

DEPARTAMENTO DE FÍSICA MODERNA
UNIVERSIDAD DE CANTABRIA

INSTITUTO DE FÍSICA DE CANTABRIA
IFCA (CSIC-UC)

ANÁLISIS COSMOLÓGICOS CON
NO-GAUSSIANIDAD PRIMORDIAL Y
MAGNIFICACIÓN DEBIDA AL EFECTO
LENTE DÉBIL

Memoria presentada para optar al título de Doctor otorgado por la
Universidad de Cantabria

por

Biuse Casaponsa Galí

Declaración de Autoría

Rita Belén Barreiro Vilas, Doctor en Ciencias Físicas y Profesor Contratado Doctor de la Universidad de Cantabria

y

Enrique Martínez González, Doctor en Ciencias Físicas y Profesor de Investigación del Consejo Superior de Investigaciones Científicas,

CERTIFICAN que la presente memoria

**Análisis cosmológicos con no-Gaussianidad primordial y magnificación
debida al efecto lente débil**

ha sido realizada por Biuse Casaponsa Galí bajo nuestra dirección en el Instituto de Física de Cantabria, para optar al título de Doctor por la Universidad de Cantabria. Consideramos que esta memoria contiene aportaciones científicas suficientemente relevantes como para constituir la Tesis Doctoral de la interesada.

En Santander, a 18 de Marzo de 2014,

Rita Belén Barreiro Vilas

Enrique Martínez González

A la Iaia Trini

Agradecimientos

Primero de todo, y aunque no esté de moda, me gustaría empezar por agradecer al Gobierno de España, por las becas recibidas al estudiar la licenciatura de física, porque cuando tienes pocos recursos, pagar las matrículas académicas no es fácil. También por ofrecer becas FPI, de la que he sido beneficiaria, y que me ha permitido formarme como investigadora cobrando un sueldo y cotizando. Por ofrecer estancias breves en centros extranjeros, donde he colaborado con expertos cosmólogos reconocidos internacionalmente. Gracias por todos estos recursos destinados a formarme como investigadora y espero que nuevos jóvenes estudiantes puedan seguir gozando de estos esfuerzos, si se desea apostar por la investigación.

Hay muchas personas sin las que su ayuda esta tesis no habría sido posible. En primer lugar a Belén Barreiro y Enrique Martínez González, mis directores de tesis, por su paciencia y ánimos. A Belén por su disponibilidad desde el primer día hasta el final, por la confianza depositada en mí, y ayudarme a salir de mis *panick attacks*. A Enrique por las ganas que pone en lo que hace y por querer abarcar tantas cosas, que aunque puede ser un “peligro”, hace que aprendamos cosas nuevas cada día.

Esta tesis tampoco habría sido posible sin la ayuda de los investigadores que me acogieron durante mis estancias en centros extranjeros: a Mike Hobson y Michael Bridges de la Universidad de Cambridge, y Alan Heavens y Tom Kitching del *Institute for Astrophysics* de Edinburgo. Quiero dar las gracias también a Patricio Vielva, por estar siempre disponible para dudas y debates sobre mi trabajo y a mis compañeros de despacho, en el cual siempre ha imperado la ley del gallinero, así que voy a ir por orden de antigüedad: a Andres Curto, por su gran ayuda al empezar mi aventura en Santander y los momentos *Will and Grace* en Cambridge. A Luis por buen compañero de despacho, de deportes y de viajes, y por sorprenderme, me gustan las sorpresas. A Raúl, por compartirlo casi todo, inicios, viajes, escuelas, observaciones, crisis, desahogos y risas. A Airam por aportar la parte teórica en el despacho, generando discusiones interesantes Edding en mano. Finalmente a David, por darle un toque de aparente normalidad y seriedad al despacho, aunque hacerse 140 Km en bicicleta en 8 horas no me parece a mí muy normal, pero sí admirable. También quería agradecer a los demás componentes del grupo por los cafés, las comilonas y la compañía durante los años de mi tesis. A Patxi, Diego, Marcos C., Marcos L-C, Quincho, Laura, Chema y los que han pasado por allí, en especial a mis italianos preferidos, Francesco Pacci y Marina Miggiaccio.

Quiero agradecer especialmente a Nuria y a Jordi, por dejarme hablarles de mi trabajo, mis dudas, mis miedos, mis alegrías y *per ser dels que m'agraden a mi, dels valents, per no fer les coses quan toquen, si no quan ens passa pel "forro", qualitat que els post-docs de "pacotilla" no tenen. Perquè sigueu feliços allà on aneu.* A David Moya, por generoso, por los cafés diarios, por los postres y por el mejor bailoteo de mis últimos 5 años en el Palacio de la Magdalena. También quiero mencionar a los otros doctorandos, Anuar, Judith, Paqui, Diego Tuccillo y Javi Brochero y a los futboleros, con los que disfruto jugando cada semana, en especial a Gerva, por aportar su grano de arena en esta tesis.

No quiero dejarme lo más importante, mi familia. Primero de todo a mis padres, *als quatre, a tu mama per donar-me no se lo que és, que em fa tirar endavant sempre i que em dona el coratge per enfretar-me al reptes més extranyes, a tu Jordi, per estar al seu costat, al nostre costat sempre, a tu papa, per confiar en mi així, a cegues, a tú Rosa, per fer-me riure i per ensenyar-me una manera diferent de prendre's la vida.* A mis hermanos *Albert, David, Sergi, Jordi, cunyades i nebots i a tu princesa, que no se què faria sense tu.* A mis amigas, las que se quedaron en Barcelona y las que andan por ahí, en especial a *l'Anna, perquè no es pot ni imaginar tot el que m'ha ensenyat* y a Regi, por empezar también sus aventuras, por valiente y guapa. Una mención muy especial a Joh por acompañarme en este viaje desde casi el primer momento que pisé la Universidad de Barcelona, y por dedicar tiempo y creatividad para dibujar la portada de esta tesis entre *paper* y *paper*, y a Jordi Salvadó por leer y comentar partes de la tesis.

Por último, porque es el más importante, a David Iehl, porque sin su apoyo, esto no habría sido posible, o sí, pero con menos estabilidad emocional (aún). Por que confías tanto en mí, que me da miedo defraudarte, aunque sé que no lo hago. Por escuchar con cara de interés todas mis charlas, y las que te quedan, sin quejarte y dándome consejos y por leerte esta tesis. Por recordar lo que es fnl y la redes neuronales y por seguir pensando en acompañarme en mis incertidumbres. *Je t'aime.*

DEPARTAMENTO DE FÍSICA MODERNA
UNIVERSIDAD DE CANTABRIA

INSTITUTO DE FÍSICA DE CANTABRIA
IFCA (CSIC-UC)

COSMOLOGICAL ANALYSES WITH PRIMORDIAL NON-GAUSSIANITY AND WEAK LENSING MAGNIFICATION

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics

by

Biuse Casaponsa Galí

Prologue

During the past few decades, large collaborations, technology innovation and a dose of creativity has led the knowledge of our Universe to grow considerably, giving rise to a standard model to explain the Universe content and evolution. The high precision cosmological observations have been and will be very important to obtain a detailed picture of our Universe, and it is the first time that those can have an impact in physics research in general. Key questions have been opened to be solved by all the scientific community, such as the mechanism for inflation and the nature of dark matter and dark energy, components of the total energy density that might require new physics to be solved.

The most notable cosmological information has been obtained from the cosmic microwave background (CMB), large scale structure distribution and supernovae Ia observations. During the next years the cosmological history should be continued with observations of weak lensing, baryonic acoustic oscillations and CMB polarization surveys, among other probes. The PhD thesis presented here is focused on two of these topics, CMB and weak lensing. The exploitation of the CMB data has obtained a very accurate picture of the early Universe, giving hints on the inflationary epoch using the non-Gaussian features in the temperature maps. The CMB consolidated the theory of the Big Bang, and has set constraints on the cosmological parameters at unprecedented levels. The picture will be completed with probes dependent on the redshift that will allow a study through different epochs of the Universe. The analysis of weak lensing signal, allows a reconstruction of the cold dark matter distribution, and has information also of the dark energy form. Although theoretically the cosmological information is very clear, the lensing effects are difficult to measure with the required precision for cosmological analysis. However, the community has made an important effort to achieve the necessary precision. In that sense, forecasts for future experiments are very encouraging.

The aim of this thesis is to contribute with new tools to the analysis of those observables in order to extract the largest amount of information possible in a simple and computationally efficient way. The thesis is organized as follows. I start with an introduction with general ideas of the standard cosmological model, followed by an overview of the CMB and the physics behind the anisotropy distribution. Also, there is an introduction of the inflationary phase and the non-Gaussian features that are transferred to the CMB. Then, in a more technical part, different estimators to detect

non-Gaussianity from the CMB maps are discussed. The last part of the introduction is devoted to gravitational lensing where I present the basic ideas and formalism of weak lensing. Chapter 2 is devoted to a Gaussianity analysis of the WMAP-7yr data with a fast wavelet estimator. Chapter 3 is an extension of the work of Chapter 2, where a neural network classifier is proposed as an alternative estimator for non-Gaussianity. In Chapter 4 a deeper research into neural networks has been done, and its implications using the binned bispectrum. Chapter 5 is focused on weak lensing, in particular on the magnification effect observed by the change on galaxy sizes. The conclusions are presented in Chapter 6 and the last chapter is dedicated to a summary of the thesis in Spanish.

Contents

1. Introduction	1
1.1. Standard cosmological model	2
1.1.1. Friedmann-Robertson-Walker metric	3
1.1.2. Cosmological parameters	3
1.2. Overview of the CMB	6
1.2.1. CMB experiments	7
1.2.2. Power spectrum	8
1.2.3. CMB polarization	11
1.2.4. Sources of the CMB anisotropies	13
1.3. Primordial non-Gaussianity	16
1.3.1. Inflationary paradigm	16
1.3.2. Dynamics of inflation	18
1.3.3. Non-Gaussianities produced by inflation	23
1.4. f_{NL} estimators	26
1.4.1. Bispectrum estimators	27
1.4.2. Wavelet estimators	29
1.4.3. Classical parameter estimation	31
1.4.4. Neural networks	34
1.5. Gravitational lensing	40
1.5.1. Cosmology with weak lensing	41
1.5.2. Lensing formalism	42

1.5.3. Weak lensing observables	44
2. Wilkinson Microwave Anisotropy Probe 7-yr constraints on f_{NL} with a fast wavelet estimator	51
2.1. Introduction	52
2.2. The HEALPix wavelet	53
2.3. Methodology	55
2.3.1. Cubic statistics	56
2.3.2. Gaussianity test and f_{NL} constraints	58
2.4. Results	59
2.4.1. Gaussianity test	60
2.4.2. Constraints on the f_{NL} parameter	60
2.4.3. Point source contribution	62
2.4.4. Local study of f_{NL}	63
3. Constraints on f_{NL} from Wilkinson Anisotropy Probe 7-year data using a neural network classifier	65
3.1. Introduction	65
3.2. Wavelet estimators	67
3.3. Neural network classifier applied to non-Gaussianity	69
3.4. The f_{NL} classification network	72
3.4.1. Training data	73
3.4.2. Network architecture	74
3.4.3. Training evolution	75
3.5. Results	75
3.5.1. Application to WMAP simulations	75
3.5.2. Application to WMAP 7-year data	78
4. Exploring local f_{NL} estimators based on the binned bispectrum	81
4.1. Introduction	82

4.2.	Regression neural network applied to non-Gaussianity	84
4.3.	Binned bispectrum	85
4.4.	f_{NL} estimators	86
4.4.1.	Approximated maximum-likelihood estimator (AMLE)	88
4.4.2.	Approximated maximum likelihood estimator with diagonal co- variance matrix (AMLED)	89
4.4.3.	Neural network estimator (NNE)	89
4.5.	Implementation	91
4.5.1.	Non-Gaussian simulations	91
4.5.2.	Binning scheme	92
4.5.3.	Inpainting	92
4.5.4.	Neural network training process	94
4.6.	Results	95
5.	Size magnification as a complement to cosmic shear	101
5.1.	Introduction	101
5.2.	Method	103
5.2.1.	Weak lensing formalism	104
5.2.2.	Estimator	105
5.2.3.	<i>lensfit</i>	105
5.2.4.	Simulations	108
5.3.	Results	110
5.3.1.	Signal-to-noise	110
5.3.2.	PSF effect	111
5.3.3.	Bulge fraction	113
5.3.4.	Most Realistic Set	117
6.	Conclusions	119
6.1.	WMAP 7-yr constraints on f_{NL} with a fast wavelet estimator	119

6.2. Constraints on f_{NL} from WMAP 7-yr data using a neural network classifier	120
6.3. Exploring local f_{NL} estimators based on the binned bispectrum	121
6.4. Size magnification as a complement to cosmic shear	122
7. Resumen en castellano	125
7.1. Introducción	125
7.1.1. Fondo cósmico de microondas	126
7.1.2. No-Gaussianidad primordial	127
7.1.3. Efecto lente gravitatoria débil	127
7.2. Métodos para acotar al parámetro f_{NL}	128
7.2.1. Cotas en el parámetro f_{NL} usando la HW en datos de WMAP-7yr	129
7.2.2. Cotas en el parámetro f_{NL} en los datos de WMAP-7yr con una red neuronal de clasificación	129
7.2.3. Explorando f_{NL} con estimadores basados en el bispectro bineado	130
7.3. Efecto lente débil usando el tamaño de las galaxias	131

List of Figures

1.1.	CMB temperature anisotropy maps	8
1.2.	Multipole range exploration for different CMB experiments	10
1.3.	Planck temperature power spectrum	11
1.4.	Quadrupolar anisotropy	12
1.5.	WMAP TE and TB cross spectra	13
1.6.	Hubble radius evolution with time	18
1.7.	Inflation potential	22
1.8.	Neuron diagram	35
1.9.	Schematic diagram of a 3-layer neural network.	36
1.10.	Gravitational lensing deflection scheme	43
1.11.	Lensing distortions	44
1.12.	Magnification and number density	48
2.1.	HEALPix wavelet	55
2.2.	HEALPix wavelet WMAP-7yr cubic statistics	60
2.3.	HEALPix wavelet cubic statistics for different f_{NL}	61
2.4.	f_{NL} histograms HEALPix wavelet	62
2.5.	HEALPix wavelet cubic statistics with and without point sources	63
3.1.	Neural network classifier performance with N_{data}	75
3.2.	Neural network classification true positive rate	76
3.3.	f_{NL} neural network estimator distribution for 1000 Gaussian	77

3.4. Neural network f_{NL} estimator bias and dispersion	78
3.5. Neural network f_{NL} estimator histograms	80
4.1. Inpainting effect shown in the masked WMAP-7yr map	93
4.2. Regression neural network performance	94
4.3. Comparison of efficiency of estimators based on the binned bispectrum .	96
4.4. Comparison of the weights obtained by a neural network and the AMLE estimator	97
5.1. Sequence of steps to obtain m and c related to the convergence field . .	111
5.2. Comparison of the estimated sizes by <i>lensfit</i> with the input size for dif- ferent S/N	112
5.3. Estimated convergence for different S/N	112
5.4. m and c values for galaxies with different S/N	113
5.5. Comparison of size estimates and input sizes for different PSF scale-lengths	114
5.6. Convergence estimates for different PSF scale-lengths	114
5.7. m and c values for different PSF scale-lengths	115
5.8. Size and convergence estimates for different bulge fractions	115
5.9. m and c values for different bulge fractions	116
5.10. Convergence estimates for a realistic set with different S/N	118
5.11. m and c values for a realistic set with different S/N	118

List of Tables

1.1. Expected values for local f_{NL} for a selection of inflationary models. . . .	26
3.1. Results obtained with neural networks and weighted least squares for the HW and the SMHW.	79
4.1. Results for noiseless full sky maps for binned bispectrum based estimators	95
4.2. Results of f_{NL} estimators based on the binned bispectrum for realistic maps.	98
4.3. Results for f_{NL} binned bispectrum based f_{NL} estimators using inpainted Gaussian realizations of set 2	99
5.1. Major characteristics of galaxies used in the size magnification analysis .	109

CHAPTER 1

Introduction

Nowadays, we know that the Universe has not always been as it looks today. The knowledge and understanding of all its stages and processes will help us to clarify why we observe it as we do, and, if we are lucky, to predict how it will look in the future. Moreover, the properties of our Universe, as its geometry, and the finite speed of light allows us to receive now particles and photons coming from very primitive epochs. This fact makes the observation of the Universe a very challenging and exciting task with a large amount of information to be disentangled.

Besides of the planets, stars, galaxies and other astrophysical objects, the Universe has been observed to be completely filled with low energy photons that maintain it at a temperature of 2.73K. This background of photons is known to come from a very early stage of the Universe, and has been named the cosmic microwave background (CMB) due to its typical wavelength now. The CMB has been a very relevant source of information about the early stages of the Universe. In fact, the observation of the CMB, together with the cosmic expansion and the abundance of light elements, has consolidated the theory of the Big Bang, and set the bases for a standard cosmological model, usually called Λ CDM model.

The name of the model is given by the main contributions on the total energy density of the Universe. Λ is a cosmological constant providing an *extra* amount of energy that would explain the current accelerated state of the Universe. At present, the energy density provided by Λ is the dominant contribution to the total energy density of the Universe, but its nature remains a mystery, and is often referred to as Dark Energy.

CDM stands for cold dark matter, which would be the second main contribution to the total energy density of the current state of the Universe. The cold dark matter only interacts with ordinary matter through gravity, and its nature remains still unknown. Those two components of the Universe are the next challenges of the modern cosmology, and new experiments as J-PAS (Benítez et al., 2009), Fermi (Atwood et al., 2009), BOSS (Schlegel et al., 2011) or Euclid (Laureijs et al., 2011), among others, shall provide some clues in the next decades. The third contribution in the total energy density is due to the ordinary matter, which represents roughly the 5% of the total.

Although it may seem that our knowledge of the Universe is shallow, the constraints on the cosmological parameters, as its composition, geometry or age are very accurately determined. In that sense, there are clear suggestions about which are the directions to follow.

1.1. Standard cosmological model

The work presented in this thesis is mainly devoted to the study of the CMB, with a last part dedicated to weak lensing. The former set the bases for the standard cosmological model while the latter was a prediction of the same model. I believe it is necessary a general introduction to the Λ CDM model before I proceed to a more detailed explanation of the two subjects of my work.

The standard cosmological model, also known as concordance model, describes a Universe that started in a singularity, where the space began its expansion. This was followed by an accelerated expansion, called cosmic inflation, and the scale factor grew exponentially in a fraction of a second. Although cosmic inflation prediction agrees very well with the observations, the details of the mechanism that created the accelerated expansion are currently under discussion. In fact, there are several proposed models of inflation (this is explained in detail in chapter 3) that are difficult to discriminate, due to the distance (in time) from us and the unusual properties of the Universe at that moment.

In the Λ CDM model the dynamics of the space-time, or the description of gravitation, are given by the general relativity theory proposed by Einstein in 1916. Most of the predictions of the theory of general relativity have been observed, as the deflection

of the light by large mass densities or the existence of black holes. One of the goals of cosmology for the following years is the detection of gravitational waves, another prediction of the theory.

Although the theory works very well on cosmological scales, a theory that explains the whole Universe and its singularities should be able to embrace quantum mechanics. There is some effort being made on that direction with a quantum gravity theory, however is a theory in development.

1.1.1. Friedmann-Robertson-Walker metric

As explained above, the theory of general relativity is the theoretical framework to describe gravitation and the properties of the space-time. The distance between two events (x^μ, x^ν) is given by:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (1.1)$$

where $g_{\mu\nu}$ is the metric tensor.

One of the main pillars of the current cosmology is that the Universe is homogeneous and isotropic at large scales. This assumption simplifies the solution of the general relativity equations and the metric for an isotropic and homogeneous Universe using spherical coordinates is the Friedmann-Robertson-Walker metric:

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi) \right], \quad (1.2)$$

where $a(t)$ is the scale factor, r is the time independent comoving distance, and θ and ϕ are the polar coordinates. Furthermore, t is the cosmic time and k describes the local spatial curvature that can take the values $k = -1, 0, 1$. Note that the speed of light has been set to $c = 1$.

1.1.2. Cosmological parameters

The beauty of having a model to explain the dynamics and content of our Universe is that it can be characterised with few parameters, which can be derived from the observations. In particular the Λ CDM model has six cosmological parameters: the baryon density, the dark matter density, the Hubble parameter, the optical depth at reionisation, the spectral amplitude and the spectral index. In this introduction I briefly describe them.

■ Hubble parameter and redshift

In 1929 Edwin Hubble found that far galaxies were departing from us at higher velocity than the closest galaxies. This discovery indicated that the space-time was not a static field, but instead the spatial part was expanding. Observing the redshift between the galaxy emitted (λ_{em}) and observed (λ_{obs}) spectral lines ,

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} , \quad (1.3)$$

Edwin Hubble found a linear relation between the proper distance and the velocity of a galaxy from us:

$$v = H_0 d . \quad (1.4)$$

where H_0 is the Hubble parameter at the current cosmological time. In general, the Hubble parameter H is defined from the scale factor:

$$H \equiv \frac{\dot{a}}{a} . \quad (1.5)$$

This quantity is useful to define relevant quantities of our Universe, as for example, the Hubble horizon, that approximately defines the connected Universe at a given time:

$$r_{HS} = \frac{c}{H} . \quad (1.6)$$

■ Content of energy

The dynamics of the Universe depends on the predominant contribution of the energy density. There are three main sources: the baryonic matter, the cold dark matter and the dark energy. Although at the end of the twentieth century it was believed that the main contribution to the energy content came from the cold dark matter, studies of type Ia supernovae found that the expansion was accelerating (Riess et al., 1998; Perlmutter et al., 1999), which could happen if the main contribution to the energy density of the Universe is given by the dark energy. The most common quantity used to define the energy density content is the density parameter Ω :

$$\Omega_i = \frac{\rho_i}{\rho_c} \quad (1.7)$$

where ρ_c is called the critical density as it is the required density for the Universe to be flat, $\rho_c = \frac{3m_p^2 H^2}{8\pi}$. There are density parameters for all the possible contributions to the energy content of the Universe; the most relevant are: cold dark matter (Ω_m), baryonic matter (Ω_b), dark energy (Ω_Λ) and the curvature (Ω_k).

The observations tell us that the curvature is consistent with 0 with a precision of less than 1%, hence, for the specific case of the Λ CDM, the Universe is assumed to be flat with $\Omega_k=0$ and then $\Omega_{total} = \Omega_b + \Omega_m + \Omega_\Lambda = 1$. Therefore, given two of the density parameters the third one is determined.

■ Initial matter perturbations

Up to date, the most accepted scenario to explain the generation of the initial matter perturbations related to the large scale structure observed nowadays, is that the Universe passed through an inflationary phase. The inflationary scenario describes a very high energy initial state with quantum fluctuations of a scalar field. Under some assumptions the Universe could have expanded exponentially between the 10^{-42} and 10^{-36} seconds of age, and this is called the inflationary epoch. While the physics behind this process is still unknown, it is widely accepted that all observations can be well accommodated within the inflationary framework. This extreme expansion would have stretched the initial quantum perturbations to cosmological scales and, from them, gravity and radiative pressure would have competed to finally obtain the actual distribution of the matter in the Universe, leaving relevant footprints in the photons of the CMB. Assuming that the gravitational field Φ acts as a quantum oscillator, the primordial fluctuations are Gaussian, and the properties of those fluctuations are encoded in the power spectrum $P_\Phi(k)$:

$$\langle \Phi(\vec{k})\Phi^*(\vec{k}') \rangle = (2\pi)^3 P_\Phi(k) \delta^3(\vec{k} - \vec{k}'). \quad (1.8)$$

One specific case, is the *scale-invariant* power spectrum (also referred as Harrison-Zel'dovich spectrum) where $k^3 P_\Phi(k)$ does not depend on k . The observations indicate that the primordial power spectrum is not exactly scale-invariant, and in order to quantify the deviations, it is commonly parametrised by a power law, $k^3 P_\Phi = A_s k^{n_s-1}$. The form of the power spectrum is derived from the power spectrum of the inflaton fluctuations that would be explained in more detail in Sec. 1.3.1, therefore the spectral amplitude A_s , and spectral index n_s , depend on the inflationary model.

■ Reionisation

The universe was reionised at late times, in such scenario the photons of the CMB interact again with the electrons and the observed anisotropies of the CMB would be affected, depending on the optical depth τ of the region. This effect would be

seen in scales within the horizon at the reionisation time. The redshift of that epoch or the optical depth would be another one of the cosmological parameters of the Λ CDM model.

1.2. The cosmic microwave background

The existence of the CMB was predicted in the middle of the twentieth century by Gamow and collaborators (Alpher et al., 1948; Alpher & Herman, 1948; Gamow, 1948). While trying to explain the observed relative abundances of the lightest elements (basically H, He, Li), they proposed the generation of these elements in a stage of the Universe with particles and photons in thermal equilibrium. A consequence of this, is that the Universe would be filled with an homogeneous and isotropic radiation with a black-body spectrum with a characteristic temperature: the adequate to allow the abundance of light elements that we observe today to be generated. Once this was stated, a prediction of the wavelength and therefore temperature of this radiation at our time, was an easy task as it only depended on the redshift. Alpher & Herman (1949) and Gamow (1956) predicted the expected value of the temperature of that radiation at the present time to be a few degrees above absolute zero. Fifteen years later an isotropic radiation with a black-body power spectrum was accidentally detected by the Bell engineers Arno Penzias and Robert Wilson with a temperature of 3K, very close to the predictions (Penzias & Wilson, 1965). Around the same time of the CMB observation, Dicke et al. (1965) developed the work started by Alpher, Hermann and Gamow, re-estimating the current Universe temperature to 3.5 K, and setting the grounds of the nowadays knowledge of the cosmic microwave background.

The observation of the CMB, together with the cosmic expansion and the relative abundance of the lightest elements, placed the Big Bang theory in front of more conservative theories as the Steady State Universe (Bondi & Gold, 1948; Hoyle, 1948; Gold, 1949). Therefore, the history of our Universe was starting to be written. It was known that at a few minutes of age the Universe was composed by an initial plasma of photons, electrons and atomic nuclei, constantly interacting with each other. When the Universe was 377,000 years old, the temperature reduced enough to allow neutral atoms to form, separating matter and radiation. At that time the photons could start to travel freely, and is commonly said that the Universe became transparent to the photons, forming a homogeneous and isotropic background, the CMB. This radiation appears to come

from a spherical surface, with radius the distance that the photons have traveled since they could freely move. This is known as the last scattering surface, as most of the photons we receive today had its last interaction with matter at that epoch.

1.2.1. CMB experiments

The detection of the CMB stimulated a large number of experiments to be conducted in order to characterise this radiation. In 1989 a confirmation of the black body spectrum with a more accurate temperature estimation of 2.726 ± 0.001 was obtained with the FIRAS spectrograph installed in the COBE (COsmic Background Explorer) satellite launched by the National Aeronautics and Space Administration (NASA) (Mather et al., 1994). With the other instrument on board of this same satellite, DMR (Differential Microwave Radiometer), there were observed for the first time small statistical fluctuations on the temperature field of the order of 10^{-5} K (Smoot et al., 1992).

At the same time, a ground-based experiment, the Tenerife experiment (Watson et al., 1992), detected a common structure between the three independent frequency channels in which the telescope operated (10, 15 and 33 GHz) (Lasenby et al., 1995). The compatibility with COBE data gave great confidence that those patterns were related to the seeds of the observed distribution of matter of the Universe. Other ground-based experiments were developed to study these small fluctuations (see Lasenby et al., 1998, for a review), from those it is worth commenting the Degree Angular Scale Interferometer (Halverson et al., 2002), that detected for the first time the E-mode polarization of the CMB. It is also worth mentioning the balloon experiments, BOOMERanG (de Bernardis et al., 2000), MAXIMA (Hanany et al., 2000) or Archeops (Benoît et al., 2003) that were conducted to avoid atmospheric contamination. Additionally, two space missions were developed, the Wilkinson Microwave Anisotropy Probe (WMAP) satellite, launched by NASA in 2001 Bennett et al. (1997), and the Planck space telescope from the European Space Agency (ESA) that started observations in 2009 (Tauber, 2004). Although these two satellites had instruments to measure the polarization of the CMB, it was not their main goal. However, there are many ground based experiments devoted to this subject, as BICEP (Keating et al., 2003), MAXIPOL (Wu et al., 2007), QUIJOTE (Rubiño-Martín et al., 2010), EBEX (Reichborn-Kjennerud et al., 2010), POLARBEAR (Kermish et al., 2012), among others (see Barreiro, 2010, for a review of polarization experiments). Moreover, there are space missions proposed to be conducted in the future (André et al., 2014).

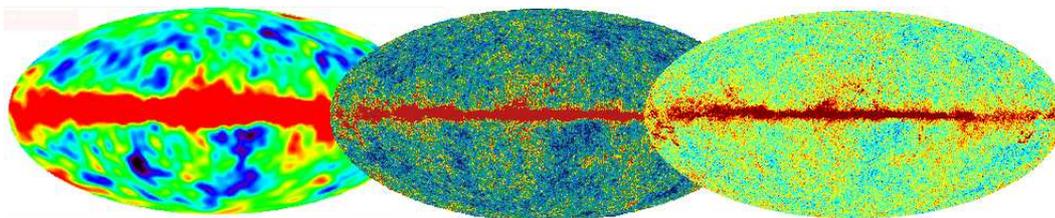


Figure 1.1 - Temperature anisotropy maps of the CMB obtained by COBE, WMAP and Planck, respectively. Images are from the website of LAMBDA (*Legacy Archive for Microwave Background Data Analysis*, <http://lambda.gsfc.nasa.gov/>).

The satellite WMAP performed full sky observations from the L2 Lagrangian point, situated at 1.5 million kilometers from the Earth. WMAP operated with 5 channels of frequency from 23 GHz to 94 GHz. The Planck satellite was launched in 2009, working at frequencies from 30 to 857 GHz and with higher resolution and sensitivity than WMAP, with the data released in March 2013. A comparison of the resolution of CMB maps of the three previously mentioned satellites is shown in fig. 1.1. Complementary ground-based experiments, such as telescope ACT (Atacama Cosmology Telescope Kosowsky, 2003) and SPT (South Pole Telescope Carlstrom et al., 2011), observe smaller regions of the sky but with higher resolution. These experiments focus on the study of the Sunyaev-Zeldovich effect and, in a second phase, on the polarization of the CMB.

1.2.2. Power spectrum

Once the observations are made, the data analyst challenge is to gather the maximum amount of cosmological information from them. Besides the importance of finding an isotropic field with a black body spectrum, allowing one to infer some general conditions of the early Universe, the CMB constitutes another source of no less enlightening information. There are small fluctuations on the temperature ($\sim 10^{-5}$ K), and its statistical distribution is related to the very primitive processes in the Universe.

Under the assumption that the temperature anisotropies distribution is Gaussian,

the information is encoded in the two first moments of the field. Therefore, it is characterized by its mean value,

$$\left\langle \frac{\Delta T}{T} \right\rangle \equiv \left\langle \frac{T - T_o}{T} \right\rangle \quad (1.9)$$

that is zero by definition and the two point correlation function

$$C(\theta) = \left\langle \frac{\Delta T(\vec{n}_1)}{T} \frac{\Delta T(\vec{n}_2)}{T} \right\rangle . \quad (1.10)$$

where θ is the separation angle between the directions \vec{n}_1, \vec{n}_2 . Due to the isotropy of the field, the two point correlation function of the CMB anisotropies only depends on the angular separation θ . The CMB photons received are emitted, or last scattered, at similar times, hence we are observing the signal over a sphere, without radial significant information. Therefore, it is very common to use the spherical harmonics to describe the CMB:

$$\frac{\Delta T(\vec{n})}{T} = \sum_{\ell=1}^{\ell=\infty} \sum_{m=-\ell}^{m=\ell} a_{\ell m} Y_{\ell m}(\vec{n}) \quad (1.11)$$

and in this case the second order moment is the power spectrum:

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = C_{\ell} \delta_{\ell \ell'} \delta_{m m'} , \quad (1.12)$$

where the dependence on m vanishes due to the isotropy of the CMB. The average on the coefficients is done over all m for a given ℓ implying larger uncertainties for lower multipoles, due to the sampling number. This is known as the *cosmic variance*:

$$\frac{\Delta C_{\ell}}{C_{\ell}} = \sqrt{\frac{2}{2\ell + 1}} \quad (1.13)$$

The power spectrum is a very valuable source of information, as it is determined by the cosmological parameter. The presence of the peaks in the power spectrum arises from the fact that the Universe was full of baryons and photons encapsulated on gravitational potential wells. The pressure of the radiation resisted the compression of the gravity resulting in acoustic oscillations. This would end at the recombination epoch when photons would travel freely and scape from the gravitational wells. Therefore, there is one fundamental mode which would have had enough time to compress once. This scale is the sound horizon (H_s) at that moment, and determines the position of the first peak. The following accoustic peaks would be related to the proportional modes to the fundamental one. Furthermore, during the acoustic oscillations there are variations of the velocities of the plasma, therefore a Doppler effect, shifted by 90 degrees to

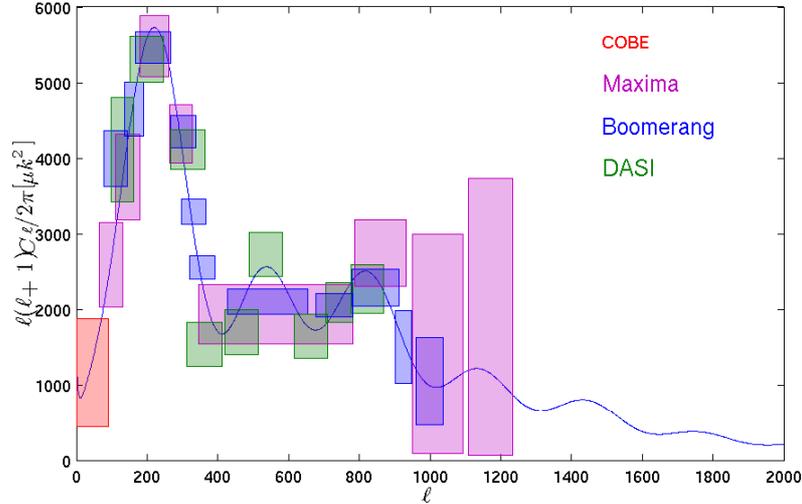


Figure 1.2 - Theoretical CMB power spectrum with the cosmological parameters derived from WMAP-7yr. Color squares are shown for different experimental data, with height roughly representing the uncertainties (data obtained from <http://background.uchicago.edu/~whu/intermediate/map7.html>).

the previous effect, can be seen also in the power spectrum (see e.g. Lineweaver, 1997; Tristram & Ganga, 2007).

After COBE first observation of the anisotropies, other experiments were proposed with their primary objective to find the first acoustic peak of the power spectrum. With Boomerang and MAXIMA balloon experiments, an entire reconstruction of the first peak was made placing it at $\ell \sim 220$ (see Fig. 1.2). The shape and position of the peak, as well as the spectral index of the power spectrum, had shown a superb agreement with inflationary predictions, what favored this model over all other alternatives to explain the initial perturbations. In fig. 1.2 it is shown the range of multipoles measured by different CMB experiments before WMAP. In 2001, the satellite WMAP was launched to study with more detail the temperature fluctuations of the CMB and characterize the full power spectrum with very high accuracy up to the third peak ($\ell \sim 800$) (Hinshaw et al., 2013). The results of WMAP have been extremely useful to consolidate the Λ CDM model and determined very accurate constraints on the cosmological parameters (Bennett et al., 2013). Planck went beyond and explored the CMB with more frequency channels and higher resolution, being able to efficiently model the foregrounds and minimise its effects, providing even tighter constraints on

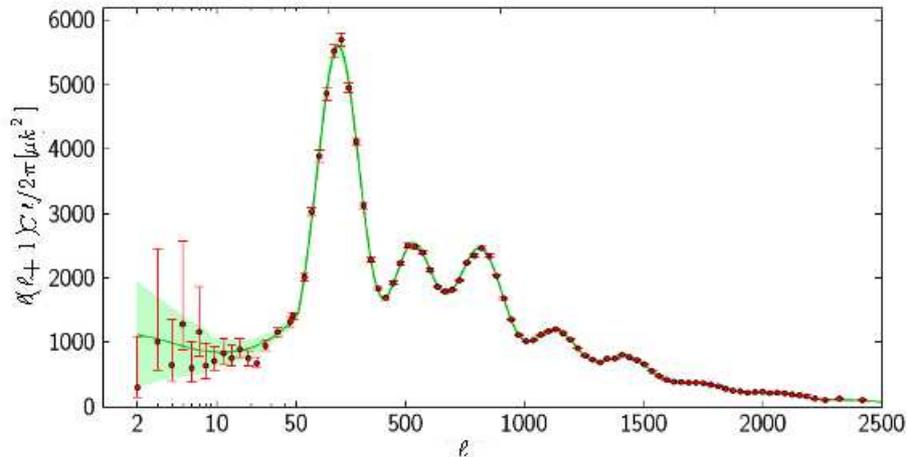


Figure 1.3 - CMB power spectrum obtained with Planck data, and published in Planck Collaboration (2013b). The cosmic variance is included in the error bars, and is indicated by the green shaded area. The Λ -CDM best fit has cosmological parameters as $\Omega_b h^2 = 0.022$, $\Omega_c h^2 = 0.120$, $\Omega_\Lambda = 0.683$, $\Omega_m = 0.317$, $\tau = 0.092$, $n_s = 0.962$, $\Delta_{\mathcal{R}}^2 = 2.21 \times 10^{-9}$.

the cosmological parameters. In Fig. 1.3 the power spectrum obtained with Planck data presented in late March 2013 is shown. The recently released Planck data has allowed the determination of the CMB power spectrum with high accuracy up to the 7th peak ($\ell \sim 2000$) (Planck Collaboration, 2013c,b). Planck data allowed a better characterisation of the foregrounds due to its large frequency range and the reconstruction of the lensing potential maps that helped to determine the cosmological parameters. Additionally, smaller scales are studied with the South Pole Telescope, within a range of multipoles from $\ell \sim 1000$ to $\ell \sim 3000$, very useful to study the interaction of the CMB photons with clusters of galaxies (Story et al., 2013).

1.2.3. CMB polarization

Another property of the CMB photons is that, as they interacted with free electrons by Thomson scattering, they are linearly polarized. The polarized component appears when the intensity field presents a quadrupolar pattern: the incident photons coming from orthogonal directions have different energies (see fig. 1.4). The quadrupolar pattern can be produced by the acoustic oscillations explained before (scalar perturbations) or by gravitational waves generated during inflation (tensor perturbations). The polarization patterns, can be described by the Stokes parameters Q and U, but

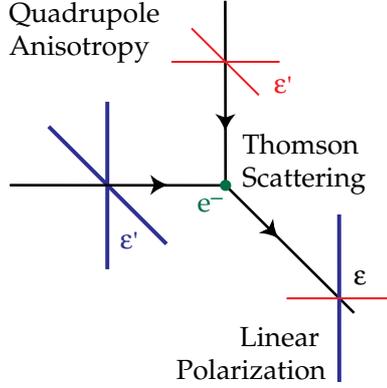


Figure 1.4 - Illustration of a Thomson scattering of the radiation with a quadrupolar anisotropy that generates linearly polarized photons. Figure obtained from Hu & White (1997).

they are usually transformed into the scalar fields E and B . The relation between the harmonic coefficients of the scalar fields E and B and the 2-spinors Q and U is (for details see Hu & White, 1997):

$$E_{\ell m} \pm iB_{\ell m} = - \int d\hat{n}_{\pm 2} Y_{\ell m}^*(\hat{n}) [Q(\hat{n}) \pm iU(\hat{n})]. \quad (1.14)$$

The power spectrum of the polarization scalar modes, as well as its correlation with the temperature, are also a source of valuable information that can break some of the degeneracies among the cosmological parameters. However, the signal-to-noise ratio of the polarisation observations is significantly lower than those obtained for temperature. The TE and TB cross spectra obtained by WMAP are shown in Fig. 1.5. Furthermore, the correlations of the temperature and polarization signals allows us to distinguish between scalar, vector and tensor modes. A valuable observable quantity, that depends on the inflationary model, can be defined as the ratio of the tensor and scalar perturbations power spectrum:

$$r = \frac{P_t(k)}{P_s(k)}, \quad (1.15)$$

evaluated at a given pivot scale (typically $k_o = 0.002 \text{ Mpc}^{-1}$). The primordial E-mode has been detected by different experiments and the B-mode due to lensing has been recently detected by the SPT experiment (Hanson et al., 2013). While finishing to write this thesis BICEP2 team claimed the first detection of primordial B-mode (BICEP2 Collaboration, 2014), a very important achievement as it resents indirect prove for primordial gravitational waves (see Hu & White, 1997, for a complete review on CMB polarization).

1.2.4. Sources of the CMB anisotropies

The CMB temperature anisotropies are divided in two main groups, primary and secondary. The primary anisotropies are those generated at the epoch of decoupling while the secondary anisotropies are generated during the travel of the photons from that epoch until now.

■ Primary anisotropies

At large scales the dominant contribution of the primary anisotropies is due to the Sachs-Wolfe effect. As commented above, after the inflationary epoch the energy density fluctuations were stretched to cosmological scales, producing wells and hills of gravitational potential at all scales. At the recombination time, the photons can freely travel and those trapped into a potential well lose energy when climbing the well, whereas the photons in less dense zones would have more energy than the average. This effect generates CMB primordial anisotropies at large scales, ($\ell \lesssim 200$). The scales larger than the sound horizon were unconnected from each other, therefore there are not acoustic peaks in these scales. Another source of primary anisotropies are the acoustic oscillations explained before, due to the competition of the gravity and the radiative pressure, that

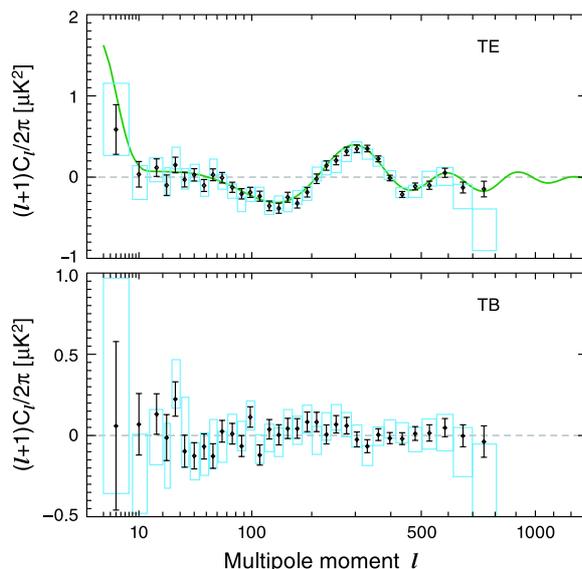


Figure 1.5 - CMB polarization TE and TB cross spectra obtained with WMAP-7y data, and published in Larson et al. (2011). The green line represents the theoretical cross power spectrum, the black points the 7-yr data with its errors, and the boxes are the results for WMAP 5-yr.

are reflected in the peaks of the power spectrum at $\ell \gtrsim 200$. There is a third main effect, that is a damping diffusion (Silk damping) present in the smaller scales ($\ell \gtrsim 1000$). This happens because the formation of neutral atoms was not instantaneous, therefore we can say that the last scattering surface has a given thickness. Those fluctuations with scales comparable to that thickness or lower would be diluted by the previously released photons. This effect is reflected on a reduction of the amplitudes of the peaks at high multipoles (see fig. 1.3).

■ Secondary anisotropies

The secondary anisotropies are generated when the photons scattered at the decoupling surface interact with gravity or matter during its travel towards us. They can also give very valuable cosmological information, complementary to the primary anisotropies, because they represent a projection of the different stages of the Universe, from the recombination epoch until now. The first interaction of those photons are with the *re-ionised* gas due to the high energetic processes during the generation of the first stars of the Universe. The CMB photons interact with the free electrons by Thompson scattering and this causes the amplitude of the anisotropies at small scales to decrease. Another secondary anisotropy comes from the fact that when a photon passes through a galaxy cluster, and interacts with the intergalactic medium, it suffers the inverse Compton scattering gaining energy from the high energy electrons in that medium. This is called the thermal Sunyaev-Zeldovich effect. A galactic cluster can also modify the energy of the incoming photon due to the peculiar movement of the cluster with respect to the reference system of the CMB, and it is called the kinetic Sunyaev-Zeldovich effect. Another relevant contribution to the secondary anisotropies is the integrated Sachs-Wolfe effect (ISW), related to the interaction of the photons with gravitational fields. This effect only occurs if the gravitational fields vary with the conformal time, which is the case for an accelerated expansion. Therefore, it depends on the amount and nature of dark energy. The ISW effect would have an impact in all scales, however its main power is concentrated in the large scales, on the left side of the first peak.

Another effect due to the interaction of the photons with gravity, is the lensing effect. The large scale structure between the last scattering surface and the observer acts as a lens deviating the photons and modifying the original anisotropies. This

effect will slightly smooth the amplitudes of the acoustic peaks and will transform the polarization E-modes into B-modes. Conversely to the previously commented effects, lensing was not detected with WMAP data, whereas Planck resolution allowed a reconstruction of the lensing map (Planck Collaboration, 2013d). Very recently the SPT experiment has detected the polarization B-mode produced by lensing (Hanson et al., 2013).

■ Contaminants

In the observed microwave maps, there are also foreground contaminants, that need to be separated to do a proper CMB analysis, but also because they might be a source of valuable information for other astrophysical studies. The most important of those contaminants is the galactic emission, which on the microwave band is basically due to synchrotron emission, generated by ultra-relativistic particles on the interstellar medium, the free-free emission, produced by the interaction of electrons and other charged particles and finally the thermal dust, that absorbs UV radiation and re-emits radiation at the far-infrared frequencies. There is also a galactic and extragalactic anomalous emission (AME) probably due to the electric dipole generated by dust grains with fast rotation. While there are several optimal methods for component separation (see for example Planck Collaboration, 2013a), the area of the galactic plane is too large and the separation of the signal becomes very complicated, therefore it is very common to mask this zone of the sky before doing any analysis. Also there are extragalactic objects emitting in the microwave range that should be taken into account, especially radio and IR galaxies, which are also usually masked. An advantage when dealing with the foregrounds of the CMB is that they are frequency-dependent, this is the main reason why multi-channel experiments are designed. Finally, the instrument introduces a certain noise level and possible systematic effects that are particular to each experiment. They need to be understood and must be taken into account before the analysis of the data. For details of CMB foreground contamination see for example Delabrouille & Cardoso (2009) and Bennett et al. (2003) and references therein.

1.3. Primordial non-Gaussianity

The standard inflationary model predicts a Gaussian distribution of the CMB anisotropies. Therefore, the study of the Gaussianity of the CMB is interesting as it gives very valuable information about this epoch. However, testing Gaussianity is a difficult task because there is an infinite number of ways to introduce a non-Gaussian signal. Moreover, if the signal is sufficiently weak it might be overlooked in a general Gaussianity test. One of the tests to check for non-Gaussianities is to look at the higher moments of the distribution, where the odd moments should vanish, whereas even moments are proportional to the second order moment. Therefore, if the distribution is not Gaussian, there is information beyond the power spectrum, that needs to be explored. The deviations from Gaussianity in the CMB can be of two types: primordial, where the non-Gaussian features were generated before the recombination epoch, or secondary, that occurred through the path that the photons traveled.

One of the main possible sources of primordial non-Gaussianity is generated during inflation. As commented before, the CMB power spectrum supported the theory that there was an inflationary epoch, but it has very little information of how this could happen. In order to gather some more information it is required to go to higher moments of the temperature anisotropy distribution, as well as, to study the polarization data, in particular the B modes.

The cosmic inflation was postulated in 1980 by Starobinsky and independently by Alan Guth in 1981 to solve three important issues of the theory of the Big Bang, the horizon, the flatness and the monopole problems (Starobinsky, 1980; Guth, 1981). Moreover, the inflation resulted to be a mechanism to generate the primordial density perturbations and hence, to explain the origin of the seeds of the large scale structure observed today. The predictions of the simplest inflationary models are that the observable Universe is flat, homogeneous and isotropic on large scales with small inhomogeneities that are almost scale-invariant and follow a Gaussian distribution (Guth & Pi, 1982; Hawking, 1982). In the following subsections the inflationary paradigm and its dynamics as well as the generation of non-Gaussianities in that period are explained.

1.3.1. Inflationary paradigm

The basic idea of inflation is that when the Universe was 10^{-42} seconds old, (setting the origin of the time coordinate in the singularity given by the Einstein equations)

there was an exponential accelerated expansion that stopped at around 10^{-36} seconds. An inflationary epoch is allowed in the Einstein General Relativity equations, hence it is *just* required a physical mechanism to produce it.

For a homogeneous and isotropic Universe, the Einstein equations yield to the Friedmann and fluid equations:

$$H^2 = \frac{8\pi}{3m_{pl}^2}\rho - \frac{K}{a^2} \quad (1.16)$$

$$\dot{\rho} + 3H(\rho + p) = 0 \quad (1.17)$$

where H is the Hubble expansion rate, a is the scale factor, K is the spatial curvature and m_{pl} the Planck mass ($m_{pl} = \sqrt{\frac{\hbar c^5}{G}}$, here and elsewhere $\hbar = c = 1$). Combining these two equations the following expression can be obtained:

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3m_{pl}^2}(\rho + 3p), \quad (1.18)$$

and an accelerated expansion ($\ddot{a} > 0$) appears naturally from this equation if

$$\rho + 3p < 0 \quad (1.19)$$

In the special case that the main contribution to the energy comes from the vacuum energy ρ_Λ and assuming that it does not change with the scale factor, eq.1.19 would be $\frac{\ddot{a}}{a} \propto \rho_\Lambda$. Hence, using eq. 1.16 it is easy to see that a grows exponentially with time, leaving the Hubble radius constant, $H = \sqrt{\frac{8\pi}{2m_{pl}^2}\rho_\Lambda}$. Therefore the comoving Hubble radius $(aH)^{-1}$ will significantly decrease during inflation, solving the *flatness problem*. Einstein equations allow any curvature of the Universe without any preferences, thus it was difficult to explain the flatness that all observations suggested, being the curvature equal to zero a very special case. Inflation naturally solves this problem, which is clearly seen if the Friedmann equation is written in terms of the ratio of the energy density to the critical density $\Omega = \frac{\rho}{\rho_c}$ where the critical density is $\rho_c = \frac{3m_p^2 H^2}{8\pi}$:

$$\Omega - 1 = \frac{K}{a^2 H^2}. \quad (1.20)$$

As the comoving Hubble radius is dramatically reduced during inflation Ω is very close to 1, leaving the effects of K negligible by the end of the inflationary phase. A second problem of the theory of the Big Bang that is solved by inflation is the *horizon problem*. The isotropy of the CMB requires that all the observed sky should be spatially connected at some point, but in a Universe without inflation some distant

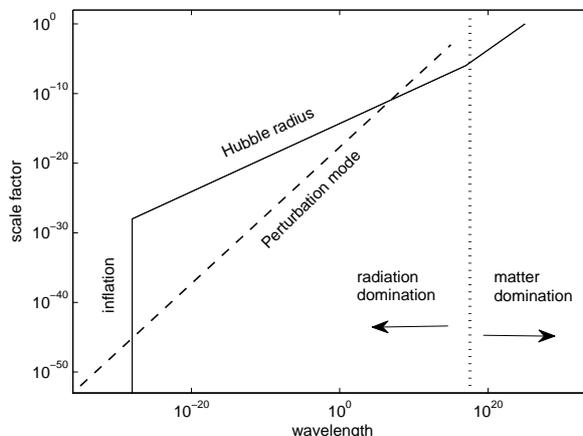


Figure 1.6 - The physical size of the Hubble radius (solid line) and size of perturbation (dashed line) as a function of the scale factor a . During inflation the Hubble radius is constant, while the perturbations grow proportional to the scale factor. When inflation ends, the Universe is dominated by radiation and the Hubble radius increases with time as $a \propto t^{\frac{1}{2}}$. Here is shown that without inflation large perturbations would have not been connected in the primitive Universe.

zones would have never been in contact. As commented above, before inflation the comoving Hubble radius was much larger, allowing zones that are disconnected after inflation, to maintain similar properties. This is more clearly seen in Fig. 1.6, where the evolution of the scale-length of the Hubble radius and the perturbation mode with the expansion of the Universe is shown. Additionally, inflation solved naturally a third problem of the Big Bang theory, related to the predicted generation of magnetic monopoles in a very hot and dense state. The early stages of the Universe should have generated a detectable amount of these particles, however they have not been observed. In an inflationary paradigm the density of monopoles is reduced by the stretching of the space to undetectable levels ($\Omega_{MO} \simeq 10^{-33}$).

1.3.2. Dynamics of inflation

The simplest inflationary models are based on one scalar field ϕ with mass m , usually referred to as the inflaton. The choice of scalar fields is not only motivated by simplicity, but also because they arise in the framework of relativistic local gauge field theories, such as the Standard Model of particle physics, where a scalar Higgs Boson is responsible for the mechanism called “electroweak symmetry breaking”, necessary for the gauge (vector) bosons of the theory (in this case the W and Z bosons) to

acquire mass. The recent detection of a possible standard model Higgs particle in the Large Hadron Collider (LHC) at CERN (CMS Collaboration Chatrchyan, 2012; ATLAS Collaboration, 2013) encouraged works on inflationary models driven by the Higgs boson (see for example Salvio, 2013; Hamada et al., 2014).

The energy and pressure density of a scalar field in a homogeneous Universe are given by:

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad (1.21)$$

$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi) \quad (1.22)$$

where $V(\phi)$ is the potential of the inflaton, that depends on the inflationary model. Using eq. 1.21 and eq. 1.22 in the Friedmann and fluid equations, and neglecting the curvature term, we obtain:

$$H^2 = \frac{8\pi}{3m_{pl}^2} \left[\frac{1}{2}\dot{\phi}^2 + V(\phi) \right] \quad (1.23)$$

and

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \quad (1.24)$$

where $V'(\phi) = dV/d\phi$. From eq. 1.21, eq. 1.22 and eq. 1.19 it can be easily seen that it is sufficient to satisfy $\dot{\phi} < V(\phi)$ to assure an accelerated expansion. The most successful inflationary models, in terms of concordance with the observations, are those in a *slow-roll* regime, where the kinetic energy is much smaller than the potential ($\frac{1}{2}\dot{\phi}^2 \ll V(\phi)$). This property induces a specific exponential expansion as the Hubble radius becomes constant in time, and the equation of state would be $p \sim -\rho$. To maintain the kinetic energy smaller than the potential during a sufficient time, the condition $|\ddot{\phi}| \ll 3H\dot{\phi}$ is imposed. Using the slow-roll conditions, eq. 1.23 and eq. 1.24 can be written as:

$$H^2 \simeq \frac{8\pi}{3m_{pl}^2} V(\phi) \quad (1.25)$$

$$3H\dot{\phi} \simeq -V'(\phi) \quad (1.26)$$

In general the slow-roll scenarios are characterized by two parameters related to the potential, ϵ and η , defined as:

$$\epsilon \equiv \frac{m_{pl}^2}{16\pi} \left(\frac{V'(\phi)}{V(\phi)} \right)^2, \quad \eta \equiv \frac{m_{pl}^2}{8\pi} \frac{V''(\phi)}{V(\phi)} \quad (1.27)$$

At the end of inflation, the inflaton remains oscillating around the minimum of the potential and the inflaton field decays into the *standard model* particles, filling the

Universe with electromagnetic radiation and increasing the temperature again. At that moment the Universe enters the radiation dominated era and its temperature increases, reheating the Universe. This process depends on the model of inflation and it is usually very fast (for a review on reheating see e.g. Bassett et al., 2006).

Perturbations in inflation

In addition to the classical motion of the inflaton field, there are quantum fluctuations that are responsible for the initial inhomogeneities of the Universe. In a simple scenario where the inflaton behaves as a quantum oscillator, the ground state of energy would have perturbations $\delta\phi$ with a Gaussian distribution. Although there can be tensor and scalar perturbations, here we are focused on scalar perturbations (see e.g. Dodelson, 2003, for detailed information on tensor perturbations). The inflaton field can then be presented in terms of a homogeneous background field with a linear perturbation:

$$\phi(\vec{x}, t) = \phi(t) + \delta\phi(\vec{x}, t). \quad (1.28)$$

This perturbation will have the main effect of having the slow-rolling behaviour $\phi(t)$ in different parts of the Universe at different times $\delta t = \frac{\delta\phi}{\dot{\phi}}$. The dynamics equation of the perturbations in Fourier space contains a gradient term that was neglected in the background field for homogeneity and isotropy reasons:

$$\delta\ddot{\phi}_k + 3H\delta\dot{\phi}_k + \frac{k^2}{a^2}\delta\phi_k = 0 \quad (1.29)$$

where the term related to the potential has been neglected, as V' can be treated as a constant for slow-roll models. Solving the differential equations the exact solution for the perturbations is :

$$\delta\phi_k = -\frac{H}{\sqrt{2k^3}}ie^{-ik\tau}(1 - ik\tau) \quad (1.30)$$

where τ is the comoving time. Therefore, during inflation the perturbation oscillates on sub-horizon scales ($-k\tau = \frac{k}{aH} \gg 1$)¹:

$$\delta\phi_k = -H\tau \frac{e^{-ik\tau}}{\sqrt{2k}}, \quad (1.31)$$

while on super-horizon scales ($\frac{k}{aH} \ll 1$) the perturbations are frozen:

$$|\delta\phi_k| = \frac{H}{\sqrt{2k^3}}. \quad (1.32)$$

¹Note that $d\tau = -\frac{dt}{a} = -\frac{da}{a^2H}$ where the minus sign is chosen because the light is moving towards us. Integrating both sides of the expression leads to $\tau = \frac{1}{aH}$.

In that limit, the power spectrum of the inflaton fluctuations, neglecting metric perturbations during inflation, is:

$$P_{\delta\phi} = \frac{H^2}{2k^3}. \quad (1.33)$$

These perturbations can now be treated as classical and they would couple with the metric perturbation field ψ . Although a detailed explanation of the perturbations of the curvature and the metric is out of the scope of this thesis (see Baumann, 2009; Kinney, 2009; Langlois, 2008; Riotto, 2002; Dodelson, 2003, and references therein are recommended for detailed information), I would like to point out the relation between the inflaton and curvature perturbations power spectra after inflation:

$$P_\psi = \frac{4}{9} \left(\frac{aH}{\dot{\phi}} \right)^2 P_{\delta\phi}. \quad (1.34)$$

Under the assumption that anisotropic stresses are small the gravitational potential Φ is equal to $-\psi$. Therefore, for a single field slow-roll inflation the initial matter perturbations power spectrum $k^3 P(k)$ depends on the slow-roll parameters ϵ and η , leaving a nearly scale-invariant power spectrum. As I will show in the following section it is feasible to obtain the temperature anisotropies of the CMB from Φ (see for example Lewis & Bridle, 2002; Zaldarriaga et al., 1997). Summarizing, we are able to trace properties of the inflationary epoch by the observation of the CMB.

Models of inflation

The modeling of an inflationary epoch is a very challenging question, as there is no possibility of studying processes at that energy level in our laboratories. After the first approach of Alan Guth in 1981, which was finally not phenomenologically viable, a large number of other models have been presented. On the one hand there are models based on the standard model of particle physics, using already studied scalar fields, as the Goldstone Bosons (Guth, 1981; Hawking, 1982; Freese et al., 1990). On the other hand, new mechanisms have been proposed with the properties required to act as an inflaton (e.g. D-brane inflation, Baumann et al., 2008). This makes the inflationary paradigm even more interesting, as it could be a confirmation of particle physics theories or to open the way for new physics mechanisms. In any case, the study of the inflationary phase would contribute to a better understanding of the largest and tiniest structures of our Universe. I will focus here on the slow-roll models of inflation, where different models arise from choices of different potentials, and therefore of the slow-roll parameters. In a broad classic sense, the slow-roll inflation models can be divided into

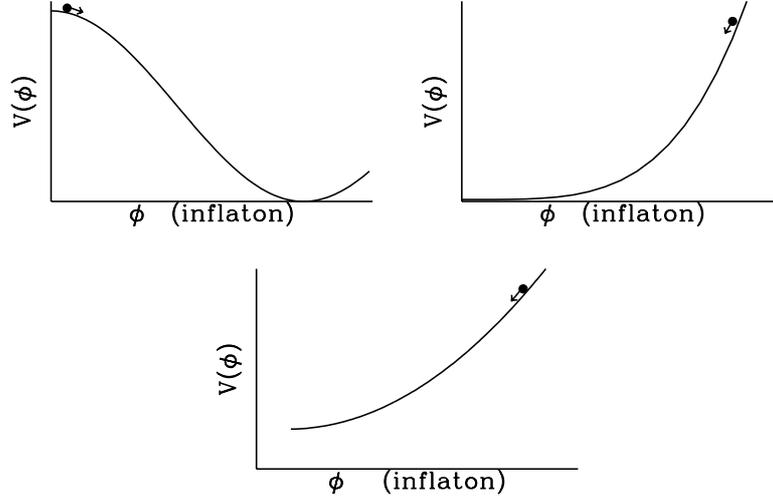


Figure 1.7 - Schematic illustration of the potential energy of the inflation for the three different main types of slow-roll inflation. On the top, the first potential is a *small-field* model, the second figure illustrates a *large-field* model and the third figure (on the bottom) is an example of *hybrid* inflation. Plot from Kolb (1999)

three main groups, *large field*, *small field* and *hybrids* models (Dodelson et al., 1997; Kolb, 1999). The typical potential for each of these models is illustrated in fig. 1.5.

The *small field* models, whose most known example is the *new inflation* scenario (Linde, 1982), arise from spontaneous symmetry breaking. They are characterised by a “red” spectral $n_s < 1$ and low contribution of the gravitational waves ($r \lesssim 0.01$). A typical potential would be of the form:

$$V(\phi) = \lambda \left[1 - \left(\frac{\phi}{\mu} \right)^p \right] \quad (1.35)$$

where λ and μ are free parameters of the potential.

The *large field* models are characterized by a scalar field displaced from the minimum of the potential by several times the Planck mass, and a slow evolution towards the minimum. These models are more difficult to fit in the particle physics framework, but have the advantage of less fine-tuning requirements, and present a mechanism to end inflation naturally. They are characterized by a “red” spectral index $n_s < 1$ (with slow roll parameters $0 < \eta < 2\epsilon$) and can give a larger contribution of the scalar-tensor ratio. The *chaotic inflation* proposed by Linde (1986) is the most studied example of this category with a potential of the form:

$$V(\phi) = \lambda \left(\frac{\phi}{\mu} \right)^p . \quad (1.36)$$

The *hybrid* models involve more than one field. Recently, inflationary models based on the supersymmetry theories have been proposed, typically with two scalar fields involved (ϕ and χ). Some of the *hybrid* models give a “blue” spectral index $n_s > 1$ which is disfavoured by the observations, whereas the *hybrid* models with expected “red” spectral index usually gives a very large contribution of tensor perturbations which is also disfavoured by observations. However, there are still valid assumptions on the free parameters that allow this type of models. An example is the spontaneous symmetry breaking of SUSY model (Dvali et al., 2004b) with a potential given by:

$$V(\phi, \chi) = \lambda \left(1 - \frac{\chi^2}{\mu^2}\right)^2 + \alpha \log\left(\frac{\phi}{\mu}\right) + \frac{g^2}{2} \phi^2 \chi^2 \quad (1.37)$$

Setting tight constraints on the spectral index and the tensor to scalar ratio would help to discriminate among all the variety of inflationary models. For more details in inflationary models and its dependency with n_s and r see Planck Collaboration (2013e). There are more exotic models of inflation that do not hold the slow-roll conditions as the *ghost* inflation, where the inflationary phase is obtained with a ghost condensate (Arkani-Hamed et al., 2004). Based on the string theory, the DBI-inflation proposes a type of inflation that occurs due to the motion of a D-brane traveling down a higher-dimensional space-time (Dvali & Tye, 1999). Another very different approach is based on the loop-quantum gravity that provides a possible mechanism to generate inflation (Ashtekar, 2009). Besides their dependence on the slow-roll parameters that can be indirectly observed by the CMB (Dodelson et al., 1997), another important feature that may help to discard some of these models is the presence of non-Gaussianity. In particular, depending on the model of inflation, different contributions to the higher moments of the CMB fluctuations distribution are expected (see for example Bartolo et al., 2004).

1.3.3. Non-Gaussianities produced by inflation

As explained in the previous section the quantum perturbations of the inflaton field couple with the metric, and are finally related to gravitational potential perturbations. The CMB temperature anisotropies are related to the gravitational potential Φ by:

$$\frac{\Delta T(\hat{n})}{T_0} = \sum_{\ell m} Y_{\ell m}(\hat{n}) \int r^2 dr \Phi_{\ell m}(r) g_{T_\ell}(r), \quad (1.38)$$

where $\Phi_{\ell m}$ is the spherical harmonic transform of $\Phi(x)$, $\Phi_{\ell m} = \int d^2 \hat{n} \Phi(r, \hat{n}) Y^*(\hat{n})$. In the above equation g_{T_ℓ} (in the literature also denoted by α_ℓ) is a mathematical

function that depends on the cosmological model which projects Φ into the temperature anisotropies through the integral over the line of sight. As a consequence, the CMB anisotropies would be Gaussian if and only if the curvature perturbations were Gaussian and its relation with the CMB anisotropies were linear.

The non-Gaussianities produced by inflation can be parametrised with the non-linear coupling parameter f_{NL} . In particular, up to the second-order the expression for the effective primordial potential can be written as:

$$\Phi(x) = \Phi(x)_L + f_{\text{NL}} [\Phi^2(x)_L - \langle \Phi_L^2(x) \rangle] , \quad (1.39)$$

that should be used in eq. 1.38 to obtain the temperature anisotropies. Note that in this case both sides of the equations are evaluated at the same location in the position space. This type of non-Gaussianity is called *local non-Gaussianity*, and the non-coupling parameter is usually referred as $f_{\text{NL}}^{\text{local}}$.

The parameter f_{NL} can be related to the amplitude of the third order moments of the CMB anisotropies. As shown below, in the spherical harmonic space f_{NL} represents the amplitude of the bispectrum. There can be different forms of the bispectrum, where the amplitude is defined by different types of f_{NL} . The most interesting forms are: the local bispectrum, where the main contribution comes from the components with two high multipoles and a low multipole, the equilateral form, where the main contribution is due to equilateral triangles, and the orthogonal form, which will be combinations orthogonal to the first two forms.

In this thesis we have studied the local form of the bispectrum (hereafter $f_{\text{NL}} = f_{\text{NL}}^{\text{local}}$), for more details on other forms see for example the review by Liguori et al. (2010). The main contribution of the local primordial non-Gaussianity is reflected in the third order moment of the distribution of the temperature anisotropies. A natural way to explore this signal is by looking at the three-point correlation function in the spherical harmonic space, the bispectrum,

$$B_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} \equiv \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \quad (1.40)$$

where the $a_{\ell m}$ are related to the primordial gravitational potential by:

$$a_{\ell m} = 4\pi(-i)^\ell \int \frac{d^3 k}{(2\pi)^3} \Phi(\vec{k}) g_{T\ell}(k) Y_{\ell m}^* \left(\frac{\vec{k}}{k} \right) , \quad (1.41)$$

where $\Phi(\vec{k})$ is the Fourier transform

$$\Phi(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\vec{x}} \Phi(\vec{k}) . \quad (1.42)$$

For the local type of non-Gaussianity the Fourier transform of the gravitational potential eq. 1.39, can be divided into two different parts:

$$\Phi(\vec{k}) = \Phi_L(\vec{k}) + \Phi_{NL}(\vec{k}) , \quad (1.43)$$

where $\Phi_L(k)$ is the linear part of the potential and $\Phi_{NL}(k)$ the non-linear part, which is defined by:

$$\Phi_{NL}(\vec{k}) \equiv f_{NL} \left[\int \frac{d^3p}{(2\pi)^3} \Phi_L(\vec{k} + \vec{p}) \Phi_L^*(\vec{p}) - (2\pi)^2 \delta^{(3)}(\vec{k}) \langle \Phi_L^2(\vec{x}) \rangle \right] . \quad (1.44)$$

Using eq.1.43 and eq. 1.44 one finds the non-vanishing components proportional to f_{NL} (see Komatsu, 2002, for details):

$$\langle \Phi_L(\vec{k}_1) \Phi_L(\vec{k}_2) \Phi_{NL}(\vec{k}_3) \rangle = 2(2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) f_{NL} P_\Phi(\vec{k}_1) P_\Phi(\vec{k}_2) . \quad (1.45)$$

Taking into account all the permutations and, as commented previously in sec. 1.1.2, that $P_\Phi \propto \frac{1}{k^{n_s-4}}$, the potential bispectrum in the limit where $k_3 \mapsto 0$ would be:

$$\langle \Phi(\vec{k}_1) \Phi(\vec{k}_2) \Phi(\vec{k}_3) \rangle = 4(2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) f_{NL} P_\Phi(\vec{k}_1) P_\Phi(\vec{k}_2) . \quad (1.46)$$

Assuming a single-field slow-roll model of inflation, Maldacena (2003) found that the amplitude of the three point correlation function depends on the spectral index n_s :

$$\lim_{k_3 \rightarrow 0} \langle \Phi_{k_1} \Phi_{k_2} \Phi_{k_3} \rangle = (2\pi)^3 \delta(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) \frac{5}{3} (1 - n_s) P_\Phi(k_1) P_\Phi(k_3) \quad (1.47)$$

when comparing eq. 1.47 and eq. 1.46 it can be seen that f_{NL} for the *squeezed* (local) shape should be of the order of $\frac{5}{12}(1 - n_s)$ for single-field inflationary models.

Creminelli & Zaldarriaga (2004) pointed out that eq. 1.47 is more general, and it is valid outside the slow-roll regime, with assuming nothing more besides that the inflaton is the only field involved. However, recent works point out that eq. 1.47 does not hold for a few non-standard single-field models and therefore could give a significant local non-Gaussian contribution. This is the case for models with super-horizon curvature perturbation (Chen et al., 2013; Namjoo et al., 2013), or non-Bunch-Davies initial states (Ganc, 2011; Agullo & Parker, 2011). Those single-field models could give a large contribution to f_{NL} .

single field	< 1	Creminelli & Zaldarriaga (2004)
DBI inflation	0.1-100	Alishahiha et al. (2004)
curvaton type models	$f_{NL} = \frac{5}{4r}$	Lyth et al. (2003)
ghost inflation	~ 100	Arkani-Hamed et al. (2004)
ekpyrotic models	$\pm \frac{5\alpha}{24\beta} \epsilon^{-1}$	Buchbinder et al. (2007)

Table 1.1 - Table of values of expected local type f_{NL} in different generation of primordial perturbations models. For the curvaton models, the f_{NL} value depends on the curvaton field density just before the decay, as a fraction of the total density (r). For ekpyrotic models f_{NL} is expected to be large, although it depends on the free parameters of the model α and β .

Besides of these exceptions, it is not expected to find a detectable f_{NL} for *large* and *small field* models but only for *hybrid* models, where the amplitude depends on the form and free parameters of the potential. Additionally to the *hybrid* models, this type of non-Gaussianity might arise also in inhomogeneous reheating (Dvali et al., 2004a) and in non-standard cosmologies as the new ekpyrotic models (Creminelli & Senatore, 2007; Koyama et al., 2007; Buchbinder et al., 2007). The expected values of f_{NL} for different models of inflation are shown in table 1.1. While with WMAP constraints ($-3 < f_{NL} < 77$ at 95% CL) most of the models in table 1.1 were still allowed, with the exception of the *ghost* inflation (Bennett et al., 2013), the recently released Planck results ($-9 < f_{NL} < 14$ at 95% CL) are able to set tight constraints on the properties of the multi-field inflationary models and ekpyrotic free parameters (Planck Collaboration, 2013f).

1.4. f_{NL} estimators

As shown in the previous section, the local non-Gaussianity produced by inflation is reflected in the third order moments. In this section some of the most common estimators are summarised as well as the estimators used in the analyses presented in this thesis.

1.4.1. Bispectrum estimators

The full bispectrum is extremely demanding, computationally speaking, what makes its use unfeasible for non-Gaussianity analysis. Different approaches have been proposed to reduce the computational demand maintaining the expected efficiency (see for example Komatsu et al., 2005; Bucher et al., 2010; Fergusson et al., 2010). In addition, there are other linear transformations, as the ones given by wavelets, whose third order moments have been used to obtain similar constraints (Curto et al., 2009a, 2011). Here are presented some of the estimators of the bispectrum.

- **Angular averaged reduced bispectrum**

The angular averaged bispectrum is obtained by assuming isotropy as:

$$B_{\ell_1\ell_2\ell_3} = \sum_{m_i} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}, \quad (1.48)$$

where the matrix denotes the Wigner-3j symbol, that ensures that the combinations of ℓ s meet the triangle condition $|\ell_i - \ell_j| \leq \ell_k \leq |\ell_i + \ell_j|$ and that $m_1 + m_2 + m_3 = 0$. Additionally, parity invariance of the angular correlation function demands $\ell_1 + \ell_2 + \ell_3 = \text{even}$. Eq. 1.48 can be expressed in a more computationally effective form, using the relation of the Wigner-3j symbols with the spherical harmonics:

$$\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^{-1} \sqrt{\frac{4\pi}{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}} \times (1.49) \\ \int Y_{\ell_1 m_1}(\hat{n}) Y_{\ell_2 m_2}(\hat{n}) Y_{\ell_3 m_3}(\hat{n}) d\hat{n},$$

The expression for the angular averaged bispectrum can then be written as:

$$B_{\ell_1\ell_2\ell_3} = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^{-1} \sqrt{\frac{4\pi}{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}} \int T_{\ell_1} T_{\ell_2} T_{\ell_3} d\Omega, \quad (1.50)$$

where $T_{\ell_i} = \sum_{m_i} a_{\ell_i m_i} Y_{\ell_i m_i}$. Finally, in Komatsu & Spergel (2001) is defined the reduced bispectrum $b_{\ell_1\ell_2\ell_3}$, in order to have a more convenient estimator to use in the flat sky approximation:

$$B_{\ell_1\ell_2\ell_3} = \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1\ell_2\ell_3}. \quad (1.51)$$

Then, using the reduced bispectrum definition and eq. 1.50, the relation between the angular averaged reduced bispectrum and T_ℓ is:

$$b_{\ell_1\ell_2\ell_3} = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^{-2} \frac{4\pi}{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)} \int T_{\ell_1} T_{\ell_2} T_{\ell_3} d\Omega \quad (1.52)$$

■ Binned bispectrum

In order to reduce the number of components of the full bispectrum Bucher et al. (2010) proposed to bin the ℓ space, in such a form as:

$$b_{I_a I_b I_c} = \sum_{\ell \in I_a} \sum_{\ell_2 \in I_b} \sum_{\ell_3 \in I_c} b_{\ell_1\ell_2\ell_3} \quad (1.53)$$

where I_i are intervals in ℓ , $I_i = [\ell_{min}^i, \ell_{max}^i]$. To be able to perform the binning before computing the full bispectrum, an alternative reduced bispectrum is defined:

$$b_{\ell_1\ell_2\ell_3} = \int T_{\ell_1} T_{\ell_2} T_{\ell_3} d\Omega. \quad (1.54)$$

This definition of the reduced bispectrum differs from the the previous expression defined in eq. 1.52 by the factor $\Gamma_{\ell_1,\ell_2,\ell_3}$:

$$\Gamma_{\ell_1,\ell_2,\ell_3} = \frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^2. \quad (1.55)$$

The binned bispectrum can then be computed from the binned maps T_{I_a} ($T_{I_a} = \sum_{\ell \in I_a} T_\ell$), allowing us to reduce the number of spherical harmonic transformations.

The number of bins is typically two orders of magnitude smaller than the number of ℓ , implying a significant reduction of the 3-point correlation components. The binned bispectrum is constructed from the binned maps as:

$$b_{I_a I_b I_c} = \sum_i^N \frac{4\pi}{N} T_{I_a}(i) T_{I_b}(i) T_{I_c}(i) \quad (1.56)$$

where N is the number of pixels of the map. The binning scheme reduces by 5 orders of magnitude the number of components of the bispectrum, reducing considerably the dimension of the problem and therefore its computational effort. The drastic reduction does not affect significantly the efficiency of the method, giving very similar constraints on the f_{NL} parameter (Bucher et al., 2010; Casaponsa et al., 2013). The performance of this method and its applications to WMAP 7-yr data have been studied in chapter 4 (Casaponsa et al.

2013). Furthermore, the binned bispectrum has been used for setting the constraints on f_{NL} in Planck data (Planck Collaboration, 2013f).

There are other methods in the literature, along the same lines of the estimators presented above. Fergusson et al. (2010) presented a similar approach as the binned bispectrum, that reduces the number of components of the bispectrum performing a separable mode expansion. Another attempt to reduce the dimensionality of the problem is the “skew-spectra” presented in Munshi et al. (2011).

1.4.2. Wavelet estimators

A large set of different wavelets have been used in the astrophysics literature. In particular, different spherical wavelets have been applied to CMB Gaussianity analysis during the last decade, including the spherical Haar wavelet (SHW, Barreiro et al. (2000)), the Spherical Mexican hat Wavelet (SMHW, Cayón et al. (2001); Martínez-González et al. (2002); Vielva et al. (2004); Mukherjee & Wang (2004); Cruz et al. (2005); Curto et al. (2009b)), elliptical SMWH McEwen et al. (2005), directional spherical wavelets (McEwen et al. 2006, 2007) and needlets (Pietrobon et al., 2009; Rudjord et al., 2009; Cabella et al., 2010). For a review on wavelet applications to cosmology see McEwen et al. (2007). Here I describe the two types of wavelets that are used in the analysis presented in this thesis, the SMHW and a new wavelet the Healpix wavelet (HW).

- **Spherical Mexican hat wavelet**

In particular for f_{NL} estimation, one of the most commonly used wavelets is the spherical Mexican hat wavelet (SMHW), which gives very close f_{NL} constraints to the ones obtained with the bispectrum. Conversely to the spherical harmonic transformation, in this case the transformation is done in the wavelet domain, maintaining part of the information of the real space while adding harmonic space information. The definition and properties of the SMHW transformation can be seen in Martínez-González et al. (2002). Basically, given a function $f(\vec{n})$ evaluated on the sphere at a direction \vec{n} , the wavelet transform is defined as

$$w(R, \vec{p}) = \int f(\vec{n}) \Psi(\vec{n}, \vec{p}, R) d\vec{n} \quad (1.57)$$

where $\Psi(\vec{n}, \vec{p}, R)$ is a continuous wavelet function, \vec{p} is the position in the sky at which the wavelet coefficient is evaluated and R is the scale of the wavelet. In

particular for the SMHW, the wavelet only depends on the polar angle θ and the scale R and is defined by:

$$\Psi(\theta; R) = \frac{1}{(2\pi)^{\frac{1}{2}} N(R)} \left[1 + \left(\frac{y}{2} \right)^2 \right]^2 \left[2 - \left(\frac{y}{R} \right)^2 \right] e^{-\frac{y^2}{2R^2}}, \quad (1.58)$$

where

$$N(R) \equiv \left(1 + \frac{R^2}{2} + \frac{R^4}{4} \right)^{\frac{1}{2}} \quad (1.59)$$

and

$$y \equiv 2 \tan \left(\frac{\theta}{2} \right). \quad (1.60)$$

As for the bispectrum, third order moments of the wavelet coefficients are constructed, which will be directly related to f_{NL} .

■ Healpix wavelet

The Healpix wavelet (HW) is a very simple statistical tool, inspired in the Haar wavelet and adapted to the HEALPix pixelization. The HW is a discrete wavelet and presents an optimal space localization, while the frequency localization is not as good as that of the SMHW. The main advantage of this wavelet is that conversely to the SMHW and to needlets, a transformation of the data into spherical harmonic space is not required. Therefore, the computational cost is significantly reduced. The wavelet functions for the HW are:

$$\begin{aligned} \Psi_{0,j,k}(x) &= \varphi_{j+1,k_0}(x) - \frac{\varphi_{j,k}(x)}{4} \\ \Psi_{1,j,k}(x) &= \varphi_{j+1,k_1}(x) - \frac{\varphi_{j,k}(x)}{4} \\ \Psi_{2,j,k}(x) &= \varphi_{j+1,k_2}(x) - \frac{\varphi_{j,k}(x)}{4} \\ \Psi_{3,j,k}(x) &= \varphi_{j+1,k_3}(x) - \frac{\varphi_{j,k}(x)}{4} \end{aligned} \quad (1.61)$$

where $\varphi(x)_{j,k}$ is the scaling function

$$\varphi(x)_{j,k} = \begin{cases} 1 & \text{if } x \in P_{j,k} \\ 0 & \text{otherwise,} \end{cases} \quad (1.62)$$

and $P_{j,k}$ is the pixel at position k at resolution j , which at the next higher resolution is divided into four daughter pixels P_{j+1,k_0} , P_{j+1,k_1} , P_{j+1,k_2} , P_{j+1,k_3} .

The wavelet decomposition of a temperature map can be written in terms of

the basis functions and a set of coefficients:

$$\frac{\Delta T}{T}(x_i) = \sum_{k=0}^{N_{j_0}-1} \lambda_{j_0,k} \varphi_{j_0,k}(x_i) + \sum_{j=j_0}^J \sum_{m=0}^3 \sum_{k=0}^{N_j-1} \gamma_{m,j,k} \Psi_{m,j,k}(x_i), \quad (1.63)$$

where N_j is the number of pixels at resolution j . $\lambda_{j,k}$ and $\gamma_{m,j,k}$ are the approximation and detail coefficients respectively (those are described in detail in chapter 2).

The third order moments of the HW are used to constrain f_{NL} of WMAP-7yr (Casaponsa et al., 2011) and would be presented in the second chapter of this thesis, as well as more details about the implementation of the HW.

It is also worth commenting other methods with a very different approach, as the Minkowski functionals that performs an analysis of the geometry of the field. Its theoretical dependence on f_{NL} is known and therefore constraints on the parameter can also be imposed (Hikage et al., 2008). Other alternative methods for f_{NL} estimation have been presented, as the use of the N-pdf of the CMB anisotropies (Vielva & Sanz, 2009), a Bayesian approach presented in Elsner & Wandelt (2010) or a goodness of fit analysis (Aliaga et al., 2005; Curto et al., 2007).

1.4.3. Classical parameter estimation

A common procedure is to construct a f_{NL} estimator using third order statistics of the form explained above. With the approximation that the cubic statistics follow a Gaussian distribution, the likelihood can be written in terms of a χ^2 :

$$\mathcal{L} \propto e^{-\frac{1}{2}\chi^2}. \quad (1.64)$$

As the bispectrum is proportional to f_{NL} , one can write $\langle B_{\ell_1 \ell_2 \ell_3} \rangle = f_{\text{NL}} \langle B_{\ell_1 \ell_2 \ell_3} \rangle^1$, where $\langle B_{\ell_1 \ell_2 \ell_3} \rangle^1$ is the value for the bispectrum at $f_{\text{NL}} = 1$. Here we use the angular averaged bispectrum, but the same reasoning could be done for the other estimators proposed above. Then the χ^2 expression is:

$$\chi^2 = \sum_{\ell_1 \ell_2 \ell_3, \ell'_1 \ell'_2 \ell'_3} \left(B_{\ell_1 \ell_2 \ell_3}^{\text{obs}} - f_{\text{NL}} \langle B_{\ell_1 \ell_2 \ell_3} \rangle^1 \right) C_{\ell_1 \ell_2 \ell_3, \ell'_1 \ell'_2 \ell'_3}^{-1} \left(B_{\ell'_1 \ell'_2 \ell'_3}^{\text{obs}} - f_{\text{NL}} \langle B_{\ell'_1 \ell'_2 \ell'_3} \rangle^1 \right), \quad (1.65)$$

For the bispectrum the scales are uncorrelated (under ideal conditions), and the off-diagonal terms of the covariance matrix C vanish. Then $C_{\ell_1 \ell_2 \ell_3, \ell'_1 \ell'_2 \ell'_3}^{-1}$ is just the inverse

of the variance: $C_{\ell_1\ell_2\ell_3,\ell_1\ell_2\ell_3} = \Delta C_{\ell_1}C_{\ell_2}C_{\ell_3}$, where Δ is a factor that takes values of 6, 2 or 1 if all multipoles are equal, two are equal or all are different respectively (see Komatsu, 2002, for details). Therefore, it can be easily found that the maximum likelihood estimator for f_{NL} would be:

$$\hat{f}_{\text{NL}} = \sum_{\ell'_1\ell'_2\ell'_3} W_{\ell'_1\ell'_2\ell'_3} B_{\ell'_1\ell'_2\ell'_3}^{\text{obs}} \quad (1.66)$$

where the weights $W_{\ell'_1\ell'_2\ell'_3}$ are:

$$W_{\ell'_1\ell'_2\ell'_3} = \frac{\langle B_{\ell_1\ell_2\ell_3} \rangle^1 / \Delta C_{\ell_1}C_{\ell_2}C_{\ell_3}}{\sum_{\ell'_1\ell'_2\ell'_3} (\langle B_{\ell'_1\ell'_2\ell'_3} \rangle^1)^2 / \Delta C_{\ell'_1}C_{\ell'_2}C_{\ell'_3}}. \quad (1.67)$$

The estimator in eq. 1.66 is unbiased as it satisfies $\langle \hat{f}_{\text{NL}} \rangle = f_{\text{NL}}$.

The error of the estimator is given by its standard deviation, which is computed with simulations, and the lower bound for a given experiment is given (from the Fisher matrix) by:

$$F_{a,a} = -\frac{\partial^2 \ln \mathcal{L}}{\partial a \partial a} \quad (1.68)$$

where a correspond to the parameter of the model, in this case f_{NL} . The lower bound on the dispersion of the parameter would be:

$$\sigma_{f_{\text{NL}}} = \frac{1}{\sqrt{F_{f_{\text{NL}},f_{\text{NL}}}}}, \quad (1.69)$$

$$\sigma_{f_{\text{NL}}} = \sqrt{\frac{(\langle B_{\ell_1\ell_2\ell_3} \rangle^1)^2}{\Delta C_{\ell_1}C_{\ell_2}C_{\ell_3}}}. \quad (1.70)$$

Dealing with non-ideal conditions

The approach defined above is very powerful for an ideal experiment (full sky and isotropic noise). However, under realistic conditions, the CMB data present two important sources of anisotropy: the noise, due to the scanning procedure of the satellite, and the mask. As explained in section 1.2.4 there are zones of the sky that can not be completely cleaned and need to be masked, typically setting those areas to a null value. This has an impact on the used statistics, which should be carefully studied. As a very simple approximation, the available sky fraction (f_{sky}) is commonly used to correct the theoretical standard deviation of the parameter:

$$\sigma_{f_{\text{NL}}} = \frac{1}{\sqrt{f_{\text{sky}}}} \sqrt{\frac{(B_{\ell_1\ell_2\ell_3})^2}{\Delta C_{\ell_1}C_{\ell_2}C_{\ell_3}}} \quad (1.71)$$

This correction is only valid for the computation of the lower bound, whereas when estimating the f_{NL} value of a map, the non-idealities would introduce correlations at different scales, that may have a large effect on the results. There are several techniques proposed to solve this problem, as for example to apodize the mask, avoiding sharp cuts at the edges of the galactic zone. Another technique is the use of inpainting, that it fills the masked pixels with some noise, with a more or less complex algorithms. These techniques might be useful but they need to be implemented carefully to prevent a contamination of the real signal.

When using the theoretical estimator for f_{NL} the correlations are not taken into account, yielding to a sub-optimal estimator. Babich (2005) and Creminelli et al. (2006) proposed an optimal estimator that included a linear term which takes into account the anisotropic effects. This term appears naturally if, instead of assuming a Gaussian distribution of the $a_{\ell m}$, the non-Gaussian signal due to f_{NL} is included in the likelihood using the Edgeworth expansion:

$$\mathcal{L}(a|f_{NL}) = \left(1 - f_{NL} \sum_{\ell_i m_i} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle_1 \frac{\partial}{\partial a_{\ell_1 m_1}} \frac{\partial}{\partial a_{\ell_2 m_2}} \frac{\partial}{\partial a_{\ell_3 m_3}} \right) \times \quad (1.72)$$

$$\times \frac{e^{-\frac{1}{2} \sum a_{\ell_4 m_4}^* C_{\ell_4 m_4, \ell_5 m_5}^{-1} a_{\ell_5 m_5}}}{\sqrt{(2\pi)^N |C|}}.$$

An estimator is optimal, unbiased and with a variance saturating the Cramer-Rao bound, if the following condition is satisfied:

$$\frac{d \ln \mathcal{L}(a|f_{NL})}{df_{NL}} = F_{f_{NL}, f_{NL}} (\widehat{f_{NL}} - f_{NL}). \quad (1.73)$$

Using eq. 1.64 and eq. 1.73, and after some calculations it is found that the optimal estimator is given by:

$$\hat{f}_{NL} = \frac{1}{N} \sum_{\ell_i m_i} \left(\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle_1 C_{\ell_1 m_1, \ell_4 m_4}^{-1} C_{\ell_2 m_2, \ell_5 m_5}^{-1} C_{\ell_3 m_3, \ell_6 m_6}^{-1} a_{\ell_4 m_4} a_{\ell_5 m_5} a_{\ell_6 m_6} \right. \quad (1.74)$$

$$\left. - 3 \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle_1 C_{\ell_1 m_1, \ell_2 m_2}^{-1} C_{\ell_3 m_3, \ell_4 m_4}^{-1} a_{\ell_4 m_4} \right),$$

where $C_{\ell m, \ell' m'} = \langle a_{\ell m} a_{\ell' m'} \rangle$. For simplicity, Creminelli et al. (2006) used the assumption of a diagonal C in the denominator, therefore the estimator would be as in eq. 1.66 plus a linear term of the form:

$$-\frac{3}{N} \sum_{\ell_i m_i} \frac{\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle_1}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} C_{\ell_1 m_1, \ell_2 m_2} a_{\ell_3 m_3}. \quad (1.75)$$

For an ideal case, one can see that the estimator for f_{NL} remains the same as in eq. 1.66. The linear term contribution is less important if correlations are taken into account, as it will be shown in chapter 4.

1.4.4. Neural networks

Besides the classical approaches for parameter estimation, there are other techniques as artificial neural networks that have been widely used in this matter (e.g. see the review of Cheng & Titterton, 1994). The first application of neural networks regarding the estimation of primordial non-Gaussianity is presented in this thesis.

Artificial neural networks were introduced during the mid twentieth century based on the first biological neural network models of the brain. The basic idea was a system composed by different neurons connected to each other that communicate through electromagnetic impulses. The human learning process was thought to be possible due to the strengthening of determined neural connections when activities were repeated. Algorithms to recreate the biological neural network were created to understand better our brain (McCulloch & Pitts, 1943; Farley & Clark, 1954; Rosenblatt, 1958). Although the simplest artificial neural networks (ANN) are not able to reproduce the complex human brain activity, they have been a very useful tool to solve complex mathematical problems with quite simple algorithms. ANN have been widely used for solving several practical problems such as pattern recognition, chaotic system predictions or classification of objects. Furthermore, artificial intelligence techniques have been introduced to astrophysical analysis in the past decades: morphological galaxy determination, photo-redshift estimations, and classification of different objects are examples of successful applications of neural networks (e.g. Baccigalupi et al. 2000; Firth et al. 2003; Ball et al. 2004 and Nørgaard-Nielsen 2012). In particular, for CMB analysis, they have been recently used to reduce the computational time of cosmological parameter estimation from the CMB power spectrum (Auld et al., 2007, 2008) and in this thesis we present the first application to a CMB non-Gaussianity analysis. Here, an overview of the neural networks and its applications to the CMB non-Gaussianity are given while the details on the implementation are left to chapter 3.

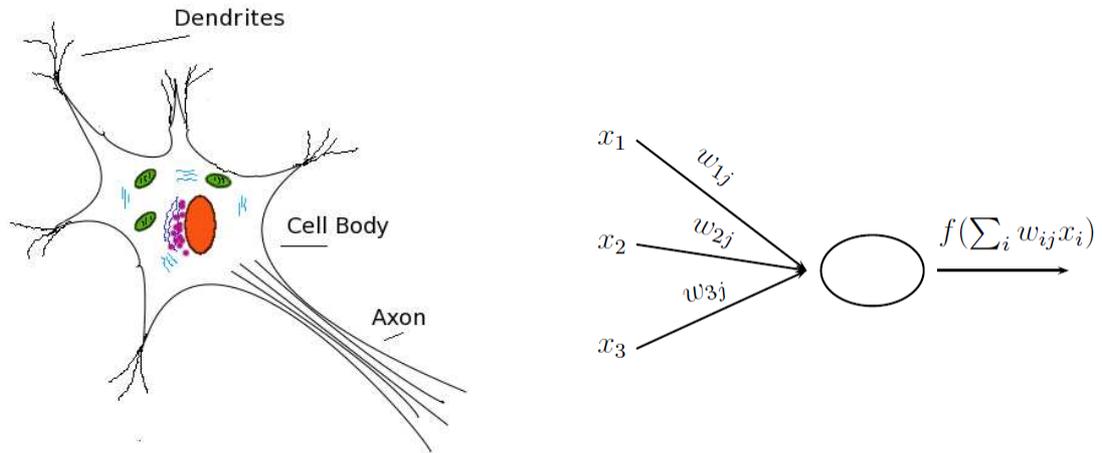


Figure 1.8 - Illustration of a schematic biological neuron on the left and the mathematical equivalent of one single artificial neuron.

ANN mechanism

An artificial neural network is a set of neurons (nodes) distributed in layers, connected to each other. The first layer represents the inputs of the problem and the last one is left to symbolise the outputs. The so-called *hidden* layers, with a number of *hidden* nodes, are those layers in between. The *job* of the neural network is to find the pattern that links the values of the inputs with the outputs. In a biological analogy, an artificial node represents the cell body, the input links would be the dendrites, and the output links would be the axon, which would connect with dendrites of other nodes. The intensity of those connections, are given by a mathematical function of the inputs (activation function), analogously to the intensity of the electromagnetic impulses of the biological network (Fig. 1.8). The strenght of the connections between the nodes (the weights value) will be given by what is called the learning process. A full network will have all the neurons connected to each other in any possible direction. However, a simpler network allowing only the forward connections, is commonly employed. A diagram of a typical *feed-forward* neural network is shown in Fig. 1.9, where all nodes are only connected to those of the following layer. The presence of a node, with the exception of the nodes in the first layer, indicates the application of an activation function. The first neural networks were designed to work with step functions, in order to activate or not the connection. Later on, it was found to be more effective the use of

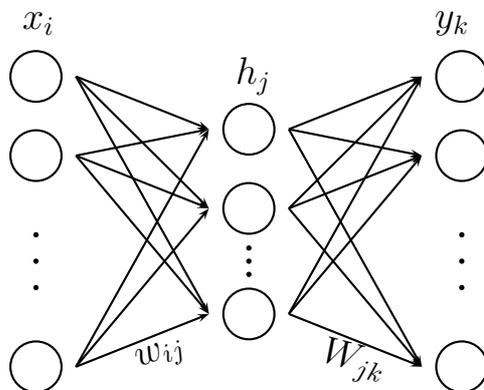


Figure 1.9 - Schematic diagram of a 3-layer neural network.

derivable functions, within a range of action. For that, sigmoidal functions are typically used, as the hyperbolic tangent.

The ANN mechanism is very simple, each node uses a linear combination of the receiving nodes and links, applies a sigmoidal function and gives an output that will be used as the input for the next layer. For example, for one hidden layer network the hidden nodes h_j are given by the activation function applied to a linear combinations of the inputs:

$$h_j = \tanh \left(\sum_i w_{ij} X_i + \theta_j \right) . \quad (1.76)$$

where w_{ij} represent the “strength” of the connection between the input x_i and the hidden node h_j . A similar process happens in the next layer: the hidden nodes would be the “inputs” for the next layer, that in this case (one hidden layer only) would be the output layer. The nodes of the last layer are usually set with a linear activation function. Then:

$$y_k = \sum_j W_{jk} h_j + \Theta_k . \quad (1.77)$$

Substituting eq. 1.76 into eq. 1.77, the outputs are left as a function of the inputs and the network parameters $\mathbf{a} = [\mathbf{w}, \theta, \mathbf{W}, \Theta]$. One sees that the final outputs are sensitive to the network architecture: how many layers there are and how populated they are.

Training the network

The basic idea of the artificial neural network is to use information that we already have, in order to learn from it and, finally, to obtain a *pseudo-model* reliable for predic-

tions. For this reason a set of known inputs and outputs (or *targets*) is used (e.g. if we want a neural network to learn how to perform an addition, the inputs would be the addends and the *targets* the result of the addition). The objective is to find the network outputs (eq. 1.77) that give the best answer for the *targets*. Hence, an optimization function \mathcal{E} is defined, such as the mean squared error, whose minimum will indicate the optimal weights. The differences between neural networks, besides the architecture and the choice of the activation function, lies on the selection of the minimization algorithm. One of the simplest methods is actualizing the network parameters using the back-propagation algorithm. In this algorithm the weights are updated through an iterative process until a desired tolerance is met. This process is very fast, however the solution is a local minimum, and for complex problems it could be far from optimal. Algorithms with a better performance as the conjugate-gradients based algorithms, simulated annealing or Powell's method among others, are commonly used (see e.g. Golden, 1996; Mackay, 2003, for complete neural network reviews.) Another approach to the neural networks is to look at it as a model fitting probabilistic problem from the Bayesian point of view. For this thesis, the algorithm to obtain the optimal weights was developed by Gull & Skilling (1999), and is based on the maximum entropy trajectory. In this case, the optimization function has a contribution given by the entropy S of the network parameters:

$$\mathcal{Q} = \mathcal{E} - \alpha S \quad (1.78)$$

where α starts from infinity and goes slowly to zero. \mathcal{E} is the optimization function (e.g. mean squared error):

$$\mathcal{E} = \frac{1}{N} \sum_l \sum_k (t_k^l - y_k(\mathbf{a}, \vec{x}^l))^2, \quad (1.79)$$

with k running from 1 to the number of outputs and l is the number of pairs input-*target*. The addition of the entropy to the optimization function helps the training process to converge smoothly to the minimum.

The error function \mathcal{E} can be interpreted as the minus log likelihood, where the data are the *targets*, and the model is given by the network parameters \mathbf{a} .

$$P(D|\mathbf{a}) \propto e^{-\mathcal{E}(\mathbf{a})} . \quad (1.80)$$

The second part of the optimization function can be interpreted in terms of a log prior probability distribution over the parameters:

$$P(\mathbf{a}|\alpha) \propto e^{\alpha S(\mathbf{a})} . \quad (1.81)$$

Following the Bayes theorem, the objective function \mathcal{Q} can be seen as the minus log posterior probability:

$$\begin{aligned} P(\mathbf{a}|D, \alpha) &= \frac{P(D|\mathbf{a})P(\mathbf{a}|\alpha)}{P(D|\alpha)} = \\ &\propto \frac{e^{-\mathcal{E}}e^{-\alpha S}}{P(D|\alpha)} \propto e^{-Q(\mathbf{a})}. \end{aligned} \quad (1.82)$$

In that sense, the network parameters that minimize \mathcal{Q} could be interpreted as the most probable parameter vector. For the neural network used here, $\mathcal{E} = \chi^2$ and S is the entropy defined as:

$$S(\mathbf{a}) = \sum_i \left(a_i - a_{o,i} - a_i \log \frac{a_i}{a_{o,i}} \right) \quad (1.83)$$

where \mathbf{a}_o is the initial guess for the network parameters. In practice the weights might take positive and negative values, therefore it is used the definition of the entropic prior for distributions with positive and negative values proposed by Hobson & Lasenby (1998):

$$S(\mathbf{a}) = \sum_i \left(\Upsilon_i - 2a_{o,i} - a_i \log \frac{\Upsilon_i + a_i}{2a_{o,i}} \right) \quad (1.84)$$

where $\Upsilon_i = \sqrt{a_i^2 + 4a_{o,i}^2}$. The advantage of using an entropic prior is basically improving for convergence. In addition, the use of the Bayes theorem allows one to obtain a posterior probability of the solution given by the network. This is very useful to decide which network architecture (number of hidden nodes and layers) to use, whereas in the frequentist approach this can only be done by training different architectures and comparing the results.

NN f_{NL} estimators

In this thesis, a new approach is proposed to obtain the f_{NL} estimator from a CMB map, avoiding large matrix estimations and inversions. Two estimators based on a neural network are tested finding similar results than with the classical approaches explained in the previous subsection (1.3.3). The inputs of the network should encapsulate in the best way the non-Gaussian signal, for this reason the third order moments defined at the beginning of this section will be used as inputs whereas the estimator of the f_{NL} parameter will be obtained by training a neural network. The f_{NL} estimators are constructed from the neural network parameters obtained in the training process. There are two main types of neural networks approaches: classification and

regression. The regression finds a relation between the inputs and the outputs, whereas the classification network relates several inputs into classes. The neural network classifier is required to find the characteristics of each class setting the boundaries between different classes. Both approaches are studied for non-Gaussianity and an overview of the estimators used are presented here, while more details are given in chapters 3 and 4.

■ Regression

A regression network is trained with the third order statistics of a map generated with a given f_{NL} . This f_{NL} value will be the required target, that is compared to the network output computed with the same inputs. To obtain the optimal NN parameters, the function \mathcal{Q} (eq. 1.78) is minimised using a conjugate-gradient algorithm. Then, with the optimal network parameters, the f_{NL} estimate for any given map can be computed. Using a neural network with one hidden layer (eq. 1.77) the estimator for f_{NL} is:

$$\hat{f}_{NL} = \sum_j W_{jk} \tanh \left(\sum_i w_{ij} S_i + \theta_j \right) + \theta_k . \quad (1.85)$$

However, as explained in chapter 3, we find that no hidden layers are required for this problem, thus the above expression is simplified:

$$\hat{f}_{NL} = \sum_j w_{ij} S_i + \theta_j . \quad (1.86)$$

■ Classification

A neural network can also be used as a classifier, where the outputs are classes. The network finds the pattern of the inputs that characterises a class. The output values would be a vector with dimension equal to the number of classes with zeros in all its components except for the class where the object belongs. In this case the outputs need to be transformed into probabilities, using a *soft-max* filter:

$$p_k = \frac{e^{y_k}}{\sum_r e^{y_r}} , \quad (1.87)$$

For the classifier, the optimisation function is also eq. 1.78, with the difference that now \mathcal{E} is typically the Kullback–Leibler divergence or the cross-entropy, that compares two distributions:

$$\mathcal{E} = \sum_t \left(- \sum_i q_i^t \log p_i(\vec{x}^t, \mathbf{a}) \right) \quad (1.88)$$

q_i^l are the training set *target* vectors, q_i is 1 if the input belongs in class i and 0 otherwise and l , as before, stands for the training pair used. For the case studied in this thesis, the classes correspond to the level of non-Gaussianity of each map. The range in f_{NL} used is divided in bins with a central value f_{NL}^c and a bin width smaller than the expected error bar. The network outputs p_i can be treated as the probability of a given input to belong to each class. Then the estimator for f_{NL} can be constructed by integrating over the classes:

$$\hat{f}_{NL} = \sum_i f_{NL,i}^c p_i \quad (1.89)$$

1.5. Gravitational lensing

As seen in the previous sections, the CMB has provided exceptional information to establish the current cosmological model, with very accurate constraints on the cosmological parameters. However, there is the need to observe the Universe in all its epochs to complete the cosmological picture. For that, the weak lensing observables are expected to be crucial.

General relativity predicts that the path of light from a distant galaxy is distorted by the gravitational potential fluctuations along the line of sight. This modification of the light paths is called gravitational lensing and is a powerful tool for probing the distribution of mass in the Universe. The variation of the light path depends on the position in the sky of the emitting object, the distance from the emitting object to the observer and on the potential along the light path (see fig. 1.10). As the Universe is in permanent evolution photons emitted at an earlier epoch will be differently deflected from those emitted later, due principally to the longer path length (for some reviews see Schneider et al., 1992; Narayan & Bartelmann, 1996; Mellier, 1999; Munshi et al., 2008). The combination of the depth information with angular information on the gravitational lensing distortion, allows for the reconstruction of the three-dimensional unbiased distribution of matter, or to perform statistical analysis to infer cosmological parameters from the dependence of the observables on the power spectrum and growth of density perturbations with redshift. Dark energy and modifications to Einstein gravity also act to modify the lensing effect by changing the distance-redshift relation in addition to the growth of density perturbations (Huterer, 2002; Munshi & Wang,

2003). Lensing effects are therefore a particularly valuable source of information for three of the important open issues in modern cosmology, namely the distribution of dark matter, the properties of dark energy and the nature of gravity.

1.5.1. Cosmology with weak lensing

The main contribution to cosmology given by the weak lensing observations rests primarily on the information provided about the dark matter and dark energy. In fact, weak lensing by itself would be, in principle, able to have more precise constraints than a CMB experiment on the parameter related to the clustering of the matter (σ_8), as well as on the parameters related to the dark energy equation of state (w_o and w_a). σ_8 represents the amplitude of the matter fluctuations, in particular the root mean square (rms) matter fluctuations today within a sphere of radius $8h^{-1}\text{Mpc}$. This cosmological parameter depends on the parameters defined in section 1.1.2, and its current value obtained with Planck is $\sigma_8 = 0.834 \pm 0.027$ at 68% CL (using only temperature).

In the standard cosmological model the dark energy is assumed to be a constant, however, there are alternatives proposed as the dynamical dark energy. These models are usually based on scalar fields allowing a connection with inflation, where there is an accelerated expansion as well. The dynamical dark energy will have an equation of state parameter w varying with the scale factor (note that for ΛCDM model $w = -1$). A typical parametrisation is:

$$w(a) = w_o + w_a(1 - a) . \quad (1.90)$$

This simple parametrisation can be useful to discriminate among different dark energy candidates, or at least discard some of the models (see Copeland et al., 2006, for a review on dark energy candidates). The dark energy parameters are probably the most weakly constrained parameters by CMB alone experiments, and this is why matter related observations are essential to reduce its errors. The most stringent constraints for w_o and w_a are obtained using CMB and barionic acoustic oscillations (BAO) combined observations: $w_o = 1.04 \pm 0.7$ and $w_a < 1.32$ both at 95% CL. (Planck Collaboration, 2013c).

Additionally, a combined analysis with CMB data will reduce the uncertainties on other cosmological parameters obtained with a CMB experiment alone. Forecasts of the cosmological parameters with a weak lensing experiment as Euclid alone and the joint analysis with CMB data can be seen, for example, in Heavens et al. (2006)

and Amendola et al. (2013), finding that a joint analysis reduces the errors of all the cosmological parameters.

1.5.2. Lensing formalism

The deflection angle produced by the lens is related to the distance from the object to the lens (D_{ds}), and the distances at which the lens (D_d) and the object (D_s) are located from us:

$$\vec{\theta}D_s = \vec{\beta}D_s + \vec{\alpha}D_{ds} , \quad (1.91)$$

or

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}\frac{D_{ds}}{D_s} . \quad (1.92)$$

The deflection angle sets the relation between the observed position of the object $\vec{\theta}$ with its true position $\vec{\beta}$.

This change of the light path is related to the Newtonian potential of the lens. For that it is defined an effective lensing potential, that is obtained from a projection of the Newtonian potential as:

$$\Psi(\vec{\theta}) = \frac{2}{c^2} \frac{D_{ds}}{D_d} D_s \int \Phi(\vec{\xi}, z) dz . \quad (1.93)$$

where $\vec{\xi}$ is the impact parameter (see fig. 1.10). The gradient of the effective lensing potential is related to the deflection angle as:

$$\vec{\nabla}_{\theta}\Psi = \frac{D_{ds}}{D_s}\vec{\alpha} . \quad (1.94)$$

This relation is only valid for the thin lens approximation, that means that the thickness of the lens is negligible compared to the distances involved, in that case the lens can be approximated by a plane. Another important relation is the one between the effective lensing potential and the mass distribution or the surface mass density of the lens ($\Sigma(\vec{\theta})$):

$$\vec{\nabla}_{\theta}^2\Psi = 2\frac{\Sigma(\vec{\theta})}{\Sigma_{cr}} . \quad (1.95)$$

Σ_{cr} is the critical surface mass density, and is defined as:

$$\Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}} . \quad (1.96)$$

The quantity $\kappa(\vec{\theta}) = \frac{\Sigma(\vec{\theta})}{\Sigma_{cr}}$ is what is called the convergence, and states the limit between strong lensing ($\kappa(\vec{\theta}) \geq 1$) and weak lensing ($\kappa(\vec{\theta}) < 1$). Therefore, eq. 1.95 can be

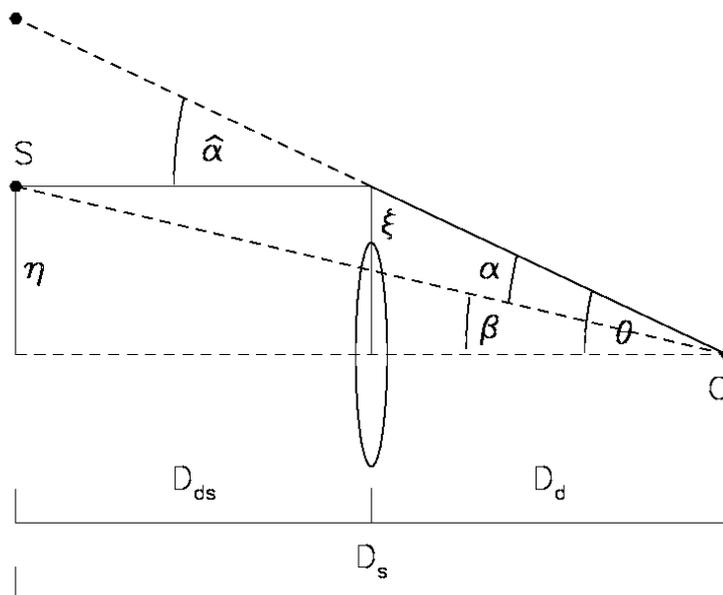


Figure 1.10 - Schematic illustration of the gravitational lensing deflection. S is the source, the middle circle represents the lens and the observer is marked with O . Figure from Narayan & Bartelmann (1996).

written as:

$$\vec{\nabla}_{\vec{\theta}}^2 \Psi = 2\kappa(\vec{\theta}) . \quad (1.97)$$

The distortions induced by gravitational lensing are described by the Jacobian matrix \mathcal{A} , that maps the true angular position of the image to the angular position of the source:

$$A_{ij} = \frac{\partial \vec{\beta}_i}{\partial \vec{\theta}_j} . \quad (1.98)$$

Using eq. 1.92 and eq. 1.94 A_{ij} can be written in terms of the Hessian matrix Ψ_{ij} of the effective lensing potential:

$$A_{ij} = \delta_{ij} - \Psi_{ij} . \quad (1.99)$$

There are two important quantities related to the second derivatives of the effective lensing potential. The first one is the convergence, mentioned before:

$$\kappa = \frac{1}{2}(\Psi_{11} + \Psi_{22}) \quad (1.100)$$

and the second one is the complex shear γ :

$$\gamma = \gamma_1 + i\gamma_2; \quad (1.101)$$

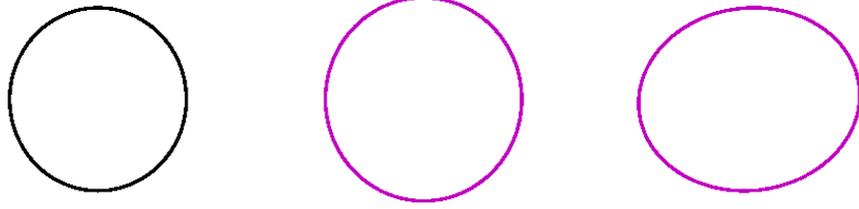


Figure 1.11 - Changes in an image due to different weak lensing fields. The first panel is without lensing, the middle panel is with $\kappa = 0.1$ and $\gamma = 0$, and the third panel has $\kappa = 0.1$, $\gamma_1 = 0.1$ and $\gamma_2 = 0.02$.

where

$$\gamma_1 = \frac{1}{2}(\Psi_{11} - \Psi_{22})\gamma_2 = \Psi_{12} . \quad (1.102)$$

Then the matrix \mathcal{A} can be written in terms of the convergence and the complex shear:

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \quad (1.103)$$

The convergence causes a magnification of the source, whereas the shear introduces anisotropic distortions that will have an impact on the shape of the source. The effects on a circular source are shown in fig. 1.11.

1.5.3. Weak lensing observables

Observationally there are three main effects on the background sources: changes in the ellipticity, magnification of the flux, and magnification of the size, the last two being directly related due to the conservation of the surface brightness in gravitational lensing. For large densities of matter the effects are all very strong, and multiple images of the background galaxies can be produced with large distortion. The study of these images has led to local reconstructions of the distribution of matter (see for example Tyson et al., 1984; Fort et al., 1988; Tyson et al., 1990; Kaiser & Squires, 1993; Mellier et al., 1993; Diego et al., 2005). When the potential fluctuations and their derivatives are small, the mapping from the source position to the image position on the sky is the identity matrix with corrections which are $\ll 1$.

In the weak lensing regime ($\kappa, \gamma \ll 1$) individual galaxy shear or convergence can not be determined observationally, but the distortions may be observed statistically using a large sample. The main cosmological information can be extracted from the convergence power spectrum P_κ . Consider the Fourier transform of $\kappa(\vec{\theta})$:

$$\kappa(\vec{\ell}) = \int e^{i\vec{\ell}\vec{\theta}} \kappa(\vec{\theta}), \quad (1.104)$$

then the 2D convergence power spectrum is:

$$\langle \kappa(\vec{\ell}) \kappa(\vec{\ell}') \rangle = (2\pi)^2 \delta_D(\vec{\ell} - \vec{\ell}') P_\kappa(\ell). \quad (1.105)$$

The Limber's equation provides an approximate relation between the spatial three dimensional correlation function $\xi(r)$ with the projected correlation function $C(\theta)$. Kaiser (1992) applied it to relate the 2D convergence power spectrum with the 3D matter power spectrum P_δ :

$$P_\kappa(\ell) = \frac{9H_0^4 \Omega_M^2}{4c^4} \int_0^\infty dr \left(\frac{g(r)}{a(r)} \right)^2 P_\delta \left(\frac{\ell}{r}, r \right). \quad (1.106)$$

where r is the radial coordinate, H_0 is the Hubble constant and Ω_M the matter energy density relative to the critical density. Furthermore,

$$g(r) = \int_r^\infty dr' n(r') \frac{r' - r}{r} \quad (1.107)$$

where $n(r)$ is the normalized radial distribution of the sources (for details see e.g. Schneider, 2005; Heavens, 2011). In the weak lensing limit the shear power spectrum is identical to the convergence power spectrum $P_\ell^\gamma \simeq P_\ell^\kappa$. Therefore, observations of both fields contain valuable cosmological information. If the fields were Gaussian we would not need to go to higher-order correlations. However, non-linearities in small scales make the weak lensing field non-Gaussian, carrying some of the information in higher-order correlations (see e.g. Bernardeau et al., 2012; Takada & Jain, 2003)

Cosmic shear

In weak lensing the most studied effect is the modification of the galaxy shape, a measure of the shear. The shape distortion has the main advantage that the intrinsic distribution of galaxy ellipticities is expected to be random, according to the cosmological principle, and therefore the average complex ellipticity is zero. In practice, in terms of changes in the ellipticity, the quantity that is directly observable is the reduced shear:

$$g = \gamma(1 - \kappa)^{-1} \quad (1.108)$$

that ignores the changes in size. This can be seen if we write the Jacobian matrix \mathcal{A} (eq. 1.103) in terms of the reduced shear:

$$\mathcal{A} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix} \quad (1.109)$$

where the factor $(1 - \kappa)$ affects the size and the distortion matrix only depends on the new quantity defined.

Weak lensing effects using galaxy ellipticities is a well-developed field and has been detected by several groups using different surveys and methods, see for example Wittman et al. (2000); Semboloni et al. (2006); Jarvis et al. (2006); Benjamin et al. (2007); Schrabback et al. (2010). In addition important efforts have been made to include and test many possible systematic effects on shape measurement (including, for example, the point spread function [PSF], instrumental noise and pixelization), and there are several algorithms that can measure shapes with varying degrees of accuracy including KSB (Kaiser et al., 1995), KSB+ (Hoekstra et al., 1998) and its variants (Rhodes et al., 2000; Kaiser, 2000), shapelets (Bernstein & Jarvis, 2002; Refregier & Bacon, 2003; Kuijken, 2006) and CHEF (Jiménez-Teja & Benítez, 2012), amongst others. A novel Bayesian model fitting approach *lensfit* was presented in Miller et al. (2007); Kitching et al. (2010). In order to test, in a blind way, the ability of methods to measure the shapes of galaxies a series of simulations have been created: STEP1, STEP2, GREAT08 and GREAT10, where several techniques have been tested and compared systematically. A summary of these methods and their performance for each of the simulated sets, is given in Heymans et al. (2006); Massey et al. (2007); Bridle et al. (2010) and Kitching et al. (2012), respectively.

Magnification

Another quantity related to the lensing observations is the magnification that, at first order, depends only on the convergence:

$$\mu = \frac{1}{\det(\mathcal{A})} = [(1 - \kappa)^2 - |\gamma|^2]^{-1} \simeq 1 + 2\kappa. \quad (1.110)$$

There are two main effects due to the magnification of background galaxies. The first of them is related to the amplification of the flux. A given galaxy is enlarged but its surface brightness remains constant, therefore the flux increases, and some faint galaxies that could not be observed in the absence of lensing can now be detected.

This changes the expected number of counts at a given magnitude (see for example Broadhurst et al., 1995; Taylor et al., 1998; Benítez & Sanz, 1999; Hildebrandt et al., 2011).

Assuming that the unlensed number density of sources with a flux larger than S and with redshift within $z \pm dz$ is $n_0(> S, z)$, thus the cumulative number density if there is a magnification effect μ is given by:

$$n(> S, z) = \frac{1}{\mu} n_0 \left(> \frac{S}{\mu}, z \right) . \quad (1.111)$$

The argument $\frac{S}{\mu}$ in the unlensed n_0 states that fainter sources can be observed if $\mu > 1$, whereas the multiplicative factor $\frac{1}{\mu}$ is related to the change of apparent solid angle. If a given region is amplified, the angular distances between galaxies are larger, and therefore the number density is reduced. From eq. 1.111 one sees that if the unlensed cumulative number density at a given flux is known, the magnification μ can be inferred from observations. If the unlensed distribution follows a power law $n_0(> S) = S^{-\alpha}$ the following expression is obtained:

$$\frac{n(> S, z)}{n_0(> S, z)} = \mu(z)^{\alpha(z)-1} , \quad (1.112)$$

and depending on the value of α the magnification effect would be easier or harder to detect. For example, on the B band $\alpha \sim 1$ while in redder bands $\alpha < 1$, and this is one of the reasons why they are preferred for weak lensing experiments.

The number of observed galaxies above a given flux may increase or decrease due to the lensing caused by the foreground galaxies. There are two competing effects, the flux increases and the number of counts would be greater than without lensing but simultaneously the solid angle covered by the images also increases, resulting to fewer observed galaxies per solid angle. In fig. 1.12 is shown that for $\mu > 1$, at low redshifts the dominant effect is the scattering of images by the lens. Meanwhile, at higher redshifts the increasing number of observable galaxies will be more important. For $\mu < 1$, that is negative values of the convergence field ($\kappa < -0.5$), the effects go in the opposite direction. Also one sees that there is a small range of redshifts where the magnification effect is cancelled. The other observable related to the magnification is the change of the source size. Near a foreground over-density, the lensed galaxies would be larger than in void regions. The area of a lensed galaxy is changed by:

$$A = \mu A_0 \simeq (1 + 2\kappa) A_0 , \quad (1.113)$$

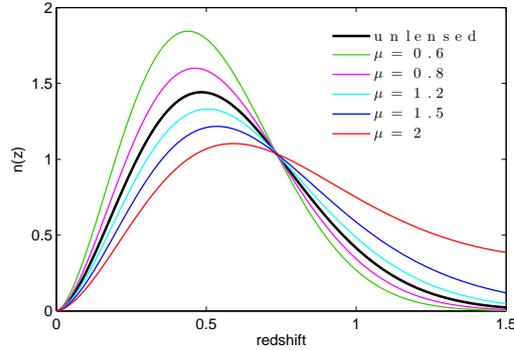


Figure 1.12 - Changes in the number density redshift relation above a given flux with the magnification. The expression for $n_0(z)$ and $\alpha(z)$ are taken from Broadhurst et al. (1995) for galaxies of magnitude $R=22.5$.

where A_0 is the unlensed galaxy area. The radius will change as $R \mapsto R_0(1 + \kappa)$. In the weak lensing limit, the power spectrum of the magnification fluctuations $(\mu - 1)$ is 4 times P_ℓ^κ , therefore, in principle, cosmological constraints could be made independently of the shear (Jain, 2002; Barber & Taylor, 2003). However the signal-to-noise ratio for the measured ellipticities is in general larger, hence the shear may carry more statistical weight. Even so a joint analysis of shear and magnification measurements will necessarily provide tighter constraints on cosmological parameters than a shear analysis alone. In particular, in van Waerbeke (2010) it is shown that the constraints on σ_8 and Ω_m can be improved up to $\sim 40\%$. Similarly combining size-magnification, galaxy densities and shear, the improvement on the precision of halo mass estimates can be $\sim 40\% - 50\%$ (Rozo & Schmidt, 2010). In Heavens et al. (2013) the improvement of including the size information is analysed for a Euclid-like experiment.

In contrast to galaxy ellipticity measurement, the size information has not been explored in detail, possibly because the complicated effects of the PSF and pixellisation were thought to be too challenging. However, there are two reasons for revisiting size magnification as a potential tool for cosmology: one is that accurate shear estimation is itself very challenging, and size could add useful complementary information; the second one is that methods devised for ellipticity estimation must deal with the PSF and pixellisation, and as a byproduct provide a size estimate, or a full posterior probability distribution for the estimated size, which is currently ignored or marginalised over. Therefore, the use of this information does not have an additional computational cost.

Intrinsic correlations

Besides instrumental and environmental issues, there can be astrophysical contaminants associated with weak lensing. In the case of shape distortion, there are intrinsic correlations that need to be taken into account. We should write the ellipticity two-point correlation function as:

$$\langle e_1 e_2 \rangle = \langle e_1^{int} e_2^{int} \rangle + \langle e_1^{int} \gamma_2 \rangle + \langle \gamma_1 e_2^{int} \rangle + \langle \gamma_1 \gamma_2 \rangle, \quad (1.114)$$

where the last term on the right is related to the power spectrum in eq. 1.106. In fact, none of the other terms vanish, since it is known that there are intrinsic alignments of nearby galaxies due to the alignment of angular momentum produced by tidal shear correlations (II correlations, see for detections Brown et al. 2002; Heymans et al. 2004; Mandelbaum et al. 2011; Joachimi et al. 2011, 2013, and for theory Heavens et al. 2000; Catelan & Porciani 2001; Crittenden et al. 2001; Heymans & Heavens 2003). In addition, as the density field might change the ellipticity of the nearby galaxies, and modifies the shape of the background galaxies, there can be correlations between density fields and ellipticities (GI correlations Hirata & Seljak, 2004; Mandelbaum et al., 2006). Intrinsic correlations have been studied in detail and it is not trivial to account for or to remove them when quantifying the weak lensing signal.

Regarding to the intrinsic correlation of sizes and its dependence on the environment, these are still open issues. In fact, the correlation of sizes and density field, it is known to play an important role in discriminating between models of size evolution; recent works find a significant correlation between sizes and the density field using around 11,000 galaxies drawn from the joint DEEP2/DEEP3 data-set (Cooper et al., 2012; Papovich et al., 2012), while earlier studies with smaller samples have been in disagreement. Using 5,000 galaxies of STAGES data-set, Maltby et al. (2010) find a possible anti-correlation between the density field and size for intermediate/low-mass spiral galaxies. According to this work, clustered galaxies seem to be 15% smaller than the field-galaxies, while they do not find any correlation for high-mass galaxies. Also for massive elliptical galaxies from the ESO Distant Clusters Survey, Rettura et al. (2010) do not find any significant correlation, while using the same data set Cimatti et al. (2012) claims a similar correlation as in Cooper et al. (2012). In Park & Choi (2009) they study the correlation between sizes and separation with late and early-type galaxies from the SDSS catalogue, at small and large scales. They compare the size of the nearest neighbour with the separation between them, and find larger galaxies at smaller

separations. This correlation is found for early-type galaxies if the separation between the galaxies is smaller than the merging scale, but not for larger separations. The size of late-type galaxies does not seem to have a correlation with the separation in any scale. We expect that further studies with larger samples will clarify the intrinsic correlations of sizes, we note that the systematics are generated from different physical processes than in the case of shear and this will affect the signal in a different way; we suggest this is a positive, and another reason why a joint analysis of ellipticity and sizes is interesting.

The following chapters are devoted to the work that has been published during my PhD. Those works were conducted before the Planck data was released, therefore the CMB analyses are performed on WMAP data. Chapter 2 is devoted to the analysis of the non-Gaussianity of WMAP-7yr data with the Healpix wavelet (Casaponsa et al., 2011). In Chapter 3 the neural networks are introduced in non-Gaussianity analysis (Casaponsa et al., 2011). Chapter 4 is an extension of the work of Chapter 3, with a more detailed discussion about the utility of neural networks as an alternative estimator for non-Gaussianity based on the paper (Casaponsa et al., 2013). In Chapter 5 the part of this thesis focused on weak lensing is presented and is based on the paper (Casaponsa et al., 2013). Finally the conclusions are drawn in Chapter 6 and Chapter 7 is left for a brief Spanish summary.

CHAPTER 2

Wilkinson Microwave Anisotropy Probe 7-yr constraints on f_{NL} with a fast wavelet estimator

This chapter is based on the published work of Casaponsa et al., 2011a. In this chapter a new method to constrain the local non-linear coupling parameter f_{NL} based on a fast wavelet decomposition is presented. Using a multiresolution wavelet adapted to the HEALPix pixelization, we have developed a method that is $\sim 10^2$ times faster than previous estimators based on isotropic wavelets and $\sim 10^3$ faster than the KSW bispectrum estimator, at the resolution of the Wilkinson Microwave Anisotropy Probe (WMAP) data. The method has been applied to the WMAP 7-yr V+W combined map, imposing constraints on f_{NL} of $-69 < f_{NL} < 65$ at the 95 per cent CL. This result has been obtained after correcting for the contribution of the residual point sources which has been estimated to be $\Delta f_{NL} = 7 \pm 6$. In addition, a Gaussianity analysis of the data has been carried out using the third order moments of the wavelet coefficients, finding consistency with Gaussianity. Although the constraints imposed on f_{NL} are less stringent than those found with optimal estimators, we believe that a very fast method, as the one proposed in this work, can be very useful, especially bearing in mind the large amount of data that will be provided by future experiments, such as the Planck satellite. Moreover, the localisation of wavelets allows one to carry out analyses on different regions of the sky. As an application, we have separately analysed the two hemispheres defined by the dipolar modulation proposed by Hoftuft et al. (2009). We do not find any significant asymmetry regarding the estimated value of f_{NL} in those hemispheres.

2.1. Introduction

As commented in the introduction of this thesis, the fluctuations of the CMB naturally arise in an inflationary scenario. The understanding of this very early stage of the history of the Universe is a challenging issue for the scientific community due to the implications on large scale structure formation and fundamental particle physics at high energies. A large number of inflationary models have been proposed in the literature (for an overview see for instance Lyth 2008) but the task of testing such scenarios is not trivial, and there is the need of new experiments and powerful statistical tools to discriminate among them. In this sense, the statistical properties of the CMB temperature anisotropies are a source of information about the processes that have generated the primordial fluctuations. In particular, the standard, slow roll, single field inflationary model predicts a nearly Gaussian distribution of the CMB temperature anisotropies, while alternative models may introduce a certain level of non-Gaussianity in the CMB. A convenient parametrization valid for a large set of non-standard inflationary models which includes the quadratic corrections of the primordial curvature perturbation is (Salopek & Bond, 1990; Gangui, 1994; Verde et al., 2000; Komatsu & Spergel, 2001):

$$\phi(r) = \phi_L(r) + f_{NL} [\phi_L^2 - \langle \phi_L^2 \rangle], \quad (2.1)$$

where ϕ_L are Gaussian linear perturbations and f_{NL} characterises the amplitude of the non-linear contribution in real space. This local form appears in non-standard multi-field inflationary models (Babich et al., 2004; Komatsu et al., 2009). For a complete review on non-Gaussianity due to inflationary models see Bartolo et al. (2004). In addition to inflationary models, there are other alternative scenarios that can be constrained, such as the ekpyrotic model where a negative value of f_{NL} is expected (Lehners, 2010). Moreover, there are other processes that can introduce deviations from Gaussianity in the third order moments (as foreground contamination, non-linear gravitational effects, topological defects, etc).

Since the quadratic parametrization was proposed, an important effort has been made to set observational constraints on local f_{NL} with a wide variety of methods including the bispectrum (Yadav & Wandelt, 2008; Smith et al., 2009; Komatsu et al., 2011), wavelet-based methods (Cayón et al., 2003; Mukherjee & Wang, 2004; Curto et al., 2009,b; Pietrobon et al., 2009; Rudjord et al., 2009), Minkowski functionals (Hikage et al., 2008) or the N-pdf (Vielva & Sanz, 2009, 2010). Most of these works find that the data are compatible with $f_{NL} = 0$, but the constraints are not yet sufficiently tight to dis-

criminate among a large set of inflationary models. The best limit for WMAP-7yr data is given by Komatsu et al. (2011) and is $-10 < f_{NL} < 74$ at the 95 per cent confidence level. These constraints have been obtained with a bispectrum estimator, which is computationally very demanding. However, Curto et al. (2011) have recently shown that an estimator based on the SMHW can provide constraints on f_{NL} as stringent as the optimal estimator, the bispectrum, while reducing considerably the CPU time. With the arrival of new data from high resolution experiments such as the ESA Planck satellite² (Tauber et al., 2010), it becomes even more important the availability of even faster and simpler methods. With this aim we present the application for CMB of a wavelet adapted to the HEALPix pixelization similar to the tool proposed by Shahram et al. (2007).

The chapter is organised as follows. In Section 2.2 we introduce the HEALPix wavelet decomposition. In Section 2.3 the method to constrain the f_{NL} parameter as well as the proposed Gaussianity test are described. Finally, we present the results of the application of this technique to the WMAP 7-yr data in Section 2.4 and conclusions are drawn in chapter 6.

2.2. The HEALPix wavelet

In this chapter we describe an application using the so-called HEALPix wavelet, described in sec. 1.4.2 of this thesis. In the previous work of (Shahram et al., 2007) a linear operator is applied to the HW to obtain wavelet coefficients corresponding to vertical, horizontal and diagonal orientations. This operation leads to a wavelet coefficients without redundancy, obtaining a number of wavelet coefficients (details plus approximation) equal to the number of original pixels. However, we have kept the HW with its intrinsic redundancy for three main reasons: first, to improve the computational time, second to obtain a wavelet decomposition as isotropic as the HW allows³, and third, because, as it is shown later, redundancy helps to improve the sensitivity in the detection of f_{NL} . Similarly to the SHW, the HW is a discrete, orthogonal wavelet, adapted to a hierarchical pixelization (such as HEALPix⁴, Górski et al. 2005), whereas

²<http://www.rssd.esa.int/index.php?project=planck>

³ It is worth mentioning that HW detail coefficients help to highlight the isotropy properties of the field as compared to the directional oriented details of the SHW. This is important because the local non-Gaussianities are expected to be isotropic.

⁴<http://healpix.jpl.nasa.gov/>

the SMHW is a continuous, non-orthogonal wavelet and does not have a hierarchical decomposition structure. The HW presents an optimal space localization, while the scale localization is not as good as that of the SMHW. However, the main advantage of the HW is that operates in the real space, with the computational cost significantly reduced.

The resolution of a HEALPix map is characterised by the N_{side} parameter, such that the number of pixels in which the sphere is divided corresponds to $N = 12N_{\text{side}}^2$. N_{side} can only take powers of two as values. The HW decomposes the temperature map at resolution J , where $N_{\text{side}} = 2^J$, in wavelet coefficient maps at all the allowed HEALPix resolutions down to the lower considered resolution j_0 . As previously commented the wavelet decomposition of a temperature map can be written in terms of the basis functions (sec. 1.4.2) and a set of coefficients:

$$\frac{\Delta T}{T}(x_i) = \sum_{k=0}^{N_{j_0}-1} \lambda_{j_0,k} \varphi_{j_0,k}(x_i) + \sum_{j=j_0}^J \sum_{m=0}^3 \sum_{k=0}^{N_j-1} \gamma_{m,j,k} \Psi_{m,j,k}(x_i), \quad (2.2)$$

where N_j is the number of pixels at resolution j . $\lambda_{j,k}$ and $\gamma_{m,j,k}$ are the approximation and detail coefficients respectively. From a practical point of view, to perform the decomposition, we start with the original resolution, i.e. $j = J$. At this resolution, the approximation coefficients $\lambda_{J,k}$ correspond to the pixels of the original temperature map. The approximation coefficients at the next resolution are simply obtained by degrading the map to the immediately lower resolution (i.e., by averaging the corresponding four daughter pixels):

$$\lambda_{j,k} = \frac{1}{4} \sum_{i=0}^3 \lambda_{j+1,k_i}, \quad (2.3)$$

On the other hand, the detail coefficients at resolution $j + 1$ are simply obtained by subtracting the approximation at resolution j from the approximation at resolution $j + 1$. Thus, the detail coefficients are defined as:

$$\gamma_{0,j,k} = \lambda_{j+1,k_0} - 4\lambda_{j,k} \quad (2.4)$$

$$\gamma_{1,j,k} = \lambda_{j+1,k_1} - 4\lambda_{j,k}$$

$$\gamma_{2,j,k} = \lambda_{j+1,k_2} - 4\lambda_{j,k}$$

$$\gamma_{3,j,k} = \lambda_{j+1,k_3} - 4\lambda_{j,k}$$

Note that the 4 in the second term on the right side appears due to the pixel area weight.

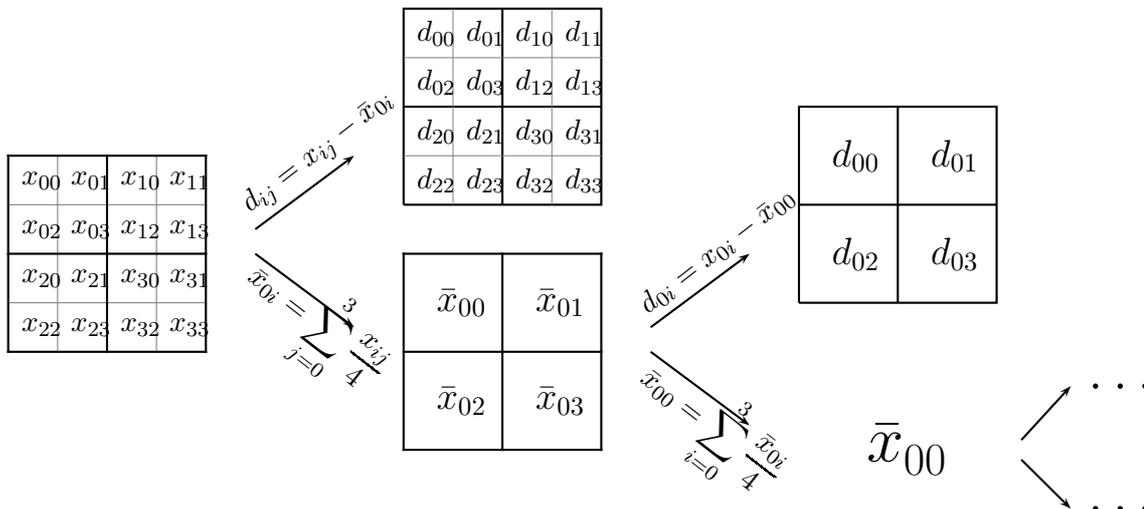


Figure 2.1 - Diagram of the construction of the approximation and detail coefficients. Approximation coefficients are computed as the average of the four daughter pixels. Detail coefficients are computed as the subtraction of that average from the original pixels and are represented by d .

A schematic diagram of how to obtain the approximation and detail coefficients is given in Fig. 2.1.

2.3. Methodology

The main purpose of this work is to constrain the parameter f_{NL} defined in Eq. (2.1) using the WMAP-7yr data⁵. For this analysis, we only consider the (foreground reduced) V and W channels, since they are less affected by foreground contamination. A single CMB map is obtained through a noise-weighted linear combination of the V and W receivers. The KQ75 mask (which covers around a 29 per cent of the sky) is applied subsequently.

In order to calibrate our estimator, we need both Gaussian and non-Gaussian simulations. To generate the Gaussian simulations, we compute the power spectrum that best fits the WMAP-7yr data accordingly to the parameters estimated by Komatsu et al. (2011). For this purpose we use the On-line tool CAMB (Lewis et al., 2000). We then

⁵The data are available at the LAMBDA web page: <http://lambda.gsfc.nasa.gov/>

apply the corresponding beam and pixel functions to simulate the data at each of the considered receivers (2 for V and 4 for W). A Gaussian noise realisation is subsequently added to the CMB maps with a variance per pixel given by $\frac{\sigma_0}{N_{obs}}$, where σ_0 is the detector sensitivity of each of the receivers and N_{obs} is the number of observations at each pixel. Finally the six maps are combined in the same way as the data.

Regarding the non-Gaussian simulations, we have used the 1000 simulations generated by Elsner & Wandelt (2009) that are publicly available⁶. The previous authors provide the harmonic coefficients for the Gaussian and non-Gaussian parts of the simulation. A non-Gaussian simulation with a given value of f_{NL} is then constructed as:

$$a_{lm} = a_{lm}^{(G)} + f_{NL}a_{lm}^{(NG)}, \quad (2.5)$$

where we have normalized $a_{lm}^{(G)}$ and $a_{lm}^{(NG)}$ ⁷ to the power spectrum that best fits the WMAP-7yr data, and that was used for the Gaussian simulations (the original simulations were obtained using the WMAP 5-yr power spectrum). Again, we construct the maps for the V and W receivers, applying the corresponding beam and pixel transfer functions and adding the appropriate level of noise. Finally a single V+W combined map is constructed for each non-Gaussian simulation.

2.3.1. Cubic statistics

In this section we define the third order moments of the wavelet coefficients that are used to constrain f_{NL} . Similar statistics have been used in other previous works (Curto et al., 2009b; Rudjord et al., 2009).

We perform the wavelet decomposition of the considered map starting at resolution $N_{side} = 512$ ($J = 9$) and down to $N_{side} = 2$ ($j_0 = 1$) (when using a higher value of j_0 we are losing efficiency whereas for $j_0 = 0$ the results are not significantly improved while the computational time increases by a 30%). We obtain 8 detail maps and 1 approximation map. In addition, we also include in the analysis the original map and the 8 intermediate approximation maps (which are obtained during the wavelet decomposition to construct the detail coefficients). Although, in principle, these additional approximation maps contain redundant information, they seem to provide additional information regarding the third order statistics, since a larger number of third order combinations can be constructed. In fact, we have tested that with the inclusion of the

⁶<http://planck.mpa-garching.mpg.de/cmb/fnl-simulations/>

⁷The amplitude of the a_{lm}^{NG} has been corrected by a factor of $\frac{3}{5}$ as indicated by the authors

approximation maps and the original map, the results are improved by a 30%. Therefore, we have a total of 18 maps for each analysed signal. The statistics are constructed as the third order moments of all the possible combinations of these 18 maps, where the coefficients are weighted to take into account the presence of a mask. In order to calculate these weights, one performs the wavelet decomposition of the considered mask (that has zeros in the masked pixels and ones in the rest). The wavelet coefficients of the mask at each detail and approximation scale are used to construct the weight $w_j(i)$ of the coefficient at position i at resolution j . This makes sense if one bears in mind how the wavelet decomposition is performed. For instance, to construct the approximation map at resolution $J - 1$ at a given position i , the four daughter pixels at resolution J have to be averaged. If the four pixels are unmasked, this corresponds to a weight of 1 in the original map and also in the approximation map at position i . However, if one of the original pixels is masked, this pixel would have a weight of zero, and the average would be done only over three pixels. Thus the weight of the corresponding approximation coefficient would be $3/4$. Therefore, this weighting scheme takes into account the fact that different coefficients contain different amount of information, depending on the considered mask. Also, contrary to the case of other wavelet estimators, this means that the mask does not need to be extended but, in fact, it is reduced when increasing the scale. This is due to the fact that a larger pixel is kept for the analysis, with the appropriate weight, if at least one of the daughter pixels was unmasked.

The third order statistics are then defined as:

$$S_{jkl} = \frac{1}{N_{l-1}} \frac{\sum_{i=0}^{N_l-1} W_{jkl}(i) \epsilon_{i,j} \epsilon_{i,k} \epsilon_{i,l}}{\sum_{i=0}^{N_l-1} W_{jkl}(i) \sigma_j \sigma_k \sigma_l}, \quad (2.6)$$

where $\epsilon_{i,j} = y_{i,j} - \mu_j$ and $y_{i,j}$ are the wavelet coefficients maps at position i at resolution j . Note that j goes from j_0 to J , k goes from j to J and l goes from k to J . μ_j and σ_j are the weighted mean and the dispersion for the map at resolution j . $W_{jkl}(i)$ is the weight associated to the wavelet coefficients at position i and scales j, k, l and is given by:

$$W_{jkl}(i) = \sqrt[3]{w_j(i)w_k(i)w_l(i)}. \quad (2.7)$$

Note that some of these statistics are redundant (linearly dependent between them), so we restrict our analysis to the set of non-redundant statistics, which gives a total of $n_{\text{stat}} = 232$ quantities.

The process for computing these statistics requires $\sim N \times n_{\text{stat}}$ number of operations, where N is the number of pixels and n_{stat} the number of statistics computed. This number is significantly lower than that of the full bispectrum that needs $N^{\frac{5}{2}}$ operations. Using the KSW algorithm presented in Komatsu et al. (2005) the number of operations is reduced to $\sim rN^{\frac{3}{2}}$, where r is the number of sampling points (of the order of 100). On the other hand, the SMHW scales as $\sim n_s N^{\frac{3}{2}}$, where n_s is the number of scales involved (of the order of 10). Thus, at WMAP resolution ($N_{\text{side}} = 512$ and $N \sim 3 \times 10^6$) we have that the method presented in this work is 10^2 times faster than the SMHW, 10^3 times faster than KSW bispectrum estimator and 10^7 faster than the general bispectrum estimator.

2.3.2. Gaussianity test and f_{NL} constraints

We first perform a Gaussianity test in order to probe whether the data is compatible with Gaussianity using the χ^2 estimator:

$$\chi^2 = \sum_{i,j=1}^{n_{\text{stat}}} (v_i - \langle v_i \rangle) C_{ij}^{-1} (v_j - \langle v_j \rangle), \quad (2.8)$$

where v_i is the vector of the third order statistics computed from the considered map (to simplify notation, hereinafter we define $v_1 \equiv S_{111}, v_2 \equiv S_{112}, \dots$). $\langle v_i \rangle$ and C_{ij} are the mean and covariance matrix of the statistics obtained with 10000 Gaussian simulations. To perform the Gaussianity test, the value of the χ^2 is computed for the WMAP data, and compared to the distribution of the estimator obtained from an additional set of 1000 Gaussian simulations.

The second analysis that has been performed is the estimation of f_{NL} from the data. As the wavelet decomposition is linear, we can obtain the wavelet coefficients from the Gaussian and non-Gaussian parts separately. Thus, the wavelet coefficients for a given value of f_{NL} are given by

$$y_i = y_i^{(G)} + f_{NL} y_i^{(NG)}. \quad (2.9)$$

Taking into account that $y^{(NG)}$ are around 4 orders of magnitude smaller than $y^{(G)}$, when we compute $\langle y^3 \rangle$ the NG high-order terms can be neglected and it can be shown that f_{NL} is proportional to the wavelet estimators, as it is also the case for other statistics (such as the bispectrum):

$$v_i = a_i f_{NL}, \quad (2.10)$$

where a_i can be computed from simulations with a simple linear regression.

In order to estimate the f_{NL} parameter, we perform a χ^2 minimisation. In particular, $\chi^2(f_{NL})$ is defined as follows

$$\chi^2 = \sum_{i,j=1}^{n_{\text{stat}}} (v_i - \langle v_i \rangle_{f_{NL}}) C_{ij}^{-1}(f_{NL}) (v_j - \langle v_j \rangle_{f_{NL}}), \quad (2.11)$$

where $\langle v_i \rangle_{f_{NL}}$ is the mean of the statistics for a given value of f_{NL} obtained from the 1000 non-Gaussian simulations and $C_{ij}(f_{NL})$ is the corresponding covariance matrix. For $f_{NL} \ll 1000$ is reasonable to use the approximation $C_{ij}(f_{NL}) \simeq C_{ij}$, where C_{ij} is the covariance matrix for the Gaussian case.

Error bars on the parameter estimation at different confidence levels are found using the Gaussian simulations. We also compute the minimum variance in a semi-analytical manner. It is well known that the diagonal of the inverse of the Fisher matrix provides an estimation of the variance of the parameters. In order to estimate the Fisher matrix, we approximate the distribution of the statistics by a Gaussian. Using this approximation and taking into account Eqs. (2.10) and (2.11), the variance from the Fisher Matrix can be written as:

$$\sigma^2 = \frac{1}{\sum_{i,j}^{n_{\text{stat}}} a_i C_{ij}^{-1} a_j}. \quad (2.12)$$

In practice, the distribution of the statistics do not follow a perfect Gaussian distribution. Therefore, this variance can be seen as a lower limit to the true underlying variance.

2.4. Results

In this section, we present the analysis of the WMAP-7yr V+W combined map. On the one hand, we analyse the compatibility of the data with Gaussianity using the cubic statistics defined in Eq. (2.6) and the estimator presented in Eq. (2.8). On the other hand, we compute the best-fit f_{NL} parameter from the data by minimizing Eq. (2.11). Error bars are set using simulations. In addition, we also present a study of the contribution of the point sources to the estimated f_{NL} value and of the variation of f_{NL} estimated from two independent hemispheres (defined by Hoftuft et al. 2009).

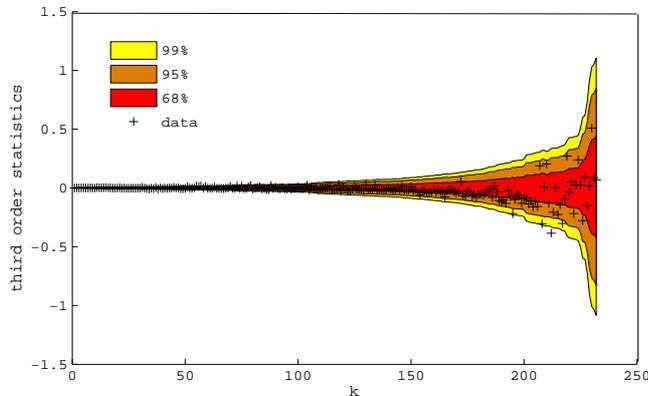


Figure 2.2 - The cubic statistics v_i from WMAP-7y V+W data are shown. Shadow areas correspond to the 68, 95 and 99 per cent confidence levels of the distribution obtained from 1000 Gaussian simulations. The statistics have been plotted from lower to higher variance.

2.4.1. Gaussianity test

As explained in Section 2.3.1, we have considered a total of 232 cubic statistics, constructed from 18 maps at 9 different scales. Fig. 2.2 shows the value of v_i for the WMAP 7-yr V+W data, after applying the KQ75 mask. The plot does not show any obvious deviation from Gaussianity. To further study the consistency of the data with Gaussianity, we also perform the χ^2 test defined in Eq. (2.8). From 1000 Gaussian simulations, we estimate the distribution of this quantity, finding a mean value of $\langle \chi^2 \rangle = 233$, very close to the number of degrees of freedom (232). The value of the dispersion is 69, larger than expected for a χ^2 distribution with the considered degrees of freedom. However, this may be explained by the fact that the distribution of the different statistics are not purely Gaussian. For the WMAP data, we find $\chi^2_{data} = 434$ with a cumulative probability of $P(\chi^2 \leq \chi^2_{data}) = 0.96$. Although the result indicates that the WMAP data is somehow in the tail of the distribution, the χ^2 value is not large enough to claim a deviation from Gaussianity.

2.4.2. Constraints on the f_{NL} parameter

We have also performed an estimation of the non-linear parameter f_{NL} . As already mentioned, for this analysis we have used the 1000 non-Gaussian simulations provided by Elsner & Wandelt (2009). In Fig. 2.3, the mean of the cubic statistics derived from

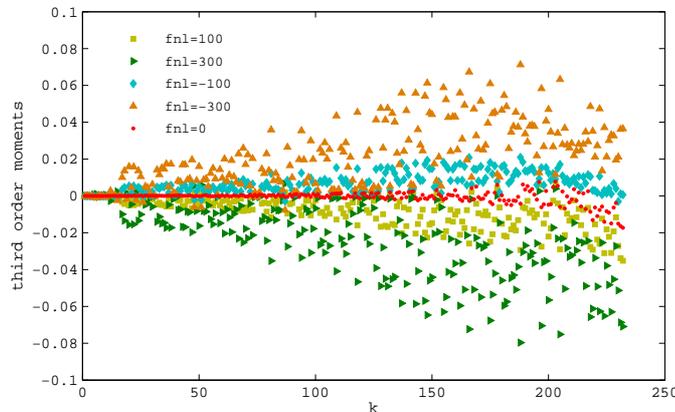


Figure 2.3 - Mean values of the cubic statistics v_i from 1000 non-Gaussian simulations with $f_{NL} = 0, \pm 100, \pm 300$.

simulations with $f_{NL} = 0, \pm 100, \pm 300$ is presented. It can be seen that, as stated in Eq. (2.10), the statistics are proportional to the value of f_{NL} .

After obtaining the cubic statistics for the WMAP-7yr data and minimising the χ^2 given by Eq. (2.11), we estimate that the best-fit value of f_{NL} is 6. Using Gaussian simulations, we find that the constraints for the parameter are $-28 < f_{NL} < 40$ at the 68 per cent confidence level and $-62 < f_{NL} < 72$ at the 95 per cent confidence level. It is also interesting to point out the agreement between the dispersion computed semi-analytically through the Fisher matrix (Eq. 2.12) and that obtained from Gaussian simulations, which are both estimated to be around 34.

Although the constraints provided by the HW are less stringent than those found with optimal estimators (such as the bispectrum or the SMHW), they are similar or even better than those obtained by other methods such as needlets (Pietrobon et al., 2009; Rudjord et al., 2009), the Minkowski functionals (Hikage et al., 2008) or the N-pdf (Vielva & Sanz, 2009). Moreover, as already pointed out, our estimator is significantly faster than all the previously mentioned methods, providing a very valuable tool, especially for future high resolution experiments such as Planck. It is also interesting to point out that we find a more symmetric constraint around zero than those obtained, for instance, by Komatsu et al. (2011) or Curto et al. (2011).

In order to study further the robustness of our estimator, we have performed some additional tests. In particular, we have estimated the mean value and dispersion of the best-fit value of f_{NL} from simulations with different values of f_{NL} . The left panel of

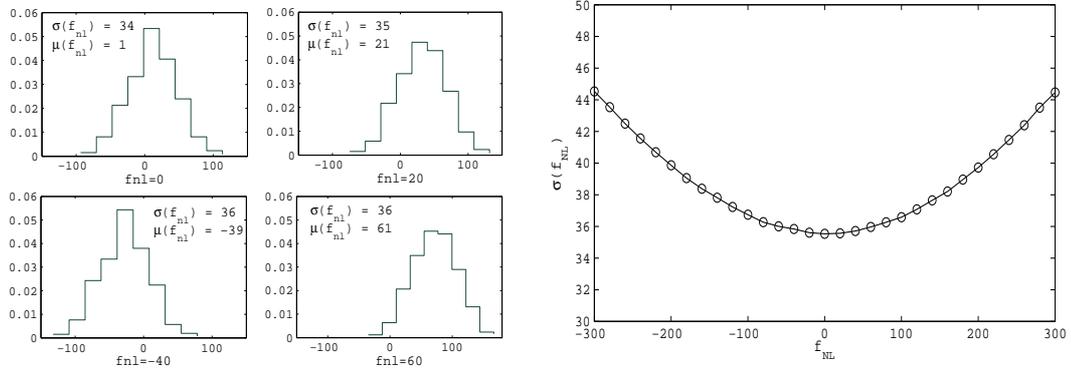


Figure 2.4 - The left panels show the histograms of the estimated f_{NL} from simulations with values of $f_{NL} = -40, 0, 20, 60$. The mean value and dispersion of f_{NL} for each considered case is indicated in the corresponding panel. In the panel of the right the behaviour of $\sigma(f_{NL})$ is shown when estimated from simulations with different values of f_{NL} .

Fig. 2.4 shows the histograms of the estimated values of f_{NL} obtained from simulations with $f_{NL} = -40, 0, 20, 60$. To carry out these tests, we have used 500 of the 1000 non-Gaussian simulations to estimate the mean value of the third order statistics $\langle v_i \rangle_{f_{NL}}$ and the remaining 500 simulations to obtain estimates of f_{NL} and construct the histograms. The mean values and dispersions of f_{NL} are given in the corresponding panels. In particular, we see that the method is unbiased, since the mean value of the estimated f_{NL} is very close to the true underlying value for all the considered cases. In addition, we also plot in the right panel of Fig. 2.4 how the dispersion of the estimator varies as a function of f_{NL} . The standard procedure to estimate this dispersion is to use Gaussian simulations but, as seen in the plot, this gives a minimum in the estimated value of $\sigma(f_{NL})$. However, for small values of f_{NL} , such as the ones found in this work, the variation is small and therefore one can safely use the value of the dispersion estimated for the Gaussian case.

2.4.3. Point source contribution

The background of unresolved point sources may introduce a bias in the estimation of f_{NL} . In order to correct this bias, we have studied the contribution to f_{NL} given by a point source background that is added to the Gaussian simulations. For that purpose, we have produced point source simulations following the procedure of Curto et al. (2009). In particular, point sources maps are simulated according to the

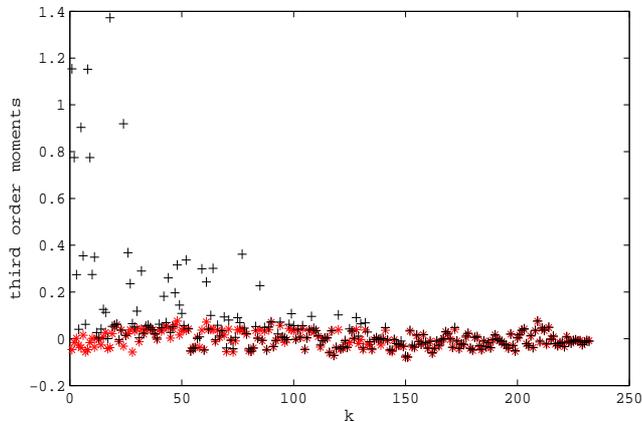


Figure 2.5 - Mean values of the v_i statistics obtained from 1000 Gaussian simulations with and without point sources. Diamonds represent CMB plus noise simulations, while crosses correspond to simulations including also the point sources. To improve the visualization, the statistics have been normalised to unit dispersion.

density distribution given by de Zotti et al. (2005) in a range of intensities between $I_{min} = 1\text{mJy}$ and $I_{max} = 1\text{Jy}$. These maps are then convolved with the corresponding beam and pixel functions and added to the simulations containing Gaussian CMB plus noise. The estimated value of f_{NL} when point sources are present is then compared to the one obtained from simulations containing only CMB and noise, finding a difference of $\Delta f_{NL} = 7 \pm 6$. Fig. 2.5 shows the effect that point sources have on the v_i statistics. As one would expect, they mainly affect the statistics involving small scales, that correspond to the ones with a lower value of k in the figure. Taking into account this result our final constraint on f_{NL} for the WMAP-7yr data is $-69 < f_{NL} < 65$ at the 95 per cent confidence level.

2.4.4. Local study of f_{NL}

Finally, we have analysed the data considering two independent hemispheres. In particular, we have considered the hemispheres associated to the dipolar modulation proposed by Hoftuft et al. (2009) where the preferred direction is pointing towards the Galactic coordinates $(l,b)=(224^\circ, -22^\circ)$. We have estimated the best-fit value and constraints on f_{NL} for the WMAP 7-yr data in both hemispheres, following the same procedure as for the full-sky. After correcting the point source contribution, the constraint found for the northern hemisphere is $-73 < f_{NL} < 119$ while for the southern hemisphere we have $-137 < f_{NL} < 62$, both at the 95 per cent confidence level. There-

fore, as it was the case for the full-sky, both hemispheres are consistent with Gaussianity (i.e, $f_{NL} = 0$). We have also tested that the results from the hemispheres are consistent between them. In particular, we have obtained the mean difference and dispersion between the f_{NL} estimates at each hemisphere for Gaussian simulations, finding values of $\langle \Delta f_{NL} \rangle = -4$ and $\sigma(\Delta f_{NL}) = 71$. For the WMAP data, we have $\Delta f_{NL} = 67$, which is perfectly consistent with the values expected from Gaussian simulations. Therefore, we do not find any assymetry for the considered hemispheres. These results are in agreement with the analysis based on needlets made by Pietrobon et al. (2010) and Rudjord et al. (2010) for the WMAP-5yr data, where several divisions of the CMB map are studied without finding a significant asymmetry. In a recent work, Vielva & Sanz (2010) have found an asymmetry in the same hemispheres studied in this work on the estimation of the f_{NL} using the N-pdf. The disagreement may be caused by the differences on the methods. While Pietrobon et al. (2010) and Rudjord et al. (2010) have worked with the same resolution as we did (6.9 arcmin), Vielva & Sanz (2010) focused on scales around 2° . Also, the non-Gaussian model used by the former works is the same as the one used in this work, whereas the model of the latter stands on the Sachs-Wolfe regime.

CHAPTER 3

Constraints on f_{NL} from Wilkinson Anisotropy Probe 7-year data using a neural network classifier

The contents of this chapter are based on the published work of Casaponsa et al., 2011b. In this chapter is presented a multi-class neural network (NN) classifier as a method to measure non-Gaussianity, characterised by the local non-linear coupling parameter f_{NL} , in maps of the cosmic microwave background (CMB) radiation. The classifier is trained on simulated non-Gaussian CMB maps with a range of known f_{NL} values by providing it with wavelet coefficients of the maps; we consider both the HEALPIX (HW) wavelet and the spherical Mexican hat wavelet (SMHW). When applied to simulated test maps, the NN classifier produces results in very good agreement with those obtained using standard χ^2 minimization. The standard deviations of the f_{NL} estimates for WMAP-like simulations were $\sigma = 22$ and $\sigma = 33$ for the SMHW and the HW, respectively, which are extremely close to those obtained using classical statistical methods in Curto et al., 2011 and Casaponsa et al., 2011a. Moreover, the NN classifier does not require the inversion of a large covariance matrix, thus avoiding any need to regularise the matrix when it is not directly invertible, and is considerably faster.

3.1. Introduction

Artificial intelligence algorithms are being used increasingly to improve the efficiency of computationally intensive data analysis. In particular, neural networks (NN) have

been successfully applied to pattern recognition, classification of objects and parameter estimation in a number of fields, including cosmology (see e.g. Auld et al., 2007).

Cosmological analysis typically involves the use of large datasets and high precision numerical tools. In particular, the study of deviations from Gaussianity in the distribution of temperature anisotropies in the cosmic microwave background (CMB) require very demanding computational methods. The simplest way to characterise such a deviation is through third order moments, as these vanish in the Gaussian case. It is now commonplace in CMB analysis to work in spherical harmonic space, where computing the three point correlation function or bispectrum can prove difficult, or indeed impossible, due to numerical instability. Some recent efforts have been applied to lessen the computational demand without reducing efficiency; see for example the KSW bispectrum estimator (Komatsu et al., 2005), or the binned estimator (Bucher et al., 2010). Other methods which have also been applied to non-Gaussianity analysis include Minkowski functionals (Hikage et al., 2008; Natoli, 2010), wavelet-based methods (Cayón et al., 2001; Mukherjee & Wang, 2004; Curto et al., 2009,b; Pietrobon et al., 2010; Casaponsa et al., 2011), a Bayesian approach (Elsner & Wandelt, 2010) and the analysis of the N -dimensional probability density function (Vielva & Sanz, 2010).

This chapter introduces an approach based on a neural network classifier which, after training on third order moments of wavelet coefficients derived from simulated Gaussian and non-Gaussian CMB realisations, can be used to estimate the presence and degree of non-Gaussianity for any given data map. We have chosen to estimate the local non-linear coupling parameter f_{NL} , which parameterises the local non-Gaussianity as a quadratic term in the primordial curvature perturbation. More precisely, f_{NL} is the amplitude of the corrections at second order of the primordial curvature perturbations (Salopek & Bond, 1990; Gangui, 1994; Verde et al., 2000; Komatsu & Spergel, 2001). This type of non-Gaussianity is predicted even in the simplest slow-roll inflationary scenario, albeit at a very low level $f_{\text{NL}} < 1$, whereas a wide range of non-standard inflationary models predict much larger typical f_{NL} values (for a more complete review see Bartolo et al. (2004), Babich et al. (2004) and Yadav & Wandelt (2010)). Estimating the value of f_{NL} from a given data map using existing methods typically has a high computational cost and usually numerical problems arise (e.g. matrix inversion). As we will show, the use of neural networks bypasses these difficulties.

In principle, one could use the pixel temperatures in the CMB map directly, or its spherical harmonic coefficients, as the inputs to the neural network classifier. Nonethe-

less, we perform a pre-processing step in which we decompose the temperature maps into their wavelet coefficients, which have shown themselves to be sensitive to non-Gaussian signals (e.g. Curto et al., 2009b, 2011; Casaponsa et al., 2011). In particular, we consider the HEALPIX wavelet (HW) and a spherical Mexican hat wavelet (SMHW), and compute third-order moments of these wavelet coefficients, the mean value of which is proportional to f_{NL} . The network is then trained so that when presented with these cubic statistics for a new (data) map, it can estimate the f_{NL} value and its error bar. We apply this method to estimate the degree of non-Gaussianity in the Wilkinson microwave anisotropy probe (WMAP) 7-year data release.

This chapter is organized as follows. In Section 3.2, we give a brief introduction to the wavelet analysis used. An overview of the type of neural network employed and our training procedure follows in Section 4.2. In Section 3.4 we explain the generation of our Gaussian and non-Gaussian simulations, and the specific characteristics of our f_{NL} classification network. We present the results of applying our classifier to simulations and to WMAP 7-year data in Section 3.5. The conclusions are summarised in chapter 6.

3.2. Wavelet estimators to train the neural network

Wavelet methods have seen increasing usage in cosmology. This has been particularly marked in CMB non-Gaussianity analyses, in which competitive results have been obtained using wavelets such as the SMHW (Cayón et al., 2003; Vielva et al., 2004; Cruz et al., 2005; Curto et al., 2011), directional spherical wavelets (McEwen et al., 2008), spherical Haar wavelet (SHW) (Tenorio et al., 1999; Barreiro et al., 2000), and recently the HEALPIX wavelet (HW) (Casaponsa et al., 2011). For a review of wavelets applied on the sphere, see, for example, McEwen et al. (2007). In essence, decomposing a CMB map into its wavelet coefficients allows one to separate the search for non-Gaussianity on different length-scales, while retaining positional information. In this section we will briefly discuss the characteristics of both the HW and SMHW and describe how we construct the statistics which are used in our analysis.

- **HEALPIX wavelet**

The HEALPIX wavelet is very similar to that presented by Shahram et al. (2007). Casaponsa et al. (2011) have used a revised version of this wavelet and perform the first cosmological application. In both papers, the central idea is the construction of a fast wavelet method adapted to the HEALPIX pixelization scheme

(Górski et al., 2005). The HW is similar to the SHW in the sense that, at each level of the wavelet transform, one produces both a high- and low-resolution map. The low-resolution map for the HW is obtained simply by averaging over 4-pixel blocks, and the high-resolution map is just the original map minus the low-resolution map. One begins with the original map at resolution $J = 9$ ($N_{\text{side}} = 512$) and performs successive wavelet decompositions until resolution $J = 2$ ($N_{\text{side}} = 2$), thereby constructing 7 sets of high- and low-resolution maps. Although the original map is fully represented by the 7 high-resolution maps plus the low-resolution map at $J = 2$, in our analysis we have used all the high- and low-resolution maps, plus the original map, since this has been shown to improve results (see sec. 1.4.2 and sec. 5 for details).

Using all these maps, the third order moments of the wavelet coefficients are computed as follows:

$$S_{jkl} = \frac{1}{N_l} \frac{\sum_{i=1}^{N_l} w_{i,j} w_{i,k} w_{i,l}}{\sigma_j \sigma_k \sigma_l}, \quad (3.1)$$

where $w_{i,j}$ is the i^{th} wavelet coefficient of the map at resolution j , σ_j is the dispersion of $w_{i,j}$, and N_l is the number of pixels in the map at resolution l (since one requires $j \leq k \leq l$). Some of these statistics are redundant (linearly dependency exists between them), so we restrict our analysis to the set of non-redundant statistics, which gives a total of 232 quantities; these are then computed for non-Gaussian simulations with a range of known values of f_{NL} .

The expected values of these statistics are proportional to the non-linear coupling parameter, and they have previously been used to estimate the best fit f_{NL} value for the data by weighted least squares parameter estimation (Casaponsa et al., 2011). In this case, the dispersion in the estimated f_{NL} value for Gaussian simulations and is found to be $\sigma(f_{\text{NL}}) = 34$, which is slightly larger than the optimal value. The main advantage of the HW is the computing efficiency; for example, the third-order statistics construction is 10^3 times faster than for the KSW bispectrum estimator (Komatsu et al., 2005) and 10^2 times faster than the SMHW (see below). This procedure (for both the HW and SMHW) does, however, include the estimation and inversion of a correlation matrix, which can be computationally demanding and, in some cases, close to singular. As we will show below, this step is avoided with the use of a NN classifier.

- **Spherical Mexican Hat Wavelet**

The spherical Mexican hat wavelet (SMHW) (Antoine & Vandergheynst, 1998; Martínez-González et al., 2002) has produced competitive results in constraining primordial non-Gaussianity (Mukherjee & Wang, 2004; Curto et al., 2009,b, 2011). It is a continuous wavelet that has better frequency localization than the HW, although the computing efficiency is lower. Curto et al. (2011) use the SMHW to constrain f_{NL} with an accuracy equivalent to the bispectrum estimators (see for example Smith et al., 2009; Fergusson & Shellard, 2009; Fergusson et al., 2010; Komatsu et al., 2011; Bucher et al., 2010). The definition of the third-order moments is the same as for the HW. In this case, however, there are more inter-scale combinations because the scales involved are not restricted by the HEALPIX pixelization. The total number of non-redundant statistics for the SMHW wavelet coefficients is 680. Using the mean values and covariances of these statistics computed from non-Gaussian simulations, Curto et al. (2011) applied a χ^2 minimisation method to obtain optimal uncertainties on the f_{NL} estimates of $\sigma \approx 21$. However, this method requires a principal component analysis to deal with the degeneracies present in the covariance matrix. As we will see, this problem is avoided with the use of the multi-class neural network classifier.

3.3. Neural network classifier applied to non-Gaussianity

Artificial neural networks are a methodology for computing, based on massive parallelism and redundancy, which are features also found in animal brains. They consist of a number of interconnected processors each of which processes information and passes it to other processors in the network. Well-designed networks are able to ‘learn’ from a set of training data and to make predictions when presented with new, possibly incomplete, data. These algorithms have been successfully applied in several areas, in particular, we note the following applications in astrophysics: Storrie-Lombardi et al. (1992); Baccigalupi et al. (2000); Vanzella et al. (2004); Auld et al. (2007) and Carballo et al. (2008); Nørgaard-Nielsen (2012).

The basic building block of an ANN is the *neuron*. Information is passed as inputs to the neuron, which processes them and produces an output. The output is typically a simple mathematical function of the inputs. The power of the ANN comes from assembling many neurons into a network. The network is able to model very complex behaviour from input to output. We use a three-layer feed-forward network consisting

of a layer of input neurons, a layer of ‘hidden’ neurons and a layer of output neurons. In such an arrangement each neuron is referred to as a node (see sec. 1.4.4 for details and figure 1.9 for an schematic illustration of an artificial neural network).

The outputs of the hidden layer and the output layer are related to their inputs as follows:

$$\text{hidden layer: } h_j = g^{(1)}(f_j^{(1)}); \quad f_j^{(1)} = \sum_i w_{ji}^{(1)} x_i + b_j^{(1)}, \quad (3.2)$$

$$\text{output layer: } y_k = g^{(2)}(f_k^{(2)}); \quad f_k^{(2)} = \sum_j w_{kj}^{(2)} h_j + b_k^{(2)}, \quad (3.3)$$

where the output of the hidden layer h and output layer y are given for each hidden node j and each output node k . The index i runs over all input nodes. The functions $g^{(1)}$ and $g^{(2)}$ are called activation functions. The non-linear nature of $g^{(1)}$ is a key ingredient in constructing a viable and practically useful network. This non-linear function must be bounded, smooth and monotonic; we use $g^{(1)}(x) = \tanh x$. For $g^{(2)}$ we simply use $g^{(2)}(x) = x$. The layout and number of nodes are collectively termed the *architecture* of the network. For a basic introduction to artificial neural networks the reader is directed to Mackay (2003).

For a given architecture, the weights \mathbf{w} and biases \mathbf{b} define the operation of the network and are the quantities we wish to determine by some *training* algorithm. We denote \mathbf{w} and \mathbf{b} collectively by \mathbf{a} . As these parameters vary during training, a very wide range of non-linear mappings between inputs and outputs is possible. In fact, according to a ‘universal approximation theorem’ Leshno (1993), a standard three-layer feed-forward network can approximate any continuous function to *any* degree of accuracy with appropriately chosen activation functions and a sufficient number of hidden nodes.

In our application, we will construct a *classification* network. The aim of any classification method is to place members of a set into groups based on inherent properties or *features* of the individuals, given some pre-classified training data. Formally, classification can be summarised as finding a classifier $\mathcal{C} : \mathbf{x} \rightarrow C$ which maps an object from some (typically multi-dimensional) feature space \mathbf{x} to its classification label C , which is typically taken as one of $\{1, \dots, N\}$ where N is the number of distinct classes. Thus the problem of classification is to partition feature space into regions (not necessarily contiguous), assigning each region a label corresponding to the appropriate classification. In our context, the aim is to classify sets of third-order statistics of wavelet coefficients

of (possibly) non-Gaussian CMB maps (assembled into an input feature vector \mathbf{x}) into classes defined by ranges of f_{NL} ; this is discussed in more detail below.

In building a classifier using a neural network, it is convenient to view the problem *probabilistically*. To this end we consider a 3-layer MLP (multi-layer perceptron) consisting of an input layer (x_i), a hidden layer (h_j), and an output layer (y_i). In classification networks, however, the outputs are transformed according to the *softmax* procedure

$$p_k = \frac{e^{y_k}}{\sum_m e^{y_m}}, \quad (3.4)$$

such that they are all non-negative and sum to unity. In this way p_k can be interpreted as the probability that the input feature vector \mathbf{x} belongs to the k th class. A suitable objective function for the classification problem is then

$$\mathcal{L}(\mathbf{a}) = \sum_l \sum_k t_k^{(l)} \ln p_k(\mathbf{x}^{(l)}, \mathbf{a}), \quad (3.5)$$

where the index l runs over the training dataset $\mathcal{D} = \{\mathbf{x}^{(l)}, \mathbf{t}^{(l)}\}$, in which the target vector $\mathbf{t}^{(l)}$ for the network outputs has unity in the element corresponding to the true class of the l^{th} feature vector $\mathbf{x}^{(l)}$ and zeroes elsewhere. One then wishes to choose network parameters \mathbf{a} so as to maximise this objective function as the training progresses. The advantage of this probabilistic approach is that we gain the ability to make *statistical* decisions on the appropriate classification in very large feature spaces where a direct linear partition would not be feasible.

One wishes to choose network parameters \mathbf{a} so as to maximise the objective function $\mathcal{L}(\mathbf{a})$ as the training progresses. This is, however, a highly non-linear, multi-modal function in many dimensions whose optimisation poses a non-trivial problem. We perform this optimisation using the MEMSYS package (Gull & Skilling, 1999). This algorithm considers the parameters \mathbf{a} to have prior probabilities proportional to $e^{\alpha S(\mathbf{a})}$, where $S(\mathbf{a})$ is the positive-negative entropy functional (Hobson & Lasenby, 1998). α is treated as a hyper-parameter of the prior, and sets the scale over which variations in \mathbf{a} are expected. α is chosen to maximise its marginal posterior probability whose value is inversely proportional to the standard deviation of the prior. Thus for a given α , the log-posterior probability is proportional to $\mathcal{L}(\mathbf{a}) + \alpha S(\mathbf{a})$. For each chosen α there is a solution $\hat{\mathbf{a}}$ that maximises the posterior. As α varies, the set of solutions $\hat{\mathbf{a}}$ is called the *maximum-entropy trajectory*. We wish to find the solution for which \mathcal{L} is maximised which occurs at the end of the trajectory where $\alpha = 0$. For practical purposes we start at a large value of α and iterate downwards until α is sufficiently small so that

the posterior is dominated by the \mathcal{L} term. MEMSYS performs this algorithm using conjugate gradient descent at each step to converge to the maximum-entropy trajectory. The required matrix of second derivatives of \mathcal{L} is approximated using vector routines only, thus circumventing the need for $O(N^3)$ operations required for exact calculations. The application of MEMSYS to the problem of network training allows for the fast efficient training of relatively large network structures on large data sets that would otherwise be difficult to perform in a reasonable time. Moreover the MEMSYS package also computes the Bayesian evidence for the model (i.e. network) under consideration, (see for example Jaynes & Bretthorst, 2003, for a review), which provides a powerful model selection tool. In principle, values of the evidence computed for each possible architecture of the network (and training data) provide a mechanism to select the most appropriate architecture, which is simply the one that maximises the evidence (although we will use a more prosaic method below for deciding on the network architecture). The MEMSYS algorithm is described in greater detail in (Gull & Skilling, 1999).

3.4. The f_{NL} classification network

To train our f_{NL} classification network we provide it with an ensemble of training data $\mathcal{D} = \{\mathbf{x}^{(l)}, \mathbf{t}^{(l)}\}$. The l^{th} input vector $\mathbf{x}^{(l)}$ contains the third-order statistics of the wavelet coefficients of the l^{th} simulated CMB map. The output classes of our network correspond to contiguous ranges of f_{NL} values. Thus, the target vector $\mathbf{t}^{(l)}$ for the network outputs has zeroes everywhere except for a unit entry in the element corresponding to the class in which the true f_{NL} value of the l^{th} simulated CMB map falls.

The N output classes of the network were defined by dividing some initial (anticipated) range of f_{NL} values into N equal-width subranges. For example, for a total range of $-30 \leq f_{\text{NL}} < 30$ and a network with just 3 output classes, input vectors constructed from maps with $-30 \leq f_{\text{NL}} < -10$ were ascribed to *class=1* with an associated target vector $\mathbf{t} = (1, 0, 0)$, maps with $-10 \leq f_{\text{NL}} < 10$ to *class=2* with $\mathbf{t} = (0, 1, 0)$, and those with $10 \leq f_{\text{NL}} < 30$ to *class=3* with $\mathbf{t} = (0, 0, 1)$. In this example, the output given by the network for some test input vector \mathbf{x} would be a 3-dimensional vector $\mathbf{p} = (p_1, p_2, p_3)$, where $\sum_k p_k = 1$ and p_k can be interpreted as the probability that the input vector belongs to class k . The discrepancy between the targets and the outputs during training can be measured by the true positive rate, which is simply the fraction

of the training input vectors for which the network assigns the maximum probability to the correct class.

From the output values p_k obtained for each map, we define the estimator of the local non-Gaussianity parameter to be simply

$$\hat{f}_{\text{NL}} = \sum_{k=1}^{n_{\text{class}}} \langle f_{\text{NL}} \rangle_k p_k \quad (3.6)$$

where $\langle f_{\text{NL}} \rangle_k$ is the mean value of f_{NL} in the k^{th} class. The statistical properties of this estimator, namely its mean and dispersion, determine the accuracy of the method.

3.4.1. Training data

The training input vectors $\mathbf{x}^{(l)}$ were generated as follows. We began with a set of 1000 non-Gaussian CMB realisations from which a_{lm}^{NG} and a_{lm}^{G} were generated by Elsner & Wandelt (2009) and normalized to the WMAP 7-year concordance model power spectrum generated by CAMB. These a_{lm} are publicly available⁸. The ultimate accuracy of the network classifier is improved, however, by the inclusion of further training data. Given the finite number of available simulations, we thus created a further set by rotation of the original maps by 90° perpendicular to the galactic plane. This rotation creates roughly 20 per cent extra map area based on the original mask; we verified that its inclusion improves the results. Using this procedure we generate a further 1000 non-Gaussian simulations. Of the 2000 non-Gaussian maps, 1800 were used for training and the remainder were set aside for testing of the networks.

For each non-Gaussian simulation used for training, sets of a_{lm} were then generated with varying f_{NL} using the following prescription

$$a_{lm} = a_{lm}^{\text{G}} + f_{\text{NL}} a_{lm}^{\text{NG}}, \quad (3.7)$$

with 20 different f_{NL} random values between -120 and 120 for the HW decomposition and between -76 and 76 for the SMHW analysis. Thus, for each non-Gaussian simulation, 20 sets of a_{lm} were generated. Hence the total number of available training data sets is 36000. Noise-weighted V+W-band WMAP realizations were then constructed as explained in Curto et al. (2009) and Casaponsa et al. (2011), and the KQ75 mask was then applied, which covers roughly 29% of the sky. A wavelet decomposition for both the HW and SMHW was performed to determine the wavelet coefficients for each

⁸<http://planck.mpa-garching.mpg.de/cmb/fnl-simulations/>

a_{lm} set, and their third-order moments computed. These statistics were provided as inputs to the neural network. Each input vector contained 232 values for the HW and 680 for the SMHW.

3.4.2. Network architecture

The architecture of our 3-layer neural networks are defined by two free parameters: the number of hidden nodes n_{hid} and the number of output classes, n_{class} , into which the f_{NL} range is divided. A further parameter, which determines the accuracy of the network classifier, is the quantity of training data n_{data} . Variation of these parameters can affect the training efficiency so it is desirable to explore this training space adequately in order to find an optimal set of parameters.

Although the MEMSYS algorithm provides routines to determine the optimal value of the number of hidden nodes using the Bayesian evidence Gull & Skilling (1999), in this application n_{hid} is determined simply by measuring training times and the accuracy of the trained networks on an independent testing set. In this example, we have found that in fact the optimal architecture contains no hidden nodes, resulting in what is effectively a linear classifier. This is not surprising, since we are effectively ‘asking’ the network to learn the mean value and dispersion of the third-order moments of the wavelet coefficients for each f_{NL} ; since the expectation value is linearly dependent on the f_{NL} , this network architecture trivially satisfies this requirement. Indeed, networks of this sort provide a simple way of obtaining the (pseudo)inverse of any matrix.

The number of output classes, n_{class} , of the network is clearly related to the total range of f_{NL} considered and size of the subranges into which this range is divided. Here we consider networks with $n_{\text{class}} = 9$ (an odd number ensures that $f_{NL} = 0$ does not lie on the boundary of a class) The range of f_{NL} was chosen *a priori* to correspond to approximately $\pm 3\sigma$, where σ is the dispersion in the f_{NL} estimates obtained previously using the standard χ^2 minimisation method. Thus, the full range was taken to be $-120 \leq f_{NL} < 120$ for the HW and $-76 \leq f_{NL} < 76$ for the SMHW, resulting in subranges per class of width 27 and 17 units, respectively. This combination fulfilled all the requirements of classification over the range of our simulated data.

The quantity of training data, n_{data} , determines the accuracy of the resulting classification network. Naturally, the network accuracy increases with n_{data} , but it typically saturates after a given number. We found that the quantity of data required saturated

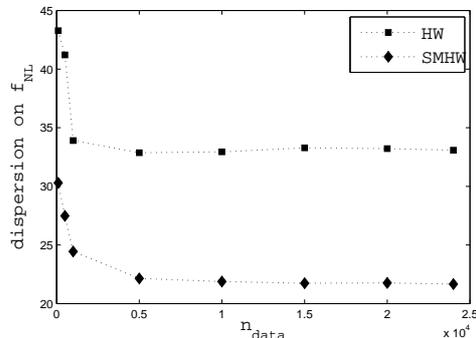


Figure 3.1 - Results of the dispersion of \hat{f}_{NL} for 1000 Gaussian simulations for different values of n_{data} .

at roughly $n_{\text{data}} \sim 10000$ (see Fig. 3.1).

3.4.3. Training evolution

Figure 3.2 illustrates the training evolution for the classification network with $n_{\text{hid}} = 0$ and $n_{\text{class}} = 9$. In the top two panels we plot the true positive rates (TPR) of the network on the training set and the test set, for the HW and SHMW respectively; in each plot, the TPR on the training set has been multiplied by a factor less than unity to highlight the divergence with the TPR for the test set. We see that this divergence occurs after ~ 100 and ~ 500 iterations of the MEMSYS optimiser for the HW and SMHW, respectively. Thus the training was terminated at this point to construct our final classification networks.

A key criterion in determining the quality of our classifiers is the dispersion of the f_{NL} values obtained in the testing set. This is plotted in the bottom two panels of Figure 3.2 for the HW and SMHW, respectively. We note that, in each case, this dispersion increases noticeable beyond the point where the TPRs on the training and testing sets diverge.

3.5. Results

3.5.1. Application to WMAP simulations

We first applied our classifiers to 1000 WMAP-7yr simulations made from Gaussian CMB maps ($f_{\text{NL}} = 0$). For the HW classifier, we obtained $\langle \hat{f}_{\text{NL}} \rangle = -1$, which

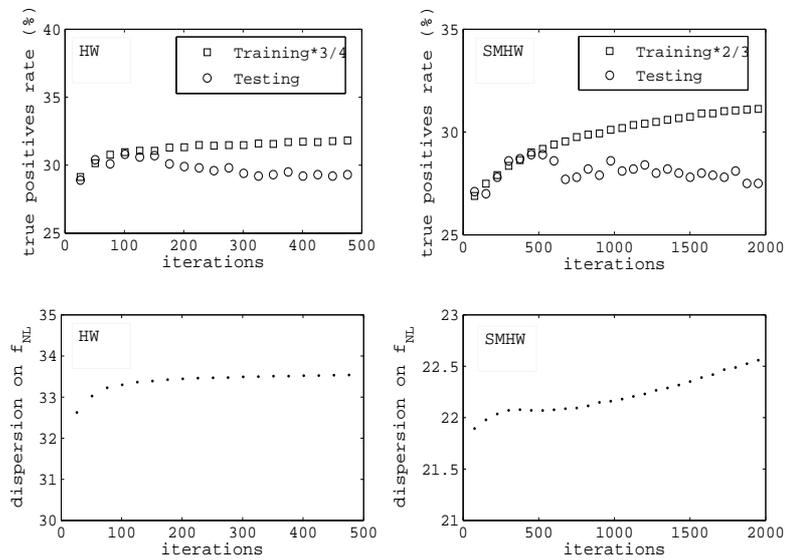


Figure 3.2 - Evolution of the true positive rate for each iteration of the training process with a neural network with $n_{\text{hid}} = 0$ and $n_{\text{data}} = 10000$. Note that the TPR of the training set have been multiplied by a factor less than unity in order to highlight the divergence of the behaviours. The bottom panels show the variation of the dispersion on the estimate \hat{f}_{NL} during the training. Left panels for HW and right panels for SMHW.

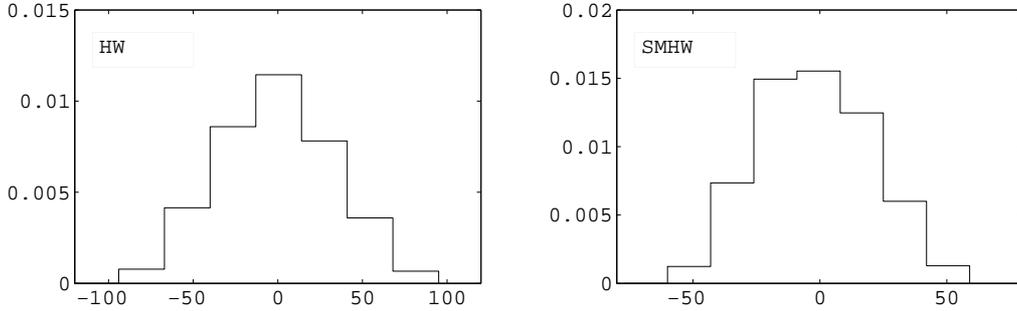


Figure 3.3 - Distribution of \hat{f}_{NL} obtained from 1000 Gaussian realizations for HW (left) and SMHW (right).

indicates the estimator is essentially unbiased. Moreover, the dispersion of the estimator $\sigma(\hat{f}_{\text{NL}}) = 33$ is extremely similar to that obtained with the weighted least-squares method ($\sigma(\hat{f}_{\text{NL}}) = 34$). The full distribution of the estimator is shown in the top panel of Fig. 3.3. For the SMHW classifier, we again found $\langle \hat{f}_{\text{NL}} \rangle = -1$, with a dispersion of $\sigma(\hat{f}_{\text{NL}}) = 22$, which is very close to the optimal value of $\sigma(\hat{f}_{\text{NL}}) = 21$. The distribution of the estimator for the SMHW is shown in the bottom panel of Fig. 3.3. The histogram bins in Fig. 3.3 have the same size and central values as those used to define the network classes. We see that the classes at extremal f_{NL} values are empty, indicating that the network placed no maps in these f_{NL} ranges. Thus for estimating f_{NL} from Gaussian or nearly Gaussian maps the range in f_{NL} used is sufficiently wide.

We next applied our estimator to sets of non-Gaussian simulations, each with a different non-zero f_{NL} value. For each true f_{NL} value, we analysed the corresponding WMAP simulations and calculated the mean and dispersion of our estimator \hat{f}_{NL} for both the HW and SMHW classifiers. The results are shown in fig. 3.4, in which we plot the mean value of \hat{f}_{NL} against the true f_{NL} value. We see that the classifiers are unbiased for $|f_{\text{NL}}| \lesssim \sigma$ with an almost constant dispersion. For larger $|f_{\text{NL}}|$ values, however, the estimator becomes progressively more biased and its dispersion decreases.

The latter behaviour is simply understood as an *edge effect* due to the finite total range of f_{NL} considered by the networks. This point is illustrated in Fig. 3.5 in which we plot the full distributions of \hat{f}_{NL} obtained for a number of representative values of the true f_{NL} . We see that for $|f_{\text{NL}}| \lesssim \sigma$, we obtain close to symmetric distribution

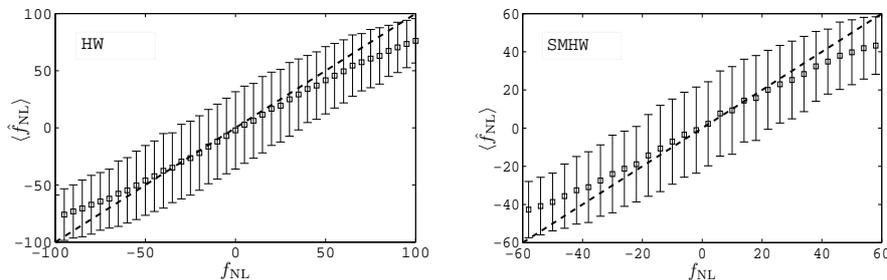


Figure 3.4 - The mean and dispersion of \hat{f}_{NL} obtained for a number of representative values of the true f_{NL} for the HW network (left) and the SMHW network (right).

centred on the true f_{NL} value, with no maps being placed in the extreme classes. As $|f_{\text{NL}}| > \sigma$, however, we see that the classifier does begin to place maps in the extreme classes, resulting in the distribution of \hat{f}_{NL} becoming severely skewed and no longer centred on the true value. Of course, if one were to encounter this behaviour in the analysis of a real data set, one could simply alter the range of f_{NL} considered by the network and retrain.

In any case, the above results show that both the HW and SMHW network classifiers produce unbiased estimates \hat{f}_{NL} provided $-\sigma < f_{\text{NL}} < \sigma$. Moreover, the dispersions on these estimators are very similar to those obtained with the classical weighted least squares (WLS) method, indicating that neural networks can produce very accurate results within the limitations described above. In the case of the SMHW, this is a particularly important result since the complexity of the covariance matrix inversion required in the standard approach is bypassed via the use of the neural network classifier. Curto et al. (2011) used a principal component analysis to reduce the covariance matrix dimension to allow inversion.

3.5.2. Application to WMAP 7-year data

Applying the neural network classifiers to real data (the V+W WMAP 7-year data map), we obtain $\hat{f}_{\text{NL}} = -12$ for the HW and $\hat{f}_{\text{NL}} = 19$ for the SMHW. Both these values lie well within the corresponding dispersion of the estimator. From the corresponding \hat{f}_{NL} distributions obtained on simulated data, we find that 95% confidence limits are

	$\hat{f}_{\text{NL,data}}$	$\sigma(\hat{f}_{\text{NL}})$	$\langle \hat{f}_{\text{NL,gauss}} \rangle$	$P_{2.5}$	$P_{97.5}$
SMHW (NN)	19	22	-1	-43	42
SMHW (WLS) Curto et al. 2011b	37	21	0	-42	46
HW (NN)	-12	33	-1	-66	63
HW (WLS) Casaponsa et al. 2011	6	34	1	-68	67

Table 3.1 - Results obtained with neural networks (NN) and weighted least squares (WLS). $\hat{f}_{\text{NL,data}}$ is the best fitting value for V+W WMAP data, $\langle \hat{f}_{\text{NL,gauss}} \rangle$ and $\sigma(\hat{f}_{\text{NL}})$ are the expected value and the standard deviation for Gaussian simulations. $P_{2.5}$ and $P_{97.5}$ represent the percentile values at 95% confidence level of \hat{f}_{NL} for Gaussian realizations.

$-78 < f_{\text{NL}} < 51$ for the HW and $-24 < f_{\text{NL}} < 61$ for the SMHW.⁹ We therefore conclude that the data are consistent with the Gaussian hypothesis. We note that the SMHW confidence limits are very similar to those obtained with the optimal f_{NL} estimator (Komatsu et al., 2011; Smith et al., 2009).

These results are summarised in Table 3.1, along with found via the weighted least squares (WLS) method. The latter results are also consistent with Gaussianity. It is worth mentioning, however, the different values of \hat{f}_{NL} obtained by the neural network and the WLS methods, for both HW and SMHW. Although all four values lie well within their corresponding dispersions, each method returns a different \hat{f}_{NL} value when applied to the same WMAP-7yr dataset. This behaviour is to be expected, however, since these are four *different* estimators of f_{NL} . Therefore, in general, they will not be equal, even when applied to the same input data. Only the statistical properties (e.g. mean, dispersion) of their sampling distributions are important.

⁹Note that the constraints are not corrected for the unresolved point sources contribution.

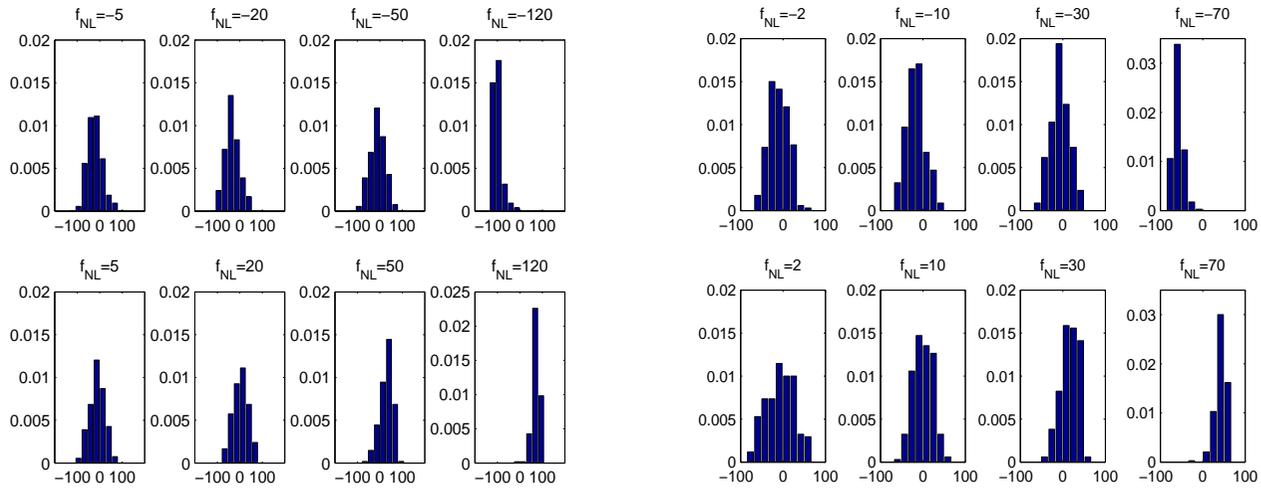


Figure 3.5 - Distribution of \hat{f}_{NL} obtained from 200 non-Gaussian realizations with representative true f_{NL} values, for HW (left) and SMHW (right).

CHAPTER 4

Exploring local f_{NL} estimators based on the binned bispectrum

This chapter is devoted also to the use of neural networks to construct non-Gaussianity estimators and is based on the published work of Casaponsa et al., 2013b. Conversely to the previous chapter, now we explore the a regression neural network, instead of a classification network, and we use other third order quantities (the binned bispectrum components) to train it. Moreover, in this work are also explored what are the requirements that make an estimator for f_{NL} based on the binned bispectrum to be as good as the optimal estimator. For that, we explore different estimators of the local non-linear coupling parameter, f_{NL} , based on the binned bispectrum presented in Bucher et al., 2010. Using simulations of Wilkinson Microwave Anisotropy Probe (WMAP)-7yr data, we compare the performance of a regression neural network with a χ^2 -minimization and study the dependence of the results on the presence of the linear term in the analysis and on the use of inpainting for masked regions. Both methods obtain similar results and are robust to the use of inpainting, but the neural network estimator converges considerably faster. We also examine the performance of a simplified χ^2 estimator that assumes a diagonal matrix and has the linear term subtracted, which considerably reduces the computational time; in this case inpainting is found to be crucial. The estimators are also applied to real WMAP-7yr data, yielding constraints at 95% confidence level of $-3 < f_{\text{NL}} < 83$.

4.1. Introduction

Cosmic microwave background (CMB) fluctuations naturally arise in inflationary models. Discriminating between different models is a difficult task, but can be addressed by observing very faint non-Gaussian signals in the high-order correlation functions of the CMB temperature anisotropies. A popular approach is to search for the local form of non-Gaussianity, where the initial curvature Gaussian perturbations are expanded up to the second order as

$$\Phi = \Phi_g + f_{\text{NL}} [\Phi_g^2 - \langle \Phi_g^2 \rangle]$$

(for more details see e.g. Bartolo et al., 2004; Babich et al., 2004).

WMAP constraints on the amplitude of the local form of non-Gaussianity have been able to rule out exotic models such as ghost inflation (Arkani-Hamed et al., 2004). New data sets, such as the recent release from Planck satellite (Planck Collaboration, 2013f), significantly reduce the uncertainties on local f_{NL} , ruling out some ekpyrotic models and imposing strong constraints on multi-field inflationary models. In fact, for single-field inflation, f_{NL} (hereafter f_{NL} is the local form) should be of the order of the spectral index (Creminelli & Zaldarriaga, 2004), given the consistency relation derived in Maldacena (2003). Recent papers show that this relation does not hold for non-vacuum initial states (Ganc, 2011; Agullo & Parker, 2011) and non-constant super-horizon modes (Chen et al., 2013), but the vast majority of single-field models should be ruled out by a detection of a larger f_{NL} value.

This type of primordial non-Gaussianity may be detected using higher-order correlation functions. The simplest of these is third-order, which is equivalent to the bispectrum in spherical harmonic space. The first derivation of the optimal estimator, in the sense of an unbiased estimator that saturates the Cramer–Rao inequality, is given in Babich (2005), assuming an isotropic field. Working with real data, however, is usually more complicated. In particular, CMB maps have anisotropic noise due to the scanning strategy and masked regions, both of which break the isotropy assumption for these theoretical estimators. The masked regions are particularly difficult to treat, as they introduce correlations among the Fourier modes, which are otherwise expected to be independent. Creminelli et al. (2006) applied the optimal estimator to real data, showing that the presence of a term proportional to the $a_{\ell m}$ is required to account for such anisotropies. In that paper the constraints are computed using an approximation to avoid numerical difficulties. Finally, this estimator was successfully

applied in its complete form to WMAP data by (Smith et al., 2009; Komatsu et al., 2011; Bennett et al., 2013, for 5th, 7th and 9th year respectively).

New imaging reconstruction techniques have recently been used to pre-process CMB maps by smoothing the contours of the masked regions. A simple approach is to apodize the mask by introducing a smooth function in the pixels surrounding the masked regions. Another approach is to fill the masked regions with a pseudo-signal, which is termed inpainting. Several techniques have been proposed in the literature for inpainting, which is a very delicate process since the signal can be distorted (Bajkova, 2005; Abrial et al., 2008; Starck et al., 2013).

Consequently, primordial non-Gaussianity analyses can be computationally demanding, and new techniques should therefore be investigated to overcome the computational cost of large matrix estimations and inversions. Here we investigate the utility of a neural network to obtain the necessary weights in the f_{NL} estimator and compare it with the direct approach via χ^2 minimization. Over the last 20 years, artificial intelligence techniques have been used in a number of areas of astrophysical analysis: morphological galaxy determination, photo-redshift estimations, and classification of different objects are examples of successful applications of neural networks (Storrie-Lombardi et al., 1992; Firth et al., 2003; Vanzella et al., 2004; Carballo et al., 2008). In particular, for cosmological analysis, they have recently been used to reduce the computational time of cosmological parameter estimation from observations of the CMB power spectrum (Auld et al., 2007, 2008). Also in CMB analysis, Casaponsa et al. (2011) used neural networks to define a new non-Gaussianity estimator and showed that networks are a valuable tool for bypassing the inversion of ill-conditioned matrices, and to avoid covariance matrix estimation in a χ^2 analysis.

The aim of the present work is to continue our earlier study of the power of the neural networks in the statistical analysis associated with cosmic microwave background (CMB) non-Gaussianity. To this end, this chapter is focused on the study of different tools, in order to identify the most robust and efficient estimator when dealing with real data. We compare three different approaches to estimate f_{NL} , based on the binned bispectrum. The first estimator is obtained by minimizing a χ^2 of the binned bispectrum components. A second approach is based on the optimal estimator, without taking into account the correlations among the binned bispectrum components, which for an isotropic field would be the same as the former. And the third method uses the weights of a regression neural network. From these approaches we construct different

estimators to account for the effects of pre-processing the data with inpainting and the presence of a the linear term.

The chapter is organized as follows. An overview of the type of neural network employed and the training procedure is given in Section 4.2. In Section 4.3 we describe the binned bispectrum. The definition of the estimators is presented in Section 4.4 followed by an explanation of the main details of the implementation in Section 4.5 and finally the results are presented in Section 3.5. The conclusions of this work are given in chapter 6.

4.2. Regression neural network applied to non-Gaussianity

In our previous application of ANN to the estimation of f_{NL} , a classification neural network was used (Casaponsa et al., 2011). Here, we instead use a *regression* network, which we find to be as useful as the classification approach, and also allows a more direct comparison with the χ^2 minimization procedure. Additionally, using a regression network has the advantage of reducing the network parameter space, making the training faster.

In Casaponsa et al. (2011), we used neural networks for which the inputs were third-order moments of two wavelet decompositions of the CMB map: the Healpix wavelet (HW) (Casaponsa et al., 2011) and the spherical Mexican hat wavelet (SMHW) (Curto et al., 2009a, 2011). We found the resulting f_{NL} estimator had the same accuracy as the standard one based on χ^2 -minimization, but was much faster to evaluate. Here, the inputs to our neural networks are the estimator for the bispectrum proposed by Bucher et al. (2010), defined in a number of bins in l -space, which reduces the dimension of the problem by a factor of 10^5 . The aim of this work is to learn a mapping from the binned bispectrum components of the (possibly) non-Gaussian CMB (assembled into an input feature vector \mathbf{x}) to the corresponding f_{NL} of the map. Therefore, conversely to the classification case, we only have one node in the output layer, reducing the dimension of the network parameter space. This is dicussed in detail below.

A suitable objective function for the regression problem is

$$\mathcal{L}(\mathbf{a}) = \frac{1}{2} \sum_n \sum_k [t_k^{(n)} - y_k^{(n)}(\mathbf{x}^{(n)}, \mathbf{a})]^2, \quad (4.1)$$

where \mathbf{a} are the network parameters (weights and biases) and the index n runs over the training data-set $\mathcal{D} = \{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}$, in which the target vector $\mathbf{t}^{(n)}$ for the network

outputs are the f_{NL} values, as explained in the next section. We need to find the network parameters \mathbf{a} that minimise this objective function as the training progresses. As in the previous work, we perform this optimisation using the MEMSYS package (Gull & Skilling, 1999), that finds the optimal solution following *the maximum-entropy trajectory*. The main advantage of using MEMSYS to the problem of network training is its computational efficiency with large data sets, and complex neural network structures. Moreover, the MEMSYS package computes the Bayesian evidence for the model under consideration, providing a tool to select the optimal network architecture, that would be the one that maximises the evidence.

4.3. Binned bispectrum

Several approaches to bispectrum analyses have been proposed in the literature, such as the KSW (Komatsu et al., 2005), Skew-CIs (Munshi & Heavens, 2010), wavelets (Curto et al., 2009a, 2011; Casaponsa et al., 2011) or needlets (Pietrobon et al., 2009; Donzelli et al., 2012) among others. Furthermore, Bucher et al. (2010) and Fergusson & Shellard (2011) presented bispectrum estimators that reduce the dimensionality of the problem without losing significant information. In particular, we use the bispectrum estimator defined in Bucher et al. (2010). The proposed method consists of joining the bispectrum components in bins, significantly reducing the computational time, but maintaining the quality of the estimator of f_{NL} . Bucher et al. (2010) show that this is the case for ideal maps, with isotropic noise and small symmetric masks. The binned bispectrum is also applied to Planck data in Planck Collaboration (2013f) to constrain primordial non-Gaussianity. Here we study with more detail its applications to realistic data, for which we used simulations with WMAP-7yr characteristics.

As a starting point, the angle-averaged reduced bispectrum is defined by

$$b_{l_1 l_2 l_3} = \int T_{l_1} T_{l_2} T_{l_3} d\Omega, \quad (4.2)$$

where $T_\ell(\vec{n}) = \sum_m a_{\ell m} Y(\vec{n})$. The binned reduced bispectrum is then

$$b_{abc} = \sum_{\ell_1 \in I_a} \sum_{\ell_2 \in I_b} \sum_{\ell_3 \in I_c} b_{\ell_1 \ell_2 \ell_3}, \quad (4.3)$$

where I_n are bins in ℓ . This definition of the reduced bispectrum, differing from the standard one by the factor $I_{\ell_1 \ell_2 \ell_3}^2$ (for details see Bucher et al., 2010; Komatsu, 2002), is convenient since one can write b_{abc} in terms of T_a , T_b and T_c which are the binned

maps:

$$T_n = \sum_{\ell_i \in I_n} T_{\ell_i}. \quad (4.4)$$

The advantage of constructing maps in ℓ -bins is that the number of transformations to spherical harmonic space is significantly reduced. Then, the resulting bispectrum estimator is faster to construct than the one based on the KSW estimator (Komatsu et al., 2005) or the SMHW (Curto et al., 2011).

4.4. f_{NL} estimators

The optimal estimator for f_{NL} , in the sense of an unbiased estimator that saturates the Cramer-Rao inequality, is obtained by performing an Edgeworth expansion of the probability distribution of the $a_{\ell m}$ for weakly non-Gaussian data (Babich, 2005; Creminelli et al., 2006; Smith et al., 2009). This estimator is found to have a cubic term and a linear term in $a_{\ell m}$. The latter term plays an important role under realistic conditions, where anisotropic instrumental noise and/or a mask is present.

The form of this estimator can also be understood using the properties of the Wick product. As demonstrated in Donzelli et al. (2012), Marinucci & Peccati (2011) and Peccati & Taqqu (2011), the Wick product of a cubic variable, which is given by

$$: x_1, x_2, x_3 := x_1 x_2 x_3 - x_1 \langle x_2 x_3 \rangle - x_2 \langle x_1 x_3 \rangle - x_3 \langle x_1 x_2 \rangle, \quad (4.5)$$

has a smaller variance than the cubic variable itself, while not affecting the mean value so long as the variables x_i are Gaussian and have a mean value of zero. Then, if we replace each cubic term in an estimator by its Wick product, it will yield an estimator with lower variance. Following this reasoning, the binned bispectrum defined in Sec. 4.3 can be replaced by its Wick product

$$\begin{aligned} : T_{I_a} T_{I_b} T_{I_c} : &= T_{I_a} T_{I_b} T_{I_c} - \langle T_{I_a} T_{I_b} \rangle T_{I_c} \\ &\quad - \langle T_{I_b} T_{I_c} \rangle T_{I_a} - \langle T_{I_a} T_{I_c} \rangle T_{I_b}. \end{aligned} \quad (4.6)$$

Note that $T_i = T_i(x)$, since there is a dependence on the pixel for anisotropic maps.

Donzelli et al. (2012) have proved that for the case of wavelet and needlet coefficients, the linear term is basically equivalent to removing the mean value of the coefficients. In order to see if this is the case for the binned bispectrum, we explore the

option of substituting $T'_n = T_n - \langle T_n \rangle$, where $\langle T_n \rangle$ is computed with the unmasked pixels. This would be less costly than estimating the correlation matrix $\langle T_a T_b \rangle$ required for the computation of the linear term.

In the following subsections, we describe three methods for choosing the weights that are used to construct the final f_{NL} estimator. In each case, estimators are constructed with and without the linear term contribution to explore its importance. Also, the performance of these estimators is tested on inpainted and non-inpainted maps, with the methodology explained in Sec. 4.5.3. In all cases the original mask M is applied again at the final stage when computing the binned bispectrum components

$$b_{abc} = \sum_{i=1}^{N_{\text{pix}}} \frac{M_i \langle T_{a,i} T_{b,i} T_{c,i} \rangle}{4\pi N_{\text{pix}}}, \quad (4.7)$$

where $N_{\text{pix}} = \sum_i M_i$. The efficiency achieved by the estimators will be compared to that defined by the Cramer-Rao inequality. The Cramer-Rao bound states that the minimum variance for any unbiased estimator is given by the inverse of the Fisher matrix information. A useful reference value in the case of partial sky coverage is obtained from the full sky estimator corrected by the fraction of the available sky. Therefore, the minimum variance for f_{NL} is estimated to be:

$$\sigma_{f_h}^2 = \left[f_{\text{sky}} \sum_{\ell_1 \leq \ell_2 \leq \ell_3} \frac{(\langle B_{\ell_1 \ell_2 \ell_3} \rangle^1)^2}{\Delta C_{\ell_1} C_{\ell_2} C_{\ell_3}} \right]^{-1} \quad (4.8)$$

where Δ takes values 1, 2 or 6 when all ℓ 's are different, two are equal, or all are the same and f_{sky} is the fraction of the sky available. For (4.8) to be used for a realistic case, the power spectrum must include the noise and the beam contribution. The beam also needs to be included in the bispectrum part. We have used WMAP-7yr characteristics, in particular the average of the two channels of 61 and 94 GHz (V and W) and the extended mask KQ75. In terms of the reduced bispectrum defined in Sec. 4.3, the angular average bispectrum $B_{\ell_1 \ell_2 \ell_3}$ is:

$$B_{\ell_1 \ell_2 \ell_3} = \sqrt{\frac{4\pi}{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}} \times \left(\begin{array}{ccc} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{array} \right)^{-1} b_{\ell_1 \ell_2 \ell_3}. \quad (4.9)$$

4.4.1. Approximated maximum-likelihood estimator (AMLE)

The standard approach in this type of analysis is to use the fact that the third-order moments are nearly Gaussian, and therefore the maximum-likelihood estimator is obtained approximately by the minimization of a χ^2 given by

$$\chi^2 = \sum_{abc,def} (b_{abc} - f_{\text{NL}} \langle b_{abc} \rangle^1) C_{abc,def}^{-1} (b_{def} - f_{\text{NL}} \langle b_{def} \rangle^1) . \quad (4.10)$$

where $\langle b_{def} \rangle^1$ is the expected value for $f_{\text{NL}} = 1$ and $C_{abc,def}^{-1} = \langle b_{abc} \rangle \langle b_{def} \rangle - \langle b_{abc} b_{def} \rangle$. From the previous equation is straightforward to show that the f_{NL} estimator for an observed map is:

$$f_{\text{NL}} = \sum_{abc,def} \frac{\langle b_{abc} \rangle^1 C_{abc,def}^{-1} b_{def}^{\text{obs}}}{\sum_{abc,def} \langle b_{abc} \rangle^1 C_{abc,def}^{-1} \langle b_{def} \rangle^1} . \quad (4.11)$$

In order to include the linear term correction, $T_a T_b T_c$ should be substituted by its Wick product (4.6), wherever it appears. The mean value of the linear term is zero, and thus it vanishes in the term of the estimator related to the model, whereas it needs to be included in the covariance matrix. Thus, the corresponding estimator is

$$f_{\text{NL}} = \sum_{abc,def} \frac{\langle b_{abc} \rangle^1 C_{abc,def}^{-1}}{\sum_{abc,def} \langle b_{abc} \rangle^1 C_{abc,def}^{-1} \langle b_{def} \rangle^1} \times \quad (4.12)$$

$$\left(\frac{1}{4\pi N_{\text{pix}}} \sum_i^{N_{\text{pix}}} T_{d,i} T_{e,i} T_{f,i}^{\text{obs}} - \langle T_{d,i} T_{e,i} \rangle T_{f,i}^{\text{obs}} - \langle T_{d,i} T_{f,i} \rangle T_{e,i}^{\text{obs}} - \langle T_{e,i} T_{f,i} \rangle T_{d,i}^{\text{obs}} \right) ,$$

where $\langle b_{abc} \rangle^1$ is estimated using the regression coefficient of a linear fit to the mean values of 1,000 non-Gaussian simulations with different f_{NL} values. In particular, this is a conservative number that ensures that the mean values have converged¹⁰. For C^{-1} we assume that it is independent of f_{NL} , which is a good approximation in the limit of weak non-Gaussianity, and it is thus estimated with Gaussian simulations ($\sim 25,000$ are necessary for convergence issues). The term $\langle T_a T_b \rangle$ is estimated with 1,000 Gaussian simulations.

¹⁰To check that convergence is reached with N simulations, we simply test that two independent sets of $N/2$ realisations give consistent results

4.4.2. Approximated maximum likelihood estimator with diagonal covariance matrix (AMLED)

The estimator proposed by Bucher et al. (2010) used the approximation of assuming a diagonal covariance matrix. In this case, the estimator simplifies significantly, since the covariance matrix does not need to be estimated or inverted, and one obtains

$$f_{\text{NL}} = \sum_{abc} \frac{\langle b_{abc} \rangle^1 b_{abc}^{\text{obs}} / \text{var}(b_{abc})}{\sum_{def} (\langle b_{def} \rangle^1)^2 / \text{var}(b_{def})} \quad (4.13)$$

where $\text{var}(b_{abc})$ is the variance of the binned bispectrum components, which is computed with simulations. Besides its computational efficiency, another advantage of this estimator is that can be obtained analytically (see Bucher et al., 2010, for details).

Strictly speaking, this estimator is optimal only for a full-sky CMB experiment with isotropic noise (although it has been shown to work well also in presence of a reduced symmetric mask). Under realistic conditions, a linear term of a similar form to that used above needs to be added, such that

$$f_{\text{NL}} = \sum_{abc} \frac{\langle b_{abc} \rangle^1 / \text{var}(b_{abc})}{\sum_{def} (\langle b_{def} \rangle^1)^2 / \text{var}(b_{def})} \left(\frac{1}{4\pi N_{\text{pix}}} \sum_i^{N_{\text{pix}}} T_{a,i} T_{b,i} T_{c,i}^{\text{obs}} \right. \quad (4.14)$$

$$\left. - \langle T_{a,i} T_{b,i} \rangle T_{c,i} - \langle T_{a,i} T_{c,i} \rangle T_{b,i} - \langle T_{b,i} T_{c,i} \rangle T_{a,i} \right)$$

As with the previous estimator, 1,000 simulations were used for the model estimation and another 1,000 to obtain $\text{var}(b_{abc})$. This implies a reduction by a factor > 10 in the number of simulations required with respect to the AMLE, as the convergence in the estimation of the variance is reached with much smaller number of realisations than that of the covariance matrix.

4.4.3. Neural network estimator (NNE)

The architecture of our 3-layer neural network is defined by three parameters: the number of input, output and hidden nodes. The first two are determined by the problem at hand; in this case the dimension of the input vector depends on the number of bins chosen and there is a single output.

In this application the number of hidden nodes (n_{hid}) is determined empirically by measuring the accuracy of the trained networks on an independent testing set. As

before, we have found that in fact the optimal architecture contains no hidden nodes, resulting in what is effectively a linear mapping between input and output. Then, for zero hidden nodes, the single network output is just a linear function of the inputs. Once the network parameters (\vec{w}, θ) are found during the training process, the estimator for f_{NL} is thus given by:

$$f_{\text{NL}} = \sum_{abc} w_{abc} b_{abc} + \theta . \quad (4.15)$$

As with the previous estimators the network is also trained including the linear term, in which case

$$f_{\text{NL}} = \sum_{abc} w_{abc} \left(\frac{1}{4\pi N_{pix}} \sum_i^{N_{pix}} T_{a,i} T_{b,i} T_{c,i} - \langle T_{a,i} T_{b,i} \rangle T_{c,i} - \langle T_{a,i} T_{c,i} \rangle T_{b,i} - \langle T_{b,i} T_{c,i} \rangle T_{a,i} \right) + \theta . \quad (4.16)$$

Comparing with the AMLE estimator, we can see that it is equivalent to a neural network with parameters

$$w_{def} \mapsto \sum_{abc} \frac{\langle b_{abc} \rangle^1 C_{abc,def}^{-1}}{\sum_{abc,def} \langle b_{abc} \rangle^1 C_{abc,def}^{-1} \langle b_{def} \rangle^1} . \quad (4.17)$$

$$\theta \mapsto 0 \quad (4.18)$$

If this were the optimal linear combination to estimate f_{NL} , the neural network would find the same result as the AMLE but avoiding all the expensive calculations required in the direct computation of this estimator (provided that we have chosen a linear combination for the NNE). Conversely, if that combination were not optimal, the network should be able to find different, more optimal, weights. For instance, for the AMLE to be optimal, the considered statistics should follow a Gaussian distribution, whereas the NNE does not make any assumptions about the intrinsic distribution of the inputs. Therefore, the neural network is expected to perform better when working with non-Gaussian statistics. In addition, the neural network does not require to assume that the covariance matrix is independent of f_{NL} . Even if this approximation works well for the current application, it may not always be the case, which would significantly complicate the calculation of the AMLE. In such cases the NNE would represent a clear advantage over the χ^2 minimization. Finally, we would also like to point out that, although for the current application a linear combination was found to be the best choice for the NNE, in a general case, this estimator is not restricted to a linear combination of the inputs, which can be useful in other problems.

4.5. Implementation

In this section the non-Gaussian simulations used for the analyses as well as some technical details required for the implementation of the estimators are described.

4.5.1. Non-Gaussian simulations

Two different sets of non-Gaussian realizations are used. A set generated with the map-making method proposed in Fergusson et al. (2010) and described also in Curto et al. (2011), and a set of publicly available realisations¹¹ generated by Elsner & Wandelt (2009). In the first method, the non-Gaussian part of the map ($a_{\ell m}^{NG}$) is taken directly from the theoretical bispectrum, while the second algorithm starts from the primordial curvature fluctuations and is therefore more precise.

The two different sets are used for the following reasons. Having a large number of independent realizations is necessary to train the network, as well as to test its performance with the number of training data. Since the first set is faster to produce, 30,000 independent realisations were generated as in Curto et al. (2011). In the analysis with the SMHW of Curto et al. (2011), they found the dispersion on f_{NL} to be slightly larger than using the simulations of set 2. In Curto et al. (2011), constraints on f_{NL} are obtained with both sets finding a discrepancy of 5%. We find similar deviations for the binned bispectrum. This is observed if the average bispectrum at the numerator in (4.8) is computed with simulations with both sets. Then, as the model of set 1 is given by an approximation, the minimum dispersion of the parameter obtained with realisations is slightly larger than using the analytical dispersion in eq. 4.8. Conversely, using realisations of set 2 we find a closer value to the analytically computed lower bound.

Hence, after proving that the NNE converges with few thousand realisations for the best performing form of the estimator, the second set is used for the final results. This is convenient to be able to compare our results with the Fisher dispersion of (4.8), and with the ones obtained with the optimal estimator (Komatsu et al., 2011), where simulations equivalent to the ones of set 2 are used.

The Gaussian and non-Gaussian harmonic coefficients of the CMB realisations, $a_{\ell m}^{NG}$ and $a_{\ell m}^G$, either generated from set 1 or set 2, are combined to obtain the non-Gaussian

¹¹<http://planck.mpa-garching.mpg.de/cmb/fnl-simulations/>

realisation with different values of f_{NL} :

$$a_{lm} = a_{lm}^{\text{G}} + f_{\text{NL}} a_{lm}^{\text{NG}} . \quad (4.19)$$

Noise-weighted V+W band WMAP-7yr realizations were then constructed as explained in Curto et al. (2009a) and Casaponsa et al. (2011), and the KQ75 mask was then applied, which covers roughly 29% of the sky.

4.5.2. Binning scheme

One is free to choose the number and size of the bins in ℓ -space for the binned bispectrum. Bucher et al. (2010) found that for $\ell_{\text{max}} = 2000$ and 64 bins the results obtained were 99.3% of the optimal value. For an application to WMAP, one has $\ell_{\text{max}} = 1024$, so the corresponding number of bins is 32. We have tested the performance of the estimators with different number of bins and find that for $n_{\text{bin}} = 28$ the results have converged. Therefore, the following results use this number of bins, which also provides a modest saving in computation with respect to 32 bins. Conversely to the exhaustive choosing of the binning scheme done in Bucher et al. (2010) estimator, here we simply use logarithmic bins. The logarithmic scale is chosen by imposing the condition that all bins have at least one ℓ .

The binned bispectrum components are computed from combinations of three binned maps $T_a T_b T_c = \sum_{\ell_1 \in I_a} \sum_{\ell_2 \in I_b} \sum_{\ell_3 \in I_c} T_{\ell_1} T_{\ell_2} T_{\ell_3}$. It can be noticed that some of the combinations $\ell_1 \ell_2 \ell_3$ might not satisfy the triangle condition ($\ell_3 - \ell_2 \leq \ell_1 \leq \ell_2 + \ell_3$). To avoid as far as possible those non-contributing combinations, we discard the binned bispectrum components where all the contained ℓ combinations do not meet the triangle condition. For that reason the components used are the ones that hold the following condition:

$$\ell_{I_c}^{\text{min}} - \ell_{I_b}^{\text{max}} \leq \ell_{I_a}^{\text{max}} \leq \ell_{I_c}^{\text{max}} + \ell_{I_b}^{\text{max}} ,$$

where $\ell_{I_n}^{\text{min}}$ and $\ell_{I_n}^{\text{max}}$ are the minimum and maximum value of ℓ of the bin I_n . Then, the binned bispectrum for $n_{\text{bin}} = 28$ consists of 1077 components, whereas the full bispectrum would have $\sim 10^8$ components.

4.5.3. Inpainting

Several inpainting methods have been developed for general imaging reconstruction (see e.g. the review by Bertalmio et al., 2000). The goal of these methods is to restore

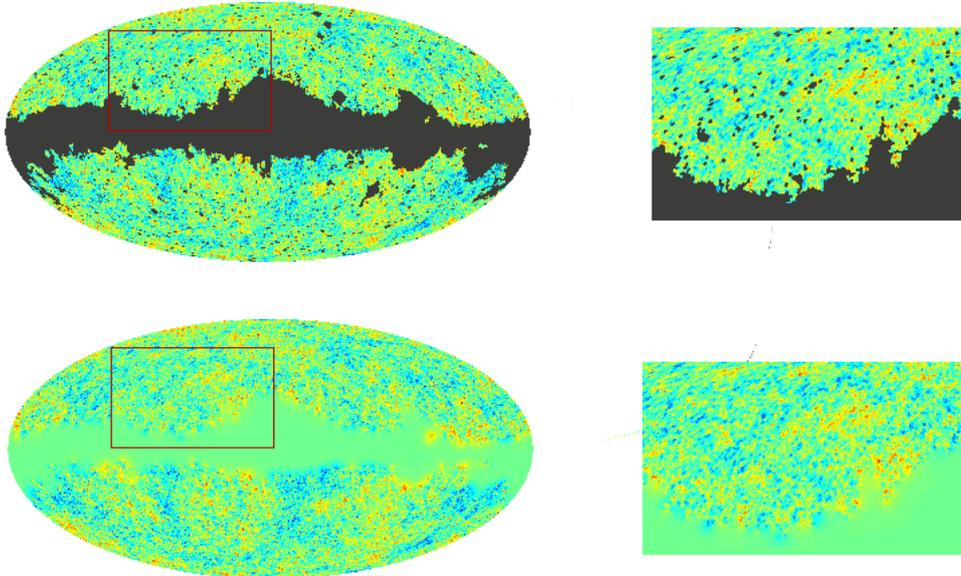


Figure 4.1 - Inpainting effect shown in the masked WMAP-7yr map. On the top the initial temperature map with the mask in dark grey and an amplified region are presented and on the bottom, the same map and region are given after inpainting.

missing or damaged regions of an image to recover the original signal as far as possible. For CMB map reconstruction, the ideal inpainting method would lead to a restored map preserving the statistical properties of the unmasked map.

Different approaches have been used to reduce the discontinuities generated by the mask edges in CMB maps, since they introduce undesirable correlations among the binned bispectrum components. As the intention here is to reduce this impact, rather than reconstruct the full map, we use a simple iterative process that averages over the direct neighbours of the masked pixels, and is based on the work of Oliveira et al. (2001).

One begins with the map $T(\vec{x})$ and the binary mask $M(\vec{x})$. Then each pixel of the masked map $T' = T \times M$ with value zero is substituted by the average of its immediate neighbours, whether masked or not, using the HEALPIX subroutine *neighbours*. The process is repeated 1,000 times, leaving the masked point sources completely inpainted and smoothing the edges of the galactic mask. The results of this process are illustrated in Fig. 4.1. We find that, in this case, the technique is more effective than simply using an apodized mask.

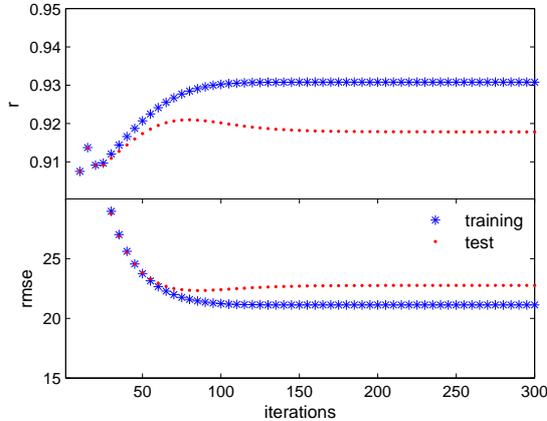


Figure 4.2 - In the top panel the Pearson correlation coefficient between true f_{NL} value and the network estimator \hat{f}_{NL} for case 3 of table 4.2 versus the number of iterations. Bottom panel is for the root mean squared error of f_{NL} at each iteration. Asterisks denote training data and dots denote validation data.

4.5.4. Neural network training process

To train our f_{NL} network we provide it with an ensemble of training data $\mathcal{D} = \{\mathbf{x}^{(n)}, t^{(n)}\}$. The n^{th} input vector $\mathbf{x}^{(n)}$ contains the binned bispectrum components, explained in Section 4.3, of the i^{th} simulated CMB map. The output target is the corresponding f_{NL} value of the i^{th} CMB simulation. Thus, for $n_{\text{bin}} = 28$ the input vector has 1077 components, and the target vector $t^{(n)}$ for the network consists of only one component. From the training set, 20 per cent of the realisations are reserved for the validation process.

The network weights are computed during the training procedure, which in this case requires only a few seconds. The performance of the network is validated during the training process using an independent set of testing data. Figure 4.2 illustrates the training evolution for the regression network with $n_{\text{hid}} = 0$ and $n_{\text{data}} = 10,000$. In the top panel we plot the correlation coefficient between the target and the network outputs on the training set and the test set. We see that a divergence occurs around 60 iterations of the MEMSYS optimiser due to over-fitting. The same behaviour is confirmed if the root mean squared error is studied (bottom panel). The network parameters use to construct our final network estimator in (4.15) and (4.16) are the ones that give a maximum value of the correlation coefficient and a minimum of the root mean squared error in the validation data set.

Estimator	σ_{fh}	σ_g	$\langle f_{\text{NL}} \rangle^{Gauss}$
AMLED	9.7	9.7	-0.2
AMLE		10.3	-0.3
NN		9.8	-0.2

Table 4.1 - Results for noiseless full sky maps of set 1. The first column is for the estimator used, second column indicates the expected dispersion for $\ell_{max} = 1024$ and in the last two columns the dispersion and mean value found for 1,000 Gaussian maps are shown.

It is worth noting that for training the neural network, we need to choose a certain range of $|f_{\text{NL}}|$ to generate the required simulations. We find that $[-220 \ 220]$ is a safe interval for training the network, without significantly biasing the results for $|f_{\text{NL}}|$ up to 30.

4.6. Results

As a preliminary check, we applied the three estimators to Gaussian full-sky maps without noise, finding very similar results in all cases (see table 4.1). In this ideal case, the AMLE should in principle coincide exactly with the AMLED, but because of the lack of correlations among the binned bispectrum components the AMLED seems to be more efficient. This is probably due to numerical uncertainties that arise in the covariance matrix estimation. The neural network is found to be nearly as efficient as the AMLED. It is also worth noticing that the estimators do not present a significant bias.

An important difference between the estimators is the total number of realisations required to converge, which is directly related to the computational efficiency. In particular, the generation of a simulation at $\ell_{max} = 1024$ takes around 1 minute and the obtention of the binned bispectrum components takes approximately 3 minutes of CPU time. For the AMLED, a few hundred realisations are sufficient to estimate the variance of the binned bispectrum. For the AMLE estimator, however, it is necessary to estimate the covariance matrix, which requires at least 25,000 Gaussian simulations. For the NNE, a few thousand realisations are required for the training process to converge. Nonetheless, it is worth noting that the number of training realisations required by the NNE does vary with the case being studied. For example, for inpainted maps

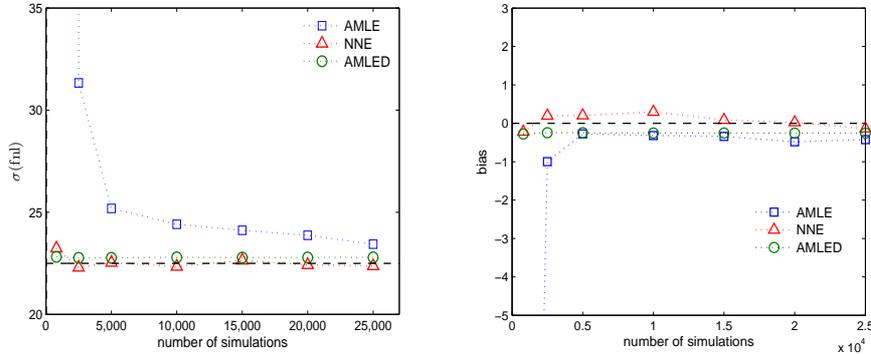


Figure 4.3 - Comparison of the efficiency (left) and bias (right) of the three estimators with respect to the number of simulations used to construct the estimator. For reference, the optimal values for the dispersion and bias (dashed black line) are also shown. Note that for the NNE, the simulations are used for the training process, whereas for the AMLE they are employed to estimate the covariance matrix. For the AMLED, they correspond to the number of simulations used to obtain the diagonal elements of the covariance matrix.

where neither the linear term is taken into account nor the mean is subtracted (case 1 of table 4.2), the NNE needs 10,000 independent simulations to converge.

In applying the three estimators to realistic simulations, based on WMAP-7yr data, larger differences are observed in the results; these are summarised in table 4.2. We find that the AMLED estimator reaches values close to the expected dispersion if and only if the linear term is subtracted and inpainting is performed. Actually, if the estimator is applied to non-inpainted maps, the dispersion worsens by a $\sim 60\%$. Of course, in the absence of the linear term, the estimator becomes highly suboptimal, giving errors of 300%. This is not the case for the other two estimators. We notice that the full covariance matrix χ^2 estimator and the neural network give similar results if instead of taking into account the linear term, the mean value of the intermediate maps is subtracted, as is the case for wavelets and needlets (Donzelli et al., 2012). This is observed in both inpainted and non-inpainted maps, comparing cases 2 and 3 and 4 and 5 respectively (see table 4.2). Indeed, these estimators appear more robust, since the improvement due to the inpainting is small. In particular, comparing cases 2 and 4, the NNE estimator without inpainting increases the dispersion only by 5% and for the full χ^2 estimator by $\sim 10\%$, while for the AMLED the results are much worse. Although similar results are found with the AMLE and the NNE estimators,

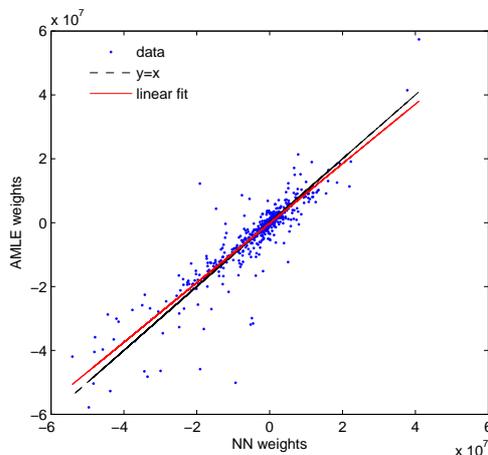


Figure 4.4 - Weights for the AMLE estimator involving the covariance matrix and the model, versus the NN weights obtained after the training process. This comparison is made when both estimators have converged presenting a linear fit slope and intercept of $a = 91, b = 2 \times 10^5$.

one important difference is the number of simulations required to construct them. As commented before, 25,000 Gaussian realizations were used to estimate the covariance matrix in AMLE. As shown in top panel of Fig. 4.3, the NNE requires dramatically fewer training realisations and also has the advantage that the average value of the binned bispectrum at $f_{\text{NL}} = 1$ does not need to be estimated. In the same figure, bottom panel, we plot the bias found for the f_{NL} estimates for 1,000 Gaussian realisations for the three estimators with the number of simulations used. One sees that the AMLE requires more realizations than the other two estimators to produce unbiased results.

All these results indicate that the neural network is a viable short cut to obtaining the necessary weights to construct the AMLE estimator. In Fig. 4.4 the weights found for the neural network are compared to those of the AMLE. Note that the weights of both estimators are very similar, validating the relation stated in (4.17). The contribution of the network parameter θ is negligible for all cases.

In terms of computational demand, the most efficient estimators are the NNE and the AMLED, with the number of simulations required at least 10 times smaller than for the AMLE. Note that for the AMLED we have used realisations to estimate the average of the bispectrum at $f_{\text{NL}} = 1$, therefore the final number of realisations employed is similar to the ones used for training the NNE.

For all three estimators, the best results are obtained when the map is inpainted and

Case	Inpainting	linear term	mean subs.	Estimator	σ_g	$\langle f_{NL} \rangle^{Gauss}$	$(\sigma_{fh} - \sigma_g)/\sigma_{fh}(\%)$
1	Yes	No	No	AMLED	107	3	300
				AMLE	32.7	-1	45
				NN	29.7	-0.3	32
2	Yes	Yes	No	AMLED	22.7	0.7	0.9
				AMLE	23.3	0.7	3.5
				NN	22.4	0.7	0.4
3	Yes	No	Yes	AMLED	31.5	0.7	40
				AMLE	24.0	0.7	6.7
				NN	23.1	0.5	2.7
4	No	Yes	No	AMLED	35.9	-0.3	60
				AMLE	24.3	0.1	9.3
				NN	23.6	0.6	4.8
5	No	No	Yes	AMLED	37.0	1.5	64
				AMLE	24.6	-0.4	8.0
				NN	23.6	0.4	4.8

Table 4.2 - Comparison of results depending on the estimator. The columns are the characteristic of the estimator, if an inpainting of the simulations is made, if the linear term is added and if the mean was subtracted on the binned intermediate maps. Next columns are $\sigma(f_{NL})$ and $\langle f_{NL} \rangle$ for 1,000 Gaussian simulations. Finally the relative error related to the minimum expected dispersion is shown in the last column.

the linear term is subtracted (see case 3 of table 4.2, indicated in bold face). For this optimal case, we compute $\langle b_{abc} \rangle^1$ with 1,000 simulations of set 2 (Elsner & Wandelt, 2009), to compare it with the expected dispersion for a WMAP-7yr characteristics, computed as in (4.8). The neural network is now trained with this set of $a_{\ell m}$. As we have seen, the NNE typically requires 2,500 independent training realisations to converge. Since only 1,000 are available, we therefore generated 10,000 simulations using the same set of $a_{\ell m}$ rotating them and adding different noise contributions. This procedure was used in Casaponsa et al. (2011) and was found to be useful when only a small number of realisations is available. In table 4.3 the final results for all of the estimators are shown. The values for WMAP-7yr data are without point sources correction, which is given in the last column of the same table. The unresolved point sources contribution to f_{NL} is obtained using the same procedure as in Curto et al. (2009a) and Casaponsa et al. (2011). As expected, by looking at the preliminary results, the tightest constraints are given by the NNE and AMLED estimators. For comparison, the WMAP-7yr map f_{NL} estimate with the optimal estimator obtained by Komatsu et al. (2011) is 42, without the point sources correction. Note that the closest value is given by the NNE. The constraints for f_{NL} with the point source contribution taken into account at 95% confidence level are $-3 < f_{\text{NL}} < 83$ to be compared with $-2 < f_{\text{NL}} < 82$ given by the optimal estimator.

We conclude that the most efficient tools are the neural network regression estimator and the AMLED estimator. The latter would be the choice if a small set of non-Gaussian simulations is available ($\sim 1,000$), or analytical models are preferred. However, the AMLED depends on a specific pre-processing of the data. Neural networks give

Estimator	σ_{fh}	σ_g	$\langle f_{\text{NL}} \rangle^{Gauss}$	f_{NL}^{map}	Δf_{NL}
AMLED	21.3	21.7	-0.2	33.4	3 ± 2
AMLE		22.4	-0.1	39.8	3 ± 2
NN		21.4	0.5	44.2	4 ± 2

Table 4.3 - Results for inpainted Gaussian realizations. Model estimated and neural network trained using Elsner & Wandelt simulations (set 2). The columns from left to right are: the estimator used, the Fisher σ computed from eq. 4.8, the dispersion and mean value of \hat{f}_{NL} for 1,000 Gaussian simulations. Followed by the f_{NL} value found for WMAP-7yr data and the contribution expected by the unresolved point sources (Δf_{NL}).

almost optimal results, without the use of inpainting, thereby avoiding the need to alter the data.

Finally, the constraints for WMAP-7yr data, with the unresolved point sources contribution included, at 95% confidence level would be $-3 < f_{\text{NL}} < 83$. These results are compatible with $f_{\text{NL}} = 0$, as found in Komatsu et al. (2011); Curto et al. (2011); Bennett et al. (2013). Note that we have used foreground reduced maps, and the foregrounds have not been marginalised over in this analysis. We note that neural networks would be a useful method to estimate jointly other forms of non-Gaussianity, such as those where the number of outputs were set to a number of different f_{NL} shapes (e.g. local, equilateral, orthogonal), but this is left for future work.

CHAPTER 5

Size magnification as a complement to cosmic shear

This chapter is based on the published work of Casaponsa et al., 2013a. In this chapter the extent to which cosmic size magnification may be used to complement cosmic shear in weak gravitational lensing surveys is investigated, with a view to obtaining high-precision estimates of cosmological parameters. Using simulated galaxy images, we find that unbiased estimation of the convergence field is possible using galaxies with angular sizes larger than the Point-Spread Function (PSF) and signal-to-noise ratio in excess of 10. The statistical power is similar to, but not quite as good as, cosmic shear, and it is subject to different systematic effects. Application to ground-based data will be challenging, with relatively large empirical corrections required to account for the fact that many galaxies are smaller than the PSF, but for space-based data with 0.1-0.2 arcsecond resolution, the size distribution of galaxies brighter than $i \simeq 24$ is almost ideal for accurate estimation of cosmic size magnification.

5.1. Introduction

In contrast to galaxy ellipticity measurement, the size information has not been explored in detail, possibly because the complicating effects of the PSF and pixellisation were thought to be too challenging. However, there are two reasons for revisiting size magnification as a potential tool for cosmology: one is that accurate shear estimation is itself very challenging, and size could add useful complementary information; the second is that methods devised for ellipticity estimation must deal with the PSF and

pixellisation, and as a byproduct provide a size estimate, or a full posterior probability distribution for estimated size, which is currently ignored or marginalised over.

In terms of signal-to-noise (S/N) of shape distortions vs magnification estimation, the relative strengths of the methods depend on the distributions of ellipticity and size. The former has an r.m.s. of around 0.3-0.4 (Leauthaud et al., 2007); for bright galaxies ($M_r < -20$), the Sloan Digital Sky Survey (SDSS) found that the size distribution is approximately log-normal with $\sigma \ln R \sim 0.3$, and for fainter galaxies $\sigma(\ln R) \sim 0.5$, where R is the Petrosian half-light radius (Shen et al., 2003); for deeper space data the dispersion is also around 0.3 (Simard et al., 2002). Thus one might expect a slightly smaller S/N for lensing measurements based on size rather than ellipticity, but not markedly so. Bertin & Lombardi (2006) proposed a method based on the fundamental plane relation (Dressler et al., 1987; Djorgovski & Davis, 1987) to reduce the observable size variance. Huff & Graves (2011) applied a similar method to measure galaxy magnification using 55,000 galaxies of the SDSS catalogue, and find consistency with shear using the same sample. Also a detection with COSMOS HST survey using a revised version of the KSB method is presented in Schmidt et al. (2012), showing reasonable consistency with shear.

Here we revisit size magnification measurement, and will show that to use size-magnification we require i) a wide area survey that enables observations of a sufficiently large sample of galaxies, this is required to overcome the intrinsic scatter, and ii) a consistently small PSF that does not destroy the size information of the observed galaxies. Both of these requirements can be met with a wide-area space-based survey, although some science may be possible from the ground. Euclid¹² (Laureijs et al., 2011) should meet these requirements (large samples will be available, and the PSF size is smaller than typical galaxies), so the size information could be considered as a complementary cosmological probe to weak lensing ellipticity measurements. One advantage of using the size information is that the magnification and distortion have different radial dependences on the spatial distribution of matter, which may be very useful to lift the so-called mass-sheet degeneracy (Bartelmann et al., 1996; Fort et al., 1997; Taylor et al., 1998; Broadhurst et al., 2005; Umetsu et al., 2011; Vallinotto et al., 2011), that occurs due to the reduced shear (or the measured ellipticities) being invariant under a transformation of the distortion matrix by a scalar multiple.

Besides the degeneracy lifting, another advantage of using magnification is that

¹²<http://www.euclid-ec.org>

combining the size magnification information with the shear will reduce uncertainties on the reconstruction of the distribution of matter (Jain, 2002; Vallinotto et al., 2011; Sonnenfeld et al., 2011).

On the measurement of the size, all of the shear estimation methods referred to above already estimate the size of galaxies when calculating the ellipticities, so we expect to measure this additional information for free, given an accurate ellipticity measurement. However, the accuracy of size information should not be taken for granted: it is important to know the uncertainties in size measurement, and how they propagate to a convergence field estimation. It is this question of how accurately one can measure the sizes of galaxies, that this chapter addresses. Amara & Réfrégier (2007); Kitching et al. (2008) have shown that to obtain an accurate determination of cosmological parameters, such as the equation of state of dark energy, the systematic errors in the measured ellipticities should be $\lesssim 0.2\%$, and we would expect similar requirements for size. Although a full study of the convergence bias at this level needs to be done, the main goal of this first work is to investigate whether unbiased measurement of size is feasible at all, and to come some basic conclusions on required image sizes and signal-to-noise.

The chapter is organised as follows. First, in Section 5.2 we will present the weak lensing quantities that are used throughout this chapter, then a definition of the estimator, and a brief comment on the method and the characteristics of the simulated images. Finally, the analysis and results are explained in Section 5.3. The conclusions are summarised in chapter 6.

5.2. Method

A good algorithm for weak lensing analysis must be able to take into account the distortion introduced by the PSF, pixelization effects and pixel noise. Another requirement is that it should be computationally fast because the statistical analysis will be made on large samples. This means that algorithm development is challenging because of the dissonant requirements of both increased accuracy and increased speed as the required systematic level decreases. Several methods have been proposed and applied to weak lensing surveys, and are described in the challenge reports of STEP, GREAT08 and GREAT10 (Heymans et al., 2006; Bridle et al., 2010; Kitching et al., 2012). These blind challenges have been critical in demonstrating to what extent methods can achieve

the required accuracy for upcoming surveys by creating simulations with controlled inputs against which results can be tested. Here we propose a very similar approach as the one presented in the GREAT10 challenge, we have used simulated galaxy images with different properties to measure the response of the size/convergence measurement under different conditions (corresponding to changes in the PSF, S/N and bulge fraction).

5.2.1. Weak lensing formalism

The distortions induced by gravitational lensing are described by the Jacobian matrix which maps the true angular position of the image to the angular position of the source (in the absence of deflections):

$$\mathcal{A}(\vec{\theta}) = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}$$

which defines the convergence field κ and complex shear field $\gamma \equiv \gamma_1 + i\gamma_2$ (for more details, see e.g. Bartelmann et al., 1996; Hoekstra et al., 1998; Munshi et al., 2008). In terms of the convergence and the shear there are two important variables, directly related to the lensing observables. The reduced shear,

$$g = \gamma(1 - \kappa)^{-1} \quad (5.1)$$

represents shape changes ignoring size. The magnification of the surface area μ is,

$$\mu = \frac{1}{\det(\mathcal{A})} = [(1 - \kappa)^2 - |\gamma|^2]^{-1}. \quad (5.2)$$

If $\kappa, |\gamma| \ll 1$ (which we assume throughout) can be approximated by

$$\mu \simeq 1 + 2\kappa.$$

The power spectrum for κ can be obtained in terms of the matter power spectrum $P_\delta(k, w)$, where w is a comoving distance coordinate which plays the role of cosmic time. For a set of sources with a distribution function $p(w)$,

$$P_\kappa(\ell) = \frac{9H_0^4 \Omega_m^2}{4c^4} \int_0^{w_h} dw \frac{g^2(w)}{a^2(w)} P_\delta\left(\frac{\ell}{w}, w\right), \quad (5.3)$$

where $g(w) = \int_w^{w_h} dw' p_w(w') \frac{w'-w}{w}$, w is the co-moving distance, w_h is the horizon distance. Ω_m and H_0 are the present matter density parameter and Hubble parameter.

Note that eq. 5.1 and 5.2 are always valid while eq. 5.3 is only meaningful for cosmic shear.

In the weak lensing limit, the power spectrum of the magnification fluctuations ($\mu - 1$) is 4 times $P_k(l)$, therefore, in principle, cosmological constraints could be made independently of the shear (Jain, 2002; Barber & Taylor, 2003), however the signal-to-noise ratio for the measured ellipticities is in general larger, hence the shear may carry more statistical weight. Even so a complementary analysis of shear and magnification measurements will necessarily provide tighter constraints on cosmological parameters than a shear analysis alone. In particular, in van Waerbeke (2010) it is shown that the constraints on σ_8 and Ω_m can be improved up to $\sim 40\%$, similarly combining size-magnification, galaxy densities and shear, the improvement on the precision of halo mass estimates can be $\sim 40\% - 50\%$ (Rozo & Schmidt, 2010).

5.2.2. Estimator

In this study, the observed galaxy size, s and the intrinsic galaxy size, s^s are defined as the half light disk radii¹³. The ratio between the magnified and intrinsic galaxy size is given by $\frac{s}{s^s} = \mu^{\frac{1}{2}}$. On the other hand, the magnification is related to the convergence, to first order, $\mu^{\frac{1}{2}} = 1 + \kappa$, therefore we can write $\frac{s}{s^s} = \mu^{\frac{1}{2}}$ then $\kappa = \frac{s}{s^s} - 1$. We can construct an estimator for κ in the weak lensing limit by assuming that, since $\langle \kappa \rangle = 0$, the mean size value is not modified by lensing, i.e., $\langle s^s \rangle = \langle s \rangle$, and replacing s^s by its expectation value:

$$\hat{\kappa} = \frac{s}{\langle s \rangle} - 1. \quad (5.4)$$

From the definition of the estimator, and the width of the s distribution, an estimate of κ from a single galaxy will be very noisy, with smaller galaxies than the mean always giving a negative $\hat{\kappa}$, while larger galaxies will produce positive $\hat{\kappa}$. What is important is to test if our estimator is unbiased over a population to a sufficient degree to be useful for real data.

5.2.3. *lensfit*

Throughout we use *lensfit* (Miller et al., 2007; Kitching et al., 2008, Miller et al., 2012) to estimate the galaxy size; we use this because: 1) it has been shown that

¹³Note that there is an erratum in the published work where is stated that the defined size is the semi-major axis

lensfit performs well on ellipticity measurement; 2) it is a model fitting code which also measures the sizes of galaxies; 3) it allows for the consistent investigation of the intrinsic distribution of galaxy sizes through the inclusion of a prior on size, and 4) it includes the effects of PSF and pixellisation. *lensfit* was proposed in Miller et al. (2007) and has been proved to be a successful tool for galaxy ellipticity shape measurements (Kitcing et al., 2008). Although model-fitting is the optimal approach for this type of problem if the model used is an accurate representation of the data, the main disadvantage is that is usually computationally demanding to explore a large parameter space. *lensfit* solves this problem by analytically marginalizing over some parameters that are not of interest for weak lensing ellipticity measurements, such as position, surface brightness and bulge fraction. The size reported by *lensfit* is also marginalised over the galaxy ellipticity.

Sensitivity correction

In the Bayesian formalism the expected value for the size of an individual galaxy can be written as $\langle s \rangle = \int s p(s|s_d) ds$, where s_d is the data and s stands for the fitted model parameter for the size explained in Sec. 5.2.4. In terms of the prior $\mathcal{P}(s)$ and the likelihood $\mathcal{L}(s_d|s)$ the expression is

$$\langle s \rangle = \frac{\int s \mathcal{P}(s) \mathcal{L}(s_d|s) ds}{\int \mathcal{P}(s) \mathcal{L}(s_d|s) ds}. \quad (5.5)$$

Individual galaxy size estimates allow errors to be assigned to each galaxy, or the full posterior can be used and the information propagated to the κ signal. Miller et al. (2007) introduced the shear sensitivity, a factor that corrects for the fact that the code measures ellipticities but that shear (a statistical change in ellipticity) is the quantity of interest. A similar correction is required for size measurement, whereby we measure the size but it is the convergence that is the quantity of interest; this correction is needed because for a single galaxy the prior information for the convergence is not known, and we assume it is zero. With a Bayesian method we can estimate the magnitude of this effect for each galaxy, a further reason to use a Bayesian model fitting code in these investigations. Consider the Bayesian estimate of the size of galaxy i and write its dependence on κ as a Taylor expansion:

$$\hat{s}_i = s_i^s + \kappa \frac{d\hat{s}_i}{d\kappa}. \quad (5.6)$$

In the simple case where the likelihood $\mathcal{L}(s_d|s)$ (for simplicity, hereafter $\mathcal{L}(s) = \mathcal{L}(s_d|s)$) is described by a Gaussian distribution with variance b^2 , with an expectation value s ,

and a prior $\mathcal{P}(s)$ that also follows a Gaussian distribution centred on \bar{s} with variance a^2 , the posterior probability will follow a Gaussian distribution with expectation value:

$$\langle s \rangle = \frac{\bar{s}b^2 + s_d a^2}{a^2 + b^2}$$

and variance

$$\sigma^2 = \frac{a^2 b^2}{a^2 + b^2}.$$

These equations illustrate that the posterior is driven towards the prior in the low S/N limit ($b \rightarrow \infty$), and thus requires correction. Differentiating the expression $s_d = s^s(1 + \kappa) + \sigma_s$, with σ_s being the systematic noise, we find that the κ sensitivity correction is:

$$\frac{d\hat{s}}{d\kappa} = \frac{a^2}{a^2 + b^2} \frac{ds_d}{d\kappa} = \frac{a^2}{a^2 + b^2} s^s, \quad (5.7)$$

substituting into eq. 5.6

$$\hat{s}_i = s_i^s + \kappa s_i^s \frac{a^2}{a^2 + b^2}$$

we find the estimator for κ will be the same as in eq. 5.4, corrected by the sensitivity factor:

$$\kappa = \left(\frac{\hat{s}}{\langle \hat{s} \rangle} - 1 \right) \frac{a^2 + b^2}{a^2}. \quad (5.8)$$

In this work we have used this approximation for simplicity but in general a normal distribution should not be assumed. A more general estimation of the κ correction can be done in the same way as with the shear and can be evaluated numerically, without the need of using external simulations.

To calculate the sensitivity correction in the general case we consider the response of the posterior to a small κ , by adding the convergence contribution in the likelihood, $\mathcal{L}(s - s^s) \mapsto \mathcal{L}(s - s^s - \kappa s^s)$ and expand it as a Taylor series:

$$\mathcal{L}(s - s^s - \kappa s^s) \simeq \mathcal{L}(s - s^s) - s^s \kappa \frac{d\mathcal{L}}{ds}.$$

We then substitute into eq. 5.5 and differentiate to obtain the analytic expression for the κ sensitivity (for more details of this applied to ellipticity measurement see Miller et al. 2007; Kitching et al. 2008)

$$\frac{ds}{d\kappa} \simeq \frac{\int (\langle s \rangle - s) \mathcal{P}(s) s^s \frac{d\mathcal{L}}{ds} ds}{\int \mathcal{P}(s) \mathcal{L}(s) ds}. \quad (5.9)$$

If the prior and likelihood are described by a normal distribution, this expression can be analytically computed and the sensitivity correction is the same as before. A similar empirically motivated correction on the estimator expression was used in eq.5 of Schmidt et al. (2012), where the factor is computed with simulations.

5.2.4. Simulations

In order to test the estimation of sizes with *lensfit* we have generated the same type of simulations used in the GREAT10 challenge (Kitching et al., 2010, 2012), but with non-zero κ . Multiple images were generated, each containing 10,000 simulated galaxies in a grid of 100x100 postage stamps of 48x48 pixels; each postage stamp contains one galaxy.

Each galaxy is composed of a bulge and a disk, each modeled with Sérsic light profiles:

$$I(r) \simeq I_0 \exp \left\{ -K \left[\left(\frac{r}{r_d} \right)^{\frac{1}{n}} - 1 \right] \right\} \quad (5.10)$$

where I_0 is the intensity at the effective radius r_d that encloses half of the total light and $K = 2n - 0.331$. The disks were modelled as galaxies with an exponential light profile ($n = 1$), and the bulges with a de Vaucouleurs profile ($n = 4$). Ellipticities for bulge and disk were drawn from a Gaussian distribution centred on zero with dispersion $\sigma = 0.3$. Both components had distributions centred at the middle of the postage stamp with a Gaussian distribution of $\sigma = 0.5$ pixels. The galaxy image was then created adding both components. The S/N was fixed for all galaxies of the image and implemented by calculating the noise-free model flux by integrating over the galaxy model, then adding a constant Gaussian noise with a variance of unity and rescaling the galaxy model to yield the correct signal-to-noise, as in Kitching et al. (2012). Finally the PSF was modelled with a Moffat profile with $\beta = 3$, with FWHM fixed for all galaxies on the image, with different ellipticities drawn from a uniform distribution, with ranges given in Table 5.1.

The different types of image were generated to study the effects of the bulge fraction (fraction of the total flux concentrated in the bulge), the S/N and the PSF separately. In summary, the main characteristics of the considered sets are:

- Set 1. Disk-only galaxies (bulge fraction = 0), negligible PSF effect (FWHM PSF = 0.01 pixels) and different S/N.
- Set 2. Disk-only galaxies (bulge fraction = 0), with S/N=20 and different sizes of PSF.
- Set 3. Negligible PSF effect, S/N=20 and different bulge fractions.

Set Name	S/N	<i>fwhm</i> PSF(pixel)	<i>e</i> PSF	B/D fraction	r_d (pix)	r_b (pix)	<i>e</i>
Set 1	[10,40]	0.01	0	0	$\langle r_d \rangle = 7, \sigma = 1.2$	-	0
Set 2	20	[0.01,10]	0	0	$\langle r_d \rangle = 7, \sigma = 1.2$	-	$\langle e \rangle = 0, \sigma = 0.3$
Set 3	20	0.01	0	[0,0.95]	$\langle r_d \rangle = 7, \sigma = 1.2$	$r_d/2$	$\langle e \rangle = 0, \sigma = 0.3$
Set 4	[10,40]	4.5	[-0.1,0.1]	0.5	$\langle r_d \rangle = 7, \sigma = 1.2$	$\langle r_b \rangle = 3.5, \sigma = 0.6$	$\langle e \rangle = 0, \sigma = 0.3$

Table 5.1 - Major characteristics of the different sets used in this analysis. In bold are marked the variables explored in each set and the range of variation. In Set 4, PSF ellipticities are drawn from a uniform distribution in the range specified. Note that r_i corresponds to the half-light radius. Last column is the galaxy ellipticity and is the same for both components, bulge and disk.

- Set 4. Bulge fraction of 0.5, FWHM of PSF 1.5 times smaller than the characteristic size of the disk, and different S/N.

To characterize the size of the galaxy the half-light radius of a circular isophote, r_d , is used. We have drawn r_d from a Gaussian distribution with expected value of 7 pixels and dispersion of 1.2 pixels, to keep disk sizes of at least 2 pixels and not larger than the postage-stamp. The galaxy sizes explored here have a somewhat smaller range ($\sigma(\ln R) \sim 0.18$) than found by Shen et al. (2003) with the SDSS catalogue, where in terms of pixels the mean value of the full sample is around 5 with $\sigma(\ln R) \sim 0.3$ (see Fig.1 of Shen et al., 2003). Therefore the sensitivity corrections are consequently larger than would be needed for real data. Besides a wider distribution of galaxies, the important change from the original images for the GREAT10 challenge is the addition of a non-zero κ -field that creates a size-magnification effect (in GREAT10 only a shear field was used to distort the intrinsic galaxy images). A Gaussian convergence field with a simple power-law power spectrum in Fourier space $P_\kappa(\ell) \sim 10^{-5}\ell^{-1.1}$ has been applied to each image. The power-law is a good approximation to the theoretical power spectrum over the scales $10 \lesssim \ell \lesssim 10000$ (see i.e. Schneider, 2005). The size of the κ -field is $\theta_{image} = \frac{2\pi}{\ell_{min}}$ and θ_{image} is set to 10 degrees, such that the range in ℓ we used to generate the power was $\ell = [36, 3600]$ where the upper bound is given by the grid separation cut-off. In real space this translates to a maximum $|\kappa|$ of around 2%, although we investigate larger $|\kappa|$ in Section 5.3.4. The κ -field is generated on the 100x100 grid, each point representing a postage stamp, and is applied to the galaxy at the same position ($s = s^s(1 + \kappa)$), neglecting the contribution of the shear. The pixel angular size is not fixed, and can be scaled to any experimental set up.

5.3. Results

Before trying to estimate the convergence field of our most realistic image, we have tested the dependence on different aspects separately: S/N, PSF size and bulge fraction. We expect these to be the observable effects that have the largest impact on the ability to measure the size of galaxies. Lower S/N will cause size estimates to become more noisy and possibly biased (in a similar way as for ellipticity, see Melchior & Viola 2012); a larger PSF size will act to remove information on galaxy size from the image and a change in galaxy type or bulge fraction may cause biases because now two characteristic sizes are present in the images (bulge and disk lengths). In order to study carefully the sensitivity of our estimator to systematic noise, PSF or galaxy properties, we started from the simplest case and added increasing levels of complexity. The number of galaxies used for the analysis is 200,000 for the first three sets and we increased the number to 500,000 for the last test to give smaller error bars.

We compare the estimated $\hat{\kappa}$ computed as in eq. 5.4 with the input field κ , and fit a straight line to the relationship to estimate a multiplicative bias m and an additive bias c :

$$\hat{\kappa} = (m + 1)\kappa + c, \quad (5.11)$$

after applying the sensitivity correction (section 5.2.3). We bin the data¹⁴ and a linear fit is done to compute m and c . This process is shown in Fig. 5.1. The error bars for the regression coefficients are given by its standard deviation, assuming that the errors are normally distributed. The regression coefficients and its errors are computed using the function *polyfit* of MATLAB software.

We now discuss each of the categories in turn.

5.3.1. Signal-to-noise

As a first approach to the problem, disk-only galaxies with a negligible PSF (FWHM=0.01 pixels) and zero ellipticity were generated to test the dependence of the bias on S/N alone, given otherwise perfect data. In Fig. 5.2 we can see that as we increase the S/N the accuracy of the size estimation grows, as expected. In Fig. 5.3 we show the estimation of the convergence field, modified by the sensitivity correction. There is a

¹⁴Note that the differences between the results before and after the binning of κ are within the error bars.

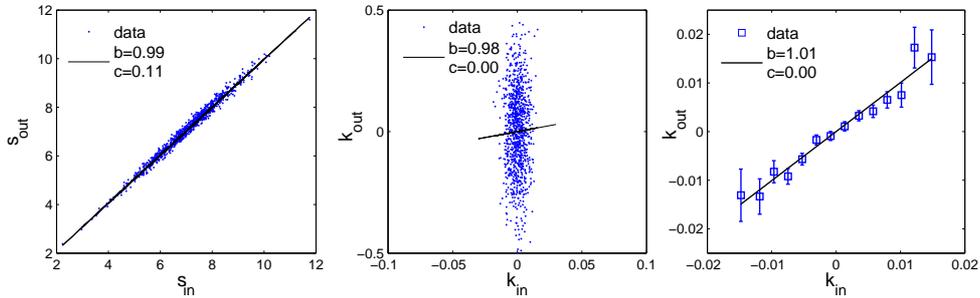


Figure 5.1 - Sequence of steps to obtain m and c values. First panel shows the *lensfit* output size compared to the input size, in the second panel the estimated κ compared to the input convergence at each galaxy, and in the third panel is shown the same plot using bins. Slope and intercept values of the fitting are shown in each plot (throughout, we fit generically $y = bx + c$, with $b = m + 1$ and $c = c$ of eq. 5.11). This is for galaxies of Set 1 with signal-to-noise 40.

clear correlation between the inputs and the outputs, and the slope is close to unity for all S/N explored. In Fig. 5.4 the estimates for m and c are shown, with and without the sensitivity correction. In this case the correction does not alter the results much except at low S/N, because the sizes are less accurately estimated. In this work the factor $a^2/(a^2 + b^2)$ is estimated by the inverse of the slope of the size estimation fitting (see Fig. 5.2). Using 200,000 galaxies for this test, the values found for m and c are consistent with zero, typically $m \simeq 0.02 \pm 0.05$, and $c \simeq (5 \pm 5) \times 10^{-4}$.

5.3.2. PSF effect

To study the uncertainties on the size estimation due to the PSF size, we generated images with different FWHM PSF values, with an intermediate signal to noise (S/N=20), maintaining the same properties as before, except that we considered here a Gaussian distribution of ellipticities with mean value of $e = 0$ and $\sigma_e = 0.3$ (per component). The size estimates are good for small PSFs, but become progressively more biased as the PSF size increases beyond the disk scale length (see Fig. 5.5). A PSF with a FWHM larger or similar to the size of the disk, tends to make the galaxy look larger, and the estimator for κ becomes biased. This effect can be seen in the slope and intercept of $\hat{\kappa}$ vs κ plot (Fig. 5.6). Fig. 5.7 shows the variation of the parameters m and c with the ratio between the scale-length of the PSF and the galaxy (ratio= r_d/PSF_{FWHM}).

We find no evidence for an additive bias, but we do find a multiplicative bias for large

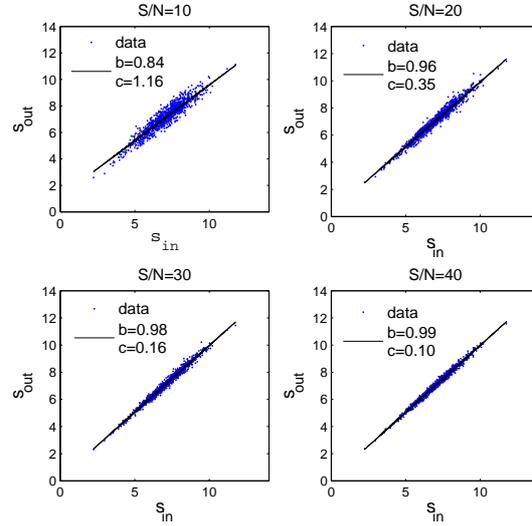


Figure 5.2 - Comparison of the estimated sizes by *lensfit* with the input galaxy size for different S/N in the range [10,40]. Disk-only circular galaxies with a negligible PSF effect are considered (Set 1). Slope and intercept of the fitting are shown (b and c , respectively). Note that the input size is the lensed one.

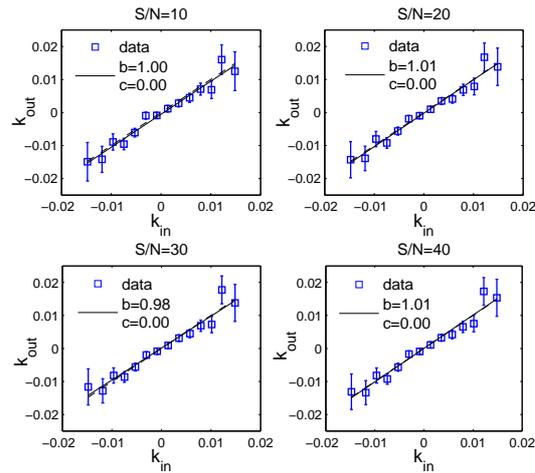


Figure 5.3 - Comparison of the binned estimated convergence and the input value for Set 1 with different S/N in the range [10,40]. Slope and intercept of the fitting are shown (b and c , respectively). For errors, see text.

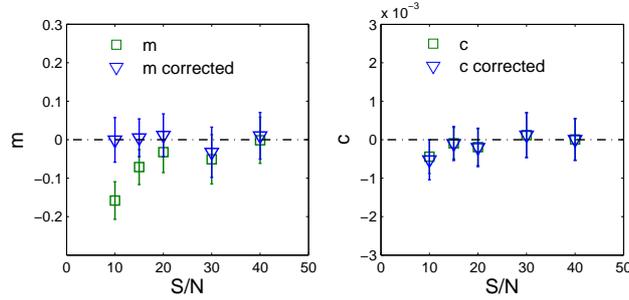


Figure 5.4 - m and c values computed with 200,000 galaxies of Set 1. Triangles are for the values obtained with the sensitivity correction and squares without it.

PSFs. With a wide size distribution, some of the smaller galaxies are convolved with a PSF larger than their size, and this could produce an overall bias in κ , but if the number of those galaxies is not very large, the effect on the global estimation of the convergence field will be correspondingly small. Similar biases exist with shear measurement for large PSFs, but the biases are larger here. For a space-based experiment, with a relatively bright cut at $i \sim 24.5$, such as planned for Euclid, the limitation on PSF size will not be dominant because the median galaxy size is 0.24 arcsec (Simard et al., 2002; Miller et al., 2012), larger than the PSF FWHM of 0.18 arcsec. For ground-based surveys, such as CFHTLenS and future experiments the situation is not so clear, the measurement will be more challenging, and large empirical bias corrections of the order of $m \simeq -0.5$ will be needed (see the first point of Fig. 5.7).

5.3.3. Bulge fraction

In this test we generated galaxy images with bulges with different fractions of the total flux, to test the response to the galaxy type. In Fig. 5.8 we show that galaxy size estimates for bulge fractions of 0.2 are much better than for galaxies with bulge fractions of 0.95. This is because for bulge-dominated models the central part of the galaxy becomes under-sampled due to a limiting pixel scale. The poor estimation of sizes is reflected in the κ estimation (bottom panels of Fig. 5.8). The parameters m and c for this set are shown in Fig. 5.9, where we can see that for bulge fraction greater than 0.8, the results are clearly biased with $m = -0.25$. For all bulge fractions the error bars are around 10%. Although for bulge-only galaxies, the κ estimates are poor most of the galaxies used for weak lensing experiments have bulge fractions lower than 0.5 (Schade et al., 1996), so that in fact the population of useful lensing galaxies is likely

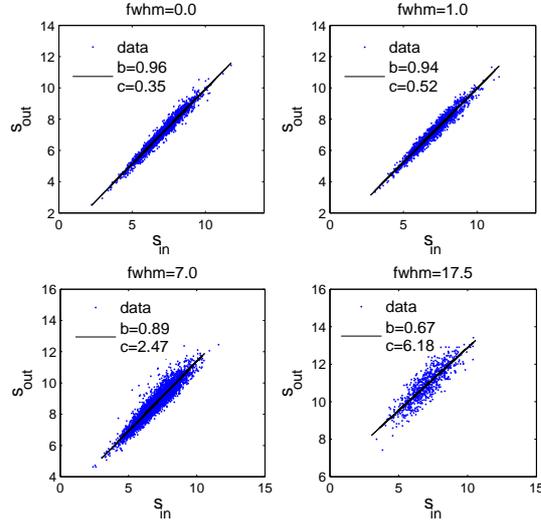


Figure 5.5 - Sizes estimates vs input sizes for four different PSF scale-lengths between 0.1 and 7 pixels. Galaxies are disks with $S/N=20$ and mean size 7 pixels. Slope and intercept of the fitting are shown (b and c , respectively).

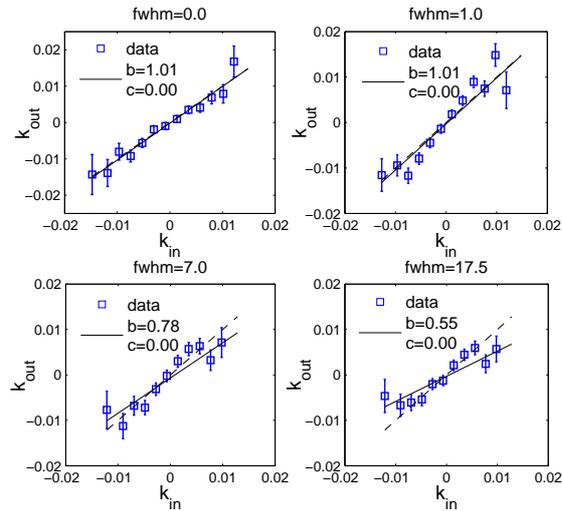


Figure 5.6 - κ estimates vs input values for four different PSF scale-lengths. Galaxies are disks with $S/N=20$ and mean size 7 pixels. Dashed line is $\kappa_{out} = \kappa_{in}$ and the solid line is the least squares fit, with slope and intercept shown in the plots. Note that $b=m+1$.

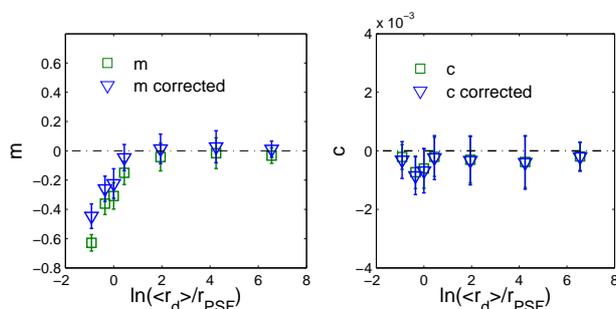


Figure 5.7 - m and c values computed with 200,000 galaxies of Set 2. Triangles represent the values obtained with the sensitivity correction and squares without it.

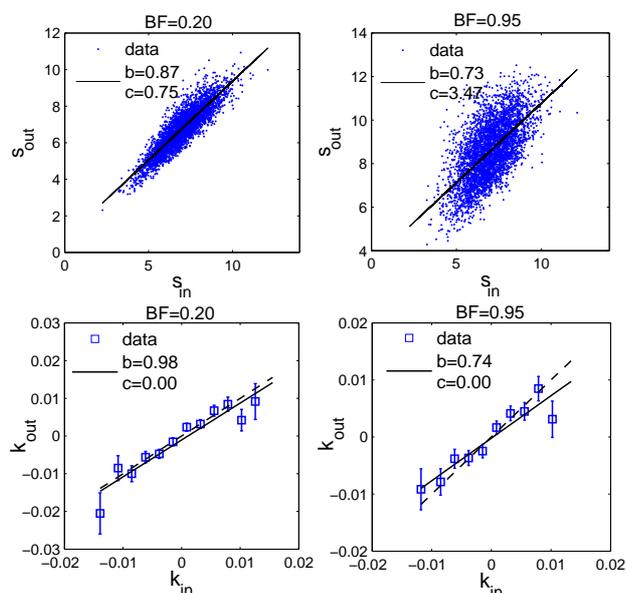


Figure 5.8 - Top panels: Sizes estimates vs input sizes for different bulge fractions in the range $[0.2, 0.95]$. Bulge+Disk galaxies with different ellipticities are used, with $S/N=20$ and negligible effect of the PSF (Set3). Slope and intercept of the fitting are shown (b and c , respectively). Bottom panels: Bulge+Disk elliptical galaxies with S/N of 20 and negligible effect of the PSF (Set 3). Dashed line is $\kappa_{out} = \kappa_{in}$ and the solid line is the fit of the output values. Note that $b=m+1$ of eq. 5.11.

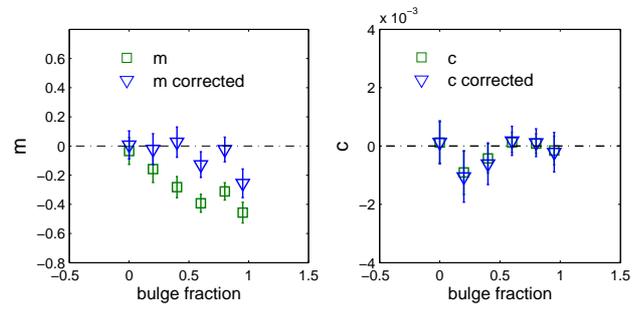


Figure 5.9 - m and c values computed with 200,000 galaxies of Set 3. Values obtained with the sensitivity correction are marked by triangles and by squares are without the correction.

to be enough to do a successful analysis.

5.3.4. Most Realistic Set

The last set includes realistic values for all the effects we investigate. We have generated galaxies with elliptical isophotes with a bulge fraction of 0.5 convolved with an anisotropic PSF with FWHM of 4.5 pixels (1.5 times smaller than the characteristic scale-length of the disk), and again we investigate the dependence on S/N. This is also a challenging test for *lensfit*, the current version (c. 2012) of which uses a simplified parameter set where the bulge scale length is assumed to be half the disk scale length. Here, we include a dispersion in the bulge scale length of 0.6 pixels around a mean value of 3.5 pixels. The analysis was done with 500,000 galaxies, to keep the error bars smaller than 10%. As expected, the size estimation for this set is poorer than in set 1; however, the errors on $\hat{\kappa}$ remain similar thanks to the sensitivity correction (see Fig. 5.10). While the theoretical κ in the studied range of ℓ has a maximum amplitude of $\sim 2\%$, non-linear density evolution increases its contribution on smaller scales (see for example Fig.17 of Bartelmann et al., 1996). For this set, the analysis was performed with higher values of κ to confirm that the method is valid for larger κ . A comparison of the original range with a larger range ($|\kappa| \lesssim 0.05$) is shown in Fig. 5.10, where no significant differences are apparent. Fig. 5.11 shows the values of m and c for this set, with and without the sensitivity correction. If we compare it with the previous plots we can see that as the galaxy population becomes more realistic, including several effects, the importance of the correction increases. Results for this set are shown in Fig. 5.11, showing unbiased results except for S/N=10, which has $m = -0.19 \pm 0.1$. For the higher S/N points, we find $|m| < 0.06$ with errorbars of ± 0.09 .

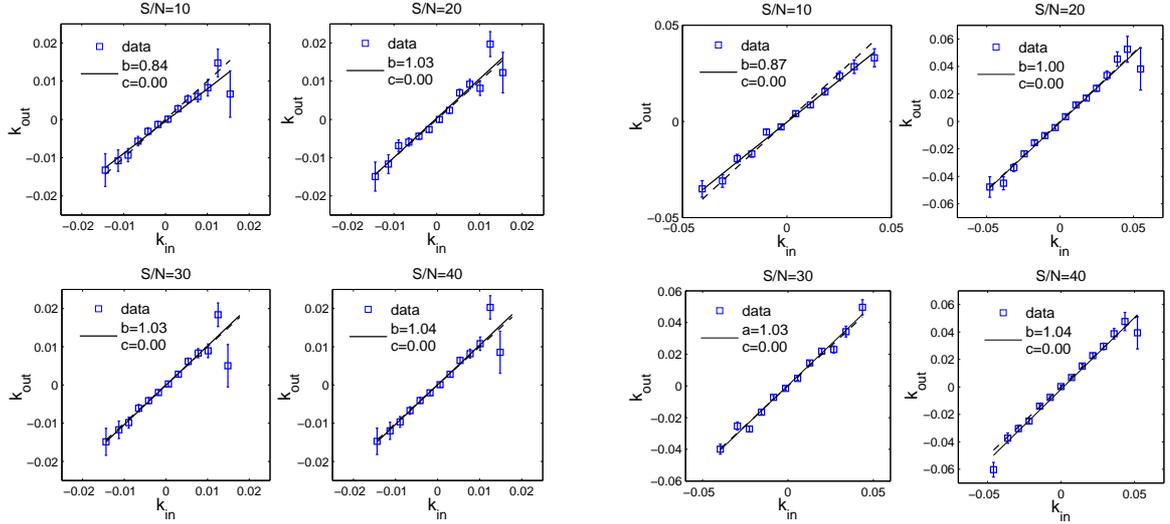


Figure 5.10 - $\hat{\kappa}$ estimates compared to the true κ values for set 4. Dashed line is for $\kappa_{in} = \kappa_{out}$ and solid line is the least squares fit, with regression coefficients shown in the plot ($\hat{\kappa} = b\kappa + c$, where $b = m + 1$). Left four plots are for a maximum value of input κ of $\sim 2\%$ and right set of plots for $|\kappa_{in}| \lesssim 0.05$.

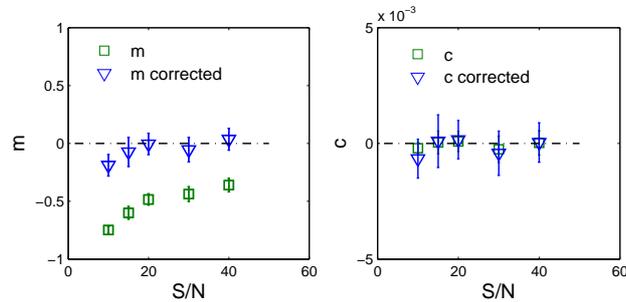


Figure 5.11 - m and c parameters for 500,000 galaxies of Set 4. Squares are raw m and c values; triangles have the sensitivity correction included.

CHAPTER 6

Conclusions

In this chapter the conclusions of the works presented in this thesis are described. High resolution cosmological data require new efficient tools in terms of accuracy and computational demand. For that, new tools are tested in this thesis, finding that fast tools, such as the HEALPix Wavelet (HW), can provide useful constraints on the f_{NL} parameter. We find that artificial intelligence techniques, in particular neural networks, can achieve a reduction of the computational time without losing accuracy. We explored their power in non-Gaussianity analysis, but they may well have wider applications in cosmology. In the weak lensing study, we find that galaxy size information in Euclid conditions is sufficiently good to estimate the convergence field, and will be a very valuable complement to the shear analysis. In the following sections more detailed conclusions for each chapter are summarised.

6.1. Wilkinson Microwave Anisotropy Probe 7-yr constraints on f_{NL} with a fast wavelet estimator

In chapter 2 a new methodology to analyse the Gaussianity of the CMB and to constrain the f_{NL} parameter using the HW has been presented. To our knowledge, the developed f_{NL} estimator is the fastest method that has been proposed up to date. In particular, for WMAP resolution ($N_{\text{side}} = 512$), it is $\sim 10^2$ times faster than the SMHW, $\sim 10^3$ times faster than the KSW bispectrum and 10^7 times faster than the general bispectrum estimator. Moreover, although the constraints imposed by our method are not as stringent as those of the optimal estimators (based on the bispectrum or on

the SMHW), they are very similar or even better than those proposed by alternative methods, such as needlets, Minkowski functionals or the N-pdf.

The method, which is based on the calculation of the third-order moments of the wavelet coefficient maps, has been applied to the WMAP-7yr V+W combined map. On the one hand, we have performed a χ^2 test to study the Gaussianity of the CMB, finding consistency with the Gaussian hypothesis. On the other hand, we have constrained the value of the local f_{NL} parameter to be $-69 < f_{\text{NL}} < 65$ at the 95 per cent confidence level, after correcting for the point source contribution. In addition, the HW gives the possibility of performing local studies of Gaussianity in the CMB map. In particular, we have analysed two independent hemispheres associated to the dipolar modulation proposed by Hoftuft et al. (2009). In this study, we do not find any significant asymmetry on the f_{NL} estimates for the two hemispheres of the WMAP data. The constraints for the northern and southern hemispheres are $-73 < f_{\text{NL}} < 119$ and $-137 < f_{\text{NL}} < 62$, respectively, at the 95 per cent confidence level.

6.2. Constraints on f_{NL} from Wilkinson Anisotropy Probe 7-year data using a neural network classifier

In chapter 3 an application of neural networks in non-Gaussianity analysis was presented. In particular a multi-class neural network classifier with third-order moments of the HW and SMHW coefficients of non-Gaussian realizations was used in order to set constraints on the local non-linear coupling term f_{NL} using WMAP 7-year data. We found that with a very simple network and with few iterations (requiring just a few secs CPU time) it is possible to produce the same results as those obtained with the weighted least squares method. This is an interesting achievement, as it bypasses any covariance matrix related computations and their associated regularisation problems. The estimation of the covariance matrix for each wavelet requires the analysis of at least 10000 Gaussian simulations which involves a huge demand in CPU time, in particular for the SMHW. The error bars on the estimation of f_{NL} computed with Gaussian simulations are $\sigma(\hat{f}_{\text{NL}}) = 33$ for HW and $\sigma(\hat{f}_{\text{NL}}) = 22$ for SMHW, which are very similar to the ones obtained in Casaponsa et al. (2011) and Curto et al. (2011) using the same statistics but a different estimator based on the weighted least squares method ($\sigma = 34$, $\sigma = 21$ for HW and SMHW respectively). The constraints for WMAP 7-year data were found to be $-78 < f_{\text{NL}} < 51$ for the HW and $-24 < f_{\text{NL}} < 61$ for

the SMHW, which are compatible with a Gaussian distribution as found in Smith et al. (2009); Curto et al. (2009b); Komatsu et al. (2011); Casaponsa et al. (2011) and Curto et al. (2011b). The results obtained with the SMHW statistics are similar to the ones found in Smith et al. (2009) and Komatsu et al. (2011), which are the most stringent ones currently available at the limit of the WMAP resolution.

6.3. Exploring local f_{NL} estimators based on the binned bispectrum

In chapter 3 an application of neural networks in non-Gaussianity analysis was presented. In particular a multi-class neural network classifier with Chapter 4 is also devoted to the use of neural networks to construct non-Gaussianity estimators. Another type of neural network was used, and also different inputs were explored. The requirement of the linear term in the estimator was studied finding that in some cases is crucial while in others its effect is less important. In this work a regression network is trained with the binned bispectrum components of non-Gaussian realizations in order to obtain constraints on the local non-linear coupling parameter f_{NL} . The results have been compared with those obtained with a maximum-likelihood estimator, using either a diagonal or a full covariance matrix. The effect of the addition of the linear term, of performing a mean subtraction and the use of inpainting, is also studied.

We find that the three estimators become close to optimal if the linear term is subtracted and inpainting is performed. We find that the linear term is absolutely necessary if a diagonal covariance matrix is used. However, its effect is very small if the full covariance matrix or the neural network is used and the mean is subtracted from the binned maps, as found for wavelets and needlets in Donzelli et al. (2012) and Curto et al. (2012). In that sense, the choice of the estimator depends on the difficulty of computing the linear term. Although the best results for all estimators are obtained when inpainted maps are used, the largest effect of this technique is seen when we use the approximated maximum likelihood estimator with only diagonal terms of the covariance matrix (AMLED), whereas the other two estimators presented in Chapter 4, the approximate maximum likelihood estimator (AMLE) and the neural network estimator (NNE), are less affected by the presence of a mask. Thus, the most robust tools are the AMLE and the NNE estimators, with the NNE displaying a clear computational

advantage, since the covariance matrix does not need to be estimated or inverted; this reduces significantly the number of simulations required. Another advantage of the neural network estimator arises from the fact that for χ^2 minimization the dependence of the covariance matrix on f_{NL} makes a full solution computationally hard, if not unfeasible, for certain problems. Conversely, the NNE bypasses such calculations, thereby simplifying the analysis.

6.4. Size magnification as a complement to cosmic shear

In chapter 5 the first systematic investigation of the performance of a weak lensing shape measurement method's ability to estimate the magnification effect through an estimate of observed galaxy sizes is presented. This test is performed by creating a suite of simulations, with known input values, and by using the most advanced shape measurement available at the current time, *lensfit*.

A full study of the magnification effect using sizes was performed testing the dependence on S/N, PSF size and type of galaxy. The requirements on biases on shear (or equivalently convergence) for Euclid such that systematics do not dominate the very small statistical errors in cosmological parameters are stringent (see Massey et al., 2012 and Cropper et al., 2012). A much larger study will be required to determine whether these requirements can be met for size, but we find no evidence for additive size biases at all, and no evidence for multiplicative bias provided that 1) the PSF is small enough ($< \text{galaxy scale-length}/1.5$), 2) the S/N high enough (≥ 15), and 3) the bulge not too dominant (bulge/disk ratio ≤ 4).

The analysis presented in chapter 5 has assumed that the statistical distribution of galaxy sizes is known, whereas in practice the size distribution depends on galaxy brightness and must be determined from observation. Gravitational lensing of a galaxy with amplification A increases both the integrated flux and area of that galaxy by A , which has the effect of moving galaxies along a locus of slope 0.5 in the relation between $\log(\text{size})$ and $\log(\text{flux})$. Thus, if the intrinsic distribution of sizes r of galaxies scales with flux S as $r \propto S^\beta$, the apparent shift in size caused by lensing amplification A is $r' \propto A^{0.5-\beta}$, resulting in a dilution of the signal compared with the idealised case investigated in this work. A similar effect occurs in galaxy number magnification, where the observed enhancement in galaxy number density N' varies as $N' \propto A^{\alpha-1}$, if the intrinsic number density of galaxies varies as $N \propto S^{-\alpha}$ (Broadhurst et al., 1995).

The value of β at faint magnitudes has recently been estimated by Miller et al. (2012), who analysed the fits to galaxies with $i \lesssim 25$ of Simard et al. (2002) and estimated $\beta \simeq 0.29$. Thus we expect this effect in a real survey to dilute the lensing magnification signal by a factor 0.42, but still allowing detection of lensing magnification. In practice, the dilution factor could be evaluated by fitting to the size-flux relation in the lensing survey.

Lensing number magnification surveys are also affected by the problem that varying Galactic or extragalactic extinction reduces the flux of galaxies and thus may cause a spurious signal (e.g. Ménard et al., 2010). Such extinction would also affect the size magnification of galaxies, but with a different sign in its effect. Thus a combination of lensing number magnification and size magnification might be very effective at removing the effects of extinction from magnification analyses.

Space-based surveys as Euclid should overcome the limitations that we have exposed in chapter 5, having a large number of galaxies, with $S/N > 10$, and importantly a PSF at least 1.5 smaller than the average disk size. The addition of the size information to the ellipticity analysis is expected to reduce the uncertainties in the estimation of weak lensing signal, and therefore improve the constraints of the distribution of matter and dark energy properties.

CHAPTER 7

Resumen en castellano

7.1. Introducción

Actualmente, la avanzada tecnología y la unión del esfuerzo de varios países, han contribuido a tener mapas del Universo observable muy precisos, que han permitido establecer las bases a un modelo cosmológico aceptado por la mayoría de la comunidad científica. Tener un modelo, nos permite probarlo o refutarlo, nos permite avanzar, explicar las diferentes etapas del Universo: como era antes y prever como será.

Durante el siglo XX se fue consolidando lo que se ha llamado modelo cosmológico estándar. Uno de los conceptos básicos del modelo es que el Universo está en expansión, propiedad que fue descubierta por Edwin Hubble al observar que las galaxias lejanas se alejaban de nosotros a más velocidad que las más cercanas. Esto implica que el espacio no es estático, que se mueve, cambiando las distancias entre los objetos que reposan en él. En el modelo cosmológico estándar la dinámica del espacio-tiempo está descrita por la Relatividad General, teoría propuesta por Einstein que ha sido validada por multitud de observaciones.

Un Universo en expansión pierde densidad y se va enfriando, entonces es lógico pensar, que en un principio era más caliente, con densidades más altas. De aquí surge la idea del Big Bang, si miráramos suficientemente atrás en el tiempo, el Universo estaría formado por un plasma de partículas elementales a muy alta densidad. Científicos de mitad del siglo XX dedujeron que el Universo se comportaría como un cuerpo negro a cierta temperatura, lo que implicaría que aún ahora recibiríamos esa radiación, a una longitud de onda mayor, debido a la expansión del Universo. Esa radiación con

espectro de cuerpo negro se detectó en los años 60, y es conocida como el fondo cósmico de microondas, confirmando ese estado caliente y denso en el Universo primitivo.

Para solventar algunos problemas de la teoría del Big Bang, se propone una época inflacionaria, donde el Universo se expandió exponencialmente. Aunque que el Universo pasara por una fase de este tipo es compatible con las observaciones, cómo se produjo está aún por esclarecer; de hecho, hay varios modelos propuestos y no es fácil comprobarlos o refutarlos. Uno de los objetivos de esta tesis es descartar algunos de estos modelos usando los datos del fondo cósmico de microondas (detalles en los capítulos 2, 3 y 4).

El modelo cosmológico estándar además tiene otros dos importantes frentes abiertos, la materia oscura y la energía oscura. Observaciones de efectos gravitatorios indican claramente que hay más materia de la que somos capaces de detectar, es decir que no emiten luz o es extremadamente débil. Con los componentes conocidos del Universo, desde partículas elementales hasta objetos astronómicos, no somos capaces de explicar la cantidad de materia oscura necesaria para generar los efectos gravitatorios que observamos. Lo que sí somos capaces de saber es cuál es la contribución de esta materia al total de energía del Universo.

La necesidad de incluir la energía oscura en el modelo cosmológico estándar aparece cuando Riess et al. (1998) y Perlmutter et al. (1999), utilizando datos de supernovas tipo Ia, encuentran que la expansión del Universo está acelerando. Para poder explicarlo se necesita un campo que genere esta aceleración, siendo esto aún un tema abierto en el que se está poniendo mucho esfuerzo para clarificarlo.

El último trabajo presentado en esta tesis (capítulo 5) está dedicado al estudio del efecto lente gravitatoria débil. Este efecto está relacionado con la materia oscura y la energía oscura, y aunque no puede determinar la naturaleza de estas dos contribuciones, si puede determinar su distribución en el espacio y en distintas etapas del Universo.

7.1.1. Fondo cósmico de microondas

El estudio de la radiación del fondo cósmico de microondas ha supuesto un importante avance para la comprensión de los distintos estados del Universo. Su detección a mediados del siglo XX supuso la aceptación de la Teoría del Big Bang por la gran

mayoría de la comunidad científica, desplazando otras teorías más estáticas del Universo. Esta radiación es isótropa, pero si se mira con detalle tiene diferencias de 10^{-5} K entre distintas direcciones, por lo que se trabaja con mapas de anisotropías. La distribución de estas anisotropías, y en particular la función de correlación a dos puntos, permite poner cotas en los parámetros del modelo cosmológico (cantidad de materia oscura, cantidad de energía oscura, ratio de expansión del Universo, etc.). Además de darnos información sobre el Universo primitivo, los fotones del fondo cósmico de microondas han atravesado distintas etapas del Universo hasta llegar a nosotros, por tanto tienen información de ese viaje. Estas desviaciones de su distribución energética y espacial inicial son lo que se denominan anisotropías secundarias.

7.1.2. No-Gaussianidad primordial

Usando las correlaciones a dos puntos de las anisotropías del fondo cósmico de microondas se obtienen los parámetros del modelo cosmológico. Si la distribución de estas anisotropías es Gaussiana, toda la información está contenida en el momento de orden dos. Pero diversos procesos del Universo primitivo, podrían introducir pequeñas desviaciones que darían una distribución no-Gaussiana. Esto es predicho por algunos modelos de inflación donde, por ejemplo, hay más de un campo involucrado. En dichos modelos se puede separar el potencial gravitatorio en dos términos uno Gaussiano y otro no-Gaussiano, y la amplitud de esa segunda parte vendría parametrizada por el parámetro f_{NL} . Analizando los momentos de tercer orden de las anisotropías del fondo cósmico de microondas se pueden obtener cotas a este parámetro. Teóricamente se puede calcular cuál sería la contribución de f_{NL} para distintos modelos de inflación, y analizando los datos podemos descartar modelos de inflación con un solo parámetro.

Tres capítulos de esta tesis están dedicados a métodos implementados para poner cotas al parámetro f_{NL} . Dichos métodos se han aplicado a los datos del satélite WMAP, siendo los resultados publicados en la revista científica de alto impacto MNRAS. Los detalles de cada método están explicados en los capítulos 2, 3 y 4 de esta tesis y los aspectos más importantes y las conclusiones están resumidos en la sección 7.2.

7.1.3. Efecto lente gravitatoria débil

Un límite importante en los estudios de la radiación del fondo cósmico de microondas, es que no hay información a distintas distancias, los fotones recibidos provienen

aproximadamente del mismo momento de emisión. Para estudiar distintas etapas del Universo podemos observar la distribución de galaxias a distintas distancias. Debido a que la luz tiene una velocidad finita, a distancias mayores observamos estados del Universo más primitivos.

La teoría de la relatividad general predice que el camino que recorre la luz de una galaxia lejana es distorsionado por los pozos de potencial que encuentra en la línea de visión. Este efecto es conocido como efecto lente gravitatoria, y en ocasiones, cuando la distorsión es producida por un pozo de potencial intenso, pueden llegar a generarse múltiples imágenes de la galaxia de fondo. Estudiando las características de las imágenes se puede deducir la distribución de masa necesaria para generalas. El efecto lente gravitatoria ha sido crucial para la confirmación de la existencia de materia oscura, y es muy útil para trazar la distribución de materia (sin distinción de su naturaleza) del Universo. Cuando el potencial gravitatorio es menor, el efecto no se puede detectar en una sola galaxia y hablamos de efecto lente débil. Aunque el tamaño y la elipticidad de la galaxia observada son distintos a los propios de la galaxia, con una sola galaxia no se puede discernir si son sus propiedades intrínsecas o han sido modificadas por el efecto lente gravitatoria. Lo que se hace es estudiar un gran número de galaxias y ver si hay patrones o correlaciones que puedan indicar que ha habido efecto lente.

7.2. Métodos para acotar al parámetro f_{NL}

Los capítulos 2, 3 y 4 de esta tesis están dedicados al desarrollo de métodos para estimar el valor de f_{NL} y sus barras de error en datos del fondo cósmico de microondas. Se ha trabajado con varios estimadores de la función de correlación a tres puntos. En particular, en el capítulo 2 se trabaja con los momentos de tercer orden construidos con una *wavelet* u ondícula con base en el espacio real. Cuando se trabaja con las anisotropías del fondo cósmico de microondas, se acostumbra a transformar la señal en armónicos esféricos, y aunque es ventajoso en varios aspectos, como proporcionar información de la escala de las anisotropías, es un proceso costoso computacionalmente. En los capítulos 3 y 4 se explora el uso de redes neuronales en análisis de no-Gaussianidad, que pueden suponer una ventaja para evitar estimar e invertir matrices grandes.

7.2.1. Cotas en el parámetro f_{NL} usando la HW en datos de WMAP-7yr

Para evitar las transformaciones al espacio de armónicos esféricos, se propone una herramienta que descompone la señal en distintas escalas, manteniendo la localización del espacio real. Además esta descomposición se ajusta al *software* específico que se usa para analizar los mapas del fondo cósmico de microondas HEALPix (Górski et al., 2005), de aquí el nombre HEALPix wavelet (HW). Se descompone el mapa inicial en mapas a distintas escalas y se construyen momentos de tercer orden combinando todas las escalas. Este proceso resulta ser varios factores de magnitud más rápido que los métodos convencionales. Las cotas en el parámetro f_{NL} que se obtienen para los datos de WMAP-7yr son $-69 < f_{\text{NL}} < 65$ al 95% de nivel de confianza. Aunque su sensibilidad al parámetro es menor que la de los métodos óptimos, supera la de otros métodos presentados en la literatura como los funcionales de Minkowsky o usando la Npdf.

7.2.2. Cotas en el parámetro f_{NL} en los datos de WMAP-7yr con una red neuronal de clasificación

En el capítulo 3 se presenta una red neuronal aplicada a análisis de no-Gaussianidad del fondo cósmico de microondas. En este trabajo se pretende estudiar las ventajas y desventajas de utilizar esta herramienta matemática de inteligencia artificial. En una red neuronal los datos de salida se relacionan con los datos de entrada mediante una combinación lineal de diversas funciones de activación (representadas por nodos o neuronas). Son herramientas que tienen un proceso de aprendizaje, por lo que se necesitan datos simulados para entrenar la red. La red neuronal tiene varios parámetros libres asociados a cada neurona, los pesos w que indican la fuerza de los enlaces entre neuronas y un sesgo b característico de cada nodo. Al proceso de encontrar estos parámetros libres se le llama entrenamiento, que básicamente es minimizar la distancia entre los datos de salida de la red y los resultados simulados.

En este primer trabajo con redes neuronales, se construye una red neuronal para clasificar un mapa según su nivel de no-Gaussianidad. Se utilizan los estadísticos cúbicos construidos en el capítulo 2 y se observa que con la red neuronal se obtienen resultados muy parecidos al método clásico. Las cotas en el parámetro f_{NL} que se obtienen con la red neuronal son $-78 < f_{\text{NL}} < 51$. La ventaja de la red neuronal es que

se necesitan menos simulaciones ya que no se necesita estimar la matriz de covarianza necesaria con el método convencional. Para el caso de la HW, reducir el tiempo de computación no es fundamental ya que la herramienta es muy rápida. Por eso se construye otra red para aplicarla con los estadísticos de la SMHW (definida en la sección 1.4.2) que es una herramienta más lenta. El interés de trabajar con la SMHW, es también porque el estimador de f_{NL} se ha comprobado que es óptimo. Se comprueba que usando la red neuronal de clasificación el estimador de la SMHW sigue siendo óptimo, reduciendo un cinco veces el tiempo de computación.

7.2.3. Explorando f_{NL} con estimadores basados en el bispectro bineado

En el capítulo 4, se usa una red neuronal de regresión, donde en vez de clasificar objetos, lo que se hace es estimar directamente el valor de f_{NL} . En este caso, los datos de entrada son el bispectro bineado del mapa, estimador que está descrito en la sección 1.4.1. Como hemos dicho, la ventaja de las redes neuronales es que se reduce el número de simulaciones necesarias y por tanto el tiempo de computación. Esto también ocurre si suponemos que los términos fuera de la diagonal de la matriz de covarianza son despreciables, aunque esta suposición solo es válida si no hay correlaciones entre estadísticos. Cuando se trabaja con los datos aparecen correlaciones debido al ruido anisótropo y al enmascarar el mapa. En Creminelli et al. 2006 se propone añadir un término lineal al estimador cúbico que tendría en cuenta estas correlaciones (detalles en la sección 1.4.3. La introducción de este término simplifica el estimador, pero es más sensible al procesado de los datos. En este trabajo se construyen varios estimadores de f_{NL} : dos basados en el estimador de máxima verosimilitud (*likelihood* en inglés), uno teniendo en cuenta los términos fuera de la diagonal de la matriz de covarianza (EMV) y otro sin tenerlos en cuenta (EMVD), y un tercer estimador basado en redes neuronales (ENN). Se estudia la necesidad de hacer un *inpainting* a los datos y el efecto del término lineal en los tres estimadores. En esta comparación se concluye que los estimadores más efectivos computacionalmente son el ENN y el EMVD, pero el segundo es más dependiente del pre-procesado de los datos (*inpainting*) que el primero. Aunque añadir el término lineal mejora los resultados para todos los estimadores, es imprescindible para que el estimador EMVD sea competitivo, mientras que para los otros dos (EMV y ENN) los resultados son cercanos al óptimo.

7.3. Efecto lente gravitatoria débil usando tamaños de las galaxias

El efecto lente gravitatoria débil, en principio, se puede estudiar con las correlaciones de la elipticidad de las galaxias o con las correlaciones entre tamaños. En general, se estudia la elipticidad, porque la señal-ruido es mayor que con los tamaños. Además se creía que la distribución intrínseca de la elipticidad de las galaxias era aleatoria y por tanto de media cero. En realidad no es así, hay ciertas correlaciones intrínsecas, que deben estudiarse y substraer a la señal.

En el proceso de estimar la elipticidad de una galaxia, se necesita estimar el tamaño (el radio o el semi-eje mayor a cierto flujo). Aunque teóricamente, la información de los tamaños mejora las predicciones en cotas de los parámetros cosmológicos relacionados con la materia y la energía oscura, no se había hecho un análisis exhaustivo de cuáles son las características necesarias, para que esta información “gratis” sea útil. Esto es lo que se pretende responder en el capítulo 5.

En este trabajo se estiman los tamaños de 200,000 galaxias simuladas con distintas características y se les aplica un campo de convergencia κ . Para estimar los tamaños se ha utilizado el código más preciso hasta el momento que se ha usado para estudios de efecto lente gravitatoria, *lensfit* (Miller et al., 2007; Kitching et al., 2008). Se construye un estimador del campo de convergencia a partir de los tamaños estimados y se compara con el campo κ aplicado a las galaxias simuladas. El análisis se realiza para estudiar la dependencia de la estimación del tamaño de una galaxia con la *point spread function* (PSF), el tipo de galaxias modificando el ratio entre bulbo y disco, y con el ratio de señal-ruido de la galaxia. Las conclusiones que se derivan del trabajo son que con las características de un satélite como Euclid se podrían usar los tamaños de las galaxias para obtener información cosmológica. Es un campo nuevo, y aún faltan por estudiar con rigurosidad las correlaciones intrínsecas entre tamaños y su dependencia con el flujo, pero los resultados son alentadores.

Bibliography

- Abrial P., Moudden Y., Starck J.-L., Fadili J., Delabrouille J., Nguyen M. K., 2008, *Statistical Methodology*, 5, 289
- Agullo I., Parker L., 2011, *Phys. Rev. D*, 83, 063526
- Aliaga A. M., Rubiño-Martín J. A., Martínez-González E., Barreiro R. B., Sanz J. L., 2005, *MNRAS*, 356, 1559
- Alishahiha M., Silverstein E., Tong D., 2004, *Phys. Rev. D*, 70, 123505
- Alpher R. A., Bethe H., Gamow G., 1948, *Physical Review*, 73, 803
- Alpher R. A., Herman R. C., 1948, *Physical Review*, 74, 1737
- Alpher R. A., Herman R. C., 1949, *Physical Review*, 75, 1089
- Amara A., Réfrégier A., 2007, *MNRAS*, 381, 1018
- Amendola L., Appleby S., Bacon D., et al. 2013, *Living Reviews in Relativity*, 16, 6
- André P., Baccigalupi C., Banday A., et al. 2014, *J. Cosmology Astropart. Phys.*, 2, 6
- Antoine J.-P., Vandergheynst P., 1998, *Journal of Mathematical Physics*, 39, 3987
- Arkani-Hamed N., Creminelli P., Mukohyama S., Zaldarriaga M., 2004, *J. Cosmology Astropart. Phys.*, 4, 1
- Ashtekar A., 2009, *General Relativity and Gravitation*, 41, 707
- ATLAS Collaboration 2013, [arXiv:1307.1427](https://arxiv.org/abs/1307.1427)
- Atwood W. B., Abdo A. A., Ackermann M., et al. 2009, *ApJ*, 697, 1071
- Auld T., Bridges M., Hobson M. P., 2008, *MNRAS*, 387, 1575

- Auld T., Bridges M., Hobson M. P., Gull S. F., 2007, *MNRAS*, 376, L11
- Babich D., 2005, *Phys. Rev. D*, 72, 043003
- Babich D., Creminelli P., Zaldarriaga M., 2004, *J. Cosmology Astropart. Phys.*, 8, 9
- Baccigalupi C., Bedini L., Burigana C., De Zotti G., Farusi A., Maino D., Maris M., Perrotta F., Salerno E., Toffolatti L., Tonazzini A., 2000, *MNRAS*, 318, 769
- Bajkova A. T., 2005, *Astronomy Reports*, 49, 947
- Ball N. M., Loveday J., Fukugita M., Nakamura O., Okamura S., Brinkmann J., Brunner R. J., 2004, *MNRAS*, 348, 1038
- Barber A. J., Taylor A. N., 2003, *MNRAS*, 344, 789
- Barreiro R. B., 2010, in Diego J. M., Goicoechea L. J., González-Serrano J. I., Gorgas J., eds, *Highlights of Spanish Astrophysics V An Overview of the Current Status of CMB Observations*. p. 93
- Barreiro R. B., Hobson M. P., Lasenby A. N., Banday A. J., Górski K. M., Hinshaw G., 2000, *MNRAS*, 318, 475
- Bartelmann M., Narayan R., Seitz S., Schneider P., 1996, *ApJ*, 464, L115
- Bartolo N., Komatsu E., Matarrese S., Riotto A., 2004, *Phys. Rep.*, 402, 103
- Bassett B. A., Tsujikawa S., Wands D., 2006, *Reviews of Modern Physics*, 78, 537
- Baumann D., 2009, (arXiv:0907.5424)
- Baumann D., Dymarsky A., Klebanov I. R., McAllister L., 2008, *J. Cosmology Astropart. Phys.*, 1, 24
- Benítez N., Gaztañaga E., Miquel R., et al. 2009, *ApJ*, 691, 241
- Benítez N., Sanz J. L., 1999, *ApJ*, 525, L1
- Benjamin J., Heymans C., Semboloni E., van Waerbeke L., Hoekstra H., Erben T., Gladders M. D., Hettterscheidt M., Mellier Y., Yee H. K. C., 2007, *MNRAS*, 381, 702
- Bennett C. L., Halpern M., Hinshaw G., et al. 1997, in *American Astronomical Society Meeting Abstracts Vol. 29 of Bulletin of the American Astronomical Society, The Microwave Anisotropy Probe (MAP) Mission*. p. 1353

- Bennett C. L., Hill R. S., Hinshaw G., et al. 2003, *ApJS*, 148, 97
- Bennett C. L., Larson D., Weiland J. L., et al. 2013, *ApJS*, 208, 20
- Benoît A., Ade P., Amblard A., et al. 2003, *A&A*, 399, L25
- Bernardeau F., Bonvin C., Van de Rijdt N., Vernizzi F., 2012, *Phys. Rev. D*, 86, 023001
- Bernstein G. M., Jarvis M., 2002, *AJ*, 123, 583
- Bertalmio M., Caselles V., Masnou S., Sapiro G., 2000, <http://math.univ-lyon1.fr/masnou/fichiers/publications/survey.pdf>, p. 259
- Bertin G., Lombardi M., 2006, *ApJ*, 648, L17
- BICEP2 Collaboration 2014, arXiv:1403.3985
- Bondi H., Gold T., 1948, *MNRAS*, 108, 252
- Bridle S., Balan S. T., Bethge M., Gentile M., Harmeling S., Heymans C., Hirsch M., Hosseini R., Jarvis e. a., 2010, *MNRAS*, 405, 2044
- Broadhurst T., Benítez N., Coe D., Sharon K., Zekser K., White R., Ford H., Bouwens R., Blakeslee J., Clampin M., Cross N. e. a., 2005, *ApJ*, 621, 53
- Broadhurst T. J., Taylor A. N., Peacock J. A., 1995, *ApJ*, 438, 49
- Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002, *MNRAS*, 333, 501
- Buchbinder E. I., Khoury J., Ovrut B. A., 2007, *Phys. Rev. D*, 76, 123503
- Bucher M., van Tent B., Carvalho C. S., 2010, *MNRAS*, 407, 2193
- Cabella P., Pietrobon D., Veneziani M., Balbi A., Crittenden R., de Gasperis G., Quercellini C., Vittorio N., 2010, *MNRAS*, 405, 961
- Carballo R., González-Serrano J. I., Benn C. R., Jiménez-Luján F., 2008, *MNRAS*, 391, 369
- Carlstrom J. E., Ade P. A. R., Aird K. A., et al. 2011, *PASP*, 123, 568
- Casaponsa B., Barreiro R. B., Curto A., Martínez-González E., Vielva P., 2011, *MNRAS*

- Casaponsa B., Barreiro R. B., Martínez-González E., Curto A., Bridges M., Hobson M. P., 2013, MNRAS
- Casaponsa B., Bridges M., Curto A., Barreiro R. B., Hobson M. P., Martínez-González E., 2011, MNRAS
- Casaponsa B., Heavens A. F., Kitching T. D., Miller L., Barreiro R. B., Martínez-González E., 2013, MNRAS
- Catelan P., Porciani C., 2001, MNRAS, 323, 713
- Cayón L., Martínez-González E., Argüeso F., Banday A. J., Górski K. M., 2003, MNRAS, 339, 1189
- Cayón L., Sanz J. L., Martínez-González E., Banday A. J., Argüeso F., Gallegos J. E., Górski K. M., Hinshaw G., 2001, MNRAS, 326, 1243
- Chen X., Firouzjahi H., Namjoo M. H., Sasaki M., 2013, EPL (Europhysics Letters), 102, 59001
- Cheng B., Titterton D. M., 1994, Statistical Science, 9, 2
- Cimatti A., Nipoti C., Cassata P., 2012, MNRAS, 422, L62
- CMS Collaboration Chatrchyan 2012, Physics Letters B, 716, 30
- Cooper M. C., Griffith R. L., Newman J. A., Coil A. L., Davis M., Dutton A. A., Faber S. M., Guhathakurta P., Koo D. C., Lotz J. M., Weiner B. J., Willmer C. N. A., Yan R., 2012, MNRAS, 419, 3018
- Copeland E. J., Sami M., Tsujikawa S., 2006, International Journal of Modern Physics D, 15, 1753
- Creminelli P., Nicolis A., Senatore L., Tegmark M., Zaldarriaga M., 2006, J. Cosmology Astropart. Phys., 5, 4
- Creminelli P., Senatore L., 2007, J. Cosmology Astropart. Phys., 11, 10
- Creminelli P., Zaldarriaga M., 2004, J. Cosmology Astropart. Phys., 10, 6
- Crittenden R. G., Natarajan P., Pen U.-L., Theuns T., 2001, ApJ, 559, 552
- Cruz M., Martínez-González E., Vielva P., Cayón L., 2005, MNRAS, 356, 29

- Curto A., Aumont J., Macías-Pérez J. F., Martínez-González E., Barreiro R. B., Santos D., Désert F. X., Tristram M., 2007, *A&A*, 474, 23
- Curto A., Martínez-González E., Barreiro R. B., 2009a, *ApJ*, 706, 399
- Curto A., Martínez-González E., Barreiro R. B., 2009b, *ApJ*, 706, 399
- Curto A., Martínez-González E., Barreiro R. B., 2011, *MNRAS*, 412, 1038
- Curto A., Martínez-González E., Barreiro R. B., 2012, *MNRAS*, 426, 1361
- Curto A., Martínez-González E., Mukherjee P., Barreiro R. B., Hansen F. K., Liguori M., Matarrese S., 2009, *MNRAS*, 393, 615
- de Bernardis P., Ade P. A. R., Bock J. J., et al. 2000, *Nature*, 404, 955
- de Zotti G., Ricci R., Mesa D., Silva L., Mazzotta P., Toffolatti L., González-Nuevo J., 2005, *A&A*, 431, 893
- Delabrouille J., Cardoso J.-F., 2009, in Martínez V. J., Saar E., Martínez-González E., Pons-Bordería M.-J., eds, *Data Analysis in Cosmology Vol. 665 of Lecture Notes in Physics*, Berlin Springer Verlag, *Diffuse Source Separation in CMB Observations*. pp 159–205
- Dicke R. H., Peebles P. J. E., Roll P. G., Wilkinson D. T., 1965, *ApJ*, 142, 414
- Diego J. M., Sandvik H. B., Protopapas P., Tegmark M., Benítez N., Broadhurst T., 2005, *MNRAS*, 362, 1247
- Djorgovski S., Davis M., 1987, *ApJ*, 313, 59
- Dodelson S., 2003, *Modern cosmology*. Amsterdam (Netherlands): Academic Press. 2003
- Dodelson S., Kinney W. H., Kolb E. W., 1997, *Phys. Rev. D*, 56, 3207
- Donzelli S., Hansen F. K., Liguori M., Marinucci D., Matarrese S., 2012, *ApJ*, 755, 19
- Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, *ApJ*, 313, 42
- Dvali G., Gruzinov A., Zaldarriaga M., 2004a, *Phys. Rev. D*, 69, 083505
- Dvali G., Gruzinov A., Zaldarriaga M., 2004b, *Phys. Rev. D*, 69, 023505

- Dvali G., Tye S.-H. H., 1999, *Physics Letters B*, 450, 72
- Elsner F., Wandelt B. D., 2009, *ApJS*, 184, 264
- Elsner F., Wandelt B. D., 2010, *ApJ*, 724, 1262
- Farley B. G., Clark W., 1954, *Information Theory, Transactions of the IRE Professional Group on*, 4, 76
- Fergusson J. R., Liguori M., Shellard E. P. S., 2010, *Phys. Rev. D*, 82, 023502
- Fergusson J. R., Shellard E. P. S., 2009, *Phys. Rev. D*, 80, 043510
- Fergusson J. R., Shellard E. P. S., 2011, *ArXiv e-prints*
- Firth A. E., Lahav O., Somerville R. S., 2003, *MNRAS*, 339, 1195
- Fort B., Mellier Y., Dantel-Fort M., 1997, *A&A*, 321, 353
- Fort B., Prieur J. L., Mathez G., Mellier Y., Soucail G., 1988, *A&A*, 200, L17
- Freese K., Frieman J. A., Olinto A. V., 1990, *Physical Review Letters*, 65, 3233
- Gamow G., 1948, *Physical Review*, 74, 505
- Gamow G., 1956, *Vistas in Astronomy*, 2, 1726
- Ganc J., 2011, *Phys. Rev. D*, 84, 063514
- Gangui A., 1994, *Phys. Rev. D*, 50, 3684
- Gold T., 1949, *Nature*, 164, 1006
- Golden R. M., 1996, *Mathematical Methods for Neural Network Analysis and Design*, Massachusetts Institute of Technology, Cambridge, Massachusetts
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
- Gull S. F., Skilling J., 1999, *Quantified maximum entropy: Mem-Sys 5 users' manual*. Maximum Entropy Data Consultants Ltd, Royston
- Guth A. H., 1981, *Phys. Rev. D*, 23, 347
- Guth A. H., Pi S.-Y., 1982, *Physical Review Letters*, 49, 1110

- Halverson N. W., Leitch E. M., Pryke C., et al. 2002, *ApJ*, 568, 38
- Hamada Y., Kawai H., Oda K.-y., 2014, *Progress of Theoretical and Experimental Physics*, 2014, 020003
- Hanany S., Ade P., Balbi A., et al. 2000, *ApJ*, 545, L5
- Hanson D., Hoover S., Crites A., et al. 2013, *Physical Review Letters*, 111, 141301
- Hawking S. W., 1982, *Physics Letters B*, 115, 295
- Heavens A., 2011, arXiv:1109.1121
- Heavens A., Alsing J., Jaffe A. H., 2013, *MNRAS*, 433, L6
- Heavens A., Refregier A., Heymans C., 2000, *MNRAS*, 319, 649
- Heavens A. F., Kitching T. D., Taylor A. N., 2006, *MNRAS*, 373, 105
- Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, *MNRAS*, 347, 895
- Heymans C., Heavens A., 2003, *MNRAS*, 339, 711
- Heymans C., Van Waerbeke L., Bacon D., Berge J., Bernstein G., Bertin E., Bridle S., Brown M. L., et al. 2006, *MNRAS*, 368, 1323
- Hikage C., Matsubara T., Coles P., Liguori M., Hansen F. K., Matarrese S., 2008, *MNRAS*, 389, 1439
- Hildebrandt H., Muzzin A., Erben T., Hoekstra H., Kuijken K., Surace J., van Waerbeke L., Wilson G., Yee H. K. C., 2011, *ApJ*, 733, L30
- Hinshaw G., Larson D., Komatsu E., et al. 2013, *ApJS*, 208, 19
- Hirata C. M., Seljak U., 2004, *Phys. Rev. D*, 70, 063526
- Hobson M. P., Lasenby A. N., 1998, *MNRAS*, 298, 905
- Hoekstra H., Franx M., Kuijken K., Squires G., 1998, *ApJ*, 504, 636
- Hoftuft J., Eriksen H. K., Banday A. J., Górski K. M., Hansen F. K., Lilje P. B., 2009, *ApJ*, 699, 985
- Hoyle F., 1948, *MNRAS*, 108, 372

- Hu W., White M., 1997, *New A*, 2, 323
- Huff E. M., Graves G. J., 2011, arXiv:astro-ph/1111.1070
- Huterer D., 2002, *Phys. Rev. D*, 65, 063001
- Jain B., 2002, *ApJ*, 580, L3
- Jarvis M., Jain B., Bernstein G., Dolney D., 2006, *ApJ*, 644, 71
- Jaynes E. T., Bretthorst G. L., 2003, *Probability Theory and the Logic of Science*.
Cambridge University Press, Cambridge, UK, June 2003
- Jiménez-Teja Y., Benítez N., 2012, *ApJ*, 745, 150
- Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, *A&A*, 527, A26
- Joachimi B., Semboloni E., Bett P. E., Hartlap J., Hilbert S., Hoekstra H., Schneider P., Schrabback T., 2013, *MNRAS*, 436, 819
- Kaiser N., 1992, *ApJ*, 388, 272
- Kaiser N., 2000, *ApJ*, 537, 555
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- Kaiser N., Squires G., Broadhurst T., 1995, *ApJ*, 449, 460
- Keating B. G., Ade P. A. R., Bock J. J., Hivon E., Holzzapfel W. L., Lange A. E., Nguyen H., Yoon K. W., 2003, in Fineschi S., ed., *Polarimetry in Astronomy Vol. 4843 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, BICEP: a large angular scale CMB polarimeter. pp 284–295
- Kermish Z. D., Ade P., Anthony A., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8452 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, The POLARBEAR experiment
- Kinney W. H., 2009, arXiv:0902.1529
- Kitching T., Balan S., Bernstein G., Bethge M., Bridle S., Courbin F., Gentile M., Heavens A., Hirsch M., Hosseini R., Kiessling A., Amara A., Kirk D., Kuijken e. a., 2010, arXiv:astro-ph/1009.0779

- Kitching T. D., Balan S. T., Bridle S., Cantale N., Courbin F., Eifler T., Gentile M., Gill M. S. S., Harmeling S., Heymans e. a., 2012, MNRAS, 423, 3163
- Kitching T. D., Miller L., Heymans C. E., van Waerbeke L., Heavens A. F., 2008, MNRAS, 390, 149
- Kitching T. D., Taylor A. N., Heavens A. F., 2008, MNRAS
- Kolb E. W., 1999, NASA STI/Recon Technical Report N, 2, 14556
- Komatsu E., 2002, arXiv:astro-ph/0206039
- Komatsu E., Afshordi N., Bartolo N., Baumann D. e. a., 2009, Astronomy, 2010, 158
- Komatsu E., et al., 2011, ApJS, 192, 18
- Komatsu E., Smith K. M., Dunkley J., Bennett C. L., Gold B. e. a., 2011, ApJS, 192, 18
- Komatsu E., Spergel D. N., 2001, Phys. Rev. D, 63, 063002
- Komatsu E., Spergel D. N., Wandelt B. D., 2005, ApJ, 634, 14
- Kosowsky A., 2003, New A Rev., 47, 939
- Koyama K., Mizuno S., Vernizzi F., Wands D., 2007, J. Cosmology Astropart. Phys., 11, 24
- Kuijken K., 2006, A&A, 456, 827
- Langlois D., 2008, arXiv:0811.4329
- Larson D., Dunkley J., Hinshaw G., et al. 2011, ApJS, 192, 16
- Lasenby A. N., Hancock S., Jones A. W., 1995, in Occhionero F., ed., Birth of the Universe and Fundamental Physics Vol. 455 of Lecture Notes in Physics, Berlin Springer Verlag, Primary Anisotropy Detections in the Cosmic Microwave Background. p. 221
- Lasenby A. N., Jones A. W., Dabrowski Y., 1998, arXiv:astro-ph/9810175
- Laureijs R., Amiaux J., Arduini S., Auguères J. ., Brinchmann J., Cole R., Cropper M., Dabin C., Duvet L., Ealet A., et al. 2011, arXiv:1110.3193
- Leauthaud A., Massey R., Kneib J.-P., Rhodes J., Johnston D. E., Capak P., Heymans C. e. a., 2007, ApJS, 172, 219

- Lehners J., 2010, *Advances in Astronomy*, 2010
- Leshno M. S. S., 1993, *Neural Netw.*, 6, 861
- Lewis A., Bridle S., 2002, *Phys. Rev.*, D66, 103511
- Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
- Liguori M., Sefusatti E., Fergusson J. R., Shellard E. P. S., 2010, *Advances in Astronomy*, 2010
- Linde A. D., 1982, *Physics Letters B*, 108, 389
- Linde A. D., 1986, *Modern Physics Letters A*, 1, 81
- Lineweaver C. H., 1997, in Valls-Gabaud D., Hendry M. A., Molaro P., Chamcham K., eds, *From Quantum Fluctuations to Cosmological Structures Vol. 126 of Astronomical Society of the Pacific Conference Series, Gold in the Doppler Hills: Cosmological Parameters in the Microwave Background*. p. 185
- Lyth D. H., 2008, in M. Lemoine, J. Martin, & P. Peter ed., *Inflationary Cosmology Vol. 738 of Lecture Notes in Physics*, Berlin Springer Verlag, *Particle Physics Models of Inflation*. pp 81–+
- Lyth D. H., Ungarelli C., Wands D., 2003, *Phys. Rev. D*, 67, 023503
- Mackay D. J. C., 2003, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, Cambridge, UK. Cambridge University Press, Cambridge, UK, October 2003
- Maldacena J., 2003, *Journal of High Energy Physics*, 5, 13
- Maltby D. T., Aragón-Salamanca A., Gray M. E., Barden M., Häußler B., Wolf C., Peng C. Y., Jahnke K., McIntosh D. H., Böhm A., van Kampen E., 2010, *MNRAS*, 402, 282
- Mandelbaum R., Blake C., Bridle S., Abdalla F. B., Brough S., Colless M., Couch e. a., 2011, *MNRAS*, 410, 844
- Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, *MNRAS*, 367, 611

- Marinucci D., Peccati G., 2011, *Random Fields on the Sphere: Applications*, Cambridge University Press, Cambridge, UK.
- Martínez-González E., Gallegos J. E., Argüeso F., Cayón L., Sanz J. L., 2002, MNRAS, 336, 22
- Massey R., Heymans C., Bergé J., Bernstein G., Bridle S., Clowe D., Dahle H., Ellis R., Erben T., Hettterscheidt M., High F. W., Hirata C. e. a., 2007, MNRAS, 376, 13
- Mather J. C., Cheng E. S., Cottingham D. A., et al. 1994, ApJ, 420, 439
- McCulloch W. S., Pitts W., 1943, The bulletin of mathematical bipphysics, 5, 115
- McEwen J. D., Hobson M. P., Lasenby A. N., Mortlock D. J., 2005, MNRAS, 359, 1583
- McEwen J. D., Hobson M. P., Lasenby A. N., Mortlock D. J., 2008, MNRAS, 388, 659
- McEwen J. D., Vielva P., Wiaux Y., Barreiro R. B., Cayón L., Hobson M. P., Lasenby A. N., Martínez-González E., Sanz J. L., 2007, Journal of Fourier Analysis and Applications, 13, 495
- Melchior P., Viola M., 2012, MNRAS, 424, 2757
- Mellier Y., 1999, ARA&A, 37, 127
- Mellier Y., Fort B., Kneib J.-P., 1993, ApJ, 407, 33
- Ménard B., Scranton R., Fukugita M., Richards G., 2010, MNRAS, 405, 1025
- Miller L., e. a., 2012, submitted to MNRAS
- Miller L., Kitching T. D., Heymans C., Heavens A. F., van Waerbeke L., 2007, MNRAS, 382, 315
- Mukherjee P., Wang Y., 2004, ApJ, 613, 51
- Munshi D., Coles P., Cooray A., Heavens A., Smidt J., 2011, MNRAS, 410, 1295
- Munshi D., Heavens A., 2010, MNRAS, 401, 2406
- Munshi D., Valageas P., van Waerbeke L., Heavens A., 2008, Phys. Rep., 462, 67
- Munshi D., Wang Y., 2003, ApJ, 583, 566
- Namjoo M. H., Firouzjahi H., Sasaki M., 2013, EPL (Europhysics Letters), 101, 39001

- Narayan R., Bartelmann M., 1996, arXiv:astro-ph/9606001
- Natoli P. e. a., 2010, MNRAS, 408, 1658
- Nørgaard-Nielsen H. U., 2012, Ap&SS, 340, 161
- Oliveira M., Bowen R., McKenna R., Chang Y., 2001, ACTA Press, pp 261–266
- Papovich C., Bassett R., Lotz J. M., van der Wel A., Tran K.-V., Finkelstein S. L., Bell E. F., Conselice C. J., Dekel A., Dunlop J. S., Guo Y., et al. 2012, ApJ, 750, 93
- Park C., Choi Y.-Y., 2009, ApJ, 691, 1828
- Peccati G., Taqqu M. S., 2011, Springer
- Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419
- Perlmutter S., Aldering G., Goldhaber G., et al. 1999, ApJ, 517, 565
- Pietrobon D., Cabella P., Balbi A., Crittenden R., de Gasperis G., Vittorio N., 2010, MNRAS, 402, L34
- Pietrobon D., Cabella P., Balbi A., de Gasperis G., Vittorio N., 2009, MNRAS, 396, 1682
- Planck Collaboration 2013a, arXiv:1303.5072
- Planck Collaboration 2013b, arXiv:1303.5075
- Planck Collaboration 2013c, arXiv:303.5076
- Planck Collaboration 2013d, arXiv:1303.5077
- Planck Collaboration 2013e, arXiv:1303.5082
- Planck Collaboration 2013f, arXiv:1303.5084
- Refregier A., Bacon D., 2003, MNRAS, 338, 48
- Reichborn-Kjennerud B., Aboobaker A. M., Ade P., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7741 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, EBEX: a balloon-borne CMB polarization experiment

- Rettura A., Rosati P., Nonino M., Fosbury R. A. E., Gobat R., Menci N., Strazzullo V., Mei S., Demarco R., Ford H. C., 2010, *ApJ*, 709, 512
- Rhodes J., Refregier A., Groth E. J., 2000, *ApJ*, 536, 79
- Riess A. G., Filippenko A. V., Challis P., et al. 1998, *AJ*, 116, 1009
- Riotto A., 2002, arXiv:hep-ph/0210162
- Rosenblatt F., 1958, The perceptron: A probabilistic model for information storage and organization in the brain, 65, 368
- Rozo E., Schmidt F., 2010, arXiv:astro-ph/1009.5735
- Rubiño-Martín J. A., Rebolo R., Tucci M., et al. 2010, in Diego J. M., Goicoechea L. J., González-Serrano J. I., Gorgas J., eds, *Highlights of Spanish Astrophysics V The QUIJOTE CMB Experiment*. p. 127
- Rudjord Ø., Hansen F. K., Lan X., Liguori M., Marinucci D., Matarrese S., 2009, *ApJ*, 701, 369
- Rudjord Ø., Hansen F. K., Lan X., Liguori M., Marinucci D., Matarrese S., 2010, *ApJ*, 708, 1321
- Salopek D. S., Bond J. R., 1990, *Phys. Rev. D*, 42, 3936
- Salvio A., 2013, *Physics Letters B*, 727, 234
- Schade D., Lilly S. J., Le Fevre O., Hammer F., Crampton D., 1996, *ApJ*, 464, 79
- Schlegel D., Abdalla F., Abraham T., et al. 2011, arXiv:1106.1706
- Schmidt F., Leauthaud A., Massey R., Rhodes J., George M. R., Koekemoer A. M., Finoguenov A., Tanaka M., 2012, *ApJ*, 744, L22
- Schneider P., 2005, arXiv:astro-ph/0509252
- Schneider P., Ehlers J., Falco E. E., 1992, *Gravitational Lenses*
- Schrabback T., Hartlap J., Joachimi B., Kilbinger M., Simon P., Benabed K., Bradač M., Eifler T. e. a. ., 2010, *A&A*, 516, A63
- Semboloni E., Mellier Y., van Waerbeke L., Hoekstra H., Tereno I., Benabed K., Gwyn S. D. J., Fu L., Hudson M. J., Maoli R., Parker L. C., 2006, *A&A*, 452, 51

- Shahram M., Donoho D., Starck J., 2007, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6701 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Multiscale representation for data on the sphere and applications to geopotential data
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
- Simard L., Willmer C. N. A., Vogt N. P., Sarajedini V. L., Phillips A. C., Weiner B. J., Koo D. C., Im M., Illingworth G. D., Faber S. M., 2002, ApJS, 142, 1
- Smith K. M., Senatore L., Zaldarriaga M., 2009, J. Cosmology Astropart. Phys., 9, 6
- Smoot G. F., Bennett C. L., Kogut A., et al. 1992, ApJL, 396, L1
- Sonnenfeld A., Bertin G., Lombardi M., 2011, A&A, 532, A37
- Starck J.-L., Fadili M. J., Rassat A., 2013, A&A, 550, A15
- Starobinsky A. A., 1980, Physics Letters B, 91, 99
- Storrie-Lombardi M. C., Lahav O., Sodre Jr. L., Storrie-Lombardi L. J., 1992, MNRAS, 259, 8P
- Story K. T., Reichardt C. L., Hou Z., et al. 2013, ApJ, 779, 86
- Takada M., Jain B., 2003, MNRAS, 344, 857
- Tauber J., Mandolesi N., Puget J.-L., et al., 2010, Astronomy & Astrophysics, in press
- Tauber J. A., 2004, Advances in Space Research, 34, 491
- Taylor A. N., Dye S., Broadhurst T. J., Benitez N., van Kampen E., 1998, ApJ, 501, 539
- Tenorio L., Jaffe A. H., Hanany S., Lineweaver C. H., 1999, MNRAS, 310, 823
- Tristram M., Ganga K., 2007, Reports on Progress in Physics, 70, 899
- Tyson J. A., Valdes F., Jarvis J. F., Mills Jr. A. P., 1984, ApJ, 281, L59
- Tyson J. A., Wenk R. A., Valdes F., 1990, ApJ, 349, L1
- Umetsu K., Broadhurst T., Zitrin A., Medezinski E., Hsu L.-Y., 2011, ApJ, 729, 127

- Vallinotto A., Dodelson S., Zhang P., 2011, *Phys. Rev. D*, 84, 103004
- van Waerbeke L., 2010, *MNRAS*, 401, 2093
- Vanzella E., Cristiani S., Fontana A., Nonino M., Arnouts S., Giallongo E., Grazian A., Fasano G., Popesso P., Saracco P., Zaggia S., 2004, *A&A*, 423, 761
- Verde L., Wang L., Heavens A. F., Kamionkowski M., 2000, *MNRAS*, 313, 141
- Vielva P., Martínez-González E., Barreiro R. B., Sanz J. L., Cayón L., 2004, *ApJ*, 609, 22
- Vielva P., Sanz J. L., 2009, *MNRAS*, 397, 837
- Vielva P., Sanz J. L., 2010, *MNRAS*, 404, 895
- Watson R. A., Gutierrez de La Cruz C. M., Davies R. D., Lasenby A. N., Rebolo R., Beckman J. E., Hancock S., 1992, *Nature*, 357, 660
- Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, *Nature*, 405, 143
- Wu J. H. P., Zuntz J., Abroe M. E., et al. 2007, *ApJ*, 665, 55
- Yadav A. P. S., Wandelt B. D., 2008, *Physical Review Letters*, 100, 181301
- Yadav A. P. S., Wandelt B. D., 2010, *Advances in Astronomy*, 2010
- Zaldarriaga M., Spergel D. N., Seljak U., 1997, *ApJ*, 488, 1