. 4.11, 4.12, and 4.13) we see a different scale dependence in the deviation from Gaussianity, reflecting the “dip” feature referred to earlier between the scales $0.2 \lesssim k \lesssim 1$ h/Mpc. Again this feature appears for all polyspectra, with maximum amplitudes of $\sim 0.02\%$, $\sim 2\%$

and $\sim 1.5\%$, for the power spectra, bispectra, and trispectra, respectively. The residual histograms for frequency f2 across polyspectra are again approximately centred at zero, with only slight skewness in some cases. The spread in the residuals is noticeably larger in the case of the bispectrum, indicating the increased variability of this statistic with increased non-Gaussianity.

Finally, in the case of frequency f3 (right panel of Figs. 4.11, 4.12, and 4.13) the long-wavelength deviations in the primordial PDF lead to a deviation from Gaussian model, which appears as a suppression across all (accessible) scales of polyspectra, except the bispectrum between the range $0.2 \lesssim k \lesssim 0.4$ h/Mpc, with a somewhat larger deviation at smaller scales. The form of the best-fit lines is very similar for the power spectra and trispectra, with maximum amplitudes of $\sim 0.4\%$, $\sim 1.5\%$ and $\sim 1\%$, for the power spectra, bispectra, and trispectra, respectively. The residual histograms are mostly centred at zero with only mild skewness for the power spectra and bispectra, but there is more significant skewness of these histograms for the trispectra, perhaps indicating that the best-fit lines are overestimating the offset from the line of equality (where $T_{ng}/T_g = 1$) for the NG3 model, and underestimating for the NG2 and NG4 models. The spread in the residuals is also larger for the power spectra and trispectra compared to the other frequencies, and much larger for the bispectra.

We should caution that the analysis of the residuals is affected by the low numbers of points that we are considering (we calculate the polyspectra using 35 values in k -space, and we have removed the first value in the bispectrum, as stated earlier).

It is of interest to note that the most extreme case that we study (that of the NG4f3 model) does not lead to the largest deviations from Gaussianity in the $z = 0$ polyspectra (at least up to the trispectrum). It is rather the NG2f3 model that exhibits the largest deviations, and we will verify that this model is apparently consistent with CMB constraints in the following section.

4.4 CMB constraints

In this section, we will check if our models are within the allowed region of parameter space as determined by constraints on the trispectrum derived from CMB observations, according to the analysis carried out in Chen et al. (2018a).

These constraints have been calculated using the full PDF reconstruction of

the primordial curvature perturbations of Eq. (2.44), so we must relate it to the asymptotic reconstruction used in this work of Eq. (2.42). For this task, we have checked that by rescaling the values of A^2 (given by table 3.1), there is a close match between both PDFs. Therefore, we have replaced A^2 of Eq. (2.44) with αA^2 . We also found that the values of α depend most sensitively on frequency. For the models corresponding to frequency f1 we used $\alpha = 33.64$, for f2 we used $\alpha = 11.56$, and for f3 we used $\alpha = 1.69$. For our most extreme case (model NG4f3) and after rescaling, we found a maximum deviation of the asymptotic form from the full reconstruction of $12\% \pm 11\%$, which corresponds to one standard deviation around the average. The deviation of the other models is smaller, and this occurs mostly in the tails of the distributions.

In Chen et al. (2018b), the following bound on the trispectrum is obtained by using the Planck observations from CMB:

$$\beta < 2.1 \times 10^{-3}, \quad (4.2)$$

where

$$\beta \equiv A^2 \frac{\sigma_\zeta^4}{f_\zeta^4} \exp\left(-\frac{\sigma_\zeta^2}{2f_\zeta^2}\right). \quad (4.3)$$

After our scaling, we can determine where our models lie in the permitted parameter space. In Fig. 4.14 (c.f. Fig. 8 in Chen et al. (2018b)), we show the position of our models together with the allowed region of the parameter space. We have set $\sigma^2 = 1$ for all of our models.

We can see that almost all of our models are well within the permitted (orange) region, albeit with large values of the A^2 parameter. There are two models that lie just within the excluded region (white): NG3f3 and NG4f3, the two most extreme cases (these have $\beta = 2.3 \times 10^{-3}$ and 3.1×10^{-3} respectively). The frequency f2 models, however, all lie within the permitted region. These models have shown interesting scale-dependent deviations from Gaussianity in the late-time polyspectra, most significantly in the bispectrum (Fig. 4.6).

It is important to keep in mind that this constraint is determined entirely from the trispectrum. If we instead consider the full PDF of temperature fluctuations in the CMB (as discussed with respect to a different NG model in Chen et al. (2018b)) it is very likely that our primordial non-Gaussian PDFs would be excluded. The

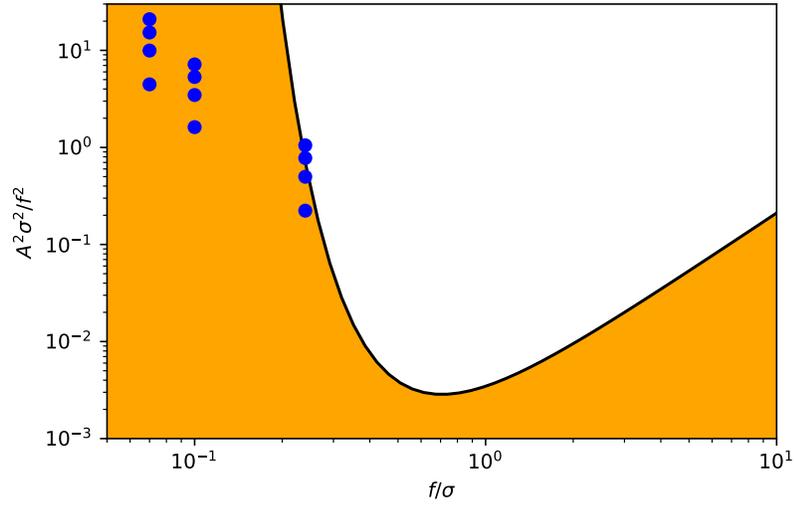


Figure 4.14: The permitted region of parameter space for the full PDF, as determined from the trispectrum, is shown in orange. The excluded region is shown in white. The models we have considered are shown as blue dots.

large amplitudes that we have studied in this work, however, allow us to investigate the feasibility of detecting this type of NG in the late-time n -point statistics. We leave a more detailed confrontation with observations for future work.

Chapter 5

Summary and Conclusions

In this work we have studied the effect of a new type of primordial non-Gaussianity, fully characterised by a probability distribution function (PDF), on large-scale structure, where the non-Gaussianity comes from an oscillatory modulation of an underlying Gaussian distribution. We have generated our initial conditions by randomly sampling from this modified PDF (using accept-reject technique). We use 12 PDFs, representing 4 levels of non-Gaussianity for 3 different frequencies. For each PDF, we generated 5 different realizations, producing 60 non-Gaussian plus 5 Gaussian realizations. Of these 12 models, the two most extreme cases were marginally excluded by the CMB constraints. These random values are convolved with the transfer function of a standard Λ CDM cosmology, to produce the initial positions and velocities using second order Lagrangian perturbation theory, which are subsequently used in both full N-body simulations as well as a mock catalogue generator called L-PICOLA. We have generated a small number of cosmological simulations with the N-body code RAMSES, where we have used the MUSIC code to generate our initial conditions. In the case of L-PICOLA, the generation of initial conditions is performed within the code, and the procedure is exactly the same.

We have focused on looking for evidence of the impact of this type of primordial non-Gaussianity on the large-scale structure by analyzing the lowest orders of the n-point correlation function in Fourier space (power spectrum, bispectrum, and trispectrum) at low redshift. First, we compared both codes across polyspectra, and although we found differences of up to 25%, 52%, and 55% in the power spectrum, bispectrum and trispectrum, respectively, at the smallest scales ($k \sim 1$ h/Mpc), we found that the normalised polyspectra is represented quite well, with differences $< 1\%$ at all scales considered in this work. Thus, we use L-PICOLA to generate all

our 65 models.

In our analysis, using the averaged normalized polyspectra, we have found scale-dependent deviations from the Gaussian model, which are at the limit of detectability. For models with frequency f_2 , we see the most significant deviations from the Gaussian case in the bispectrum. In the rest of the models, even considering the most extreme ones (with the longest wavelength in the non-Gaussian modulation), we only find deviations from the Gaussian case at the sub-percent level, which are within the sample variance of the different realizations.

The most significant result has been provided by the NG4f2 model, which has produced a deviation from Gaussianity of up to 2%. The variance indicates that the signal is at the detectability threshold, but it could be measured with sufficient precision. In addition, this signal is a scale-dependent suppression, which is present in all polyspectra: power spectra, and in the equilateral configurations of the bispectra and trispectra. On the other hand, we have found a systematic effect in all polyspectra of all models. First, produced by the variation of the amplitude of the non-Gaussianity, which translates as an elevation and/or suppression of the polyspectra in different scale ranges at low redshift. And second, we have seen that varying the frequency of the non-Gaussianity produces a systematic increase in the scatter of the data concerning the fitting curve at low redshift. We expect this feature to manifest itself at other redshifts as well.

We hope that a broader search, considering more geometrical configurations in the polyspectra, and a wider parameter space for this type of non-Gaussianity, can give us more information and evidence of deviations from Gaussianity.

As for the limitations of our study, several aspects can be improved to obtain more precision and information in our results. First, the size of the cosmological volume used in this work is limited to a 500 Mpc/h box, while the particle (mass) resolution is 256^3 particles. By improving these parameters, we would achieve a better reconstruction of the clustering, thus achieving accuracy in the measured statistics by reducing the variance. Second, we have limited ourselves to using, in general, symmetric configurations for the bispectrum and the trispectrum, and we have performed only some preliminary calculations considering deviations from the equilateral case for the bispectrum, where we have not found a significant difference in the measurement. Therefore, a more complete analysis of asymmetric configurations in the n -point correlation function is certainly necessary.

Finally, our limitations in resolution led to a lack of structure at small scales, so we have not been able to measure the mass function of halos, where it is expected that the enhanced probability for certain ranges of density peaks could lead to an oscillatory modulation in the mass function of halos in a certain mass range. There are other possible tests of non-Gaussianity beyond polyspectra, such as scale-dependent bias, Minkowski functionals, and topological measures that can also be explored.

While our results suggest this type of non-Gaussianity will be very challenging to detect, further study of the parameter space, using more observables, may uncover unexpected and novel tests of these models, helping to shine a light on the inflationary landscape.

Bibliography

- Abazajian, K. N., Adshead, P., Ahmed, Z., Allen, S. W., Alonso, D., Arnold, K. S., Baccigalupi, C., Bartlett, J. G., Battaglia, N., Benson, B. A., Bischoff, C. A., Borrill, J., Buza, V., Calabrese, E., Caldwell, R., Carlstrom, J. E., Chang, C. L., Crawford, T. M., Cyr-Racine, F.-Y., De Bernardis, F., de Haan, T., di Serego Alighieri, S., Dunkley, J., Dvorkin, C., Errard, J., Fabbian, G., Feeney, S., Ferraro, S., Filippini, J. P., Flauger, R., Fuller, G. M., Gluscevic, V., Green, D., Grin, D., Grohs, E., Henning, J. W., Hill, J. C., Hlozek, R., Holder, G., Holzappel, W., Hu, W., Huffenberger, K. M., Keskitalo, R., Knox, L., Kosowsky, A., Kovac, J., Kovetz, E. D., Kuo, C.-L., Kusaka, A., Le Jeune, M., Lee, A. T., Lilley, M., Loverde, M., Madhavacheril, M. S., Mantz, A., Marsh, D. J. E., McMahan, J., Meerburg, P. D., Meyers, J., Miller, A. D., Munoz, J. B., Nguyen, H. N., Niemack, M. D., Peloso, M., Peloton, J., Pogosian, L., Pryke, C., Raveri, M., Reichardt, C. L., Rocha, G., Rotti, A., Schaan, E., Schmittfull, M. M., Scott, D., Sehgal, N., Shandera, S., Sherwin, B. D., Smith, T. L., Sorbo, L., Starkman, G. D., Story, K. T., van Engelen, A., Vieira, J. D., Watson, S., Whitehorn, N., and Kimmy Wu, W. L. (2016). CMB-S4 Science Book, First Edition. *arXiv e-prints*, page arXiv:1610.02743.
- Achúcarro, A., Atal, V., Germani, C., and Palma, G. A. (2017). Cumulative effects in inflation with ultra-light entropy modes. *JCAP*, 2017(2):013.
- Amendola, L. and Tsujikawa, S. (2010). *Dark Energy: Theory and Observations*.
- Angulo, R. E., Foreman, S., Schmittfull, M., and Senatore, L. (2015). The one-loop matter bispectrum in the Effective Field Theory of Large Scale Structures. *JCAP*, 2015(10):039.
- Assassi, V., Baumann, D., Pajer, E., Welling, Y., and van der Woude, D. (2015). Effective theory of large-scale structure with primordial non-Gaussianity. *JCAP*, 2015(11):024.

- Baldauf, T., Mersilli, L., Mirbabayi, M., and Pajer, E. (2015). The bispectrum in the Effective Field Theory of Large Scale Structure. *JCAP*, 2015(5):007.
- Bartolo, N., Komatsu, E., Matarrese, S., and Riotto, A. (2004). Non-Gaussianity from inflation: theory and observations. *Physics Reports*, 402(3-4):103–266.
- Bennett, C. L., Hill, R. S., Hinshaw, G., Nolta, M. R., Odegard, N., Page, L., Spergel, D. N., Weiland, J. L., Wright, E. L., Halpern, M., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., and Wollack, E. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Foreground Emission. *Apjs*, 148(1):97–117.
- Berger, M. J. and Olinger, J. (1984). Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations. *Journal of Computational Physics*, 53(3):484–512.
- Buchert, T., France, M. J., and Steiner, F. (2017). Model-independent analyses of non-Gaussianity in Planck CMB maps using Minkowski functionals. *Classical and Quantum Gravity*, 34(9):094002.
- Celoria, M. and Matarrese, S. (2018). Primordial Non-Gaussianity. *arXiv e-prints*, page arXiv:1812.08197.
- Chen, X. (2010). Primordial Non-Gaussianities from Inflation Models. *Advances in Astronomy*, 2010:638979.
- Chen, X., Palma, G. A., Riquelme, W., Scheihing, B. H., and Sypsas, S. (2018a). Landscape tomography through primordial non-Gaussianity. *Phys. Rev. D*, 98(8):083528.
- Chen, X., Palma, G. A., Scheihing H., B., and Sypsas, S. (2018b). Reconstructing the Inflationary Landscape with Cosmological Data. *Phys. Rev. Letters*, 121(16):161302.
- Cicoli, M. (2013). Axion-like Particles from String Compactifications. *arXiv e-prints*, page arXiv:1309.6988.
- Colless, M., Peterson, B. A., Jackson, C., Peacock, J. A., Cole, S., Norberg, P., Baldry, I. K., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R., Collins, C., Couch, W., Cross, N., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Sutherland, W., and Taylor, K. (2003). The 2dF Galaxy Redshift Survey: Final Data Release. *arXiv e-prints*, pages astro-ph/0306581.

- Desjacques, V., Seljak, U., and Iliev, I. T. (2009). Scale-dependent bias induced by local non-Gaussianity: a comparison to N-body simulations. *MNRAS*, 396(1):85–96.
- Enqvist, K. and Väihkönen, A. (2004). Non-Gaussian perturbations in hybrid inflation. *JCAP*, 2004(9):006.
- EUCLID (2021). <https://sci.esa.int/web/euclid/>.
- Giannantonio, T. and Porciani, C. (2010). Structure formation from non-Gaussian initial conditions: Multivariate biasing, statistics, and comparison with N-body simulations. *Phys. Rev. D*, 81(6):063530.
- Gil-Marín, H., Noreña, J., Verde, L., Percival, W. J., Wagner, C., Manera, M., and Schneider, D. P. (2015). The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - I. Bias and gravity. *MNRAS*, 451(1):539–580.
- Gil-Marín, H., Percival, W. J., Verde, L., Brownstein, J. R., Chuang, C.-H., Kitaura, F.-S., Rodríguez-Torres, S. A., and Olmstead, M. D. (2017). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies. *MNRAS*, 465(2):1757–1788.
- Gordon, C., Wands, D., Bassett, B. A., and Maartens, R. (2000). Adiabatic and entropy perturbations from inflation. *Phys. Rev. D*, 63(2):023506.
- Groot Nibbelink, S. and van Tent, B. J. W. (2000). Density perturbations arising from multiple field slow-roll inflation. *arXiv e-prints*, pages hep-ph/0011325.
- Groot Nibbelink, S. and van Tent, B. J. W. (2002). Scalar perturbations during multiple-field slow-roll inflation. *Classical and Quantum Gravity*, 19(4):613–640.
- Grossi, M., Dolag, K., Branchini, E., Matarrese, S., and Moscardini, L. (2007). Evolution of massive haloes in non-Gaussian scenarios. *MNRAS*, 382(3):1261–1267.
- Grossi, M., Verde, L., Carbone, C., Dolag, K., Branchini, E., Iannuzzi, F., Matarrese, S., and Moscardini, L. (2009). Large-scale non-Gaussian mass function and halo bias: tests on N-body simulations. *MNRAS*, 398(1):321–332.
- Groth, E. J. and Peebles, P. J. E. (1977). Statistical analysis of catalogs of extragalactic objects. VII. Two- and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies. *ApJ*, 217:385–405.

- Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D.*, 23(2):347–356.
- Hahn, O. and Abel, T. (2011). Multi-scale initial conditions for cosmological simulations. *MNRAS*, 415(3):2101–2121.
- Hikage, C., Coles, P., Grossi, M., Moscardini, L., Dolag, K., Branchini, E., and Matarrese, S. (2008). The effect of primordial non-Gaussianity on the topology of large-scale structure. *MNRAS*, 385(3):1613–1620.
- Hockney, R. W. and Eastwood, J. W. (1988). *Computer simulation using particles*.
- Howlett, C., Manera, M., and Percival, W. J. (2015). L-PICOLA: A parallel code for fast dark matter simulation. *Astronomy and Computing*, 12:109–126.
- Jeong, D. (2010). *Cosmology with high ($z \gtrsim 1$) redshift galaxy surveys*. PhD thesis, University of Texas at Austin.
- Kazanas, D. (1980). Dynamics of the universe and spontaneous symmetry breaking. *APJ*, 241:L59–L63.
- Kravtsov, A. V., Klypin, A. A., and Khokhlov, A. M. (1997). Adaptive Refinement Tree: A New High-Resolution N-Body Code for Cosmological Simulations. *Apjs*, 111(1):73–94.
- Liguori, M., Sefusatti, E., Fergusson, J. R., and Shellard, E. P. S. (2010). Primordial Non-Gaussianity and Bispectrum Measurements in the Cosmic Microwave Background and Large-Scale Structure. *Advances in Astronomy*, 2010:980523.
- Maldacena, J. (2003). Non-gaussian features of primordial fluctuations in single field inflationary models. *Journal of High Energy Physics*, 2003(5):013.
- Marsh, D. J. E. (2016). Axion cosmology. *Phys. Rev.*, 643:1–79.
- Matarrese, S. and Pietroni, M. (2007). Resumming cosmic perturbations. *JCAP*, 2007(6):026.
- Palma, G. A. and Riquelme, W. (2017). Axion excursions of the landscape during inflation. *Phys. Rev. D.*, 96(2):023530.
- Palma, G. A., Scheiing Hitschfeld, B., and Sypsas, S. (2020). Non-Gaussian CMB and LSS statistics beyond polyspectra. *JCAP*, 2020(2):027.

- Peebles, P. J. E. (1980). *The large-scale structure of the universe*.
- Pietroni, M. (2008). Flowing with time: a new approach to non-linear cosmological perturbations. *JCAP*, 2008(10):036.
- Planck Collaboration, Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Benabed, K., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bond, J. R., Borrill, J., Bouchet, F. R., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Casaponsa, B., Challinor, A., Chiang, H. C., Colombo, L. P. L., Combet, C., Crill, B. P., Cuttaia, F., de Bernardis, P., de Rosa, A., de Zotti, G., Delabrouille, J., Delouis, J. M., Di Valentino, E., Diego, J. M., Doré, O., Douspis, M., Ducout, A., Dupac, X., Dusini, S., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Fantaye, Y., Fergusson, J., Fernandez-Cobos, R., Finelli, F., Frailis, M., Fraisse, A. A., Franceschi, E., Frolov, A., Galeotta, S., Galli, S., Ganga, K., Génova-Santos, R. T., Gerbino, M., González-Nuevo, J., Górski, K. M., Gratton, S., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Handley, W., Hansen, F. K., Herranz, D., Hivon, E., Huang, Z., Jaffe, A. H., Jones, W. C., Jung, G., Keihänen, E., Keskitalo, R., Kiiveri, K., Kim, J., Krachmalnicoff, N., Kunz, M., Kurki-Suonio, H., Lamarre, J. M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Le Jeune, M., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Lindholm, V., López-Caniego, M., Ma, Y. Z., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Marcos-Caballero, A., Maris, M., Martin, P. G., Martínez-González, E., Matarrese, S., Mauri, N., McEwen, J. D., Meerburg, P. D., Meinhold, P. R., Melchiorri, A., Mennella, A., Migliaccio, M., Miville-Deschênes, M. A., Molinari, D., Moneti, A., Montier, L., Morgante, G., Moss, A., Münchmeyer, M., Natoli, P., Oppizzi, F., Pagano, L., Paoletti, D., Partridge, B., Patanchon, G., Perrotta, F., Pettorino, V., Piacentini, F., Polenta, G., Puget, J. L., Rachen, J. P., Racine, B., Reinecke, M., Remazeilles, M., Renzi, A., Rocha, G., Rubiño-Martín, J. A., Ruiz-Granados, B., Salvati, L., Savelainen, M., Scott, D., Shellard, E. P. S., Shiraishi, M., Sirignano, C., Sirri, G., Smith, K., Spencer, L. D., Stanco, L., Sunyaev, R., Suur-Uski, A. S., Tauber, J. A., Tavagnacco, D., Tenti, M., Toffolatti, L., Tomasi, M., Trombetti, T., Valiviita, J., Van Tent, B., Vielva, P., Villa, F., Vittorio, N., Wandelt, B. D., Wehus, I. K., Zacchei, A., and Zonca, A. (2020). Planck 2018 results. IX. Constraints on primordial non-Gaussianity. *A&A*, 641:A9.
- Sato, K. (1981). Cosmological baryon-number domain structure and the first order phase transition of a vacuum. volume 99, pages 66–70.

- Scoccimarro, R. (1998). Transients from initial conditions: a perturbative analysis. *MNRAS*, 299(4):1097–1118.
- Sefusatti, E. (2005). *Probing fundamental physics with large-scale structure: From galaxy formation to inflation*. PhD thesis, New York University, New York, USA.
- Sefusatti, E., Crocce, M., and Desjacques, V. (2010). The matter bispectrum in N-body simulations with non-Gaussian initial conditions. *MNRAS*, 406(2):1014–1028.
- Sefusatti, E., Crocce, M., Scoccimarro, R., and Couchman, H. M. P. (2016). Accurate estimators of correlation functions in Fourier space. *MNRAS*, 460(4):3624–3636.
- Starobinsky, A. A. (1980). A new type of isotropic cosmological models without singularity. *Physics Letters B*, 91(1):99–102.
- Tassev, S., Zaldarriaga, M., and Eisenstein, D. J. (2013). Solving large scale structure in ten easy steps with COLA. *JCAP*, 2013(6):036.
- Teyssier, R. (2002). Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *A&A*, 385:337–364.
- Vera C. Rubin Observatory (2021). <https://www.lsst.org/>.
- Verde, L. and Heavens, A. F. (2001). On the Trispectrum as a Gaussian Test for Cosmology. *ApJ*, 553(1):14–24.
- Villaescusa-Navarro, F. (2018). Pylians: Python libraries for the analysis of numerical simulations.
- Wagner, C. and Verde, L. (2012). N-body simulations with generic non-Gaussian initial conditions II: halo bias. *JCAP*, 2012(3):002.
- Wagner, C., Verde, L., and Boubekur, L. (2010). N-body simulations with generic non-Gaussian initial conditions I: power spectrum and halo mass function. *JCAP*, 2010(10):022.
- Watkinson, C. A., Majumdar, S., Pritchard, J. R., and Mondal, R. (2017). A fast estimator for the bispectrum and beyond - a practical method for measuring non-Gaussianity in 21-cm maps. *MNRAS*, 472(2):2436–2446.

Weltman, A., Bull, P., Camera, S., Kelley, K., Padmanabhan, H., Pritchard, J., Raccanelli, A., Riemer-Sørensen, S., Shao, L., Andrianomena, S., Athanassoula, E., Bacon, D., Barkana, R., Bertone, G., Boehm, C., Bonvin, C., Bosma, A., Brügger, M., Burigana, C., Calore, F., Cembranos, J. A. R., Clarkson, C., Connors, R. M. T., Cruz-Dombriz, Á. d. l., Dunsby, P. K. S., Fonseca, J., Fornengo, N., Gaggero, D., Harrison, I., Larena, J., Ma, Y. Z., Maartens, R., Méndez-Isla, M., Mohanty, S. D., Murray, S., Parkinson, D., Pourtsidou, A., Quinn, P. J., Regis, M., Saha, P., Sahlén, M., Sakellariadou, M., Silk, J., Trombetti, T., Vazza, F., Venumadhav, T., Vidotto, F., Villaescusa-Navarro, F., Wang, Y., Weniger, C., Wolz, L., Zhang, F., and Gaensler, B. M. (2020). Fundamental physics with the Square Kilometre Array. , 37:e002.

Zel'Dovich, Y. B. (1970). Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations. *A&A*, 500:13–18.