

NEEDLET MULTIPLE TESTING FOR POINT SOURCE DETECTION IN PLANCK CMB ANISOTROPY MAPS

Javier Carrón Duque

Supervisors:

Prof. Nicola Vittorio
Prof. Domenico Marinucci
Dr. Alessandro Buzzelli



Needlet Multiple Testing for Point Source Detection in Planck CMB Anisotropy Maps

Javier Carrón Duque^{1,2}

Supervisors:

Nicola Vittorio¹ Domenico Marinucci³ Alessandro Buzzelli¹

September 2018

Version 1.0

Abstract

The Cosmic Microwave Background is one of the most important sources of information about the Early Universe, since it is composed by photons emitted when the Universe was only 377 000 years old. Observing it, however, entails several difficulties. One of them is the presence of unresolved point sources in the foreground. In this work, we implement a recent algorithm that takes advantage of needlet filtering and multiple testing Statistics. We create a flexible and complete programme and apply it to simulations and Planck anisotropy maps. In the simulations, the algorithm is able to recover essentially all point sources with an intensity above 3 to 4σ with very low percentage of false detections. On some of the Planck maps, the algorithm detects a population of maxima incompatible with the assumed gaussianity of the Cosmic Microwave Background.

This work is a Master Thesis conducted at University of Rome Tor Vergata within the Erasmus Mundus Joint Masters Programme AstroMundus.

Based on observations obtained with Planck (<http://www.esa.int/Planck>), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada.

¹Physics Department, University of Rome Tor Vergata

²Institute of Astro- and Particle Physics, University of Innsbruck

³Mathematics Department, University of Rome Tor Vergata

*To my father Alberto,
who always fostered my curiosity and
made me the person I am today.*

Contents

Contents	iv
List of Figures	vi
List of Tables	vii
List of Equations	viii
Acknowledgements	ix
Introduction	x
1 Cosmological Framework	1
1.1 Standard Cosmological Model	2
1.1.1 Dynamics and content of the Universe	2
1.1.2 History of the Universe	6
1.1.3 Parameters of the Λ CDM model	8
1.2 Cosmic Microwave Background	11
1.2.1 Physics of the CMB	11
1.2.2 Power spectrum of the CMB	14
1.2.3 Satellite observations of the CMB	19
1.3 Foregrounds	21
1.3.1 Diffuse Galactic Radiation	23
1.3.2 Point Sources	24
1.3.3 Algorithms for component separation	25
2 Methodology	33
2.1 Needlets	34
2.1.1 Definition	34
2.1.2 Filter function b and parameters	35
2.1.3 Filtering and reconstructing	39
2.2 Maxima and multiple testing	40
2.2.1 Maxima distribution	40
2.2.2 p -values and multiple testing	41
2.2.3 Summary of the algorithm	44

3 Code and Implementation	45
3.1 Main algorithm	48
3.1.1 Parameters of the main algorithm	48
3.1.2 Steps of the main algorithm	49
3.1.3 Effects of the parameters in the main algorithm	52
3.2 Simulations	54
3.2.1 Parameters of the simulations	54
3.2.2 Steps of the simulations	55
3.2.3 Effect of the parameters in the simulations	58
3.3 Planck data	59
3.3.1 Parameters for the Planck data	59
3.3.2 Steps for the Planck data	60
3.3.3 Effect of the parameters for the Planck data	62
4 Results and Discussion	63
4.1 Simulations	63
4.1.1 Sensitivity of the algorithm	64
4.1.2 Noise	67
4.1.3 Use of masks	70
4.2 Planck maps	70
4.2.1 Results common to all maps	73
4.2.2 Comparison of the maps	74
5 Conclusions and Future Work	77
Bibliography	80
A Code repository	84
B Table of results	85
Acronyms	89
Symbols	90

List of Figures

1.1	Pie chart of the component of the Universe	5
1.2	History of the Universe	9
1.3	Temperature anisotropies of the CMB with confidence mask	11
1.4	Black Body spectrum of the CMB	13
1.5	Visibility function	14
1.6	CMB maps reconstructions with $\ell = 2, 4, 10$ and 100	15
1.7	Angular Power Spectra of the CMB	16
1.8	Simulated Power Spectra for different cosmological parameters	18
1.9	Comparison between COBE, WMAP and Planck	20
1.10	CMB intensity maps in the nine Planck bands	22
1.11	Contribution of foregrounds at different frequencies	24
1.12	Luminosity distribution of Point Sources	26
1.13	Spectral index of the detected Point Sources	27
1.14	Temperature maps for the four algorithms	29
1.15	Power spectra differences between algorithms	30
1.16	Confidence masks	31
2.1	Filter function b for standard needlets	36
2.2	Profile of standard needlet	37
2.3	Filter function b for Mexican needlets	38
2.4	Profile of Mexican needlet	38
2.5	Theoretical distribution of maxima	42
3.1	Illustration of the HEALPix standard	47
3.2	CMB temperature map and β -map	50
3.3	Comparison of the C_ℓ for a β -map with prediction	51
3.4	Example of the multitempling approach	53
3.5	Maxima distribution for 50 maps	57
4.1	Sensitivity of the algorithm	65
4.2	Effect of α in the sensitivity	66
4.3	Sensitivity on the β -maps	66
4.4	Angular power spectrum of the noise	68
4.5	Effect of the noise in the simulations	69
4.6	Creation of new maxima by using masks	71
4.7	Intensity distribution for PS candidates	75
4.8	PS candidate in SMICA maps outside the confidence mask	76

List of Tables

1.1	Value of cosmological parameters as determined by Planck	21
4.1	Candidates for CMB maps without sources	64
4.2	Short table of results for Planck data	72
4.3	Behaviour of the results with j and α	73
4.4	Maps with maxima more intense than $37.5\sigma_\beta$	74

List of Equations

(1.1)	Equation of state of a fluid	3
(1.2)	Friedmann–Lemaître–Robertson–Walker metric	3
(1.3)	First Friedmann equation	3
(1.4)	Hubble parameter and Hubble constant	4
(1.5)	Definition of the density parameters Ω_X	4
(1.6)	Dependence of Ω_X with a	4
(1.7)	First Friedmann equation (Simplified)	5
(1.8)	Sum of the density parameters Ω_X	5
(1.9)	Second Friedmann equation	5
(1.10)	Cosmological redshift	6
(1.11)	Planck’s Law of Black Body Radiation	12
(1.12)	Spectral radial density	12
(1.13)	Evolution of the Black Body Radiation	12
(1.14)	Baryon to photon ratio	13
(1.15)	Temperature decomposition into spherical harmonics	14
(1.16)	Definition of $a_{\ell m}$	15
(1.17)	Definition of the angular power spectrum	16
(1.18)	Model of observed data for COMMANDER algorithm	26
(1.19)	Model of observed data for SEVEM algorithm	28
(2.1)	Definition of needlet function	34
(2.2)	Addition theorem of spherical harmonics	35
(2.3)	Alternative expression for needlets	35
(2.4)	b equally sampling of multipoles	35
(2.5)	Definition of $b(\cdot)$ for Mexican needlets	36
(2.6)	Definition of needlet coefficient β_{jk}	39
(2.7)	β -map reconstruction with a_{lm}	39
(2.8)	Spherical harmonics coefficients of the β -map	39
(2.9)	Angular power spectrum of the β -map	40
(2.10)	Squared sum of the β -maps	40
(2.11)	Needlet reconstruction formula	40
(2.12)	Definition of C' and C''	41
(2.13)	Definition of κ_1 and κ_2	41
(2.14)	Theoretical distribution of maxima in a gaussian map	41
(2.15)	General definition of p -value	42
(2.16)	p -value for testing gaussianity in the maxima	42
(2.17)	Multiple testing criterion	43
(3.1)	Total number of pixels in HEALPix	47

Acknowledgements

I want to take this opportunity to express my gratitude towards the people that have supported me and this project.

First, I want to thank my supervisors, as the project would have not been possible without their invaluable help. The frequent meetings with them were very insightful and motivating, I was always leaving the meetings eager to work towards the next step. My first supervisor was Prof. Nicola Vittorio. His wise supervision and extensive knowledge on Cosmology and Cosmic Microwave Background Radiation was crucial to keep pushing the work into relevant and interesting challenges. My second supervisor was Prof. Domenico Marinucci. His mathematical expertise on the use of needlets and Statistics was proven extremely helpful for the development of the project. The team was completed by Dr. Alessandro Buzzell. His help with more technical issues and very frequent discussions about the implementation and partial results were always very helpful. I want to express my most sincere gratitude towards the three of them, as they made the work dynamics one of my favourite part of this project.

I also want to sincerely thank PhD. Lucía Menéndez-Pidal de Cristina for very kindly proofreading this entire document and providing me with many corrections for mistakes that I had overlooked. She also provided me with very useful suggestions that have improved the quality of this document.

Finally, I want to express my gratitude towards the AstroMundus Consortium and everyone that made it possible. Their master programme, organised by five universities in four different countries, has been a great opportunity for me, both academically and personally. I have had the opportunity to know different Universities and countries and learn from very good professors and researchers.

Introduction

In this work we will explain, implement, and apply an algorithm to detect point sources on spherical maps. We will implement this code so it can be applied to any map, and we will test its power on simulations. Then we will try to detect point sources in the Cosmic Microwave Background (CMB) maps reported by the Planck Collaboration.

In [Chapter 1](#) we will give an introduction to the cosmological framework in which this work is embedded. We will start by explaining the main characteristics of the Standard Cosmological Model. We will talk about the dynamics, contents, and evolution of the Universe, as well as the physical meaning of its main parameters. Then, we will discuss the CMB, starting by how it was formed. We will also explain one of the most common tools to study it, the angular power spectrum, and the main experiments performed to observe it. We will finish this chapter by discussing the contribution of the foregrounds: diffuse galactic radiation and point sources. We will discuss the physics behind these components and how they affect the observations, as well as the main techniques used to remove foregrounds from the temperature maps by the Planck Collaboration.

In [Chapter 2](#) we will develop the mathematical tools required for the detection algorithm. We will start by explaining the concept of needlet: its definition, kinds, and advantages. We will also explain the characteristics of the needlet filtering of a map and its implementation. In the second part of the chapter we will develop the statistical basis of the algorithm. We will talk about the distribution of maxima in a random gaussian map as the primordial CMB. We will introduce then the concept of p -value and how we will use it in a multiple testing framework: we will not test each individual maximum but the whole maxima population at once. We will also give a brief summary of the algorithm with the steps we will follow in its implementation.

In [Chapter 3](#) we will focus on the code and the implementation of the algorithm. We will structure the code, and this chapter, in three parts. In the first part we will talk about the implementation of the main algorithm, which extracts candidates to point sources from a given map. The second part will be focused on the generation of simulated CMB maps and how we will use them to test the main algorithm. In the last part we will explain the use of real Planck data and how we will analyse the results. All the three part will be divided in a brief discussion of the input parameters, the main steps of the code, and the consequences of changing the values of the parameters. In this way we will try to give an intuitive feeling of the behaviour of the code in order to better interpret the results for a variety of cases.

In [Chapter 4](#) we will apply the code to real CMB maps, like the temperature anisotropy maps reported by Planck. We will start by explaining the results about the algorithm obtained from simulated maps. We will then talk about the sensitivity of the algorithm to point sources, as well as the effect of the noise and masks over this sensitivity. After that, we will discuss the results obtained from real Planck maps: what are the main results present in most maps and how the different characteristics of the maps can affect the results.

In [Chapter 5](#) we will conclude the work by summarising its most important aspects. We will discuss the main points that we extract from the work and the different ways in which it can be extended.

After the main matter of the book, we can find the [Bibliography](#) and two appendices. [Appendix A](#) contains information about the online code repository where we can find the code and other useful resources, like the most recent version of this document or the complete table of results. In [Appendix B](#) we will give an almost complete version of this table, that contains all the results obtained from the Planck maps, as well as a brief explanation of its meaning. The book will finish with a list of acronyms and commonly used symbols.

For any question, please contact us at Javier.Carron-Duque@student.uibk.ac.at.

❖ Chapter One ❖

Cosmological Framework

Modern Cosmology started developing as a physical science after the discovery of General Relativity (GR) in 1916 by Albert Einstein, which provided a mathematical formulation of the geometry and the dynamical laws of space-time. At that time, it was thought that the Universe was static, but Alexander Friedmann found later that this was only one of the possible solutions of General Relativity (1922). In fact, his solutions admitted universes in expansion or contraction, and with different geometries. They were compatible both with General Relativity and the so-called *cosmological principle*:

The properties of the Universe on a large enough scale must be the same for all observers. In particular, the Universe must be homogeneous (same observable properties at any location) and isotropic (same observable properties in any direction) on a large enough scale.

The cosmological models used to explain the Universe were mostly theoretical, as the assumption of the cosmological principle, since the observational tools needed to check these theories were not yet available. One model, proposed by George Lemaître (1931), postulated that the Universe started with a very quick expansion similar to an explosion. During the first moments, the Universe was in a state of very high energy compressed in a very small volume. This is now known as the *Big Bang* model. He also provided one of the first observational breakthroughs when he confirmed that there was a relation between the redshift of a galaxy and its distance to us (G. Lemaître 1927). This means that the further away a galaxy is, the faster it moves *away* from us, reinforcing the idea that the Universe may be in fact expanding. This relation is now widely known as the *Hubble Law*, who popularised it (1929).

The *Big Bang* model of the Universe did not become widely accepted until the discovery of the CMB, one of the most critical predictions of the model (Penzias and Wilson 1965; Dicke et al. 1965). Since then, the focus has moved from knowing the correct model that describes the Universe to establish the exact values of the parameters that describe it (see Peebles 2011, for a historical review of the last half of the century). Ground and satellite observations during the last decades have allowed the scientific community to obtain precise measurements and establish the *concordance model*, which we will explain in the next section. This has marked the beginning of the so-called *high precision cosmology*.

In this chapter we will develop the current cosmological framework where this work is embedded. In [Section 1.1](#) we will explain the most widely accepted model to explain the Universe, the Lambda Cold Dark Matter (Λ CDM) model. We will introduce its basic concepts, such as the components of the Universe and how they will affect its evolution. Then, we will give a brief summary of the history of the Universe as explained by this model. We will end the section by explaining its main parameters.

In [Section 1.2](#) we will talk about the CMB, since this work is mainly dedicated to the analysis of CMB maps. We will start describing the physical processes involved in its creation and propagation. Then we will explain some of its main features through one of the most important tools used to analyse it, the power spectrum, that we will also introduce. The last part of the section is dedicated to a brief discussion of the CMB observations that have been performed from space in the last decades.

In [Section 1.3](#) we will focus on the foregrounds that are measured when observing the CMB. Understanding and correctly removing them is crucial for its analysis. We will explain first the diffuse contamination coming from our galaxy and then the concentrated contamination coming from sources that are not resolved, since in this work we will study a technique to detect this last component. Finally, we will end by talking about the main aspects of some standard algorithms used to remove the foreground from the CMB observations.

1.1 Standard Cosmological Model

The Standard Cosmological Model is a widely accepted model that reasonably reproduces most of the cosmological observations to date throughout a wide range of scales and most of the lifetime of the Universe. According to it, the Universe obeys GR with a positive Cosmological Constant Λ , and it contains a significant amount of Cold Dark Matter. We will explain what this means later in this section. This model is sometimes also called *Lambda Cold Dark Matter (Λ CDM) model*, referring to the main components of the Universe, or *concordance model*, since it explains a wide variety of observations. It is a relatively simple model, based on only six independent parameters that are fitted to the observations, plus a series of parameters that are fixed according to theory. For more in-depth texts about Cosmology, see Liddle ([2015](#)) and Vittorio ([2017](#)).

1.1.1 Dynamics and content of the Universe

In order to study the characteristics and evolution of the Universe, the model considers all the matter and energy within it as a perfect fluid formed by several distinct components. According to GR, the energy density and pressure of this fluid determines the geometry of the Universe and, therefore, its evolution. It is then convenient to consider the relation between the energy density and pressure of the different fluid components, since they will affect the Universe in different ways. This receives the name of *equation of state*:

$$p = \omega \rho c^2 \tag{1.1}$$

where p is the pressure, ρ is the density, c is the speed of light in vacuum, and ω is a parameter that varies depending on the fluid. The different components considered are:

- **Matter:** this component constitutes all the usual baryonic matter (planets, stars, gas, dust, black holes...) as well as the cold dark matter, which does not interact with light and is only observed through its gravitational effects. The contribution to the current Universe is around 31% of the total energy density. Both kinds are composed by non-relativistic particles, so the pressure of the fluid is zero. This means that $\omega = 0$.
- **Radiation:** this component is formed by all photons, mainly CMB photons, plus all relativistic particles such as neutrinos. The contribution to the current Universe is negligible, but it was the main component in the Early Universe, as we will see in [Section 1.1.2](#). The parameter of the equation of state for this component is $\omega = 1/3$.
- **Dark Energy:** when calculating the GR field equations with a Cosmological Constant Λ , the term that contains Λ can be regarded as an additional component of the fluid. In this case, it receives the name of Dark Energy. It is the main component in the current Universe, contributing to around 69% of the energy density. The parameter of the equation of state for Dark Energy is $\omega = -1$, meaning that the pressure is negative. We will see later that this component will drive the Universe accelerated expansion.

Another key aspect of the model to consider is that the Universe can expand, its scale is not constant. The evolution of the size of the space itself can be studied through the *scale factor*, $a(t)$. In the current time, it is defined as $a(t_0) = 1$, meaning that the physical distance between two points in the Universe at time t is $a(t)$ times the distance today. More formally, the metric describing the Universe depends on $a(t)$. The complete expression is called Friedmann–Lemaître–Robertson–Walker metric ([1922](#); [1931](#); [1935](#); [1937](#)):

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

where (t, r, θ, ϕ) represent the space-time spherical coordinates, c is the speed of light in vacuum, $cd\tau$ is the proper distance, and k is the curvature of the Universe: 0 for a flat Universe, negative for a hyperbolic (open) Universe and positive for a spherical (closed) Universe. We are not going to study this metric in detail, but we have to mention an important consequence of it, the Friedmann equations, which describe the evolution of the Universe through the behaviour of the scale factor $a(t)$. The first Friedmann equation one can be expressed as:

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2} \quad (1.3)$$

where G is the newtonian gravitational constant and the rest of the parameters are the same as before. It is common to define the *Hubble parameter* and the

*Hubble constant*¹ as

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \quad (1.4a)$$

$$H_0 \equiv H(t_0) = \dot{a}(t_0) \quad (1.4b)$$

It is also common to define the density parameters $\Omega_X(t)$ for every component plus curvature as:

$$\Omega_M = \frac{8\pi G}{3H^2} \rho_M \quad (1.5a)$$

$$\Omega_R = \frac{8\pi G}{3H^2} \rho_R \quad (1.5b)$$

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H^2} \quad (1.5c)$$

$$\Omega_k = -\frac{kc^2}{a^2 H^2} \quad (1.5d)$$

where the temporal dependences have been dropped for simplicity.

Now, we can intuitively study the dependence of these parameters on the scale factor $a(t)$. We want to use these parameters to simplify the [equation \(1.3\)](#), so we will study the behaviour of $H^2 \Omega_X$. For the matter component, the energy density ρ_M is proportional to the number of particles per volume. The number of particles does not vary, but the volume increase as a^3 . Therefore, $H^2 \Omega_M \sim \rho_M \sim n/V \sim a^{-3}$.

The radiation fluid behaves differently, since it is composed of relativistic particles. Here, the energy density is proportional to the number of photons (or neutrinos) per volume times the frequency (or velocity) of the particle: $H^2 \Omega_R \sim \rho_R \sim n\nu/V \sim a^{-1}/a^3 \sim a^{-4}$. This can also be seen as a factor of a^3 due to the volume expansion and an additional factor a due to elongation of the wavelength, which makes the photon less energetic.

The two remaining density parameters are more straightforward to analyse. The Dark Energy density is constant since Λ does not depend on $a(t)$. Therefore $H^2 \Omega_\Lambda \sim a^0$. On the other hand, the curvature k does not change, so looking at the expression in [equation \(1.5d\)](#) we have that $H^2 \Omega_k \sim -k/a^2$. Comparing these values with the values at the present day $\Omega_{0,X}$, we have that:

$$H(t)^2 \Omega_M(t) = H_0^2 \Omega_{0,M} a^{-3} \quad (1.6a)$$

$$H(t)^2 \Omega_R(t) = H_0^2 \Omega_{0,R} a^{-4} \quad (1.6b)$$

$$H(t)^2 \Omega_\Lambda(t) = H_0^2 \Omega_{0,\Lambda} \quad (1.6c)$$

$$H(t)^2 \Omega_k(t) = H_0^2 \Omega_{0,k} a^{-2} \quad (1.6d)$$

We can now simplify the first Friedmann equation ([1.3](#)) in two ways. First, we have an equation for the evolution of $a(t)$ that only depends on current parameters:

$$H(a) \equiv \frac{\dot{a}}{a} = H_0 \sqrt{\Omega_{0,M} a^{-3} + \Omega_{0,R} a^{-4} + \Omega_{0,\Lambda} + \Omega_{0,k} a^{-2}} \quad (1.7)$$

¹Because of historical reasons, it is common to use the parameter h (dimensionless Hubble constant) defined as $H_0 = 100h \frac{km/s}{Mpc}$. See also Croton ([2013](#))

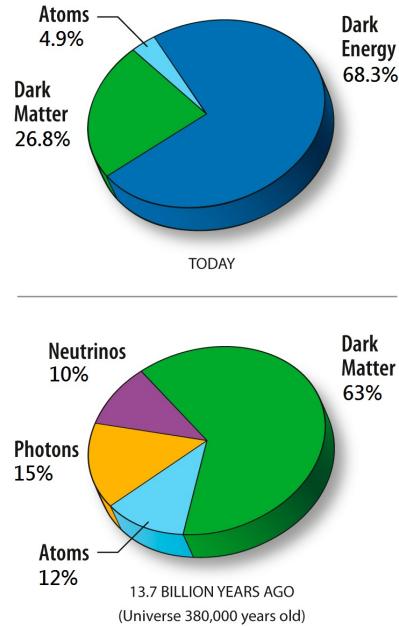


Figure 1.1: Percentage of the total energy density of the Universe due to each component. On the top, values for today as observed by the Planck satellite. On the bottom, values during the recombination, when the CMB is produced. Image Credit: adaptation of original by NASA/WMAP with values from Planck.

from where we can see that the Universe does expand (or contract), since $\dot{a} = 0$ is unstable against small perturbations. We also note that the different components will dominate the behaviour of $a(t)$ at different moments of the expansion of the Universe.

On the other hand, we also obtain by construction a useful constrain on the values of $\Omega_{0,X}$:

$$\Omega_{0,M} + \Omega_{0,R} + \Omega_{0,\Lambda} + \Omega_{0,k} = 1 \quad (1.8)$$

where we also know that $\Omega_{0,R}$ is negligible. This means that we are able to know the geometry of the Universe by measuring the sum of matter plus dark energy density parameters. In particular, if they add up to 1, we live in a flat Universe ($\Omega_{0,k} = 0 \Rightarrow k = 0$). This appears to be the case, as measured by the Planck satellite (see figure 1.1, or table 1.1 later for the exact parameters).

We have talked about the first Friedmann equation until now, but the field equations of GR also imply the second Friedmann equation regarding the second derivative of $a(t)$, which can be simplified and written as:

$$\frac{\ddot{a}}{a} = H^2 (\Omega_\Lambda - 2\Omega_M - 4\Omega_R) \quad (1.9)$$

from where we can extract the acceleration of the scale factor $a(t)$ and, therefore, the expansion of the Universe. As we mentioned before, we can see that the Dark Energy is the only component accelerating this expansion. Matter and radiation, having a negative sign, will decelerate the expansion (see Frieman et al. 2008, for a review on the accelerating Universe).

An important consequence of the expansion of the Universe, from an observational point of view, is the cosmological redshift z . When a photon is emitted, the expansion of the Universe will elongate its wavelength, shifting it redder and making it lose energy². The redshift z of a photon is defined as the ratio between the difference in wavelength and the wavelength of emission. If we call λ_e and λ_o to the wavelengths at emission (t_e) and detection (t_o), respectively, we have that the cosmological redshift obeys the following expression:

$$1 + z = \frac{\lambda_o}{\lambda_e} = \frac{a(t_o)}{a(t_e)} \quad (1.10)$$

It is common to do a slight abuse of terminology and mention the Universe at redshift z when referring to the Universe when it had $a = 1/(1+z)$. That is, the instant of the Universe whose photons would be observed at redshift z today.

1.1.2 History of the Universe

We are now prepared to talk about the history of the Universe according to the Λ CDM model. Since the current day parameters are reasonably accurate (see [table 1.1](#) for the last determination of them), it is possible to use the [Equation \(1.7\)](#) to rewind the scale factor a and the known laws of Physics to study the Universe at earlier epochs. Doing that, we find an initial singularity called *Big Bang* that marks the beginning of the Universe as we know it³. However, studying this singularity and the very first instants following the Big Bang has been proven to be tricky: during the Planck time (the first 10^{-43} second of the Universe), the laws of Physics as we know them may not apply. Additionally, it has been conjectured that during the first 10^{-32} second, the Universe undergoes a process of extremely quick expansion called *Cosmic Inflation* ([Guth 1981](#)).

These problems are usually treated by extensions of the Λ CDM model (or alternative theories), while the model itself can confidently predict the behaviour of the Universe since $\sim 10^{32}$ seconds after the Big Bang onwards. We are going to explain now the main points of the history of the Universe:

- **First second:** during the first second after the Big Bang, the four fundamental forces of the Universe, which are unified at the beginning, separate: gravity, strong, weak, and electromagnetic forces. The Universe is filled with an extremely hot and dense plasma formed by elementary particles, which can be created and annihilated in particle-antiparticle pairs. As the Universe expands, the density and temperature of this plasma will decrease. By the end of the first second, the temperature drops enough to stop the production of pairs and the antiparticles are

²The energy lost by a photon due to redshift is not transmitted elsewhere or dissipated, it is truly lost. This does not contradict the laws of Physics, since the conservation of energy only happens when the metric has temporal invariance. This is the case in our everyday experience, where the expansion of the Universe is negligible.

³We note that some authors use this term differently. Some authors use it to refer to the initial singularity, like we will do. Other authors use the term to refer to some period of time from the singularity to around 3 minutes later, sometimes excluding the first instants. Other authors use it to refer to the entire history of the Universe.

annihilated soon after. Quarks form protons and neutrons, and neutrinos stop interacting with the rest of particles (*neutrino decoupling*).

- **First second – 3 minutes:** after the annihilation of the antiparticles, the plasma is vastly dominated by the contribution of the photons. The baryonic matter within the plasma is constituted by protons, neutrons, and electrons. Neutrons start decaying into protons through β -decay, which has a well-defined time rate of $886\text{ s} \sim 15\text{ min}$. It is worth noting that the plasma is very dense and charged particles are abundant (protons and electrons). These charged particles interact effectively with photons, making the plasma extremely opaque. The baryonic matter and the radiation component are then coupled. Dark matter starts to form structures by falling into the potential wells created by the radiation component.
- **3 – 20 minutes:** after 3 minutes, at a temperature of the plasma of around 10^9 K , the pressure and temperature of the plasma start allowing fusion burning (*Primordial Nucleosynthesis*), effectively confining a fraction of the neutrons in helium (He) nuclei. Neutrons that are not forming He will decay into protons. The amount of He generated in this way can be calculated from values obtained from Nuclear Physics and does not significantly depend on any cosmological parameter or free parameter. This constitutes a solid evidence for the theory. Other elements such as deuterium or heavier nuclei are also produced in less amount. The measurement of the abundance of these elements can provide information regarding the conditions of the plasma during this stage. After 20 minutes, the nucleosynthesis is finished and now a plasma of ionised hydrogen (H) and He with traces of other elements will cool down together with the dominant radiation component. Both components are coupled due to the large cross section σ_T of ionised particles and photons.
- **47 000 yr:** after the nucleosynthesis finishes, the baryonic plasma just cools down with the expansion along with the radiation field while dark matter forms structures in the potential wells. There are no changes in these processes until around 47 000 years after the Big Bang. As we explained before, the energy density ρ of matter and radiation evolves in a different way with the scale factor a : a^{-3} and a^{-4} respectively. That means that the radiation energy density decreases a times faster than the matter energy density. 47 000 years after the Big Bang ($z \sim 3000$), called *equivalence time*, both components have the same energy density. This marks the transition from a *radiation-dominated* Universe to a *matter-dominated* Universe. From this point onwards, the baryonic matter and radiation fall into the dark matter potential wells, which are now dominating. Baryonic matter is still ionised and therefore coupled with radiation, which means that the Universe is still opaque.
- **377 000 yr:** as the Universe keeps expanding, the temperature of the plasma decreases. 377 000 years after the Big Bang ($z \sim 1080$), it drops below 3000 K. At this point, the photons can not keep the H in the plasma ionised, so the protons and electrons recombine to form neutral H. This is the phase of *recombination*. Additionally, neutral atoms have much smaller cross section with photons, which means that the radiation component

stops interacting with the baryonic component and start propagating freely through the Universe. This is called *photon decoupling* and it has two very important consequences: first, the Universe becomes transparent; second, we can observe photons whose last interaction with matter was in this epoch. This is the origin of the CMB, which we will explain in more detail in [Section 1.2](#).

- **377 000 yr – 200 Myr:** after the decoupling, the Universe becomes transparent but there are still not stars or new sources of light, only the CMB and a faint 21 cm H emission from the plasma. This is why this period is called the *Dark Ages*. The temperature keeps decreasing (passing through a temperature range where liquid water could be possible, Loeb 2014) and by the end of this period, baryonic matter starts to form the first structures.
- **200 Myr – 1 000 Myr:** the period starts with the formation of the first generation of stars ($z \sim 20$), called *population III* stars. Some early dwarf galaxies and quasars are also formed in this period ($z \sim 11$). These sources are thought to emit large amounts of energy which could re-ionise the H medium, as it is still today. This process is called *reionisation*. Structures are formed in a hierarchical manner, smaller objects forming first. After this period, structures continue to interact, merge and grow under the influence of gravity in the same manner that it does today. Baryonic matter follows the dark matter distribution, forming the Large Scale Structure with the imprint of the effects of the early Universe.
- **9 800 Myr:** By this point ($z \sim 0.4$), the Universe looks very similar as it is today. However, the matter energy density keeps decreasing due to the expansion of the Universe, while dark energy, that was not relevant at earlier times, stays constant. 9800 Myr after the big Bang, both terms contribute the same, starting the epoch of the *dark-energy-dominated* Universe.
- **13 800 Myr:** Present day Universe.
- **Future:** since our Universe is now dominated by dark energy, the Universe will continue expanding at an increasing rate. Additionally, the H that form the stars will get depleted as more stars die, forming degenerate stars or black holes. In around 10^{14} yr, no more stars will be formed, ending the so-called *Stellariferous Era*. After this point, the predictions are highly speculative and depend on the accurate values of physical constants, the decay or not of the proton, or the laws of Physics beyond the Standard Model (of Particle Physics). One possibility is that the Universe would be formed by degenerate stars and black holes that would eventually evaporate. All baryonic matter would turn into leptons and photons, in which case the Universe could reach then a maximum of entropy in a state called *Heat Death* of the Universe.

1.1.3 Parameters of the Λ CDM model

Before we finish this section, we are going to briefly explain the parameters used by the Λ CDM model. We are going to follow here the parametrisation

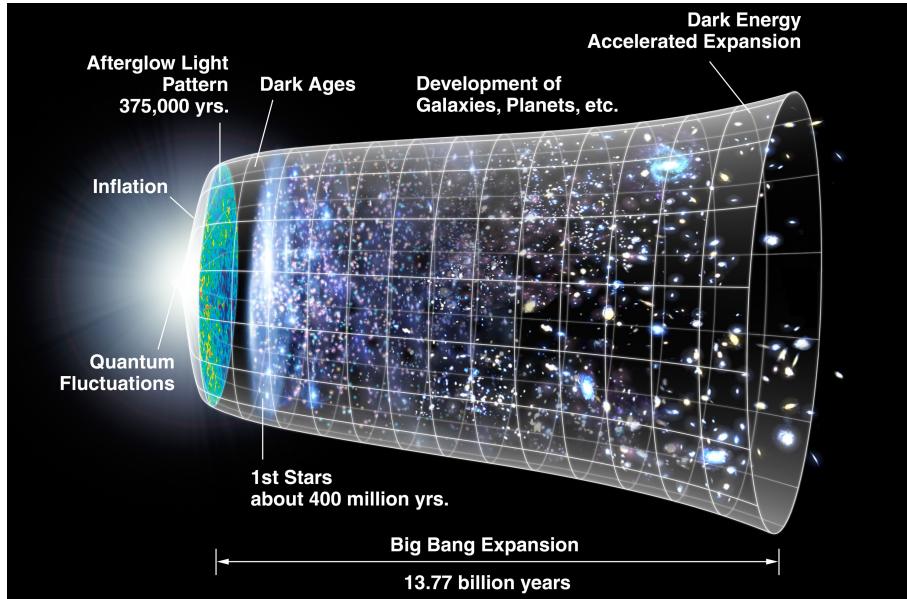


Figure 1.2: Graphic representation of the main stages of the history of the Universe. Image Credit: adaptation of original by NASA/WMAP.

used by the Planck Collaboration (2018a, especially section 3), since we will be working with Planck data throughout this project. As we said before, there are *6 independent parameters*:

- **Baryonic density parameter, Ω_b .** Baryonic matter is composed by all the usual matter: stars, dust, gas, planets, black holes... Unlike dark matter, baryonic matter interacts with light and can be observed directly, provided the necessary tools.

For years there was a discrepancy between the amount of baryonic matter that was inferred from theoretical predictions like CMB or primordial nucleosynthesis with the amount of observed baryonic matter. This is known as the *Missing Baryon Problem* and it has been claimed solved recently when the missing $\sim 30\%$ of the matter was found in the form of diffuse hot strands between galaxies by two independent teams (Tanimura et al. 2017; de Graaff et al. 2017).

- **Cold dark matter density parameter, Ω_c .** Dark matter is the most abundant kind of matter, but it does not interact electromagnetically so it can only be observed indirectly by its gravitational effects.

Cosmological cold dark matter is necessary to explain the characteristics of the perturbations today (galaxies, clusters, large scale structure...). Dark matter is not coupled with the radiation component in the Early Universe, meaning that perturbations in dark matter, unlike the ones in baryonic matter, can start growing before recombination. This correctly predicts the size of the perturbations today. Additionally, it needs to be *cold* dark matter instead of *hot* dark matter. Otherwise, any perturbation

below a large scale would be erased. This would mean that perturbations of galactic size would have been erased and the only structures present today would be Large Scale structures.

It is important to note that this *cosmological* cold dark matter is perfectly compatible with the *astrophysical* cold dark matter that can be observed in galaxy clusters or velocity curves in galaxies and are therefore considered to be the same.

- **Angular scale of acoustic oscillations, θ_{MC} .** In the plasma before recombination, a positive density perturbation would produce a positive pressure perturbation, since the radiation obeys the equation of state $p = \rho c^2/3$. This pressure produce an adiabatic expansion of the perturbation until it becomes a negative density perturbation, re-starting the cycle. These acoustic oscillations take place until the photon decoupling, when the fluid becomes mostly pressureless. The oscillations that were peaking at this point (maximum overdensity) will collapse faster due to gravity. The larger one correspond to the sound horizon, which peaks for the first time at recombination. The angular scale of this perturbation is easy to obtain from the CMB and influences the Large Scale Structure of the Universe at later times.
- **Reionisation optical depth, τ .** As mentioned before, the reionisation takes place when the radiation from the first stars, galaxies, or quasars are strong enough to ionise the neutral H in the Universe. Its cross section with photons increases, making the Universe less transparent. These processes are complex and they are not completely understood yet. By using this parameter, we can just express the amount of power absorbed by this ionised gas, disregarding the exact instant or reason why it happens.
- **Curvature fluctuation amplitude, A_s .** The amplitude of the perturbations is not predicted by the theory but it is left as a free parameter. It is measured as the intensity of the *curvature* fluctuations (related with *density* perturbations). The parameter is usually expressed at the pivot comoving scale of $k_0 = 0.05 \text{ Mpc}^{-1}$.
- **Scalar spectral index, n_S .** The primordial perturbations are generated in all scales. The power spectrum of these perturbations is usually modelled as a power law with index n_S . Scale invariance corresponds to $n_S = 1$. The Inflationary Theory predicts a value slightly under 1, compatible with the latest observations.

It is important to note that the Λ CDM is a phenomenological theory which does not provide any information regarding the value or nature of these parameters, they are just free parameters to be fitted to the observations. A possible future theory of everything would have to explain their values from basic physical principles.

Using the six parameters above, it is possible to obtain the values of other cosmological parameters. Some of the most important are: Hubble constant H_0 , energy density parameter of matter (Ω_M) and dark energy (Ω_Λ), fluctuation amplitude at present day σ_8 , redshift of equivalence (z_{eq}) and reionisation (z_{re}), and age of the Universe. The most current values for all these parameters

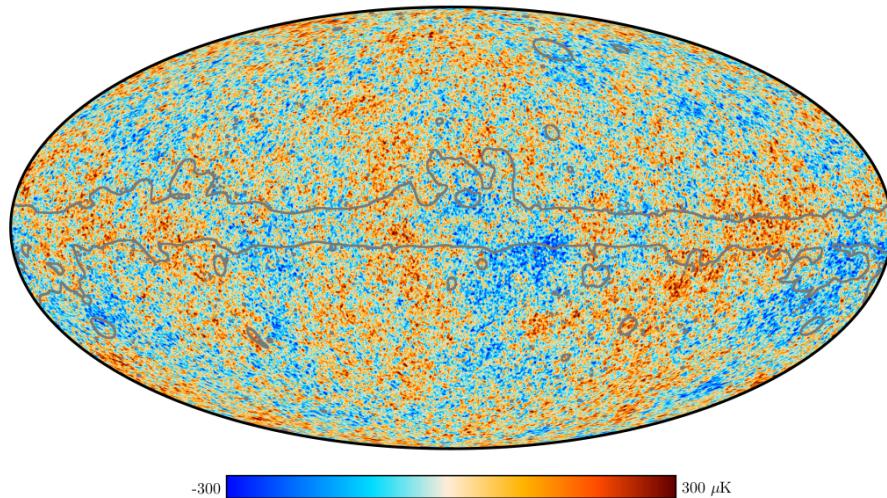


Figure 1.3: Image of the CMB temperature anisotropies as measured by the Planck Collaboration, with the SMICA algorithm (explained in [Section 1.3.3](#)). The grey line approximately denotes the confident mask. Image credit: Planck Collaboration.

can be found in [table 1.1](#) in [Section 1.2.3](#), as calculated by Planck Collaboration ([2018c](#)).

1.2 Cosmic Microwave Background

The Cosmic Microwave Background (CMB) is a collection of photons emitted in the recombination era. This means that their last scattering with matter was around $z \sim 1080$, when the Universe was only 377 000 yr old. We detect it as a very nearly isotropic Black Body at 2.725 K and therefore peaking at microwave frequency. The temperature anisotropies of the CMB, however, are one of the most important sources of information regarding the Early Universe, since it allows us to know the conditions of the Universe at that time and study the conditions leading to it.

It was detected for the first time by Arno Penzias and Robert Woodrow Wilson ([1965](#)) while testing a radio antenna for satellite communication experiments. For this discovery they were awarded the Nobel Prize in Physics in 1978. An image of the anisotropies of the CMB can be seen in [figure 1.3](#), as measured by the Planck Collaboration.

1.2.1 Physics of the CMB

Before the recombination, the Universe was filled by a plasma of ionised H and He (and small traces of heavier elements). The cross section of this charged particles with photons is large, σ_T , so both components are coupled. This means that photons interact frequently with matter and the Universe is opaque.

Additionally, the frequency of interactions guarantees local⁴ thermodynamic equilibrium. The photon population follows then a thermal distribution, also known as *Black Body Radiation*, given by the Planck's Law:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1} \quad (1.11)$$

where h is the Planck constant, c is the speed of light in vacuum, k_B is the Boltzmann constant, ν is the frequency, T is the temperature of the Black Body, and $B_\nu(T)$ is the spectral radial density. This last quantity is the power per solid angle per area normal to propagation per frequency and can be expressed in terms of the number density of photons and their energy:

$$\frac{dN}{d\nu} = \frac{4\pi}{c} \frac{B_\nu(T)}{h\nu} \quad (1.12)$$

This photon distribution corresponds to the distribution during recombination. The photons that we measure are mostly unaffected by matter since then, but the Universe has expanded more than 1000 times. Let us study what effect does this expansion have on the photon energy distribution. We know that the number density of any particle evolves like a^{-3} . On the other hand, the cosmological redshift causes that $\nu \sim a^{-1}$. This means that if a initial distribution have (N, ν) at scale factor a , the distribution today ($a_0 = 1$) will have $(Na^3, \nu a)$. Writing the evolution of the photon number density we get:

$$\begin{aligned} \frac{dN_{obs}}{d\nu_{obs}} &= \frac{d(Na^3)}{d(\nu a)} = a^2 \frac{dN}{d\nu} = \\ &= a^2 \left[\frac{4\pi}{c} \frac{1}{h\nu} \right] \left[\frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1} \right] = \\ &= \left[\frac{4\pi}{c} \frac{1}{h(\nu a)} \right] \left[\frac{2h(\nu a)^3}{c^2} \frac{1}{e^{\frac{h(\nu a)}{k_B(Ta)}} - 1} \right] = \\ &= \left. \frac{dN}{d\nu} \right|_{T_{obs}=Ta} \end{aligned} \quad (1.13)$$

We then conclude that the observed photon distribution of the CMB today corresponds exactly to a Black Body with $T_{obs} = T_{\text{CMB}} a(t_{\text{CMB}})$. We can measure the temperature observed today with great accuracy to obtain $T_{obs} = 2.275$ K. On the other hand, it is possible to calculate the actual temperature of the plasma at recombination, with which we would be able to calculate the scale factor and redshift at which it happened. In figure 1.4 it can be seen the energy spectrum of the CMB as measured by COBE (Mather et al. 1990), matching perfectly a Black Body spectrum.

The recombination (and subsequent decoupling) happens due to the inability of the photons to keep the plasma ionised when the temperature drops below a

⁴Only local equilibrium is guaranteed here because distant regions of space may be causally disconnected. Nonetheless, observations of the CMB reveal that the thermodynamic equilibrium appears to be global. This would be the case if these regions were connected at earlier times but a very fast expansion drove them apart, such as in the Inflationary Theory.

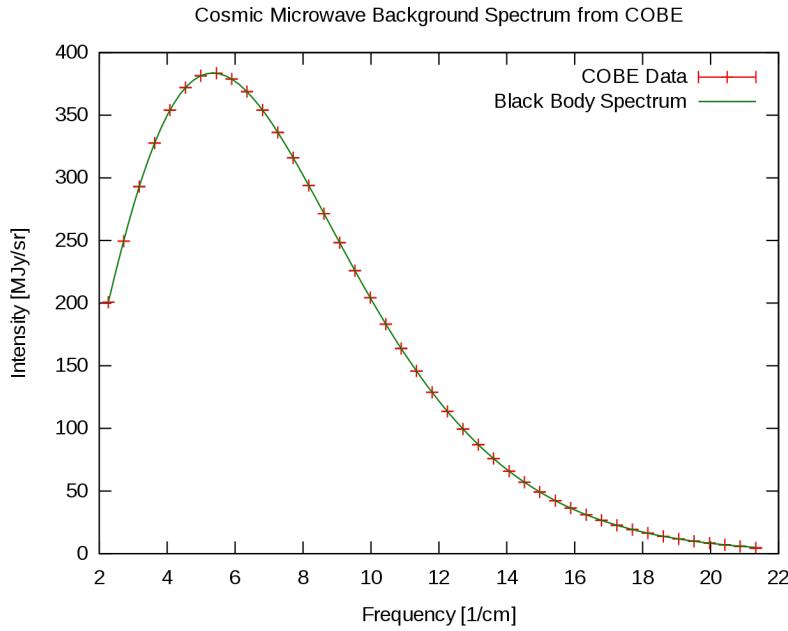


Figure 1.4: CMB spectrum as measured by Cosmic Background Explorer. The error bars on the data points (red) are thinner than the width of the line of the Black Body spectrum (green).

threshold. Naively, we could think that it will happen when the temperature of the photons is lower than 13.6 eV (1.6×10^5 K), ionisation energy of H. However this does not take into account that the number of photons n_γ vastly surpasses the number of baryons n_B :

$$\eta \equiv \frac{n_B}{n_\gamma} \approx 10^{-9} \quad (1.14)$$

Even at temperatures much lower than 13.6 eV, there will be a small fraction of the photons that have enough energy to ionise H according to the Black Body distribution. But given that the number of photons is so large, this small fraction can contain enough photons to ionise all the H. Therefore, we are more interested in the temperature at which $n_\gamma(E > 13.6 \text{ eV}) = n_B = \eta n_\gamma$.

At this temperature ($T \approx 5900$ K) the Universe is still completely ionised. More detailed calculations show that the production of neutral H is maximum at around $T = 3000$ K. This corresponds to a redshift $z = 1080$. It is important to notice that the transition of the plasma from ionised H to neutral H is fast, but not instantaneous. Similarly, the photons that we detect from the CMB do not all come from the same redshift. Different photons will have their last scattering process at different times. This is illustrated in Figure 1.5 through the visibility function $V(z)$: the fraction of detected photons that are last scattered at redshift z (see Chluba and Sunyaev 2006).

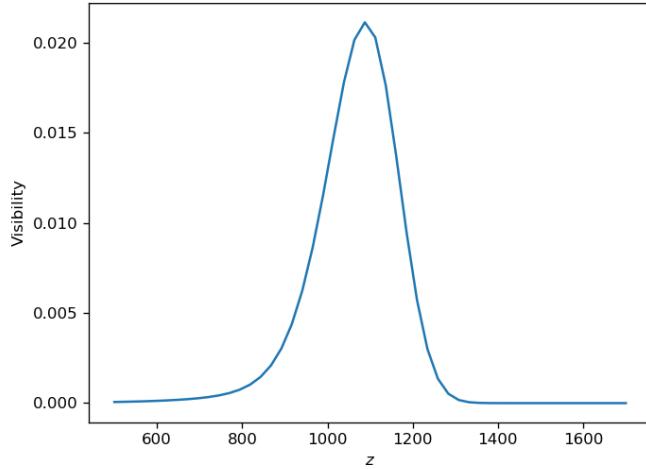


Figure 1.5: Visibility function as a function of redshift for the CMB photons: fraction of photons that are last scattered at redshift z . Calculated using the software CAMB, see [Section 3.2](#).

1.2.2 Power spectrum of the CMB

The importance of the study of the CMB does not only lay on spectral distribution of the photons, but also in the small temperature anisotropies that we can find in different parts of the sky. The CMB have a very uniform temperature of 2.725 K, but we can find some temperature differences at the level of one part in 1000.

The most important anisotropy is a dipole anisotropy, at the level of 3.3 mK. It is produced by the movement of the reference system of the observer with respect to the reference system of the CMB. With it, we can calculate that the velocity of the Earth is $(368 \pm 2) \text{ km s}^{-1}$ with respect to the CMB rest frame and the Local Group is moving at $(627 \pm 22) \text{ km s}^{-1}$. When this dipole is removed, the anisotropies are at the order of *hundreds* of μK , which is one part in 10 000.

The CMB is observed by us on the spherical sky, so the natural way to describe it is decomposing it into spherical harmonics. The idea behind it is analogous to a Fourier Transform, but on a spherical surface. Let $\hat{\gamma}$ be a direction on the sky, then we can write the temperature field T as:

$$T(\hat{\gamma}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{\gamma}) \quad (1.15)$$

where $Y_{\ell m}$ are the spherical harmonics⁵ of degree ℓ and order m , and $a_{\ell m}$ are complex number coefficients that can be obtained by multiplying the

⁵Spherical harmonics are defined as

$$Y_{\ell m}(\theta, \varphi) = \sqrt{\frac{(2\ell+1)}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} P_{\ell m}(\cos \theta) e^{im\varphi}$$

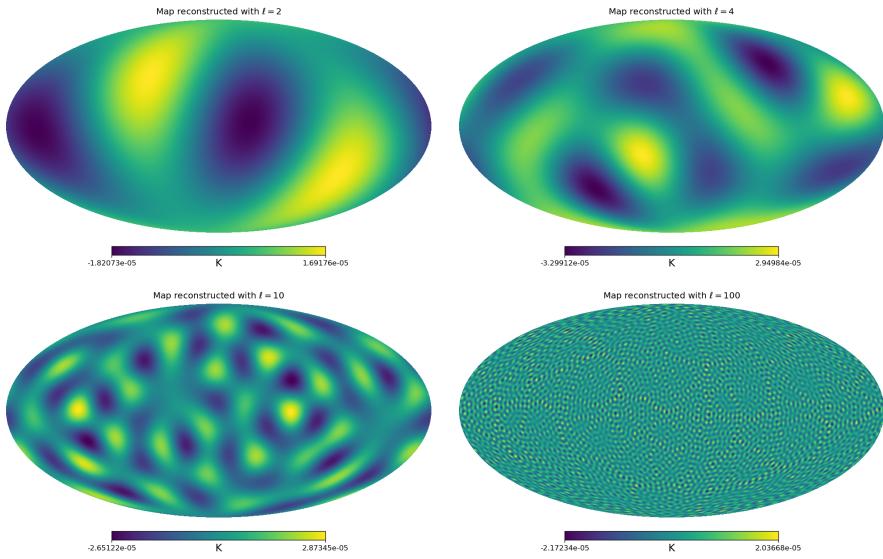


Figure 1.6: CMB temperature anisotropy map reconstructed with only $\ell = 2$ (top left), $\ell = 4$ (top right), $\ell = 10$ (bottom left) and $\ell = 100$ (bottom right). It can be seen that higher multipoles correspond to smaller scales.

temperature and the corresponding spherical harmonic and integrating over the entire sphere:

$$a_{\ell m} = \int_{\mathbb{S}^2} T(\hat{\gamma}) \overline{Y_{\ell m}}(\hat{\gamma}) d\Omega \quad (1.16)$$

where $\overline{Y_{\ell m}}$ is the complex conjugate of $Y_{\ell m}$.

The parameter ℓ is called the *multipole* and reflects the angular scale of the contribution to T . The monopole $\ell = 0$ corresponds to a uniform value on the sphere, $\ell = 1$ corresponds to the dipole, $\ell = 2$ is the quadrupole, and in general the multipole ℓ corresponds to the 2^ℓ -pole. On the other hand, m holds the direction information and takes integer values between $-\ell$ and ℓ , meaning that there are $2\ell + 1$ values of m for each multipole ℓ . In Figure 1.6 we can see several examples of CMB temperature maps reconstructed with only certain multipoles. More precisely, this is $\sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{\gamma})$ for multipoles $\ell = 2, 4, 10$ and 100 . We note in the figure that each ℓ presents spots of different scales, as we mentioned before. As a rule of thumb, the approximate scale of the spots for the multipole ℓ is $180^\circ/\ell$.

In order to compare the contributions of different multipoles, we can find the average of all the $a_{\ell m}$ for each ℓ . Because the values can be positive or negative, the average is done in a quadratic way. This magnitude is called *angular power*

where $P_{\ell m}$ are the associated Legendre polynomials:

$$P_{\ell}^m(x) = \frac{(-1)^m}{2^\ell \ell!} (1-x^2)^{m/2} \frac{d^{\ell+m}}{dx^{\ell+m}} (x^2 - 1)^\ell$$

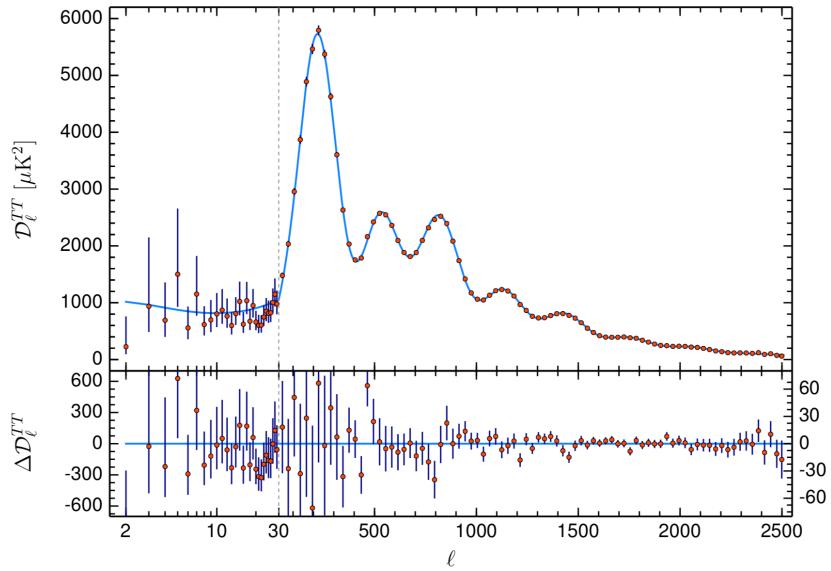


Figure 1.7: On the top, the angular power spectrum of the CMB as measured by Planck Collaboration (2018c). The y-axis corresponds to $D_\ell \equiv \frac{\ell(\ell+1)}{2\pi} C_\ell$. On the bottom, the difference with respect to the fit to the predicted by the Λ CDM model. Image credit: Planck Collaboration.

spectrum:

$$C_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \overline{a_{\ell m}} a_{\ell m} \quad (1.17)$$

We will not go into details, but this magnitude is also directly related to the angular covariance function, which tells us how correlated are two points separated by a certain angle.

We have to mention that this value, measured on a particular map, is an estimator of the physical C_ℓ . As the CMB map is a gaussian random map, the values of $a_{\ell m}$ are also random variables: two different realisations of the CMB map (with the same underlying C_ℓ) will present variations in $a_{\ell m}$ and in the measured C_ℓ . The observed CMB map is only one realisation of all the possible maps, so we have to use this map to infer the properties of the physical *ensemble average*.

Now we are going to talk briefly about the contributions of different processes to the angular power spectrum at different scales. In figure 1.7 we have the latest determination of the CMB angular power spectrum by Planck Collaboration (2018c). The behaviour is usually separated in three regimes:

- **Large angular scales, $\ell \leq 30$:** we mentioned before that $\ell = 0$ and $\ell = 1$ are the monopole and the dipole, respectively. For $2 < \ell \leq 30$, the anisotropies are mainly caused by the *Sachs-Wolfe effect*. During the recombination, there are density perturbations that cause some regions to have a deeper gravitational well. The photons present in these regions

will have to climb this potential well after the decoupling. Because of this, the photons that come from these regions will be less energetic than the photons coming from regions with underdensity. This energy difference is perceived in the CMB as a temperature difference. The effect is related with the total mass in a perturbation, making it the dominant effect at large scales.

The shape of the angular power spectrum in these part of the spectrum is directly related to the distribution of perturbations at large scale. If they are scale invariant, we would expect to see a completely flat region of the angular power spectrum when plotting the quantities in [figure 1.7](#). As we notice, this part is indeed almost flat.

There are two additional effects that are worth mentioning. The first one is the possibility of primordial gravitational waves as the ones predicted by inflationary theories. Although less prominent than the Sachs-Wolfe effect, primordial gravitational waves would contribute in the same way. The second effect is a *secondary* anisotropy, meaning that the physical origin is not during recombination, but later, during the path of the photons. In this case, the *integrated Sachs-Wolfe* and the *Rees-Sciama* effects are caused by the evolution of the gravitational well of large galaxy clusters. The difference in potential energy from the moment a photon enters it and the moment it leaves the cluster will produce a change in the energy of the photon.

Experimentally, we can see in [equation \(1.17\)](#) that the measure of C_ℓ at this scale involves an average over a small number of $a_{\ell m}$. For each ℓ there are only $2\ell + 1$ different values of m . At low ℓ , this means that the random nature of $a_{\ell m}$ can have a significant impact on the value of C_ℓ for a single realisation. This is due to the fact that we are considering large scales and the sky surface is limited to $4\pi \text{ sr}$, so we have a reduced number of independent measurements and therefore an increased uncertainty. In other words, the specific realisation of the CMB map in our Universe can moderately deviate from the physical value of the ensemble average. We say then that this multipole region is affected by *cosmic variance*.

- **Intermediate angular scales, $30 < \ell \leq 1000$:** this scale is dominated by the *acoustic peaks*. Before the recombination, density perturbations are linked to pressure perturbations since the radiation follows a equation of state $p = \rho c^2/3$. A positive density perturbation tends to collapse due to gravity, but when the density grows, so do the pressure and temperature in an adiabatic compression. The pressure now produce an outwards force that neutralise the gravity. In this way, the perturbations follow an oscillatory behaviour. At the recombination, the photons are free to escape the perturbations, halting the pressure and allowing the perturbations to collapse by gravity. However, the state of the perturbations at the recombination leaves an imprint on the photons, since the temperature is different for different perturbations.

The period of this oscillatory behaviour is larger the larger is the size of the perturbation. At very large scale, perturbations do not have time to compress once before recombination happens. As the scale decrease, we will find one specific size that have time to compress exactly once. This

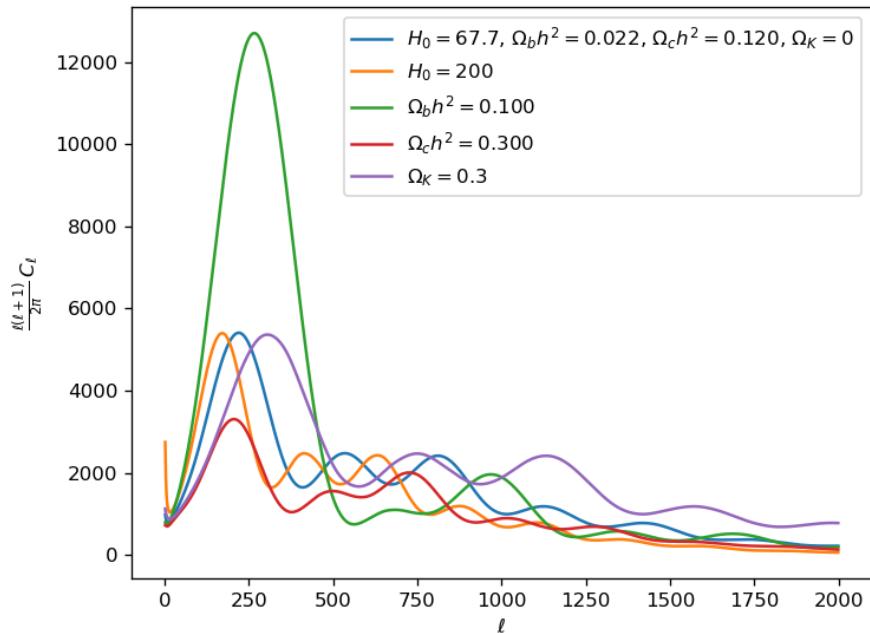


Figure 1.8: Simulated power spectra for Universes with different cosmological parameters. The blue line corresponds to the values measured by Planck. Other colours correspond to an increase of a single parameter, the parameters not indicated in the legend remain the same as in the blue line. It can be seen that the location and amplitude of the acoustic peaks are very sensitive to variations of the cosmological parameters. Units for H_0 are $\frac{\text{km}}{\text{s Mpc}}$, other constants are adimensional. Simulation done with the software CAMB, see [Section 3.2](#).

scale is the one that produce the largest signal, as they are the largest perturbations that are maximally compressed. This produces the main peak at $\ell \sim 200$, $\theta \sim 1^\circ$. Smaller scales will have time to compress and expand once, producing the second peak; even smaller scales will compress twice (third peak) and so on.

Additionally, the compression and expansion of the perturbation creates a movement of the plasma that produces Doppler effect on these photons. The effect is smaller than the temperature variation but operates out of phase with it: the maximum velocity is when the perturbation is going from compressed to rarefied. This explains that the troughs are not zero, since there is a contribution from Doppler effect.

Acoustic peaks on the power spectrum are very sensitive to some cosmological parameters, as it can be seen in [figure 1.8](#).

- **Small angular scales, $1000 < \ell$:** the power spectrum in this region still presents some acoustic peaks, but at lower intensities. This damping is produced mainly by two effects. The first one is due to the fact that not all photons come exactly from the same state of the Universe, the same

redshift z (see [figure 1.5](#)). This means that, for very small perturbations, it is possible that there are several of them in our line of sight. We no longer detect photons coming from a single perturbation, but several of them together. These perturbations can be at different state of the oscillation, which erase the temperature signal that we describe before. This produces a damping of the signal that increases with ℓ .

The other effect is a viscosity effect between the radiation and the baryonic plasma. For very small perturbations, a significant amount of the photons are able to escape. Since photons and baryons are coupled before the recombination, this can erase the perturbation. This is called *Silk damping* and its effect also increases with ℓ .

An additional secondary effect is due to the reionisation phase, since ionised H present from $z \lesssim 10$ can scatter photons coming from the CMB, blurring the signal. Scattered photons do not have time to travel large distances, so the effect is limited to the small scale.

From an experimental point of view, the high end of the power spectrum is limited by the size of the pixel. Additionally, when the scale of the multipole approaches the size of the pixel, the determination of individual $a_{\ell m}$ can be dominated by instrumental noise.

1.2.3 Satellite observations of the CMB

Since its discovery, the CMB has provided useful data to constrain cosmological models and know more about the Early Universe. First, the measurement of its temperature and now, the measurements of the anisotropies. Among the multiple experiments done and designed to observe the CMB, there are ground (like BICEP in the South Pole), balloon (like BOOMERanG), and satellite experiments⁶. Especially important are the three multi-band satellite observations, in which we are going to focus: *COBE*, *WMAP*, and *Planck*.

Cosmic Background Explorer (COBE) was the first of the three, observing from 1989 to 1993 and operated by NASA. One of the main results provided by COBE is the measurement of the nearly perfect black-body spectrum, as seen before on [figure 1.4](#) in [Section 1.2.1](#). The other main result is the discovery of faint anisotropies in the temperature map, although the limited angular resolution did not allow to resolve structures under 7° , such as the first acoustic peak (see [Section 1.2.2](#)). It can be considered the starting point of precision cosmology.

Wilkinson Microwave Anisotropy Probe (WMAP) operated from 2001 until 2010 by NASA. The angular resolution was massively improved with respect to COBE, reaching $15'$, enough to observe several acoustic peaks and set constraints on the cosmological parameters.

Planck is the most recent one and the last of this kind, as there are not plans to launch a successor that can observe the full sky with better resolution. It was operated by ESA from 2009 to 2013. It had an angular resolution of $5'$, which allows it to extent the power spectrum obtained by WMAP. This allows the calculation of more precise constraints on the cosmological parameters. The

⁶In this section we will only talk about the multi-band satellite observations, but a curious reader can find a more complete list of CMB experiments in https://en.wikipedia.org/wiki/List_of_cosmic_microwave_background_experiments

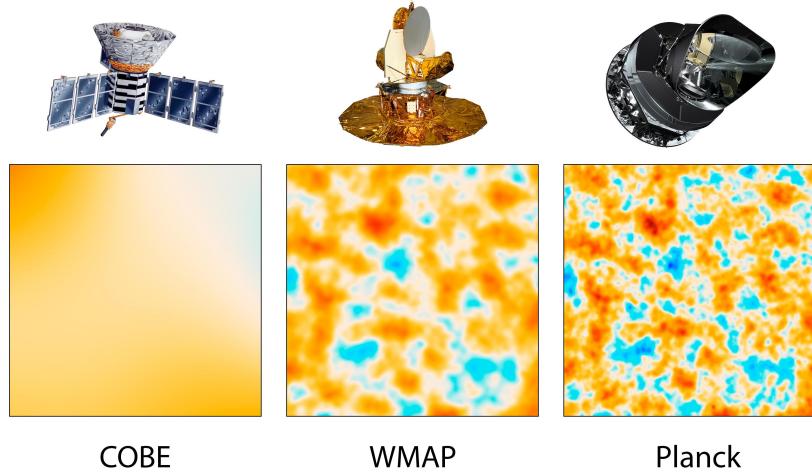


Figure 1.9: Comparison between the temperature anisotropies measured by COBE, WMAP and Planck. The images correspond to a 10 square degrees of the sky. Image credit: NASA/JPL-Caltech/ESA.

second data release of Planck was on 2015, although the final data release has been released on July 2018, during the preparation of this work. We will analyse both releases in this work.

A comparison between the angular resolutions of the three satellites can be seen in [figure 1.9](#). This difference in angular resolution imply a difference in the maximum multipole ℓ for which the power spectrum can be calculated with confidence. For COBE, it was $\ell \lesssim 30$; for WMAP, $\ell \lesssim 1100$; and for Planck, $\ell \lesssim 2500$.

In this work, we are going to use only Planck data. The CMB temperature map reported by the Planck team can be seen in [figure 1.3](#). We will see in [Section 1.3.3](#) that there are different ways to produce this map, producing slightly different results. Lastly, the values for the cosmological constants reported by Planck final data release can be found in [table 1.1](#). The best fit and the 68% confidence regions are reported as by Planck Collaboration ([2018a](#)).

When observing the light of the CMB, Planck can measure not only the intensity, but also its linear polarisation. One standard way to describe this polarisation is by using the *Stokes* parameters: I for the total intensity, Q and U for the linear polarisation (encoding intensity and direction) and V for circular polarization. The latter is not measured by Planck as no circular polarisation is expected from the CMB. Maps for the Q and U parameters are reported by the Planck Collaboration. This polarisation can be express as a curl-free field called E and a divergenceless field called B , by analogy to the electromagnetic field.

These maps can be analysed in a similar way to the intensity map, but they have a different aspect and power spectrum, being dominated by different physical processes. In particular, theory predicts that B -modes could have been produced during Inflation. There are several future missions, like LiteBIRD, that will observe the CMB in polarised light. Although polarisation maps are a

Parameter	Planck best fit	Planck 68% region
$\Omega_b h^2$	0.022383	0.02237 ± 0.00015
$\Omega_C h^2$	0.12011	0.1200 ± 0.0012
$100\theta_{MC}$	1.040909	1.04092 ± 0.00031
τ	0.0543	0.0544 ± 0.0073
$\ln(10^{10} A_S)$	3.0448	3.044 ± 0.014
n_S	0.96605	0.9649 ± 0.0042
$H_0 [\text{km s}^{-1} \text{ Mpc}^{-1}]$	67.32	67.36 ± 0.54
Ω_Λ	0.6842	0.6847 ± 0.0073
Ω_M	0.3158	0.3153 ± 0.0073
σ_8	0.8120	0.8111 ± 0.0060
z_{rec}	7.68	7.67 ± 0.73
z_{eq}	—	3402±26
Age [Gyr]	13.7971	13.797 ± 0.73

Table 1.1: Value of cosmological parameters as determined by the Planck Collaboration (2018a). On the top, the six independent parameters of the Λ CDM model, as explained in Section 1.1.3. On the bottom, another seven important cosmological parameters derived from the former. It is reported both the best fit and the 68% confidence region.

very interesting aspect of the study of the CMB, we postpone the application of our techniques on them and we will focus on intensity maps in this work.

1.3 Foregrounds

Measuring the CMB is not as simple as pointing a telescope to the sky in the microwave frequency. Unfortunately for the study of CMB, there are other components closer to us that emit in the same range of the electromagnetic spectrum. These components, however, do not emit with the same spectral distribution, so their effect can be reduced by observing in a wide range of the spectrum and extracting only the Black Body behaviour of the CMB. In order to do this, Planck observes in nine different bands. Three bands with the Low Frequency Instrument (LFI): 30 GHz, 44 GHz and 70 GHz, with bandwidth $\Delta\nu/\nu = 0.2$; and six bands with the High Frequency Instrument (HFI): 100 GHz, 143 GHz, 217 GHz, 353 GHz, 545 GHz and 857 GHz, with bandwidth $\Delta\nu/\nu = 0.33$. The intensity maps for all the individual bands can be seen in figure 1.10. It is apparent that different components have different spectral behaviour, being more prominent in some specific bands.

It is interesting to note that CMB observations are also useful in order to study the foregrounds themselves. For example, this kind of experiment could lead to the first direct observations of the Planet 9 in our own Solar System, as suggested by Cowan et al. (2016), or to the detection of Oort clouds of stars in the neighbourhood of the Sun, as suggested by Baxter et al. (2018).

Due to its mission design and being a satellite, Planck completely avoid certain kinds of foreground. These include the Earth atmosphere, thermal radiation coming from the ground, radio frequency interferences and the effect

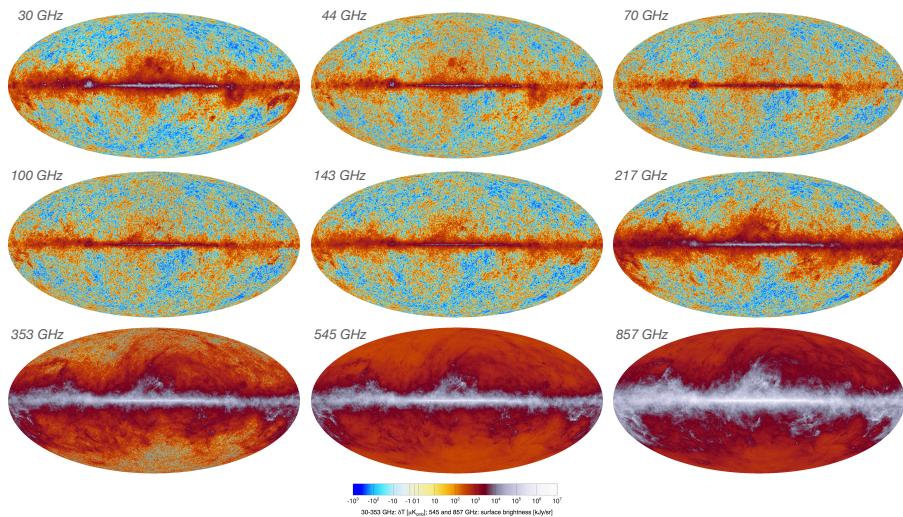


Figure 1.10: CMB intensity maps in the nine frequency bands sampled by Planck. We note the difference in the galactic foregrounds at different frequencies. Image credit: Planck Collaboration.

of the Sun, Moon, and other Solar System objects. It has to be noted that thermal emission from the components of Planck was reduced by cooling them with liquid He. The depletion of this coolant was the reason that motivated the end of the observation, since the thermal emission from the satellite, even if it is only of *a few K*, would be enough to render the observation of the CMB impossible.

Once these problems have been solved, there are two kinds of foregrounds that can not be avoided. The first one is the diffuse galactic emission, since we have to observe from within the galaxy. This kind of foreground is dominant in the galactic plane and lower galactic altitudes, the regions where most of the emission is located. It affects large scales and it causes the uncertainty of the CMB to be higher in this region. This emission highly depends on the frequency, as it can be seen in [figure 1.10](#).

The second kind of foregrounds is due to extragalactic sources. From the telescope until the CMB, we can find the whole observable Universe. Of course, there are objects in the Universe that can emit significantly in this spectral region. In particular, different types of galaxies can affect the observation of the CMB. In this case, we are talking about an isotropic distribution of typically⁷ unresolved sources. Therefore, this contribution is observed as a foreground of Point Sources (PS). Their effect is limited to very small scales but is not restricted to a specific region.

⁷Although most galaxies are unresolved, some very close galaxies can be resolved. This is the case of M31, Andromeda. This galaxy is studied in the context of Planck data in Planck Collaboration (2015). Images of the galaxy in different bands can be found in figure 2 of that reference.

1.3.1 Diffuse Galactic Radiation

Galactic radiation in this range of frequencies can be due to several distinct processes. In this section we are going to list and briefly explain these processes, but we refer the literature for more information, such as Ichiki (2014) and Dickinson (2016), who also treat the effect of this radiation on polarisation maps. The main processes involved in the diffuse galactic radiation are:

- **Synchrotron radiation:** it is produced when charged particles interact with a magnetic field. In this case, cosmic ray electrons interact with the large scale magnetic fields of our galaxy. Electrons inside a magnetic field will spiral around the lines of the field, which produces the emission of photons. The energy distribution of the electrons is well described by a power law, which produces a temperature signal $T_{syn} \propto \nu^\gamma$ with temperature spectral index $\gamma \sim -3.0$ at ~ 10 GHz and spatial variations of $\Delta\gamma \sim -0.2$. Therefore, its importance is mainly at the lowest frequency bands. Although its effect is secondary in temperature maps, its high degree of polarisation (10 to 40% at high latitudes) makes it the main effect at low frequency for polarisation maps.
- **Free-free radiation:** also known as thermal bremsstrahlung, is produced by free electrons that are accelerated by ions, usually protons, in an ionised gas. This process has a well-understood effect on the temperature maps, with $T_{f-f} \propto \nu^\gamma$ and temperature spectral index $\gamma \sim -2.1$. This means that it is important at low frequencies, and it could be the main component of the foreground for the 70 GHz frequency band in Planck, the band with less foreground contribution. It can be difficult to separate from other components from Planck data alone, but it is known that this effect is correlated with the optical H α emission.
- **Thermal dust radiation:** interstellar dust is composed mainly by graphites, silicates and polycyclic aromatic hydrocarbons. This dust has a temperature of around 15 K in the outer parts of the galaxy, which increases towards the galactic centre up to ~ 19 K. We therefore expect to see thermal emission from this dust. The exact emission depends on the shape of the dust, but it is well described by $I = \nu^\gamma B_\nu$, where B_ν is the Black Body spectrum and $\gamma \sim 1.6$. Its effect is then more important at higher frequencies, from $\nu \gtrsim 70$ GHz, where it is in fact the main component of the foreground.
- **Spinning dust radiation:** some dust particles can be naturally rotating, since there are barely any forces that stop them. The smallest particles, of the order of 10^{-9} m, can reach a rotation frequency of the order of GHz. If these particles have an electric dipole moment, they will emit by electric dipole radiation. They peak at around 30 GHz, just in the lowest frequency band observed by Planck, and the contribution there is significant. At higher frequencies, the effect of the spinning dust radiation decreases sharply.
- **CO molecular cloud radiation:** initially underestimated, the radiation coming from the rotational transition of carbon monoxide (CO) can be an important foreground in the 100 GHz band. The molecular CO radiation

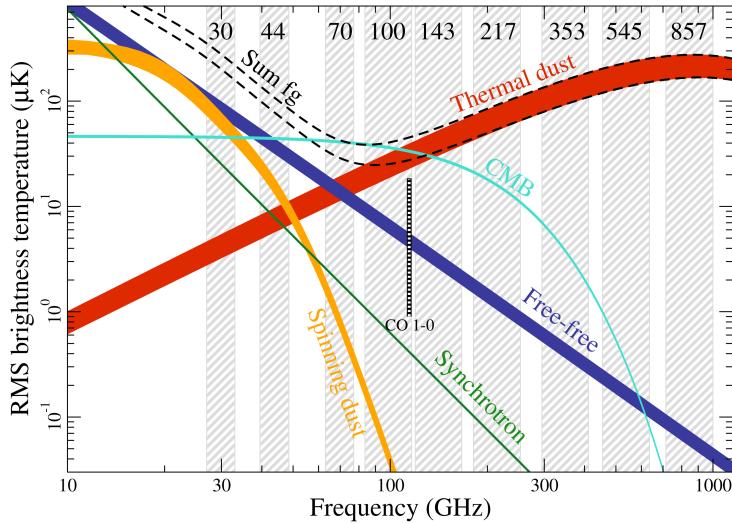


Figure 1.11: Contribution of different diffuse galactic foregrounds as a function of the frequency. The comparison is done for scales of 1° outside the galactic plane. As we explain in the text, thermal dust emission dominates at high frequencies, while the other contributions dominate at low frequencies. The CMB is best observed in the 70 GHz and 100 GHz bands. Image credit: Planck Collaboration.

is sharply localised at the energy of the rotational transitions. The first one, $J = (0 - 1)$, is at 115 GHz. The second and third ones, although less important, also fall in the 217 GHz to 353 GHz bands. CO clouds are associated with the dust distribution, but the CO clouds themselves are more localised and can contribute to the power spectrum at $\ell \gtrsim 1000$.

It can be noted that the foreground components, being different in nature, will affect the observations in different ways. Just as an example, we could ask the relative contribution to the temperature map of each component and the CMB for a given scale of 1° . We can find this comparison in figure 1.11, where we can clearly see the frequency dependence of the different components mentioned in this section.

1.3.2 Point Sources

Aside from the diffuse galactic foreground, the CMB is also contaminated by a foreground of extragalactic Point Sources (PS). These have an effect mostly at small scales (high multipoles), but if there is a large number of PS, they can also increase the overall signal at lower multipoles. They are unresolved, so their size is comparable to the effective beam, up to $\sim 5'$ for Planck.

Detecting them, however, is not as straightforward as in the visible frequencies. Here, the CMB itself can have random fluctuations that look like a PS. Therefore, the detection of PS is always probabilistic, we have to set a criterion for the detection.

The Planck Collaboration, in the second data release (Planck Collaboration 2016), published a catalogue with the PS detected on the frequency maps. For each band, there is a catalogue with a list of points identified as PS at a confidence of 80%, with flags for higher confidence values. Additionally, for each of the 6 bands in the HFI, there are other catalogue for points whose confidence could not been determined, mostly at low galactic latitudes. Three other catalogues are provided using the joint information of several band: objects at high z , regions with Sunyaev-Zeldovich effect, and galactic Cold Clumps.

The nature of these PS is diverse, but they are mainly of two kinds:

- Radio galaxies: this is the name given to galaxies which are very luminous in the radio part of the spectrum. They have two twin jets created by an Active Galactic Nuclei (AGN), a compact central region dominated by a supermassive black hole. These jets have free electrons that interact with the strong magnetic fields present there. This creates a significant amount of synchrotron radiation, as in the case of the galactic radiation. These objects are very luminous and can be detected at large distances, making them useful for some cosmological studies.

In our case, this kind of PS is mainly present at low frequencies, as the galactic synchrotron radiation, since their spectral index is negative. The distribution of sources per luminosity bin of these sources can be seen in figure 1.12, where it is also compared to previous estimates.

- Dusty galaxies: some galaxies have a significant amount of dust that, like in our galaxy, emits thermal dust radiation. These sources are generally closer than the previous kind, as they are not as luminous. The spectral index of these objects is positive and, therefore, dusty galaxies are present mainly at higher frequencies.

One way to distinguish between the two types is, as we mentioned, the spectral index. In figure 1.13 we can see the approximate spectral index of the PS present in several consecutive bands. We can see that the most part of the sources are radio galaxies for frequency bands up to $\sim 217\text{ GHz}$ and dusty galaxies for higher frequencies, as we expected. We also note a difference between the lowest frequencies, where there is a known steepening of the spectral index for radio sources, even if both are produced by synchrotron radiation.

1.3.3 Algorithms for component separation

We have seen that the CMB foreground can be very complex and dominate over the signal in some regions of the sky. Therefore, it is very important to be able to separate the radiation coming from each component. This is called *component separation* and it is an entire field on its own. The Planck Collaboration apply four different independent algorithms to extract the CMB from observations. Here we are going to briefly describe the main aspects of each one. We refer curious readers to Planck Collaboration (2018b, in particular section 2 and appendices A to D), where they describe the algorithms used for the last release, as well as the Planck Legacy Archive internal wiki (2018e).

- **COMMANDER:** this algorithm (Eriksen et al. 2004) uses parametric Bayesian Statistics to extract the contribution of each foreground in each pixel. If

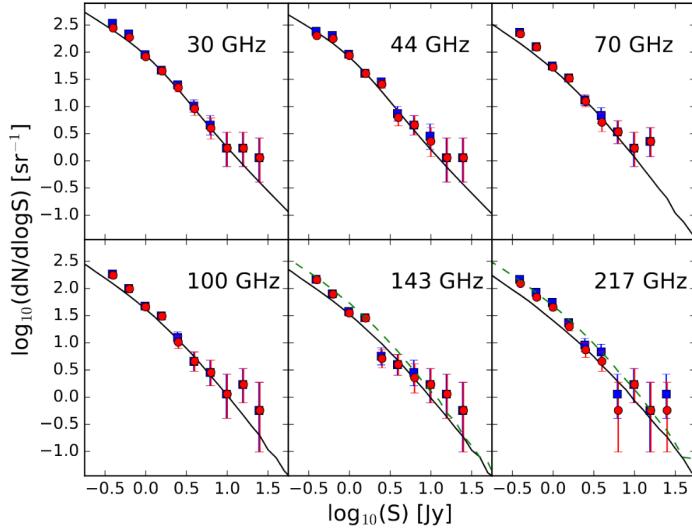


Figure 1.12: Number of detected sources per luminosity bin of radio galaxies for each frequency band. The distribution is approximately linear in log-log scale. Blue squares are detections at minimum 4σ outside the galactic plane, and red circles are sources estimated with a low frequency counterpart. Lines correspond to number counts models by Tucci et al. (2011) for radio sources only, which dominate at these frequencies. Image credit: Planck Collaboration (2018d).

we call $\mathbf{d}_\nu(p)$ to the observed data at frequency ν and pixel p , we have the following generic model:

$$\mathbf{d}_\nu(p) = g_\nu \sum_{C=1}^{N_C} F_\nu(\gamma_C) T(p) \mathbf{a}_C + \mathbf{n}_\nu(p) \quad (1.18)$$

where N_C is the number of components considered, each one (labelled by C) with an amplitude at certain frequency \mathbf{a}_C , $T(p)$ is a map projector operator, $F_\nu(\gamma_C)$ is an operator that scales the intensity depending on the frequency ν and the spectral parameters γ_C of the component, g_ν is a calibration parameter for each frequency, and $\mathbf{n}_\nu(p)$ is the instrumental noise.

This model depends on a reduced number of both global parameters and parameters for the amplitude and characteristics of each component. This makes it a perfect target for the Bayesian determination of these parameters. The diffuse components considered are only four: the CMB, a general low-frequency power-law component, thermal dust radiation and CO emission. Ultimately, the goal is to accurately determine the parameters corresponding to the CMB in order to reconstruct the CMB temperature map.

For the 2018 data release, the entire code has been rewritten. One of the most significant differences is that it operates in spherical harmonics

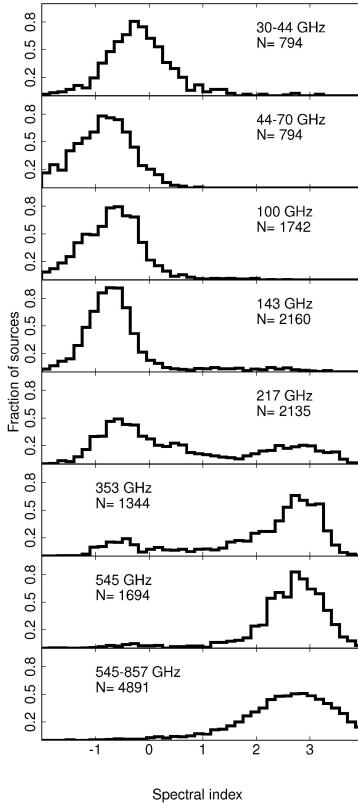


Figure 1.13: Histogram of the spectral index of the PS detected in several bands. At lower frequencies, the dominant spectral index is negative, signal that the mechanism involved is synchrotron radiation. At higher frequencies, most sources have positive spectral index, signal that the driving mechanism is thermal dust radiation. Image credit: Planck Collaboration (2015).

space instead of pixel space, which allow different components to have different instrumental beaming. The priors for the parameters are chosen to be minimally informative when possible.

- **NILC** (Needlet Internal Linear Combination): this algorithm (Basak and Delabrouille 2013) takes advantage of needlets to minimise the contribution of large scale diffuse components and very small scale extragalactic sources or instrumental noise. We will explain needlets in detail later, in Section 2.1, as they are central to our work as well. For now, it suffices to mention that needlets allow a very localised filtering of the map, both in pixel and multipole space, for different scales.

The algorithm filters all the frequency maps with needlets of different scales, knowing that the effect of filtering the CMB can be calculated. Assuming the CMB signal is uncorrelated with the foregrounds, a linear combination of the different frequencies can be obtained. This is done by minimising the variance of the map and forcing the underlying CMB to be the same for all filters, also behaving like a perfect Black Body with

frequency. The variance in one pixel is computed not only using that pixel, but a small gaussian window whose size depend on the scale of the needlet. In this way, the effect of large diffuse components and noise is reduced.

It has to be noted that there is a preprocessing of the data before applying the algorithm. On the one hand, all frequency maps are convolved or deconvolved so they all have the same angular resolution of $5'$. On the other hand, a small fraction of the sky (0.2%) in the galactic plane is removed and refilled using the Planck Sky Model (Delabrouille, Betoule, et al. 2013). This is done to erase pixels where the value is too high.

- **SEVEM** (Spectral Estimation Via Expectation Maximisation): this algorithm (Fernández-Cobos et al. 2012) tries to clean the foreground of the frequency maps using templates derived from the maps themselves. If $\mathbf{d}_\nu(p)$ is again the observed data at frequency ν and pixel p , we have that:

$$\mathbf{T}_\nu(p) = \mathbf{d}_\nu(p) - \sum_{n=1}^N \alpha_n \mathbf{t}_n(p) \quad (1.19)$$

where $T_\nu(p)$ is the cleaned map at frequency ν and pixel p , n labels the templates $t_n(p)$, of which there are N , and α_n is a coefficient to be found to minimise the variance outside a mask.

In this case, there are a total of four templates, three of them are the subtraction of the consecutive Planck channels (30 – 44, 44 – 70 and 545 – 353), where both are convolved with the beam of the other; and the other template is obtained just from convolving the 857 band with the 545 beam. The maps that are cleaned are the 70 GHz, 100 GHz, 143 GHz and 217 GHz bands. The final map is a combination of the last two to avoid the effect of CO in the 100 GHz band and the lower resolution of the 70 GHz band.

We have to note that the maps are additionally processed to reduce the contamination from PS. In order to do that, a Mexican-Hat-Wavelet algorithm is run on the frequency maps used to calculate the templates, and the detected PS are removed, inpainting the holes. This is repeated in the final maps, once the templates have been subtracted.

- **SMICA** (Spectral Matching Independent Component Analysis): this algorithm (Delabrouille, Cardoso, et al. 2003) combines the contribution of the different frequency maps in a ℓ -dependent way to extract the best fit for the CMB temperature map. It does that by working in the harmonic space. First, all the auto- and cross-power spectra are estimated, and then this information is combined with the expected behaviour of a Black Body to obtain the weight of each frequency map for each multipole ℓ .

In this case it is assumed perfect calibrations for all frequencies, but the emissivities and power spectra of the foregrounds are freely fitted by the algorithm, as well as their possible correlations. However, different foregrounds are not reconstructed separately since the coupling between them is significant.

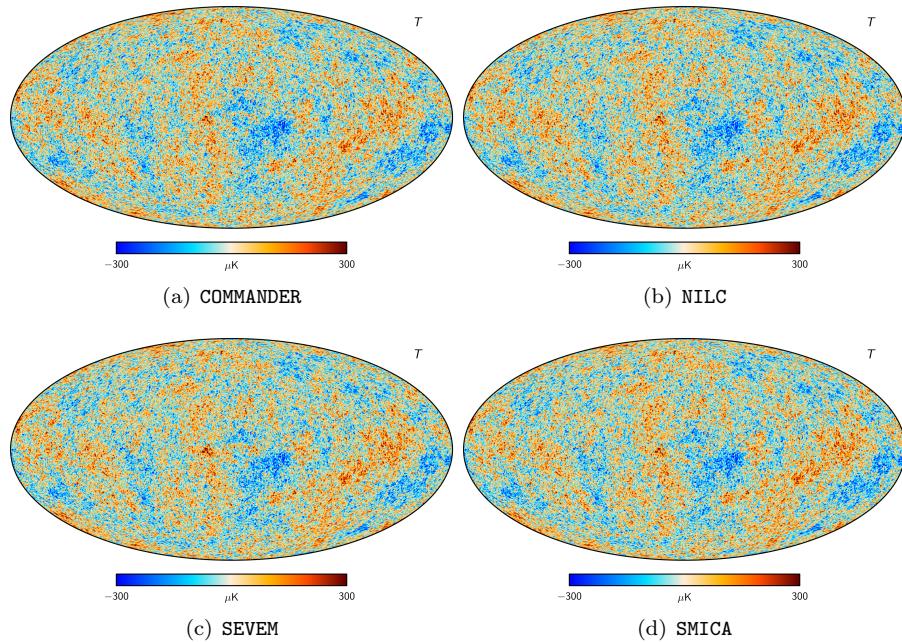


Figure 1.14: Temperature anisotropies maps produced by the four different algorithms. Seeing the differences with the naked eye is almost impossible: they are practically identical even if the algorithms are based on very different approaches. The differences are most notable in the galactic plane and in very small scales. Image credit: Planck Collaboration.

For the last Planck release, the algorithm produces two temperature maps that are merged later. One map is extracted from the HFI frequency maps and is intended to describe better the cleanest region of the sky in intermediate and small scales. The other map is intended to describe better the whole sky at all scales, using all HFI and LFI frequencies. In this way they avoid the possible effect that the foreground can have in the weights of each frequency when reconstructing regions dominated by the CMB.

The preprocessing for this algorithm also includes the removal of sources. In particular, PS found in the Planck Catalogue at 5σ or more are removed, and any point above 50σ is likewise removed. Additionally, a small Galactic mask is added. All pixels within these masks are removed and inpainted to avoid effects on the spherical harmonic coefficients.

The final temperature maps for the four different algorithms can be seen in figure 1.14. From these maps we can obtain the power spectra. The ratio between the values of C_ℓ and the average value between different algorithms can be seen in figure 1.15.

As we have seen, the four algorithms are based on very different techniques. In order to analyse the CMB, it is recommended to do it in the regions where the values are similar between different algorithms and where additional processing

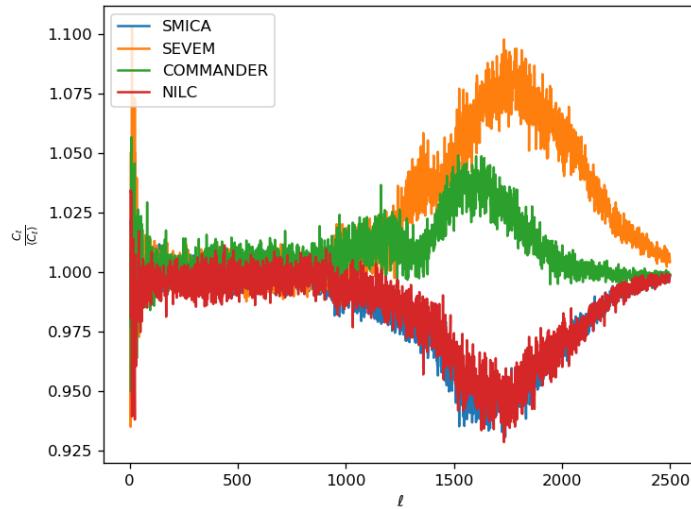


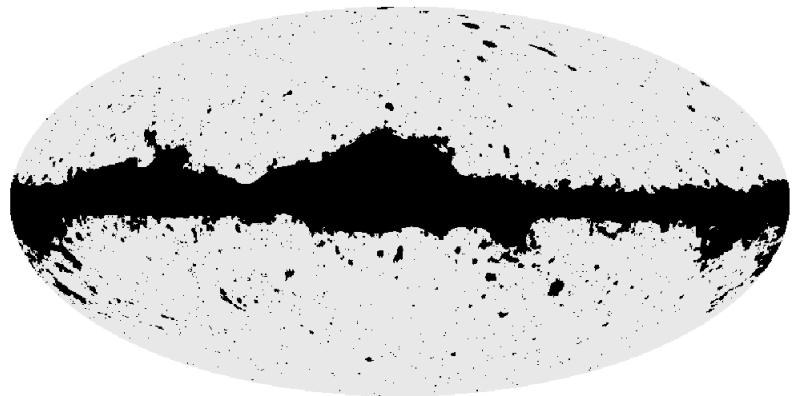
Figure 1.15: Comparison between the power spectra obtained for the maps generated by different algorithms. In the y-axis, it is represented the ratio of the power spectrum of the algorithm and the mean of the power spectra for the four methods. The differences become evident from $\ell \sim 1000$, but the differences are always within 10% of the mean value. It is important to note that the value of C_ℓ in this range is very small, so any variation will become amplified.

(as the removal of PS and inpainting) has not taken place. In order to avoid these regions, the Planck Collaboration also provide a *confidence mask* in their last data release, which includes a fraction os the sky of 78.0%, free of problematic areas. In order to do that, they limit the standard variation between maps to $3 \mu\text{K}$ at large scales ($80'$ smoothing) and $10 \mu\text{K}$ at small scales ($10'$ smoothing). This takes into account the emission in the galactic plane and compact sources, respectively. This is then augmented with the masks used internally by the algorithms, such as the PS mask in **SEVEM** and **SMICA** or a goodness of fitness mask by **COMMANDER**.

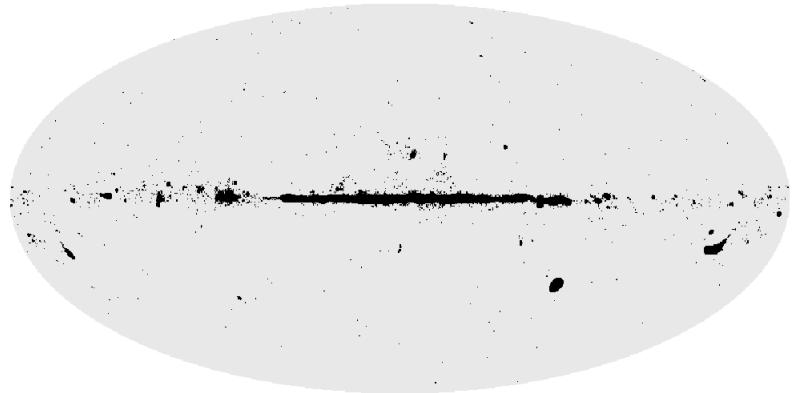
An additional mask is provided where the difference of any pair of maps is $100 \mu\text{K}$ plus the inpainting mask of **SMICA**. This mask excludes a 2% of the sky and it is the mask used to create *inpainted* temperature maps. Both masks can be seen in figure 1.16.

Scope of this work

In this work, we will analyse the Planck temperature maps produced by the different algorithms to search for significant deviations of a random gaussian map, expected of a purely CMB map. In particular, we are looking for PS that are not totally cleaned. We will use needlet techniques, as they are very efficient to study well localised points without being affected by contamination at a different location or scale. We use a multiple testing approach as we want to study significant deviations of the map as a whole, not point by point: if a maxima of the map have a high value that we expect to find only once on the



(a) Common Mask



(b) Inpainting Mask

Figure 1.16: Confidence masks to exclude regions of CMB maps that could be contaminated by foreground. On the top, the common mask, covers 22% of the sky. It is determined by the regions where the maps present diverging values (see text) or the confidence of a good CMB reconstruction is low. On the bottom, the inpainting mask, covers 2% of the sky. It is determined by large discrepancies between algorithms and areas that have been inpainted by **SMICA**. This mask is used to generate the inpainted maps for all algorithms.

map, that point is not a significant deviation of a gaussian map; but if we find a thousand points like that, the map as a whole is probably not a gaussian map.

We will discuss the results for the Planck temperature maps. In particular, we will discuss the maps from the four different algorithms in the second data release and in the last data release.

❖ Chapter Two ❖

Methodology

The goal of this work is to detect PS in the maps of the CMB temperature anisotropies. In order to do that, we use needlets and multiple testing. On the one hand, needlet techniques are useful tools to analyse spherical maps. They are very localised in space, so we can analyse different regions in a completely independent way, unlike spherical harmonics, whose value is given by the behaviour on the entire sphere. Additionally, they are very concentrated in multipole space so we can study different scales in independent ways, unlike in pixel space, where contributions of all scales are summed. These two properties are very useful in our case, since PS have a certain characteristic scale, comparable the resolution of the telescope; but also allows us to study them without having to worry about the behaviour of the map in the galactic plane or where the reconstruction of the CMB maps is not confident enough.

On the other hand, we will use multiple testing techniques to test if the maps differ from a random realisation of a gaussian map in a statistically significant way. The idea behind it is clear: a map can be very different from a random gaussian map even if each of its points are perfectly compatible with gaussianity. A very simple example would be a map composed only by points with values $1 \mu\text{K}$ and $-1 \mu\text{K}$, mean $0 \mu\text{K}$ and standard deviation $\sigma = 1 \mu\text{K}$: no value exceeds 1σ difference with the mean, each point is compatible with a gaussian distribution; but the map clearly does not follow such distribution, it only have two different values. In particular for our case, we want to test the distribution and intensity of the maxima, to test if the maxima have as a whole higher intensities than expected.

In this chapter, we will discuss and formalise these ideas. In [Section 2.1](#) we will introduce the concept of needlet and its advantages. We will formally define them and discuss their main properties. We will talk about the differences between standard and Mexican needlets and why we will use the later. To end the section, we will talk about the filtering with needlets and how to reconstruct the original map. In [Section 2.2](#) we will talk about the maxima in the maps and their theoretical distribution. Then we will discuss the concept of p -value, very useful in hypothesis testing. We will explain how we take into account the global distribution of maxima to determine which points, if any, are candidates to be actual PS. We will end the section with a very brief summary of the procedure, with focus on the implementation.

2.1 Needlets

Needlets are functions defined over the sphere that we will use to filter the CMB maps in order to extract the signal of a certain scale, like a convolution. They pose several advantages with respect to spherical harmonics or other wavelets techniques:

- They do not rely on tangent plane approximation, they are defined over the sphere and take advantage of its manifold structure.
- They are easily calculated from the spherical harmonics coefficients, which are already implemented in standard software for spherical images processing. This makes their computation very convenient.
- The reconstruction formula to recover the signal from the coefficients is simple and uses the same mathematical expression of the needlets: they appear in both the direct and inverse transformations: [equations \(2.6\)](#) and [\(2.11\)](#).
- They are at least quasi-exponentially concentrated in pixel space (faster than any polynomial).
- They are localised on multipole space. Depending on the kind of needlet, they can be exactly localised in a finite number of multipoles.
- Needlets for different scales are (asymptotically) uncorrelated. This means that different scales can be studied almost independently with very low correlation between them.

Needlets were introduced by Narcowich et al. (2006) in Mathematics and latter applied to the analysis of cosmological data sets by Baldi et al. (2009). A general discussion of the use of needlets for CMB analysis can be found in Marinucci et al. (2008) or Pietrobon et al. (2010).

2.1.1 Definition

We will start by formally defining the needlets functions. Let $\hat{\gamma}$ be a direction in the sky (equivalently, a point in the sphere). The needlet of frequency j (an integer number) and centred at the direction ξ_{jk} is defined as:

$$\psi_{jk}(\hat{\gamma}) = \sqrt{\lambda_{jk}} \sum_{\ell} b\left(\frac{\ell}{B^j}\right) \sum_{m=-\ell}^{\ell} \overline{Y_{\ell m}}(\hat{\gamma}) Y_{\ell m}(\xi_{jk}) \quad (2.1)$$

where the set $\{\xi_{jk}\}$ denote the so-called *cubature points* where the needlet will be centred, at frequency j (this set can be different for each frequency). The spherical harmonics are denoted by $Y_{\ell m}$ (see [footnote 5](#) for the formal definition), ℓ is the multipole and m runs between $-\ell$ and ℓ . B is a *bandwidth* parameter that determines how localised the needlet will be in multipole space. The function $b\left(\frac{\ell}{B^j}\right)$ is a filter that determines the weights for the multipoles considered for the needlet (we will talk about it in [Section 2.1.2](#)). Lastly, λ_{jk} is the weight for the point ξ_{jk} .

Before we proceed further, a word of caution may be in place. The parameter j receives the name of *frequency* of the needlet because it is in some way

analogous to the frequency of a Fourier decomposition. The idea is similar to the multipole: the higher j is, the smaller it is the needlet in pixel space. This would also increase the rate at which the filtered maps vary with position: they will oscillate faster. This parameter must not be confused with the *frequency of the light*, which we always denote ν .

In this work, we will make some simplifications. First of all, the cubature points, as it is usual, will be taken as the centres of the pixels as established by the Hierarchical Equal Area isoLatitude Pixelization (HEALPix) standard (see [Chapter 3](#)). We will work always with the same resolution, regardless of frequency j . Therefore $\{\xi_{jk}\} \equiv \{\xi_k\}_{HEALPIX}$, where k runs over the $N_p = 12 nside^2$ number of pixels. We will set the weights of all pixels to the same value of $1/N_p$.

In some cases it is also common to apply the addition theorem of spherical harmonics to have:

$$\sum_{m=-\ell}^{\ell} \overline{Y_{\ell m}}(x_1) Y_{\ell m}(x_2) = \frac{2\ell + 1}{4\pi} P_{\ell}(\langle x_1, x_2 \rangle) \quad (2.2)$$

where P_{ℓ} is the Legendre polynomial⁸ of degree ℓ . Needlets can be expressed then as:

$$\begin{aligned} \psi_{jk}(\hat{\gamma}) &= \frac{1}{\sqrt{N_p}} \sum_{\ell} b\left(\frac{\ell}{B^j}\right) \sum_{m=-\ell}^{\ell} \overline{Y_{\ell m}}(\hat{\gamma}) Y_{\ell m}(\xi_k) = \\ &= \frac{1}{\sqrt{N_p}} \sum_{\ell} b\left(\frac{\ell}{B^j}\right) \frac{2\ell + 1}{4\pi} P_{\ell}(\langle \hat{\gamma}, \xi_k \rangle) \end{aligned} \quad (2.3)$$

The last expression will be useful to calculate the profile of the needlet in the pixel space, as we will see later in [figures 2.2](#) and [2.4](#).

2.1.2 Filter function b and parameters

In the definition of needlets, there is one important aspect on which we have not elaborated yet: the filter function $b\left(\frac{\ell}{B^j}\right)$. This function determines the multipoles that are taken into account to calculate the needlet and in which proportions. The parameter B is fixed and the frequency j runs over the integers. There are three requisites that the function b has to fulfil:

- b has to be infinitely differentiable in the interval $(0, \infty)$.
- Among all frequencies, each multipole has to be equally sampled (except maybe the lower ones). Formally:

$$\sum_j b^2\left(\frac{\ell}{B^j}\right) \equiv 1 \quad \forall \ell > B \quad (2.4)$$

⁸The analytical definition of the Legendre polynomial is

$$P_{\ell} = \frac{(-1)^{\ell}}{2^{\ell}} \frac{d^{\ell}}{du^{\ell}} (1 - u^2)^{\ell}$$

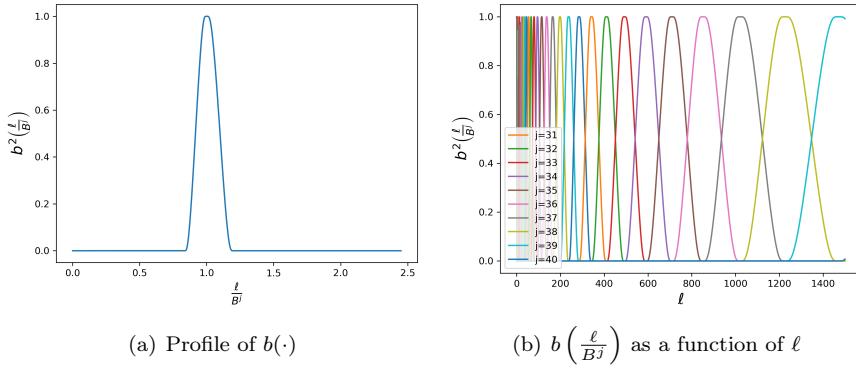


Figure 2.1: Representation of the filter function b for a standard needlet. On the left, we can see b as a function of its argument. We note that the support is $(\frac{1}{B}, B)$. In this case, it is defined for $B = 1.2$. On the right, we can see $b\left(\frac{\ell}{B^j}\right)$ in multipole space. For simplicity, only frequencies higher than $j = 30$ are labelled.

- $b^2(\cdot)$ has to be close to 0 outside the region $(\frac{1}{B}, B)$. Therefore, $b\left(\frac{\ell}{B^j}\right)$ contributes mainly in the region $\ell \in (B^{j-1}, B^{j+1})$. A more restrictive version impose that the support of b has to be exactly this interval, but we will not use this condition.

According to these restrictions, there are multiple ways to define b . In our work, we will use Mexican needlets, which are better localised in pixel space than the standard needlets, but we will explain both of these types:

- **Standard Needlet:** The definition of b for this kind is a four-steps algorithm to construct it that we are not going to reproduce here, it can be found in Marinucci et al. (2008). The final result of $b(\cdot)$ and $b\left(\frac{\ell}{B^j}\right)$ can be seen in figure 2.1.

This filter function b fulfils the restrictive version of the condition on its support, so it enjoys better properties in the multipole space. In exchange, it has worse properties in pixel space. A one-dimensional representation of the needlet (as a function of the angle with respect to the centre) can be seen in figure 2.2. We note that there are several lobes around the central peak. These can produce additional maxima around a peak of intensity during the filtering.

- **Mexican Needlet:** in this case the filter function $b(\cdot)$ is defined as:

$$b(u) = K u^2 e^{-u^2} \quad (2.5)$$

where K is a constant so that $\sum_j b^2\left(\frac{\ell}{B^j}\right) = 1$ for all ℓ , which can be done because the value is constant with ℓ .

This function peaks at 1 and decays quickly, as it can be seen in figure 2.3. It does not have a finite support, but most of the contribution is close to the maximum and therefore in the region $(\frac{1}{B}, B)$ for a B not too small.

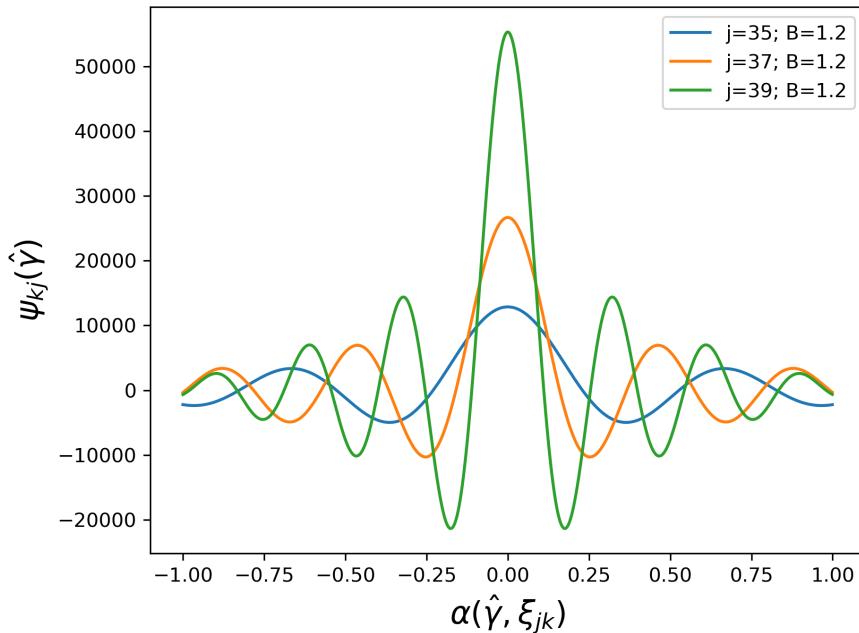


Figure 2.2: Profile of the standard needlet in pixel space. We note that the localisation is quasi-exponential. This means that the localisation of the needlet is better than any polynomial. In the x-axis, we have the angle between the direction $\hat{\gamma}$ and the centre of the needlet ξ_{jk} , in degrees; in the y-axis, the value of the needlet as a function of this angle. Needlets are chosen with $B = 1.2$ and frequencies $j = 35, 37, 39$: the higher the frequency is, the narrower the needlet is.

The smaller B is, the wider it is the region of significant overlap between needlets of different j , in multipole space (and larger the correlation between similar frequencies).

The advantage with respect to standard needlets is their better properties in pixel space. They decay much faster than the standard needlets: if we look at the profile of the Mexican needlet in figure 2.4, we see that they have fewer side lobes than the standard needlet (figure 2.2). This is desirable in our case since filtering with Mexican needlets will introduce fewer residual maxima. Since our algorithm will look for maxima in the maps, we would like to reduce false signals. Still, we will take these residual signals into account.

Mexican needlets were introduced for the first time in Mathematics by Geller and Mayeli (2009).

There are other kinds of needlets, such as the Bernstein needlets, but Mexican needlets appears to have the best characteristics to study the CMB, such as the localisation in pixel space (Scodeller et al. 2011). In this work, we will be using Mexican needlets for the filtering of the maps.

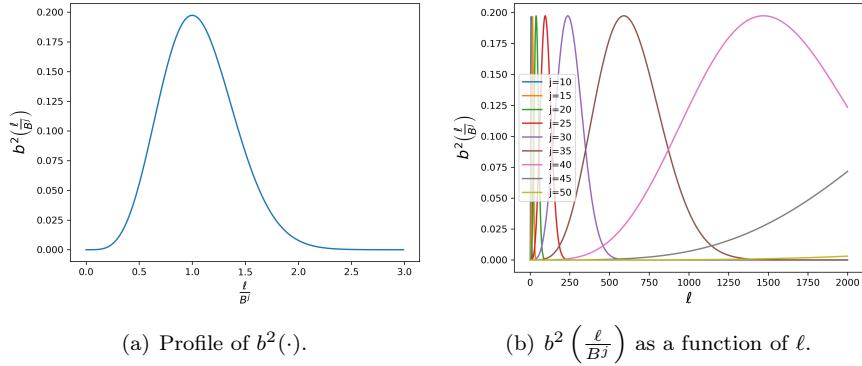


Figure 2.3: Representation of the filter function b for a Mexican needlet. On the left, we can see b as a function of its argument. We note that the maximum is at 1, which is always between $\frac{1}{B}$ and B . On the right, we can see $b\left(\frac{\ell}{B^j}\right)$ in multipole space. For clarity sake, only frequencies multiple of 5 are represented, from 10 to 50. Different frequency bands can overlap significantly, unlike in the standard needlet case, where all bands are plotted. The value of B used here is 1.2.

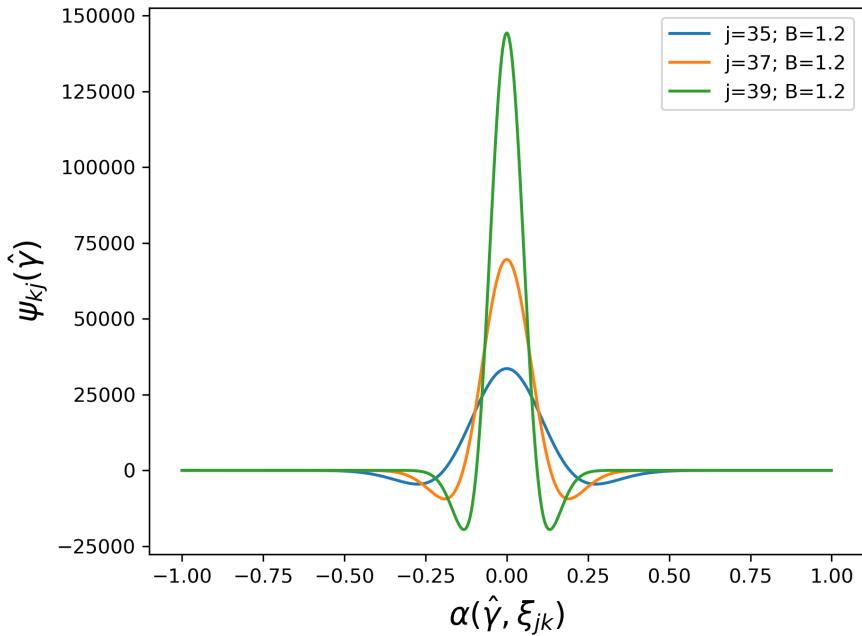


Figure 2.4: Profile of the Mexican needlet in pixel space. We note that the localisation in this case is gaussian, more than quasi-exponential as in the standard needlet, figure 2.2. In the x-axis, we have the angle between the direction $\hat{\gamma}$ and the center of the needlet ξ_{jk} , in degrees; in the y-axis, the value of the needlet as a function of this angle. Needlets are chosen with $B = 1.2$ and frequencies $j = 35, 37, 39$: the higher the frequency, the narrower the needlet.

2.1.3 Filtering and reconstructing

To complete our discussion on needlets, we are now going to talk about how to use them to filter the map and extract features of a certain scale. Let $T(\hat{\gamma})$ a function defined over the sphere, such as the temperature of the CMB. The needlet coefficient β_{jk} of frequency j and pixel ξ_k is defined as:

$$\beta_{jk} = \int_{\mathbb{S}^2} T(\hat{\gamma}) \psi_{jk}(\hat{\gamma}) d\Omega \quad (2.6)$$

which can also be interpreted as a map $\beta_j(\xi)$, called β -map. It is just the result of filtering the map with needlets centred at each pixel.

It will be very useful to link these values with the $a_{\ell m}$ coefficients. Substituting the expression of the needlet, [equation \(2.3\)](#), in the previous expression, [equation \(2.6\)](#):

$$\begin{aligned} \beta_{jk} &= \frac{1}{\sqrt{N_p}} \sum_{\ell} b\left(\frac{\ell}{B^j}\right) \sum_{m=-\ell}^{\ell} \left[\int_{\mathbb{S}^2} T(\hat{\gamma}) \overline{Y_{\ell m}}(\hat{\gamma}) d\Omega \right] Y_{\ell m}(\xi_k) = \\ &= \frac{1}{\sqrt{N_p}} \sum_{\ell} b\left(\frac{\ell}{B^j}\right) \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\xi_k) = \\ &= \sum_{\ell} \sum_{m=-\ell}^{\ell} \left[\frac{1}{\sqrt{N_p}} b\left(\frac{\ell}{B^j}\right) a_{\ell m} \right] Y_{\ell m}(\xi_k) \end{aligned} \quad (2.7)$$

where the square brackets in the first line is the spherical harmonics decomposition of $T(\hat{\gamma})$, that is $a_{\ell m}$. The last line is exactly the spherical harmonics reconstruction with coefficients

$$\widetilde{a}_{\ell m} = \frac{1}{\sqrt{N_p}} b\left(\frac{\ell}{B^j}\right) a_{\ell m} \quad (2.8)$$

where we use \widetilde{X} to denote the quantity X corresponding to the β -map instead of the temperature map.

This gives us a very convenient way to implement the extraction of the filtering of the maps, which does not require a computationally expensive calculation point-by-point:

1. Extract the spherical harmonics coefficients $a_{\ell m}$ from the temperature map to be filtered $T(\hat{\gamma})$.
2. Multiply the coefficients by a factor that depends on ℓ but not on m : the filter function $b\left(\frac{\ell}{B^j}\right)$.
3. Multiply the result by $\frac{1}{\sqrt{N_p}}$.
4. Reconstruct a map using these coefficients $\widetilde{a}_{\ell m}$.

Additionally, from [equation \(2.8\)](#), we have an expression for the angular power spectrum of the β -map (\widetilde{C}_{ℓ}) from the one of the temperature map (C_{ℓ}):

$$\widetilde{C}_{\ell} = \frac{1}{N_p} b^2 \left(\frac{\ell}{B^j} \right) C_{\ell} \quad (2.9)$$

Considering needlets as we have defined them in [equation \(2.1\)](#), we have that they are not a base, since they have redundancy. However, they are a *tight frame*⁹ of $L^2(\mathbb{S}^2)$. In our case, this implies:

$$\sum_{j,k} \beta_{jk}^2 = \sum_{\ell=1}^{\infty} \frac{2\ell+1}{4\pi} C_\ell \quad (2.10)$$

which is a result of the conditions on b and the normalisation factor. It leads to a much more important result, the reconstruction formula. If we have the values for all the β_{jk} , we can reconstruct the original function $T(\hat{\gamma})$ simply with:

$$T(\hat{\gamma}) = \sum_{j,k} \beta_{jk} \psi_{jk}(\hat{\gamma}) \quad (2.11)$$

This formula is completely analogous to the case of spherical harmonics in [equations \(1.15\)](#) and [\(1.16\)](#). It is not shared by other wavelets systems and it is one of the general advantages of the use of needlets.

2.2 Maxima and multiple testing

The goal of this work is to find candidates to be PS in the CMB temperature maps. The CMB map is a purely gaussian map or, at least, very close to it¹⁰. This means that any contribution different from the background itself will contribute to non-gaussian effects. Hence, we are going to search for this non-gaussianity. In order to do that, we are going to filter the map, find local maxima in the β -map, and use multiple testing to select which of these maxima, if any, is a candidate of being a PS.

Although these techniques can be applied to any scale, we will focus in scales comparable to the resolution of the telescope, since this is the approximate scale of PS. It is also advised not to go to smaller scales since the instrumental noise and pixelisation can cause artificial effects to arise at high multipoles.

We are going to apply the Smoothing and TEsting of Maxima (STEM) algorithm introduced in Schwartzman et al. ([2011](#)). It was analysed in the context of CMB data analysis in Cheng, Cammarota, et al. ([2016](#)), both theoretically and numerically.

2.2.1 Maxima distribution

First of all, we have to filter the temperature map that we want to analyse with a needlet of a suitable frequency j to extract the information about the

⁹A tight frame is a countable set of functions $\{e_i\}$ such that any square integrable function f on the space, in this case, $L^2(\mathbb{S}^2)$ preserves its norm:

$$\sum_i \langle f, e_i \rangle^2 = \int_{\mathbb{S}^2} f(\hat{\gamma})^2 d\Omega$$

¹⁰Non-gaussianity of the CMB is a very interesting field on its own and it could imply some faint effects from the inflationary epoch. See, for example, Bartolo et al. ([2004](#))

location and intensity of fluctuations of the desired size. This removes most of the contributions of both larger anisotropies and pixel-size effects such as instrumental noise.

Once we have extracted the β -map, we want to extract its local maxima. This is done by a simple search of the pixels that have more intensity than all of its nearest neighbours. We are going to analyse now the distribution of these maxima.

For any gaussian map with an angular power spectrum \widehat{C}_ℓ , the normalised theoretical distribution of maxima is known. Its derivation can be found in Cheng and Schwartzman (2015), although we follow the formulation of Cheng, Cammarota, et al. (2016); here we are only going to present the solution. If we define:

$$\begin{aligned} C' &\equiv \sum_{\ell=1}^{\infty} \frac{2\ell+1}{4\pi} \widehat{C}_\ell \left[\frac{\ell(\ell+1)}{2} \right] \\ C'' &\equiv \sum_{\ell=1}^{\infty} \frac{2\ell+1}{4\pi} \widehat{C}_\ell \left[\frac{(\ell-1)\ell(\ell+1)(\ell+2)}{8} \right] \end{aligned} \quad (2.12)$$

where the quantities in square brackets are the first and second derivatives of the Legendre polynomials $P_\ell(\cdot)$ evaluated at 1. Now, we define:

$$\kappa_1 = \frac{C'}{C''}; \quad \kappa_2 = \frac{[C']^2}{C''} \quad (2.13)$$

We can write the theoretical distribution of maxima as the following formula:

$$\begin{aligned} f(x) = & \frac{2\sqrt{3+\kappa_1}}{2+\kappa_1\sqrt{3+\kappa_1}} \left\{ [\kappa_1 + \kappa_2(x^2 - 1)] \phi(x) \Phi \left(\frac{\sqrt{\kappa_2}x}{\sqrt{2+\kappa_1-\kappa_2}} \right) + \right. \\ & + \frac{\sqrt{\kappa_2(2+\kappa_1-\kappa_2)}}{2\pi} x e^{-\frac{(2+\kappa_1)x^2}{2(2+\kappa_1-\kappa_2)}} + \\ & \left. + \frac{\sqrt{2}}{\sqrt{\pi(3+\kappa_1-\kappa_2)}} e^{-\frac{(3+\kappa_1)x^2}{2(3+\kappa_1-\kappa_2)}} \Phi \left(\frac{\sqrt{\kappa_2}x}{\sqrt{(2+\kappa_1-\kappa_2)(3+\kappa_1-\kappa_2)}} \right) \right\} \end{aligned} \quad (2.14)$$

where ϕ and Φ denote the standard normal probability density and probability distribution, respectively. In figure 2.5 it can be seen the values of this function for typical parameters of the CMB temperature maps and the β -maps with which we will work, both normalised to their standard variation.

It has to be noted that this is the normalised maxima distribution for a *continuous* gaussian map. In our case, we will unavoidably work with pixelated maps, where the distance between pixels is finite, as well as the number of pixels. This will introduce a small effect in the maxima distribution. However, the effect can be minimised using a large total number of pixels.

2.2.2 p -values and multiple testing

The p -value is a useful tool commonly used in statistical hypothesis testing. In a general form, if we want to test a high-end value x of a continuous distribution

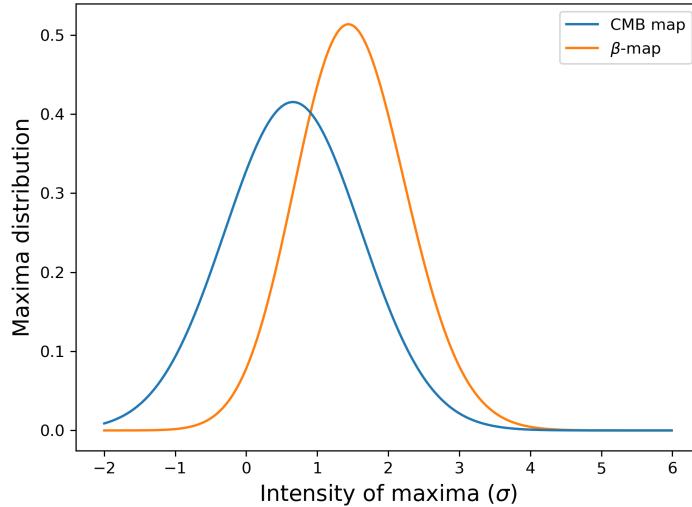


Figure 2.5: Normalised maxima intensity distribution for a typical CMB map (blue) and for the β -map resulting of filtering it with a needlet $B = 1.2, j = 39$ (orange). The intensity of both maps is normalised by their respective standard variation. The y-axis corresponds to the function f described in equation (2.14).

X under a null hypothesis H_0 , the p -value is defined as:

$$p\text{-value}(x) \equiv \Pr(X \geq x | H_0) \quad (2.15)$$

where \Pr denotes probability. In words, it is the probability of finding values greater or equal to the given value *if* the hypothesis is correct. There are analogous definitions for testing the low values or both ends at the same time, but in our case it is more natural to test the higher end, since they are the most affected values by undetected PS. In general, a very low p -value means that the measured value is highly unlikely if the hypothesis is correct, so it is possible that it is not correct.

Coming back to our case, we have just seen that we have an analytical expression f for the distribution of maxima on a gaussian map. We can then calculate the p -values of the intensity of the maximum x under the null hypothesis $H_0 \equiv$ “The map is a perfectly gaussian map”:

$$p\text{-value}(x) = \Pr(f(X) \geq x | H_0) = \int_x^{\infty} f(y) dy \quad (2.16)$$

It is important to note that x here refers to the intensity value of a *maximum*. We are not testing all the points in β -map, only the ones that are local maxima. That is why we use the distribution f of the *maxima* of the map described in equation (2.14).

We calculate this value for every detected maxima in the β -map. The maxima with higher values for the intensity will report lower p -values, since they are less likely to occur under the hypothesis. If we were doing single

testing, we would just set a threshold of how low this probability has to be in order to reject the hypothesis. We would say that maxima with p -values lower than certain threshold are too unlikely to be normal fluctuations of the CMB and, therefore, they must be PS.

However, this approach does not take into account the information of the whole set of maxima, only the one being tested at a time. To understand how this could be a problem and how we are going to proceed, we give the following example:

- We have a map on which we have applied all the procedure explained until now. We have a set of maxima of the β -map and their p -values. The most intense maximum has a low p -value of 10^{-5} . The probability of this happening is 1 every 10 000 times. However, we have a total number of local maxima of $N_{max} = 10^5$, so we expect to obtain around one maxima being at least this intense. The maxima is (probably correctly) not reported as a candidate of PS but it is considered a normal fluctuation of the CMB.
- We have a very similar map to the previous one, but with one difference. Now, we have 100 maxima with p -value= 10^{-5} out of the same total number of maxima $N_{max} = 10^5$. We would expect to obtain around one maxima this intense, but we have 100, so it seems very likely that most of them are actually due to PS and not random fluctuations of the CMB. However, if we filter only by threshold, none of these maxima will be reported as a candidate of PS, as a maximum like them is statistically expected to occur once.

Instead of setting a threshold on the p -values, we are going to follow the Benjamini-Hochberg procedure for multiple testing (Benjamini and Hochberg 1995). According to it, we sort the p -values from smallest (less likely) to largest (more likely). We call p_i to the i -th smallest p -value and x_i to the intensity of the associated maximum: $p\text{-value}(x_i) = p_i$.

The idea behind the procedure is the following: for the i -th maximum x_i , we want to know how many maxima we would expect to be at least as intense as x_i in a gaussian map. If this number is more than i , then some of the points could be coming from a non-gaussian map, in this case, PS. In order to control the effect of the normal fluctuations of the CMB, there is a significance level parameter $\alpha \in (0, 1)$. We will report the first i maxima as PS candidates if the expected number of maxima as intense as x_i is lower than $\alpha \cdot i$. For example, if $\alpha = 0.01$, the number of maxima greater than x_i has to be 100 times larger than the expected number of maxima for its intensity. Since this is done for all detected maxima, the reported candidates are determined by the largest i that fulfil the condition, if any.

We can formalise this procedure as follows. Let p_i be the i -th smallest p -value, N_{max} the total number of maxima detected, and $\alpha \in (0, 1)$ a fixed significance value. We define

$$I \equiv \max (i \mid p_i N_{max} < i\alpha) \quad (2.17)$$

Then, the largest I maxima are reported as candidates of PS. We note that this is equivalent to plotting all the p -values in order as a function of i , tracing a

line with slope $i\alpha/N_{max}$ and selecting the last point below the line. Its value of i is then I , and all points up to it, including itself, are reported as PS candidates. An example of this can be seen later in [figure 3.4](#).

This procedure has a good control on the fraction of false detections among the candidates (as a function of α), as well as the fraction of detected sources among the total number of sources. The properties are mathematically proven in Cheng, Cammarota, et al. ([2016](#)), where the authors, also using Mexican needlets, confirm their results with computer simulations as well.

2.2.3 Summary of the algorithm

We write here a brief summary of the algorithm that we have described in this section in order to have a quick global vision of all the steps and how they are connected:

1. We filter the temperature map we want to study with a suitable needlet using [equation \(2.8\)](#) to obtain the β -map.
2. We find all the local maxima in the β -map.
3. We calculate the p -values for all the maxima using [equation \(2.16\)](#) and [equation \(2.14\)](#).
4. We sort the p -values obtained from lower to higher and find I as defined in [equation \(2.17\)](#).
5. The I most intense maxima (lowest p -values) are reported as PS candidates.

This will be the structure we will follow when we implement the algorithm in the next chapter.

❖ Chapter Three ❖

Code and Implementation

The code needed to obtain the results has been written from scratch and trying avoid the use external code as a black box as much as possible. This is done in order to have a better control of how the parameters affect the output and not to introduce unwanted effects of which we may be unaware.

We develop the code in `Python 3.5` using `Jupyter Notebooks` as a developing environment (Kluyver et al. 2016). In this chapter, we will explain the main steps and technical details of the code, but we will not explain it line by line. However, the notebooks with the complete code explained here can be found in detail in the online repository explained in [Appendix A](#).

Prerequisites

Aside from a working installation of the programming language itself, there are some basic packages that are also considered prerequisites, as the code rely heavily on them. The use of mathematical functions and constants is provided by `numpy` (T. E. Oliphant 2006) and `scipy` (Jones et al. 2001). Also `astropy` (Collaboration 2018) is loaded to import FITS images as the ones provided by the Planck Collaboration. Lastly, `HEALPix` (Gorski et al. 2005) and its Python libraries `HEALPy` are used to deal with full-sky images, as it is the standard adopted by the Planck Collaboration to represent their spherical maps. The basis of this standard will be explained below.

In order to treat and visualise the results, we use `pandas` (McKinney 2010) and `Matplotlib` (Hunter 2007). Finally, the Python interface for `CAMB` (Lewis 2011) can be used to generate CMB simulations from input cosmological parameters, although its use can be avoided if the user prefers.

We note that all the packages used are very common in the scientific community and widely tested, while we code everything related with the use of needlets, the maxima analysis, the detection of PS and in general all the steps described in this chapter.

The code should be able to be run on a standard laptop in a reasonable amount of time: the analysis of a single map of `nside = 1024` (see explanation of `HEALPix` below) should take less than a minute, time scaling with the square of this quantity. The whole set of results presented in [Section 4.2](#) and [Appendix B](#) should take around a couple of hours at full resolution.

Structure

In this work we want to use different *cleaned* CMB temperature maps to analyse the possible existence of residual PS. Before we apply the algorithm to the actual maps, we want to check its behaviour on simulated maps. Therefore, the code has been structured in three natural ways:

- **Main algorithm:** this is the main part of the code, it takes an input map, processes it and reports a list of candidates to be PS. This part implements the theory explained in [Chapter 2](#). The other parts will wrap this one to use it on different maps and scenarios.
- **Simulations:** in order to test that the code for the main algorithm works well and quantify its sensitivity to PS, we run some simulations. This part generate synthetic CMB maps compatible with the values of the cosmological parameters, insert a set of artificial PS and then it uses these new maps as input of the main algorithm. When it has the list of candidates, this part check how well they correlate to the injected artificial points. This will be our sandbox to test the effect of different contributions.
- **Analysis:** this part is the one that manages the actual Planck maps. It will load all the maps that we are analysing in a consistent way, and loop over all the parameters that we want to explore. After the candidate list is generated for each map, this part will check their location on the sky, whether they are inside some mask, and if they are known points in the PS Planck catalogues.

HEALPix standard

In most cases in Astrophysics and in our daily experience, photographies and images are rectangular. However, the study of the CMB has to be done in the entire sky, so we have to deal with *spherical* images. This problem is not exclusive of this field, it arises naturally when the surface of a sphere has to be studied as a whole, without the possibility of applying tangent plane approximations. An example can be found in Earth Science, when they have to study the behaviour of some parameter on the surface of the Earth, such as the altitude, the magnetic field, the concentration of some substance...

HEALPix, which stands for Hierarchical Equal Area isoLatitude Pixelization is a standard for spherical images designed to facilitate the spherical harmonics analysis. Images from Planck and WMAP are reported using this standard, so it is the one in which we will be working.

In this standard, the surface of the sphere is divided first into 12 spherical squares, which are the base pixels. Now, each of these squares can be further divided, but only in subsequent halves. The number of pixels in one side of a square is called `nside` and has to be of the form 2^n , where n is the number of times the square has been divided. This means that each of the 12 squares will be $\text{nside} \times \text{nside}$ pixels. The total number of pixels of the sky is then

$$N_{pix} = 12 \cdot \text{nside}^2 = 12 \cdot 2^{2n} \quad (3.1)$$

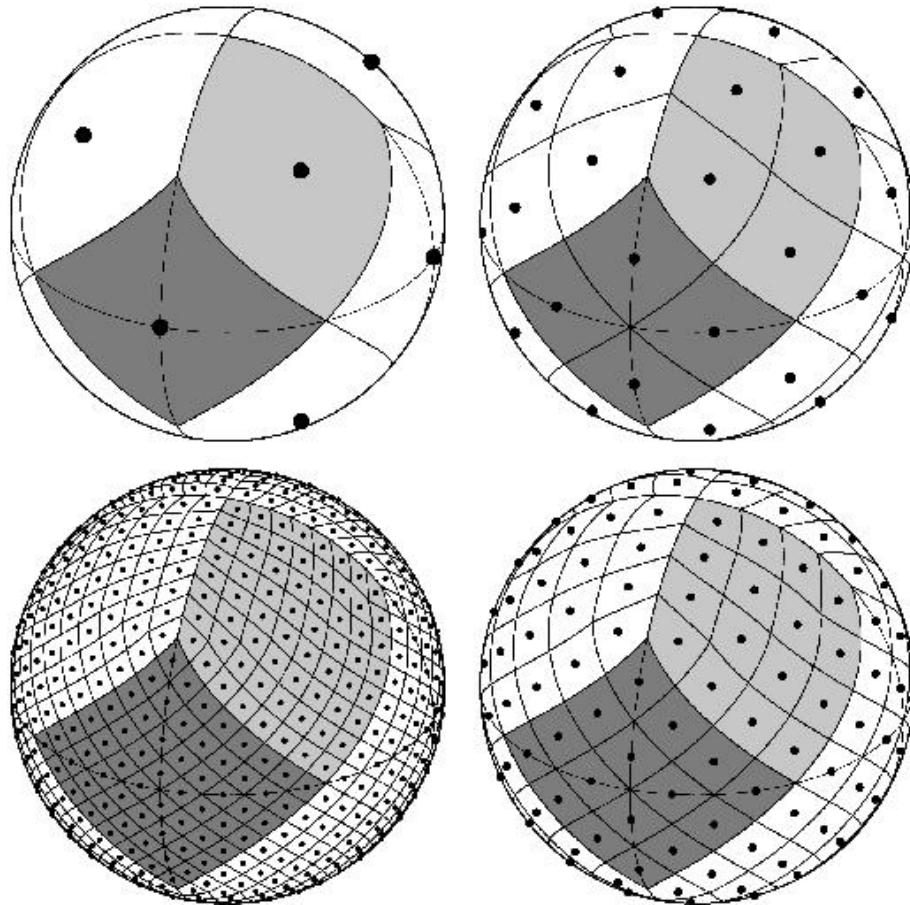


Figure 3.1: Pixels corresponding to the four lowest possible resolutions in HEALPix. Starting by the top left and in clockwise order, `nside` = 1, 2, 4, 8. Image credit: NASA/JPL.

for some non-negative integer n . In this way, all pixels for a given resolution have the same area on the sphere. An image of the pixels corresponding to the four lowest possible resolutions is given in [figure 3.1](#).

Notes

In this chapter we will follow a similar structure to the structure of the code. In [Section 3.1](#), we will explain the main algorithm in charge of extracting the list of candidates to PS from an input map. In [Section 3.2](#) we will explain the code used to produce synthetic CMB maps and PS to test the algorithm. Finally, in [Section 3.3](#), we will explain the code that analyses the real Planck maps and compare the results with the Planck catalogue of PS.

Every section will start with the list of parameters that can be passed to that part of the code, a brief explanation of what they are, and their default values. We distinguish between important parameters that the user may want to change, indicated with a bullet symbol (\bullet) and more technical parameters

that the user may not want to change on a first try, indicated with a circle (\circ). Then, we will explain the main steps of the code. After this, we will comment on how different parameters can affect that part. In this way, we (and the user) can control the effects of changing a parameter in the algorithm, which also allow us to choose them in a more sensible way.

There will be some small snippets of code to illustrate how a certain step is implemented. These lines of code are usually simplifications of the actual code to illustrate the approach to a certain problem or the exact functions used, while also being completely functional. The actual code is based on the one shown here, but modified for efficiency or to be able to deal with more complex situations, such as several maps in a loop. The complete code can be found in the online repository, as detailed in [Appendix A](#).

3.1 Main algorithm

We are going to explain now the code used to implement the main algorithm, using the steps described in [Section 2.2.3](#). This part of the code is the one responsible of taking a CMB map and extracting the list of candidates to PS according to the procedure explained in [Chapter 2](#).

3.1.1 Parameters of the main algorithm

This part admits a series of parameters that the user can change to adapt the algorithm to their specific needs. We are going to start by listing and explaining these parameters. We also indicate their default value by a “=” sign in the name.

- **tmap**: this is the input temperature map, with $nside = \text{nside}$ ¹¹. We are going to call it **tmap** since **map** is a reserved word in Python.
- $B = 1.2$: this parameter indicates the width of the needlet in multipole space, which is going to be used to filter the map. It corresponds to B in [Section 2.1](#). We are using Mexican needlets, so this parameter does not exactly indicate the support of the filter. Instead, it is connected to the distance between needlets of different j in multipole space.
- $j = 39$: this parameter is the frequency band j of the needlet that is going to be used to filter the map. A higher frequency implies a higher range of multipoles and smaller size. It has to be chosen in a way so that the needlet has a size comparable with the size of the PS. If it is larger, we may step into a region of very high ℓ , dominated by noise and with pixelisation effects. If it is too small, the needlet will be too large and we lose sensitivity. It has to be noted that this value will depend on the value of B . The needlet peaks at B^j , which means that the default values correspond to a peak around $\ell = 1225$.
- $\text{alpha} = 0.01$: this parameter is the significance level parameter α used for the Banjamini-Hochberg procedure. We use a conservative value where

¹¹In general, we use the **typewriter font** for computer parameters that we use in the code and *italic font* for theoretical variables in the algorithm.

we only consider a maximum as a candidate if we find 100 times more maxima than expected for its intensity.

- **folder:** the programme needs a folder to store the filtered map and the list of maxima.
- **`lmax = 3 · nside - 1`:** this is the maximum multipole ℓ considered for the calculation of the β -map and, in general, all calculations that theoretically require an infinite sum. Due to the limited resolution of a map, the maximum ℓ for which C_ℓ can be calculated is $3 \cdot \text{nside} - 1$. We are working with maps of `nside = 2048`, corresponding to a value `lmax = 6143`. It is important to consider that the needlet has to be negligible at higher ℓ when considering its parameters.
- **`interpolation = True` and `step = 0.05`:** as we will see later, the calculation of the p -values is computationally very expensive if we calculate them for each of the 10^5 to 10^6 maxima. Instead, we have the option to interpolate and to select the step, in units of the standard deviation of the map. The default values speed up the programme while losing very little accuracy, but the user can choose to increase or decrease the step if needed.

3.1.2 Steps of the main algorithm

The algorithm is implemented following the next steps:

Filtering of the map In order to filter the map with the needlet, we are not going to do a convolution point by point. Instead, we are going to follow the procedure described in [Section 2.1.3](#), which is computationally more efficient.

First, we extract the $a_{\ell m}$ of the map with the built-in `HEALPix` function:

```
import healpy as hp
alm=hp.map2alm(tmap)
```

Now, `alm` is multiplied by the filter function `b`, which is previously defined by the function `b1f(l, j, B)`. The result is then normalised. This can be done by:

```
b1=b1f(np.arange(lmax+1),j,B)
beta_alm=hp.almxfl(alm,b1)/np.sqrt(12.*nside**2.)
```

And finally, we reconstruct the map, which now is the β -map and save the map in the folder:

```
betas=hp.alm2map(beta_alm, nside)
hp.write_map(folder+'/'+'betas.fits', betas)
```

With this, the filtering with the needlet is finished. We can see an example of the result in [figure 3.2](#), comparing also with the original map. In this case, the temperature map is filtered with a needlet of $B = 1.2$, $j = 39$, which highlights the small scales, comparable with PS. The figure shows the same portion of the sky of $5^\circ \times 5^\circ$ for both cases.

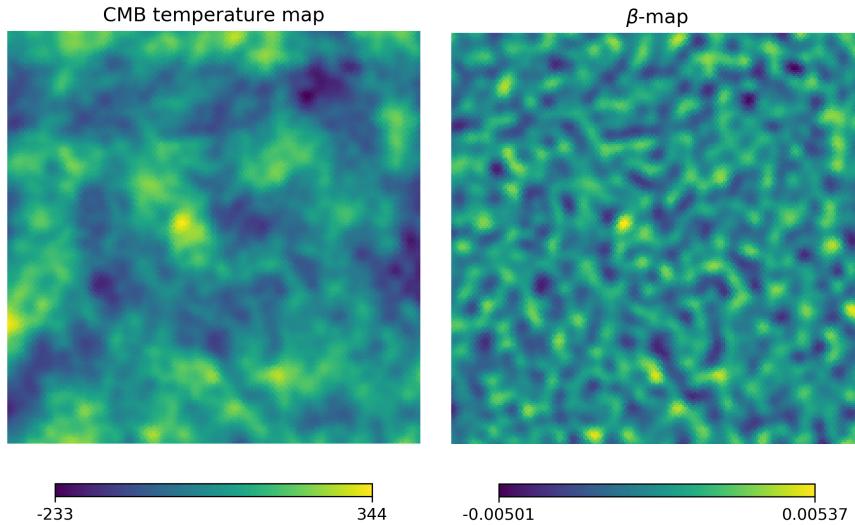


Figure 3.2: Comparison between a CMB temperature map (on the left) and the β -map resulting of filtering this map with a needlet $B = 1.2, j = 39$ (on the right). We note that the small scale clumps in the temperature map get accentuated in the β -map, especially at the scale of the needlet used. The field of view for both images corresponds to $5^\circ \times 5^\circ$.

Detection of the maxima This is the only part that is not implemented in Python. HEALPix provides a tool to find maxima in spherical images, but it is not implemented yet in the Python library, so it has to be run from the terminal. Fortunately, this can be automated to run from within the code:

```
file = open(folder + '/hotspot.par', 'w')
file.writelines(['infile=' + folder + '/betas.fits\n',
                 'extrema_outfile=' + folder + '/pixlminmax.fits\n',
                 'maxima_outfile=' + folder + '/maxima.dat\n',
                 'minima_outfile=' + folder + '/minima.dat'])
file.close()

! hotspot {folder}/hotspot.par
```

where the `!` symbol indicates the code to execute that line in the terminal. Now, we only have to load the list of maxima, which has two columns: number of pixel and intensity of the maximum. We normalise the intensity in terms of the standard deviation of the β -map.

```
maxfile=np.loadtxt(folder + '/maxima.dat')
maxi=(maxfile[:,0], maxfile[:,1]/np.std(betas))
```

Distribution of the maxima Although not required to run the algorithm, we check the distribution of the maxima at this point. For that, we plot the histogram to the maxima and compare with the theoretical distribution that can be found in [equation \(2.14\)](#).

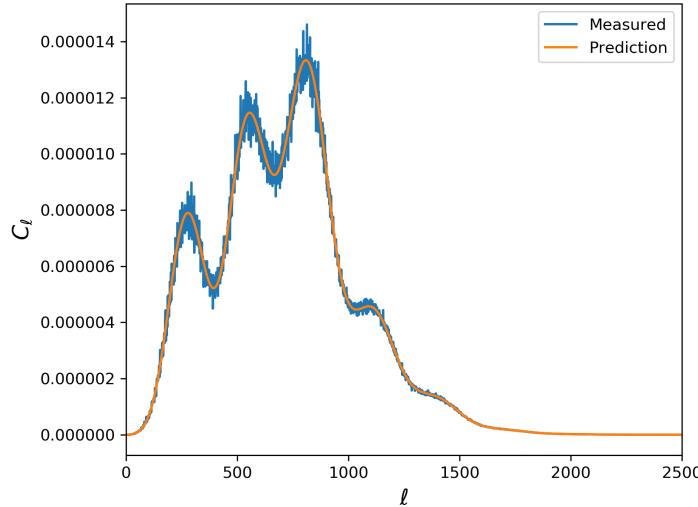


Figure 3.3: Comparison of the C_ℓ for a β -map and the prediction as expressed in [equation \(2.9\)](#), multiplying the C_ℓ of the temperature map and the filter function b . The used needlet has $B = 1.2$ and $j = 39$.

In order to obtain the theoretical distribution of the maxima in the β -map, its C_ℓ have to be introduced. We do this in two different ways: directly calculating the C_ℓ for the β -map and with the C_ℓ of the initial map (or the theoretical values if it is a simulated map), multiplying by the filter, as in [equation \(2.9\)](#). Doing so is a way to test that all the theory and procedures until this point seem to be correct.

An example of this step for a simulated map can be found in [figure 3.3](#). In this figure we can see the comparison between the C_ℓ of the β -map and the prediction. The needlet used has $B = 1.2$ and $j = 39$. We observe that the C_ℓ of the map follow perfectly the behaviour of the prediction, albeit with significant dispersion. The effect of this difference in the values of the prediction for the maxima are negligible, as both C_ℓ are almost indistinguishable. We will find the histogram of the maxima with the expected distribution for the theoretical formula later in [figure 3.5](#), when we talk about simulated maps.

Calculation of p -values The next step is to calculate the p -value for each detected maximum. In order to do that, we have to integrate the theoretical distribution f (we remember that the definition of p -value is the probability of having a maximum equal or greater than a certain value). The integration is done numerically with the `numpy` general purpose integrator `quad`, keeping an upper bound on the relative error of at most 10^{-4} .

```
def pval(x,k1,k2):
    result=integrate.quad(f,x,np.inf, args=(k1,k2))
    return(result)

vpval=np.vectorize(pval)
```

where f is the function f of the theoretical distribution and k_1, k_2 are κ_1 and κ_2 , respectively.

Additionally, we do not calculate the p -values for every point, since there is a large amount of them. Instead, we take the largest and smallest maxima and calculate the exact p -value for points in this interval with a `step = 0.05` (units of standard deviation of the β -map) by doing the integral in these points. With these exact values, and due to the smoothness of the function, we interpolate for all maxima.

```
sortmax=np.array(maxi)[maxi[1].argsort()]
xp=np.arange(sortmax[0]-step,sortmax[-1]+step,step)
fp=vppval(xp,k1=k1,k2=k2)
pvalues=np.interp(sortmax,xp,fp[0])
```

We have to note that for some maps with very intense points (like *not inpainted* temperature maps in the galactic plane) may saturate the numerical precision of the computer while computing its p -value. This is expected, since these very extreme values are almost impossible to occur by chance in a gaussian map. In order to avoid numerical errors while integrating, the function `vppval` returns 0 if the intensity of the maximum is more than 37.5 times the standard deviation of the map, where the p -value is approximately 10^{-304} .

Determination of PS candidates From these detected maxima, we have to extract the list of candidates to PS. We will do this according to the multiple testing algorithm described in [Section 2.2.2](#).

In order to proceed, we first sort all the p -values, from lower to higher. If we multiply this by the total number of maxima, we have an estimate of the expected number of maxima with this intensity or more. We want to find the maxima for which this quantity is larger than its position in the ordered list times the confidence value:

```
spv=np.sort(pvalues)
Nmax=len(spv)
vec=np.arange(1,Nmax+1)*alpha
index=np.argwhere(vec >= spv*Nmax)[-1][0]+1
canlist=sortmax[-index:]
```

The candidates to be a PS are then the `index` most intense maxima. Their location and intensity are stored in the `canlist` variable. We can find in [figure 3.4](#) the typical behaviour of the quantities compared in the last line of the code for a simulated CMB with some sources. Lines for different `alpha` are plotted. The line of p -values should roughly follow the `alpha=1` line if there were no sources in the map.

3.1.3 Effects of the parameters in the main algorithm

Changing B and j has obvious effects on the production of the β -maps. Decreasing the multipole region of the needlet means going to larger scales and vice-versa. If the needlet multipole region is so high that it overpasses `lmax`, then the needlet profile will have more side lobes, which would introduce artificial maxima around the peaks in the filtering phase.

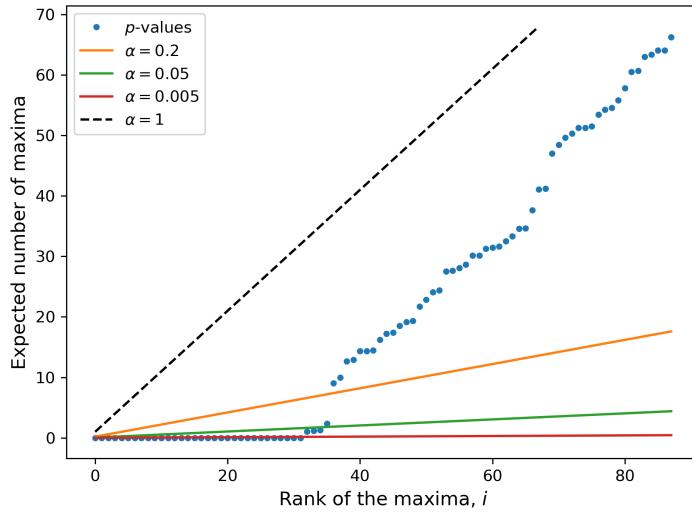


Figure 3.4: Example of the Benjamini-Hochberg procedure for multiple testing. In the x-axis, we have the rank i of the p -values, sorted from lower (less likely) to higher (more likely). The value of the p -value times the total number of maxima is plotted for each maximum as blue points. This is an estimate of the number of maxima expected to be at least as intense in a gaussian CMB map. The three coloured lines represent the threshold for 3 different values of α . The last value below the line determines the number of points reported as candidates to PS. In dashed black line, corresponding to $\alpha = 1$, we have the expected behaviour for a purely gaussian map. Typically, this is the behaviour of the maxima once the candidate points are removed.

We also note that decreasing the size of the needlet (for example, by increasing the frequency of the needlet j) will increase the total number of maxima significantly. Even if the total number of maxima changes, the distribution of the maxima closely follows the theoretical distribution. However, there is a systematic effect that slightly increases the lower part and decrease the higher part. This is especially noticeable close to the maximum of the distribution, as we will see later in [figure 3.5](#). We believe this small effect is due to pixelisation at small scales. In particular, we note that this effect increases when the needlet size approach the resolution of the map. Therefore, we recommend the rule of thumb $\ell_{need} \lesssim nside$, where this effect is negligible. ℓ_{need} here refers to the multipole region where the needlet is mostly present: $\ell_{need} \sim (B^{j-1}, B^{j+1})$.

For the computation of the p -values, the user can choose not to do the interpolation, but in this case the calculation will take several orders of magnitude more. Instead, the step between the points where it is exactly calculated can be reduced. The default value of 0.05 should provide a very fast execution while having a relative error at less than 10^{-4} .

The parameter `alpha` for the significance level of the candidates (α) will impact the characteristics of the candidate list. A low value of the parameter means that all the candidates are extremely likely to be actual PS, but also less amount of candidates will be reported by the algorithm, losing sensitivity at

lower intensities. We will normally use a value of `alpha` = 0.01 as a conservative balance of both effects, testing the results also with a more sensitive but more prone to error `alpha` = 0.05 and a more robust but less sensitive `alpha` = 0.002. We will explore this effect in detail later in [Section 4.1.1](#).

3.2 Simulations

In this section we are going to explain the implementation of the simulations done in order to test the main algorithm described in [Section 3.1](#). In this part, we will simulate a number of CMB realisations, introducing artificial PS when needed, and explore how the detection algorithm behaves.

3.2.1 Parameters of the simulations

We are going to start by describing the input parameters that this part will use to test the algorithm. The input parameters of the previous section can still be given, as they will be passed to the main algorithm (such as `B`, `j`, and `alpha`). The new parameters are:

- `nmaps` = 10: this is the number of maps that will be simulated. We always create a set of them to avoid statistical fluctuations that could affect our tests. Increasing this quantity will produce more robust results, but it will take longer to compute. Because of the way the algorithm is coded, this number has to be at least 2.
- `nside` = 2048: this will determine the resolution of the map. We use 2048 since this is the default resolution of Planck maps. Additionally, this number allows us to reach multipoles in the region of $\ell \sim 1000$ without suffering from pixelisation effects. This quantity is used in the context of the HEALPix standard, so the parameter has to be of the form 2^n for a natural n . Computational time increase roughly as $\sim n_{\text{side}}^2$.
- `fwhm_min` = 5: this is the full width at half maximum of a gaussian profile used to convolve the image. This is done in order to simulate the observational beam of the telescopes. The unit is arcminutes, and the default value is set to match the nominal resolution of the most recent Planck maps. It can be adjusted to match the values of other telescopes if needed.
- `usesources` = `True` if we want to introduce artificial sources in the map and check whether they are detected. In this case, we can set the parameters of the sources with the following input parameters:

```
nsources=200
Imin=0.
Imax=200.
```

Every PS is introduced as a high intensity value in a single pixel and then convolved along with the map. In this way, the artificial PS have a gaussian profile of the size of the beam. There are a total of `nsources` sources, introduced with intensities following a uniform distribution between `Imin`

and I_{max} . These values are normalised to a pixel of 1 arcminute², so if the pixel has half the area, it will have double the intensity¹². They are in units of the map standard deviation, after convolving with the beam.

- o Cosmological parameters and initial power law of the (spatial) power spectrum:

```
cosmo=dict(H0=67.3, omhb2=0.022, omch2=0.120,
            tau=0.078, omk=0)
inipow=dict(ns=0.966, r=0)
```

These parameters are used to simulate CMB temperature maps compatible with the desired values of cosmological constants. The simulation is done through the Python interface of **LAMBDA CAMB**. If preferred, the user can directly give the C_ℓ of the maps they want to simulate.

- o **rho** = 3: this parameter set the angular tolerance to declare a positive detection. After the list of candidates is produced by the algorithm, we will compare the location of these points with the location of artificial sources. If the difference is less than **rho** pixels, the detection is considered positive. Technically, this is transformed into an angular distance **angtol_min**, in minutes, that the user can modify instead of **rho**.

3.2.2 Steps of the simulations

This part is implemented in the following steps:

Simulation of CMB temperature maps We start by creating the C_ℓ of the maps using **CAMB** (Lewis 2011). This software takes as input some cosmological parameters defined by the user. With that, the programme produces the C_ℓ of the CMB map corresponding to a Universe with those parameters.

In our code we incorporate some of the Planck values of the main parameters as the default value. However, if the user prefers, they can provide the desired C_ℓ to the code, instead of generating it from the cosmological parameters. In case the user wants to have more control over the options and parameters of the calculations, the web interface for **LAMBDA-CAMB**¹³ is very flexible and accepts a wide variety of options to produce the C_ℓ , which can be inputted in the variable **cls**.

In any of these cases, the C_ℓ are then multiplied by a filter to take into account the convolution beam. The standard deviation **cmb_sigma** is calculated at this point:

```
ls = np.arange(lmax+1)
gbl = hp.gauss_beam(fwhm, lmax)
cmb_sigma = np.sqrt(np.sum(cls*(2.*ls+1.)*
                           gbl**2./(4.*np.pi)))
```

¹²This is done in order to be able to compare results of maps with different resolutions if needed. Otherwise, the signal of a small pixel (high **nside**) would be more diluted than the signal of a large pixel (low **nside**) when they are convolved by the 5' beam.

¹³The web interface for **CAMB** can be found in the **LAMBDA** webpage from NASA: https://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm.

After this, `nmaps` different maps are generated compatible with the calculated (or given) C_ℓ . Each one is generated with:

```
tmap = hp.synfast( cls , nside , fwhm=fwhm )
```

which uses a random seed and automatically does the beam convolution.

Simulation of artificial PS The user can choose to whether or not introduce artificial PS. If they do, the code will randomly select `nsources` different directions on the sky in an isotropic way. Then, on a zeroes map of the same size, the pixels corresponding to these directions are assigned a value for the intensity. These values follow a uniform distribution between `Imin` and `Imax`, divided by the area of the pixel as explained before:

```
phi = np.random.rand(nsources)*2.0*np.pi
theta = np.arccos(2.*np.random.rand(nsources)-1.0)
vecsou = hp.ang2vec(theta,phi)
pixsou = hp.vec2pix(nside,
                     vecsou[:,0],vecsou[:,1],vecsou[:,2])

pixarea = hp.nside2pixarea(nside,degrees=True)*3600.
souinten=((np.arange(nsources)/nsources)*
           (Imax-Imin)+Imin)*cmb_sigma/pixarea
sources = np.zeros(hp.nside2npix(nside))
sources[pixsou] = souinten
sources = hp.smoothing(sources, fwhm=fwhmsources)
```

Now, `sources` is a map that contains only the artificial sources and is zero anywhere else. This map has been convolved by the same beam as the map to mimic the finite optical resolution of the telescope. The directions in the sky of the sources, `vecsou`, are stored for later comparison with detections. The `sources` map is then added to all the synthetic CMB maps produced in the previous step.

Application of the main algorithm The detection algorithm explained in [Section 3.1](#) is now applied to `tmap+sources`, for each different simulated `tmap`. Additionally, an extra map is passed to the algorithm without adding the sources, to check anomalies and for visualisation and comparison purposes.

If the user has chosen not to add artificial sources, all the maps are given to the algorithm skipping the previous step. This may be useful to check the statistical properties of the algorithm on large amounts of maps. For example, in [figure 3.5](#) it can be seen the distribution of maxima for 50 different maps without sources, compared to the theoretical distribution, and with the 68 and 90% confidence regions.

Comparison of the detected and injected PS For every map, we have now a list of candidates to be PS. This list consists of the pixel number and intensity of the maxima. We convert the pixel into its coordinates and then compare with the coordinates of the artificial sources that we had injected into the image. If there is any artificial source at less than `rho` pixels from the detection, we will claim the detection to be successful. Otherwise, we will declare it a false detection.

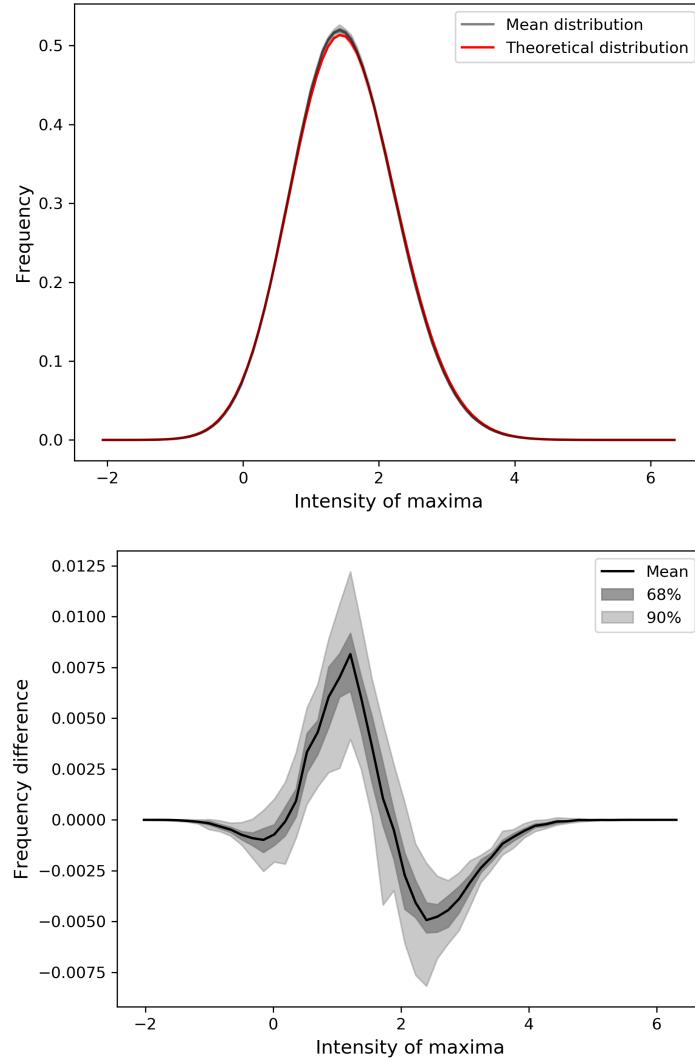


Figure 3.5: Maxima distribution for 50 simulated CMB maps, after filtering with a needlet $B = 1.2$, $j = 39$. On top, we have the mean frequency and the theoretical distribution. The mean frequency is represented in black, with 68 and 90% confidence regions. The theoretical distribution is represented in red. On the bottom, we have the difference between the mean and the prediction, with the same confidence regions. We note the difference of magnitude in the y-axis. It can also be noted the small dispersion of the distribution and how close it is to the prediction. The small discrepancy could be caused by pixelisation effects, as the theoretical distribution is calculated for a continuous map.

```

sortmax = maxi[maxi[1].argsort()]
canlist = sortmax[0,-index:]
vecdet = np.array(hp.pix2vec(nside,
                             canlist.astype(int))).T
dist = scipy.spatial.distance.cdist(
      vecsou, vecdet, 'cosine')
where = np.argwhere(np.abs(dist) < 1.-np.cos(angtol))

detect = np.unique(where[:,0])
detectinten = souinten[detect]

```

In this way, `detect` is the list of successful detections and `detectinten` is a list of the intensities that the artificial sources corresponding with the detections have. There are two quantities that can be measured at this point: the power of detection (number of successful detections over number of artificial sources); and the false detection rate (number of false detections over number of detections). If the source distribution that we inject is realistic, these numbers can give an estimate of what we expect for the actual map. On the other hand, introducing a uniform distribution is useful to know how sensitive is the algorithm. In this way we can know down to which intensity we are able to detect PS. We will study this in [Section 4.1](#).

3.2.3 Effect of the parameters in the simulations

We are going to start discussing the effect of the `cls` of the simulated maps. A change of this parameter can be due to a direct input from the user or a change in the cosmological parameters, but we are not going to discuss here the specific effect of these parameters. The important aspect to take into account here is the amplitude of the `cls` in the multipole region of the needlet, $\ell_{need} \lesssim 1000$, compared to the amplitude of the sources. If `cls` is large in this region, the maxima from CMB fluctuations in this scale are more intense and, therefore, PS are more difficult to detect. This is the expected effect in polarised light, for example, whose C_ℓ have a very different shape than the ones from the intensity. Unfortunately for source detection, we can not change the cosmological parameters of the Universe, so we will be using the C_ℓ for the Planck cosmological parameters.

On the topic on generating artificial PS, we have to mention the importance of the beam smoothing. We define their intensity in terms of their total signal, before convolving with the beam. When the convolution is done, the intensity of the central pixel of this source is greatly reduced. If the width of this beam `fwhm_min` is too large, the sources will become too diluted and their peak intensity may be too low to be detected by the algorithm. By contrast, if this source is convolved by a very narrow beam, the intensity will be still high and this maximum will be more easily detectable. This is, of course, a consequence of the fact that it is easier to detect PS when we have better angular resolution. Computationally, this means that we may need to adjust the maximum intensity `I_max` of the sources according to the width of the beam. The default value works well with `fwhm_min = 5'`.

Less important is the choice of `nmaps`. For robust simulations, we choose values up to `nmaps = 100` different maps, but all the results are very close

to more modest `nside` = 10. If the user just wants to quickly explore the behaviour of some aspect of the simulation, even lower numbers will give good estimates of the actual values, in our experience. The modification of `nside`, does not significantly modify the results as long as the resolution is good enough for the corresponding multipole region. As mentioned before, we recommend $\text{nside} \gtrsim \ell_{\text{need}} \Rightarrow \text{nside} \gtrsim B^{j+1}$ to avoid any possible pixelisation effect.

Finally, there are a couple of points to take into account when modifying `rho` or, alternatively, `fwhm_min`. It is too optimistic to expect all the detections to happen in the exact same pixels as the source was introduced, since a CMB map has been added. That is why we have to impose a tolerance criterion on the distance between the detection and the injection of the point. We find that increasing this parameter more than ~ 3 pixels does not heavily influence the results, while lower values significantly reduces the reported detections. This is hinting that this distance may be approximately the maximum error that can be induced by introducing the CMB map. If we increase `rho` or the number of sources `nsources` too much, we are at risk that all (or a significant part of) the sphere are closer than `rho` pixels from a artificial source. During calculations, we always modify the two parameters together to avoid this possibility.

This scheme is also suitable to test the influence of other effects, such as instrumental noise, galactic masks, or PS masks. These effects are introduced before passing the maps to the detection algorithm. We will see the results of these considerations in [Section 4.1](#).

3.3 Planck data

In this last section of the code and implementation, we are going to explain the part of the code in charge of applying the detection algorithm to maps of real temperature anisotropies of the CMB and how the results are obtained.

3.3.1 Parameters for the Planck data

Most of the input parameters are the same that for the other parts of the code. However, this part accepts a list for some parameters, in which case it will run the detection algorithm for each option and for each map. The parameters are the following:

- `jlist = np.array([38, 39, 40])` and $B = 1.2$: this parameters determine the needlet to be used. Each option for `j` is executed independently for each map.
- Files and folders: there are several parameters that determine the location of files and folders needed to execute the algorithm. These are:
`folder` determines the folder where the code is executed and β -maps are stored, as before.
`mapsfolder` determines the folder that contains all the CMB temperature maps that we want to analyse.
`cataloguesfolder` determines the folder with the catalogues of known PS that we want to compare with the list of PS candidates.

`masksfolder` determines the folder that contains the masks to be used to check the location of the points. We note that this files will not be used to obtain the list of candidates, only to check if these points are outside the confidence region of the maps.

- `alphalist = np.array([0.05, 0.01, 0.002, 0.])`: this is the list of values for `alpha` that will be used during the multiple testing explained in [Section 2.2.2](#). We note that all the processing of the map and *p*-values are calculated only once, and then the list of candidates is extracted for each different value of `alpha`.
- `angtol = 3`: this parameter, similar to what it does in the previous part, denotes how close a candidate has to be to a known source in the catalogue to be declared a match. It is expressed in arcminutes.
- `nside = 2048`: although the original resolution of the Planck maps that we are using is 2048, the user can decide to use a lower resolution to speed up the execution of the code. We recommend avoiding increasing the resolution, as the information contained in the map will be the same but the computation time will increase significantly.

3.3.2 Steps for the Planck data

In order to apply the algorithm and obtain results from the Planck CMB maps, we follow the next steps:

Loading the maps Planck temperature maps are publicly available in the Planck Legacy Archive¹⁴. Here, we can obtain the maps corresponding to the four different algorithms in `.fits` format. In the code, we define a function `loadmap` that will take the name of the file and convert it to a variable similar to the `tmap` in previous sections. In order to do that, the function opens the file and converts the map to the correct resolution given by `nside`, if necessary. The HEALPix standard admits two ways of ordering the pixels, *ring* and *nested*. The programme also changes the ordering of the pixels if they are numbered in the other way (we work in ring scheme). Additionally, each file can contain several maps such as temperature, polarisation or inpainted maps. The function loads only the target map.

Aside from getting the map, we also store the algorithm used to extract it, the version (second or third release), and for the final release, whether the map has been inpainted or not. In this way, each algorithm will have three maps: second version, third version, and third version inpainted.

Application of the main algorithm The detection algorithm is now applied to each map as explained in [Section 3.1](#). For each map, the whole algorithm is applied independently for each `j` in `jlist`. The step that consists in the application of the multiple testing is then computed for all different `alpha` in `alphalist`. The data of each computation are also stored: `B`, `j`, `alpha`, number of candidates, location and intensity of each candidate, and `nside`. The last one is stored even if we are working always with the same value, so all the relevant information about the map is stored together.

¹⁴The Planck Legacy Archive is publicly accessible at <http://pla.esac.esa.int/pla>.

Checking the location of the candidates We use the confidence masks provided by Planck. For the second data release, they provide masks covering several fractions of the sky such as 40, 70 and 90%. For the last data release, they provide the two confidence masks explained in [Section 1.3.3](#), which can be seen in [figure 1.16](#). The first of them is a confidence mask with a 78% of the sky, common to all four algorithms. The second is an inpainting mask, with 98% of the sky, covering only the most problematic regions of the sky, which are removed in the inpainted maps. In order to compare both versions with the same criteria, we will use the two masks provided in the final data release for all maps.

For each candidate, we check if they are inside these masks or if they are in a confidently extracted region of the sky. This gives us an idea of the kind of location where the candidate has been detected. If it is inside the inpainted mask, the source is known and taken into account for scientific analysis; in the inpainted maps it has been supposedly removed and the area has been inpainted. If it is in the common confidence mask, the candidate is located on a reportedly problematic area on the sky, probably dominated by the foregrounds or with very different values for different algorithms. Finally, if the candidate is not in any of these masks, it is located in a region of the sky where the extraction of the CMB map is assumed to be accurate. This would mean that the cleaned maps are not compatible with gaussianity in certain points, where they could be contaminated by PS. The number of sources outside each mask for each map is also stored.

Checking the catalogues of PS After we have located all the candidates in the sky, we want to check if these candidates are known sources present in the catalogues. For that, we check the Planck catalogues of PS explained in [Section 1.3.2](#). We check the sources for all frequencies in a similar way that we checked the artificial sources while doing the simulations in [Section 3.2](#): we have an angular tolerance of `angtol` arcminutes (or `rho` pixels) and if there is any source closer than that from a candidate, we consider that the candidate is a real PS. We store the number of candidates that match a source in the catalogue, both for the entire map and for the regions outside each mask.

Display of the results After computing all these data for every map, all results are organised in a `pandas` table and stored locally to avoid losing or inadvertently modify the results. This table has one row for each time that a list of candidates is computed: $4 \text{ algorithms} \times 3 \text{ versions} \times \#(\text{jlist}) \times \#(\text{alphalist})$. With the default values, this means $4 \times 3 \times 3 \times 4 = 144$ rows.

The columns will contain the information of the map, the parameters of the algorithm, and the results. The results for the default values can be found in [Appendix B](#). Since the table is too large to be easily analysed as-is, we recommend to take advantage of its `pandas` structure to show only certain columns, and only the rows that we are interested to compare, as well as the possibility to sort the table using the information within it. For example, if we want to see the results for $j = 39$ and $\alpha = 0.01$, sorted by algorithm and version, we can use the code:

```
index_to_show=[ 'Map' , 'version' , 'inpainted' , 'B' , 'j' ,
    'nside' , 'alpha' , 'detections' , 'Cat_PS' ,
```

```
'outside_common_mask', 'outside_inpainting_mask']
order=[ 'Map', 'version' ]
query='j==39&alpha==0.01'
results.sort_values(order).query(query)[index_to_show]
```

where `results` is the table containing all the results.

3.3.3 Effect of the parameters for the Planck data

Some of the parameters such as `nside`, `B` or `angtol` are the same than before. They influence the extraction of the candidate lists in a similar fashion that they did in previous parts. The obvious difference is the change of the input maps. A curious reader that wants to try maps different from the Planck ones has to be careful with the process of loading the maps, since different sources may pack the information differently. An effort has been made to make the loading of the maps as general as possible (it accepts different resolutions, ordering of the pixels, or dimensions of the map array), but we recommend the user to check that the map is correctly loaded before proceeding with the algorithm. Additionally, the values for the algorithm, version, and inpainted or not will show incorrect results unless manually modified.

A further comment must be made in regard to the default values of `alphalist`. One of these values correspond to `alpha = 0`. If we remember the criterion for the extraction of the candidate list, in [equation \(2.17\)](#), we see that $\alpha = 0$ means that the only p -values selected are the ones strictly equal to 0, which are mathematically impossible. Indeed, this is a tool to detect the number of candidates whose p -value saturates the precision of the algorithm. This happens when $x > 37.5$ times the standard deviation of the map, corresponding to $p\text{-value}(x) \lesssim 10^{-304}$. Although these values are uncommon, they are present in some not inpainted maps, and in this way we can control their quantity and location

❖ Chapter Four ❖

Results and Discussion

In this chapter we are going to present the results obtained using the tools described in [Chapter 2](#) and implemented in [Chapter 3](#). That is, we present the results for the algorithm to detect PS with a needlet filtering and multiple testing. As explained in the last chapter, we use this algorithm in two distinct ways: to test the procedure with simulated CMB maps and to obtain PS candidates in real Planck data.

First, we developed a framework to test the algorithm against a large set of simulated CMB temperature maps, with the possibility to introduce artificial PS. This allows us to test the behaviour of the algorithm. We will study its sensitivity and power through different metrics to understand the limits and strengths of the procedure. This scheme also allows us to study the effects of different aspects that could affect the algorithm, such as instrumental noise or the use of masks. We will present the results derived from these explorations in [Section 4.1](#).

The second way we use the algorithm is to apply it to the real Planck CMB temperature maps. We obtain several results for each map, aside from the lists of candidates themselves, such as number of candidates in each region (inside or outside of each mask) or the number of candidates found in some of the Planck PS catalogues. On the one hand, we can study how these results vary with parameters of the algorithm. On the other hand, we can study how they vary with the algorithm used to clean the map or the version of the map itself. We will present our results and explore these questions in [Section 4.2](#).

4.1 Simulations

We are going to start by exploring the results obtained from the simulations to understand the behaviour of the algorithm. First of all, we have to test if there are candidates to PS reported in a simulated gaussian map, without added sources. We expect the algorithm to report no candidates, within a reasonable statistical margin. In order to do that, we simulate a total of 50 maps, filter them with a needlet with $B = 1.2$ and $j = 39$, and apply the multiple testing approach with different confidence parameters α . In [table 4.1](#), we show the results of these simulations. As expected, almost no candidate to PS is found, although a small number is indeed reported for higher values of α .

Therefore, for a real map without PS, we expect the algorithm not to report any candidate in most of the cases, even for high values of α . In the case that

α	Maps with candidates	Total number of candidates
0.05	3 /50	4
0.01	1 /50	1
0.002	0 /50	0

Table 4.1: PS candidates obtained for a total run of 50 CMB maps without added artificial sources. As expected, very few candidates are reported and the number increases with α .

the algorithm reports a number of candidates by random fluctuations, we expect this number to be negligible (1–2). In any case, both effects can be reduced by choosing a suitable value of α .

4.1.1 Sensitivity of the algorithm

We have just seen that the algorithm behaves as expected on maps without sources. The next step is to test its behaviour on maps with artificial sources. In order to do that, we will inject 200 artificial PS in the way described in [Section 3.2](#). We will run the algorithm on 10 different maps and then check the candidates obtained. For every candidate, we save its intensity and whether or not it corresponds to one of the artificial sources injected.

We are going to focus on two aspects. The first one is which fraction of the sources are we correctly detecting for each intensity bin. Or, in other words, down to what intensity we recover most of the injected sources. This would be the limit on the intensity for which we expect to be complete on the PS candidate list.

The other aspect we want to know is what kind of false detections the algorithm can do. That is, the reported candidates that do not correspond to any injected source: we want to know which fraction of the candidates corresponds to false detections in each intensity bin.

In [figure 4.1](#) we can see these results for a needlet filtering with $B = 1.2$ and $j = 39$, and a confidence value $\alpha = 0.01$. The intensity of the maxima is measured on the unfiltered temperature map, with units of its standard deviation σ . It can be seen that the algorithm detects essentially every artificial source over 4σ and more than half over 3σ . Additionally, the number of false detections is very small and almost negligible over 2σ .

One may argue that real data are never as good as an ideal case and may introduce unforeseen effects that reduce the efficiency of the algorithm. However, even if the efficiency is slightly affected by unforeseen effects, these results indicate that the algorithm reports the majority of the real PS down to very low levels of the intensity and with a small amount of false detections.

In [figure 4.2](#) we can see the same kind of results, but with different values of the confidence parameter α . In the previous case, we were using the standard value $\alpha = 0.01$. In this figure we see the result for $\alpha = 0.05$ and $\alpha = 0.002$, 5 times lower and higher than the previous one. One effect is the change in sensitivity: the intensity for which we detect most of the sources is lower when α is higher. The second effect is on the incorrectly detected candidates: with higher α , the algorithm makes a higher number of mistakes. This is what

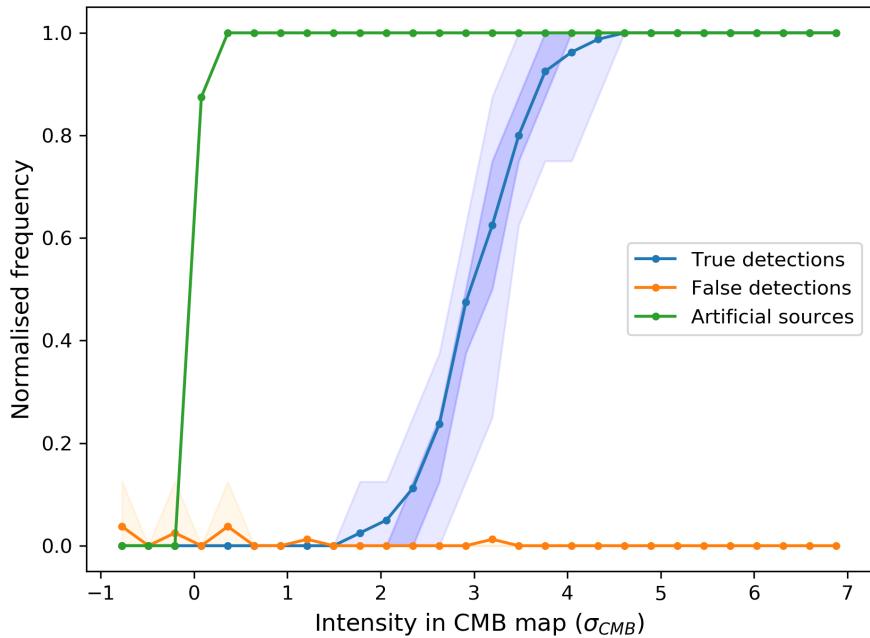


Figure 4.1: Sensitivity of the algorithm. In the x-axis, we have the intensity of the candidates to PS, referred to the *unfiltered* CMB temperature maps, in units of its standard deviation. In the y-axis, we have the frequency of detection of the maxima in each bin, normalised so that the amount of artificial sources is 1. In green, we have the artificial sources, introduced uniformly between 0 and $7\sigma_{CMB}$. In blue, we have the true detections: number of artificial sources correctly detected. In orange, we have the false detections: reported candidates that do not correspond to any artificial source. For both true and false detection, we plot the average and the confidence regions where 70 and 90% of the maps lay. The intensity for the artificial sources is the one with which they were generated, the one for the false detections is measured on the map. The algorithm has been applied with 200 sources, a needlet filter of $B = 1.2$ and $j = 39$, and $\alpha = 0.01$.

was expected, since a more generous (higher) α will imply a higher amount of candidates: this means both more correctly and incorrectly reported candidates. In any case, the algorithm with any of these values of α exhibits very good properties.

These previous results are reported with the intensity referring to the CMB temperature map, before the filtering is applied. Since the detection is done on the β -maps, it may be interesting to see similar plots with the intensity on these maps. This plot can be seen in figure 4.3. The first difference is the noisy profile of the detections. This is because this intensity is the one measured on the filtered map after adding a gaussian CMB map, instead of the intensity of the artificial PS itself. More relevant is the difference in scale: while PS are generated with intensities between 0 and $7\sigma_{CMB}$, the intensity of these points in the β -maps is up to $14\sigma_\beta$. Filtering the maps implies then a significant boost

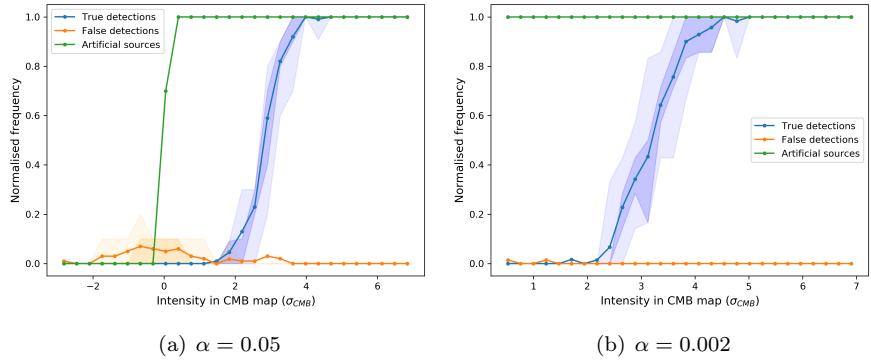


Figure 4.2: Same graphs as in figure 4.1, but with different values of the confidence parameter α . On the left, 0.05. On the right, 0.002. It can be observed that the intensity for which the algorithm correctly detects most of the sources is higher when we use a lower value of α . On the other hand, This lower value of α also implies that almost no source is incorrectly reported.

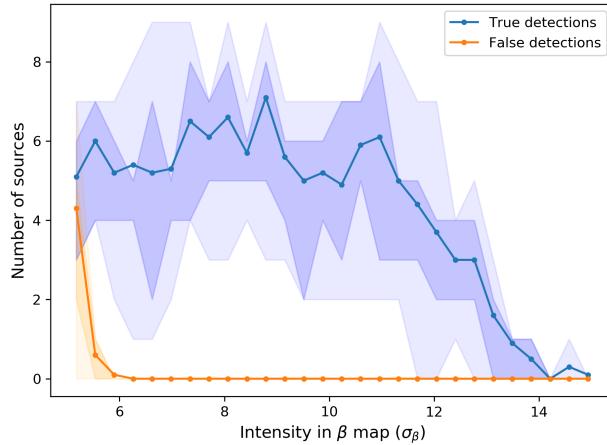


Figure 4.3: Sensitivity of the algorithm on the *filtered* maps. In the x-axis, we have the intensity of the candidates to PS, measured directly on the β -maps. In the y-axis, we have the number of detections. Detections that correspond to artificial PS are plotted in blue, while detections that do not, are represented in orange.

of these signals. This plot is obtained for $\alpha = 0.05$; lower values imply that a small part of the lower end of the graph is dismissed (the lower cut for the intensity is higher). At $\alpha = 0.002$, the lowest intensity is around $6\sigma_\beta$. As we saw before, this will significantly reduce the false detections but will also reduce the true detections.

Moving to another aspect, the total number of PS candidates (or in the simulation case, artificial PS injected) also has an effect on the sensitivity. Having a large number of sources will increase the sensitivity and the incorrectly

reported candidates in the same way that increasing α . The effect, however, is small and can be neglected for most applications.

Lastly, the smoothing beam (or, similarly, the telescope resolution) also plays an important effect. If the change is relatively small, we observe that the sensitivity of the algorithm increases when the beam becomes larger. That is, a wider beam will allow us to better detect maxima at lower intensities. This may seem counter-intuitive at first sight, but we also have to remember that a larger beam means that a PS signal will be much less concentrated. Therefore, the *peak* intensity of the source will be much smaller even if the total signal is the same. Although the peak intensity of the detected maxima become a little bit lower, the total fraction of detected sources will decrease, as expected. On the other hand, if the change of beam is too large, it may be necessary to change the needlet frequency j to match the new size before doing any kind of analysis.

4.1.2 Noise

When we deal with actual data, there is one additional factor that we have not mentioned yet: instrumental noise. In this case, this is translated in noise for individual pixels. We want to analyse whether this noise can introduce significant differences in the results of the algorithm.

One thing to consider is that PS, although small, are several pixels wide. Because of this reason, the needlets chosen are, likewise, several pixels wide. This means that filtering will take into account the value of tens of pixels, which makes it less susceptible to noise in individual pixels.

In other words, noise acts predominantly at very high multipoles ℓ (corresponding to pixel sizes) and we filter at slightly lower multipoles (corresponding approximately to the optical resolution). This can be seen in [figure 4.4](#), where we can see a comparison between the estimated C_ℓ corresponding to noise and the C_ℓ of the temperature map. It can be seen that noise dominates at $\ell \gtrsim 2000$, larger than the main region we are filtering. The profiles of the Mexican needlets that we are using ($B = 1.2$ and $j = 38, 39$ and 40) are also plotted to illustrate the working region.

Therefore, we have that, theoretically, the algorithm should be mostly unaffected. Noise may have some slight impact of the quality on the detection, but the list of candidates should remain mostly the same. In order to test this, we introduce some noise in simulations with sources. Noise for the Planck maps can be computed in several ways. We use two different different noise maps: one map directly provided by the Planck Legacy Archive for the second data release for 100 MHz and the other is obtained using the difference of two CMB temperature maps computed with different halves of the data. The latter is a common way to estimate noise, where the halves are specifically divided using even and odd observation rings, a way of dividing the data with little bias.

In [figure 4.5](#) we can see an example of a detected source with and without adding a noise component to the CMB temperature map, and its corresponding β -map. We see that the effect that noise have on the filtered map is minimal. In general, we find that the lists of PS candidates are indeed unaffected except for very few sources in the limit of the detection. Therefore, we can conclude that instrumental noise does not seem to affect the results in any significant way, motivating the absence of noise in the main algorithm for simulations.

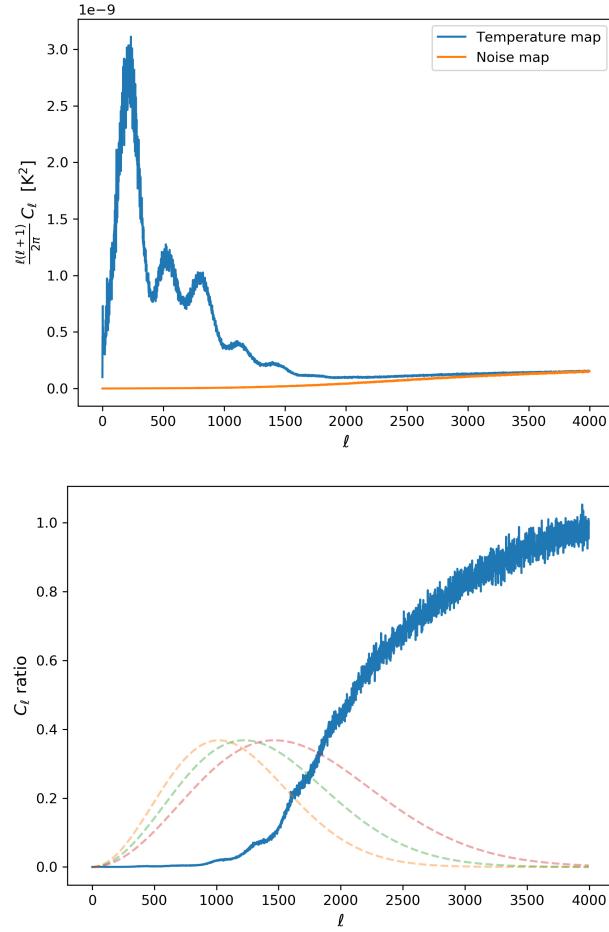


Figure 4.4: Angular power spectrum C_ℓ of the noise. On the top image, we can see the contribution of the noise and the total C_ℓ . On the bottom, we have the ratio between the two, which approximately represents the fraction of the C_ℓ due to the noise. Overlapped with this image we can see the filter profiles b of the Mexican needlets for $B = 1.2$ and $j = 38, 39$ and 40 . The map used is the one extracted with the SMICA algorithm. The noise has been estimated with the semi-difference of the odd and even ring maps (see text).

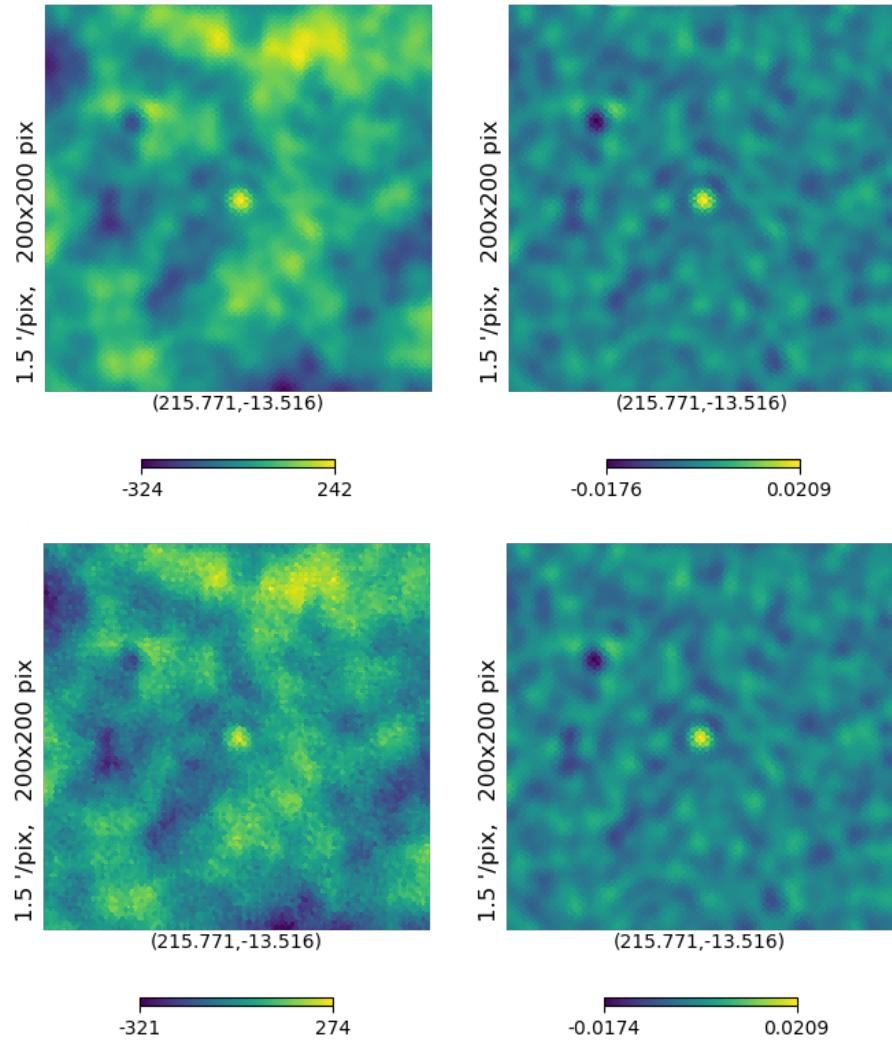


Figure 4.5: Example of a simulation without noise (on top) and with noise (on the bottom). They are the same simulation with the same artificial PS injected at the centre. It can be seen that the introduction of random noise in the temperature maps (on the left) have a minor impact on the filtered β -maps (on the right). The maximum intensity for the source is 2.30 and 6.98 in the temperature and β -map, respectively, without noise and 2.32 and 6.84 with noise. All intensities are given in units of the standard deviation of the maps without noise.

4.1.3 Use of masks

One of the main advantages of using needlets is their localisation properties. This means that the result of filtering depends only on the values close to the centre of the needlet. Furthermore, we are using small scale needlets, due to the high frequencies that we are using ($j \sim 39$). Therefore, the result of filtering the map only depends significantly on values closer than a fraction of degree.

This means that we can consider the possibility of using masks before applying the filtering. In this case, we make all the pixels inside the mask equal to 0. This is usually done when we know that some region does not consist only on CMB data, but it has a known PS or strong contamination from the galaxy. In this way, this contaminated data will not be reflected on the final result and we would avoid to detect the sources that are already known and masked. Since the needlets used act only on a very small scale, the effect of removing this points will leave the β -map unaffected except for points very close to the mask ($\lesssim 5'$).

While testing this possibility, we found mostly what we expected: that the value of the needlets is indeed unchanged for most of the map, its effect being very localised near the masked points. However, the maxima of the map do change in a non-obvious way. Making some areas equal to 0 will produce that some extended regions get cut and reported as a maximum, while the actual maximum was inside the mask. Two examples of this effect can be seen in [figure 4.6](#). This produces a new population of maxima that are not easy to treat or to check with catalogues. We decided not to take this approach as it introduces unnecessary complications for our purpose.

An alternative approach could be to first apply the filter in the unmasked CMB map and obtain the candidates to PS in the usual way. After this, once we have the list of candidates, we check the location of them and disregard the ones located inside the mask. This approach avoids the creation of new maxima by modifying the maps and makes it easier to properly locate and correlate the detected candidates with actual catalogues.

4.2 Planck maps

We are going to explore now the results obtained from applying the algorithm to real Planck CMB temperature maps. We use the pipeline explained in [Section 3.3](#). We analyse the maps reported by the four different algorithms, in the second data release (2015) and final data release (2018), which was made publicly available when our pipeline was almost completely programmed. For the final data release, we analyse both the inpainted and not inpainted versions. All maps are filtered with Mexican needlets with $B = 1.2$ and three different values for j : 38, 39 and 40. The significance parameter is set to the standard $\alpha = 0.01$ for the standard analysis, but the algorithm is also computed for $\alpha = 0.05$, 0.002 and 0, the last one being a proxy to calculate the maxima that saturates the computer precision. All maps are treated at their original resolution of `nside = 2048`.

For each map, the algorithm is run and several results are stored. First, the total number of candidates reported. Their location and intensity are also stored but will not be shown on the tables due to their length. With this,

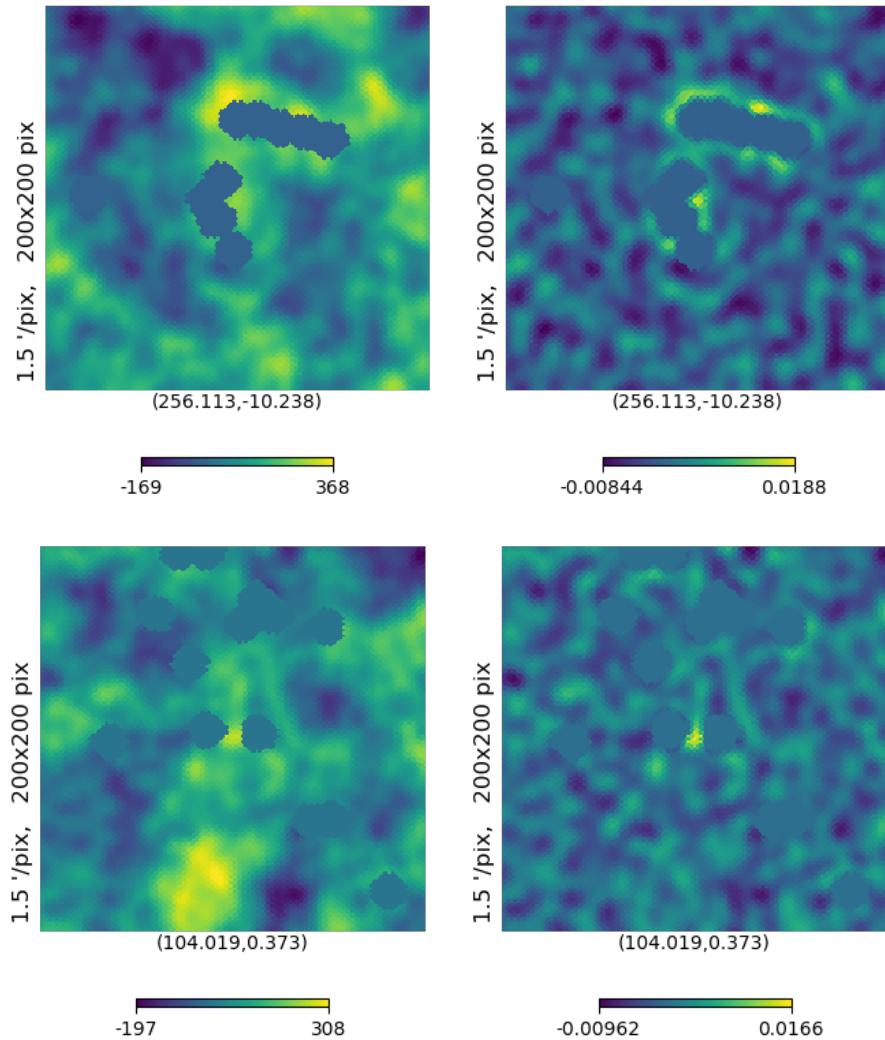


Figure 4.6: Example of new maxima generated because of the use of masks. In this case, A CMB temperature map is masked on the location of known PS, meaning that the values in a patch around the maxima are made 0 before applying the filtering. On the left, we can see the temperature map. On the right, we can see the filtered β -map. We note that points that were not PS are reported as such in the filtered maps. The intensity of these points (at the centre of the images), in units of the standard deviation of their maps, are 2.76 and 2.57 in the temperature maps and 5.46 and 5.48 in the β -maps.

Map	version	inpainted	detections	mask C	mask I	cat	cat C
commander	2.01	False	55	0	2	53	0
commander	3.00	False	6335	37	717	2927	11
commander	3.00	True	70	1	12	8	0
nilc	2.01	False	1222	6	10	753	1
nilc	3.00	False	1240	8	15	755	1
nilc	3.00	True	0	0	0	0	0
sevem	2.01	False	7064	39	690	4056	13
sevem	3.00	False	5894	31	261	3424	10
sevem	3.00	True	3	0	1	0	0
smica	2.01	False	558	2	10	69	0
smica	3.00	False	318	2	16	197	0
smica	3.00	True	11	1	3	7	0

Table 4.2: Small portion of the results table. Here, the three versions for the four algorithms have been filtered with a needlet $B = 1.2$ and $j = 39$. Also, α is fixed to 0.01. The columns are as follow. Map: algorithm used to extract the cleaned temperature map. Version: data release used, 2.01 corresponds to 2015 and 3.00 to 2018. Detections: number of candidates reported. Mask C: number of candidates *outside* the common confidence mask (78% of the sky). Mask I: number of candidates *outside* the inpainting mask (98% of the sky). Cat: number of candidates present in a catalogue. Cat C: number of candidates *outside* the common confidence mask that are present in a catalogue.

we calculate the amount of the candidates outside the inpainting mask and common confidence mask. Lastly, we check how many candidates are in the Planck PS catalogues, both in the total map and outside the common confidence mask. These catalogues are reported for each frequency, here we check all the frequencies at the same time. The catalogues are checked separately for the 80% confident detection (PCCS catalogues), for where confidence can not be determined (PCCS_excluded catalogues), and for both together. We note that some sources may be present in both catalogues for different frequencies, especially for the lower frequencies LFI bands, were there is only one catalogue per band.

We can find the table with the results for the 144 times the algorithm has been applied in [Appendix B](#) a digital version of the `pandas` table is also provided in the online repository, see [Appendix A](#). Each execution of the algorithm is located on a different row. The columns store the information of the parameters of the algorithm, the map used and the results obtained. A portion of this table is shown in [table 4.2](#), only for $j = 39$ and $\alpha = 0.01$, and showing the results only for the global catalogue. More specific tables will be shown when discussing particular effects.

As a general observation, we note that the algorithm reports a significant amount of PS candidates for most of the maps. Some of them are sources already present in the Planck catalogues that, apparently, could not be completely removed for the final maps. However, a fraction of them are sources that are not intense enough to be present in these catalogues.

Map	version	inpainted	j	alpha	detections	mask C	mask I
smica	3.00	False	38	0.01	169	1	6
smica	3.00	False	39	0.01	318	2	16
smica	3.00	False	40	0.01	744	5	79
smica	3.00	False	39	0	4	0	0
smica	3.00	False	39	0.002	268	1	8
smica	3.00	False	39	0.01	318	2	16
smica	3.00	False	39	0.05	403	11	49

Table 4.3: Example of the behaviour of the results with j and α for the SMICA maps from the last data release. On the top, varying j with fixed α . On the bottom, varying α with fixed j .

4.2.1 Results common to all maps

We are going to discuss first the results that are common to all twelve maps. That is, how the parameters influence the results and what are the main repeated behaviours.

We start by focusing on the needlet frequency j . Our standard value is $j = 39$, which can detect a significant amount of PS candidates. This needlet filters multipoles around $\ell = 1225$, corresponding to sizes similar to the nominal optical resolution of the telescope. A lower value $j = 38$, peaking around $\ell = 1020$ is tested, but the larger size of the needlet implies a worse detection of PS. The fraction of the detected candidates is 0.60 ± 0.14 .

A higher value of $j = 40$ is also used, corresponding to a multipole filter peaking at $\ell = 1470$. This needlet detects a higher amount of candidates, but the smaller size means that it is also slightly more sensitive to noise. We have to remember that, although the multipole peak is not in the noise dominated region, the filter will also include a significant contribution from higher multipoles, as we have seen in figure 4.4. With this needlet frequency $j = 40$, the algorithm detects 2.2 ± 1.2 times more sources with respect to $j = 39$.

In the top part of table 4.3 we can see this difference between multiple values of j for SMICA maps for the last data release, without inpainting.

Another important aspect, as discussed before, is the value of the confidence parameter α . An example of the effect of modifying this parameter can be seen in the bottom part of table 4.3, where the results are shown for all four values of α used in the SMICA maps for the last data release, without inpainting. With respect to $\alpha = 0.01$, a lower value of 0.002 means a fraction of detections of 0.79 ± 0.17 . On the other hand, a higher value of 0.05 means the number of detections is 1.39 ± 0.32 times larger. Naturally, these are average numbers that vary from map to map. They are given here just to give an idea of the magnitude of the difference that differences in the parameters can imply.

It is also interesting to see the case of $\alpha = 0$. As we explained before, this is not an actual value of the significance parameter but a proxy to study the maxima so intense that they saturate the precision of the computer when calculating the p -value. This corresponds to maxima of intensity higher than 37.5 times the standard deviation of the β -map and a p -value smaller than 10^{-300} . We can see in table 4.4 all the maps with this kind of saturating points.

Map	version	inpainted	j	alpha	detections	mask C	mask I
sevem	2.01	False	39	0	192	0	0
commander	3.00	False	39	0	907	0	55
sevem	3.00	False	39	0	160	0	0
smica	3.00	False	39	0	4	0	0

Table 4.4: Maps that present at least one detection with intensity larger than 37.5 times the standard deviation of the β -map. At this intensity, the precision of the computer is saturated when calculating their p -values. These maps also present detections at $j = 38$ and 40. These maxima are only detected in not inpainted maps and only inside the common confidence mask.

We note that they are only detected in not-inpainted maps, as the inpainting completely remove such high intensity points. They are mainly located close to the galactic plane. All of them are located inside the common confidence mask (22 % of the sky) and almost all of them are located inside the inpainting mask (2 % of the sky). We also note that **COMMANDER** presents a significantly higher number of these intense maxima, possibly due to the fact that this is the only one of the four algorithms that does not preprocess the maps to remove PS, as we explained in [section 1.3.3](#). The only algorithm that does not present these maxima in the last version is **NILC**.

Some words must be said about the population of PS candidates. We would like to know what kind of intensity distribution they follow. As a representative example, we choose the last version of the **COMMANDER** maps, both inpainted and not inpainted. The intensity distribution of the candidates for these maps can be found in [figure 4.7](#). They approximately follow a power-law distribution, which reminds us the power-law distribution of the physical PS and, particularly, radio galaxies (as we saw in [figure 1.12](#)). However, the comparison is not straightforward, as that distribution applies for the fluxes of the objects, while the one for the candidates here is for intensities measured in the filtered β -map. Furthermore, that distribution only applied for radio galaxies, while the candidates can also contain dusty galaxies. All things considered, the intensity distribution for the PS candidates is more similar to one expected from physical PS than one expected from random gaussian fluctuations.

4.2.2 Comparison of the maps

We are now going to analyse the differences between the results for different maps, as illustrated in [table 4.2](#). In particular, we are going to discuss differences between the algorithm used to extract the cleaned map, the version (data release of the map), and whether it is inpainted or not. We discuss the results for the parameters $j = 39$ and $\alpha = 0.01$, but other choices report similar results.

- **SEVEM**: this algorithm has the higher number of detected candidates for the 2.01 version (7064). The results for this algorithm follow the expected behaviour: the 3.00 version has a lower amount of candidates (5894). If we count how many of these candidates are outside the confidence mask, we have 31 points for the new version. This is a small fraction of the total amount, although it is still a high number of candidates. If we

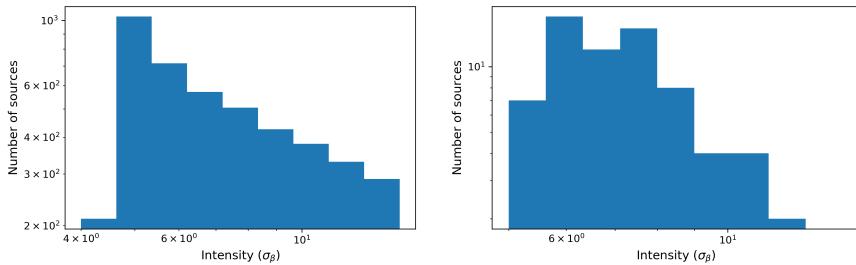


Figure 4.7: Intensity distribution, in logarithmic scale, of the PS candidates found by the algorithm on COMMANDER maps. On the left, distribution for candidates in the not inpainted map. On the right, distribution for candidates in the inpainted map. Both seem to approximately follow a power law, albeit the inpainted map presents a more noisy distribution due to the lower amount of candidates.

look at the inpainted map, which has the very intense points removed, the algorithm reports almost no candidates for the whole map (only 3), and none of them is outside the confidence mask, as expected.

- **NILC:** this algorithm reports a lower number of candidates overall. For the 2.01 version, it reported 1222 candidates and this quantity remains similar for the new version, 1240. As before, a small fraction of them are outside the confidence mask (6 and 8, respectively). The most relevant result for this algorithm, however, is that it is the only one that reports no candidates at all for the inpainted map. In this case, the inpainted map is compatible with a random gaussian map under the criteria used during this work.
- **SMICA:** this algorithm has the lowest number of candidates for a (not-inpainted) 3.00 map, at 318 candidates. This number is also lower than the previous 558 for the 2.01 version. Very few of them are outside the confidence mask, only 2. For this and the **SEVEM** algorithms, improvements for the new version have produced maps with a significantly lower amount of PS candidates, compared with the second version.

If we look at the inpainted map, the number of detections is 11 and, more interestingly, 1 of them is outside the confidence mask. This is one of the points also detected in the not inpainted map. This particular candidate can be seen in [figure 4.8](#), both in the temperature map and in the filtered β -map. This candidate has a typical signal common for most of the candidates and resembling the profile of a PS.

- **COMMANDER:** the results for this algorithm present a particular behaviour. It has the lower amount of candidates for the 2.01 version (55), but the highest for the 3.00 version (6335). This high number is due to the fact that it is the only algorithm that does not preprocess the frequency maps to remove possible PS before applying the algorithm. This produces a more robust procedure in exchange of higher levels of contamination to the map. This effect is mainly located in the galactic plane, although 37

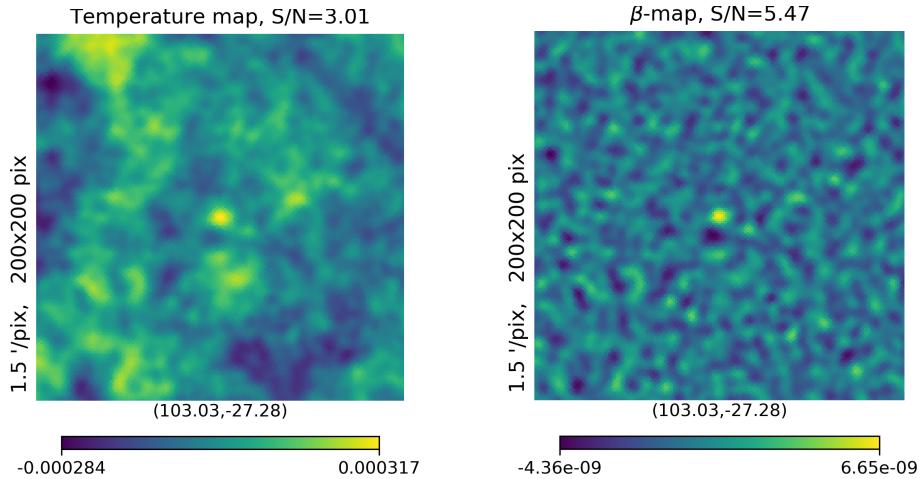


Figure 4.8: Example of a candidate reported by the algorithm. This PS candidate is detected in the SMICA maps, both inpainted and not-inpainted. On the left, we see the CMB temperature map. On the right, we see the filtered β -map corresponding to this map. It is located outside the confidence common mask, so it is in a region where the contamination is expected to be low. It can be seen that the size and shape match the expected from a physical PS.

of the candidates are located outside the confidence mask. As expected, the result for the inpainted map is lower (70), but still higher than for other algorithms. These points are mainly located in the galactic plane, but one of them is located outside the confidence mask.

Regarding the presence of the candidates in the Planck catalogues, about half of them can be found in at least one of them. However, more than two thirds of the candidates outside the confidence mask are not present in the catalogues, which can contribute to their presence in the cleaned maps.

❖ Chapter Five ❖

Conclusions and Future Work

In this chapter we are presenting a brief summary of our conclusions and the most important aspects of this work. We are also outlining several directions in which this work could be expanded, as there are several lines of research where this approach could be applied. We start by listing what we consider to be the main points of this work:

- We have provided a brief overview of the Λ CDM model, the currently accepted cosmological model, to understand the evolution of the Universe and the importance of an accurate determination of the cosmological parameters. After that, we have briefly explained the theory behind the CMB and how we observe it, showing that it is an important gateway to understand the Early Universe. Additionally, we have explained the main difficulties and sources of contamination that arise when studying it. We have sketched the idea of the different algorithms that extract cleaned CMB maps, as these are the maps that will be analysed by our programme. This is done in [Chapter 1](#).
- We have explained the concept of needlet and their main characteristics. The fact that they are well localised makes them an ideal tool to analyse PS. We have discussed the standard and Mexican needlets, and the differences between them. We have chosen the latter due to their better localisation properties. Then, we have explained the algorithm used to extract the possible PS from a map, using a multiple testing technique based on the Benjamini-Hochberg procedure on the maxima on the map. This algorithm has been previously analysed in Cheng, Cammarota, et al. ([2016](#)), where they proved it has good mathematical properties. The description of needlets and the algorithm is done in [Chapter 2](#).
- We have implemented the needlet filtering and the algorithm to extract the candidates to PS from a map. In order to do that, we have written a code from scratch in `Python 3`. We have tried to develop this pipeline as flexible as possible, accepting a wide variety of parameters as input. Doing so makes this piece of code one of the most important results of this work. We have used it in a Planck temperature maps as an example, but it can be used to analyse other kinds of maps. This piece of code is explained in [Section 3.1](#).
- In order to test the code and understand how different effects of parameters may affect its results, we have developed a framework to generate CMB

artificial maps and feed them to the algorithm. The maps are generated with **CAMB** according to an input set of cosmological parameters. They can be convolved with a gaussian beam and a set of artificial PS can be likewise introduced. This part of the code is discussed in [Section 3.2](#).

- We have analysed a large number of simulated maps, some of them with artificial sources, in order to obtain the sensitivity of the algorithm. Without artificial PS, the algorithm reports no candidates in almost every case. With sources, we recover most of them at intensities higher than 3 to 4 times the standard deviation of the map. We have observed that the algorithm reports some false sources, mostly at intensities below 2 times the standard deviation. The results depend slightly on the confidence parameter α . These results are discussed in [Section 4.1.1](#).
- Using simulations again, we have tested the effect of noise and masks. We have concluded that noise is negligible in our multipole range, although it would become important for smaller needlets (higher j). Masks do not significantly affect the result of filtering the map except very close to the masks. However, they can alter the detection of sources and the location of known maxima. Therefore, we do not include masks in our algorithm, we only use them to check the location of candidates. These two effects are discussed in [Sections 4.1.2](#) and [4.1.3](#).
- We have developed a third piece of code to take the files with the Planck maps and extract the selected map with the correct size and format. This code also store the characteristics of the map and parameters of the algorithm, to make the later analysis easier. After running the detection algorithm, this part of the code store all the relevant results such as the number and location of candidates to PS. Although this programme is specific for Planck maps, it can be easily adapted to other maps if the user wants. This is advised as it automatises the process of loading the maps and stores the results in a compact way. This part of the code is explained in [Section 3.3](#).
- We have run the algorithm on a set of cleaned Planck CMB maps, including second and final data release (2015 and 2018, respectively). For the final data release, we have included both inpainted and not inpainted maps. All the used maps are temperature anisotropy maps. We have observed that the algorithm reports candidates for most maps. Some of the candidates to PS are found outside the confidence mask. We note that, since the algorithm is based on multiple testing, this does not mean that each candidate is incompatible with a gaussian map. Instead, it means that the candidate *population* is incompatible with the maxima population found in a gaussian map. We discuss the results about Planck temperature maps in [Section 4.2](#).

There are several ways in which the work explained here could be applied or extended. As for now, we list four aspects that could be useful for the future, and that can be faced at the current state of the project. We hope to keep working in the following directions:

- Until now, we have worked on *temperature* maps. However, this techniques can be readily applied to any kind of spherical map that we suspect to be contaminated by PS. In particular, we plan to apply this algorithm to *polarisation* maps, both the *E* and *B* modes. They have a different aspect than temperature maps, dominated by perturbations at smaller scales, which may make the detection more difficult. However, we estimate that it could be possible to detect highly polarised sources with this algorithm.
- Likewise, this algorithm could be eventually applied to *frequency* maps, before they are combined to obtain the temperature (or polarisation) maps. In this case, we have the additional problem of diffuse galactic radiation contaminating the background, plus all the population of physical PS, not just a few. Ideally, the algorithm could be applied to different frequency bands and then combined, which may increase the sensitivity. We note that as for now we do not use the spectral information of the CMB observations.
- We have mentioned that masks are not used in the algorithm except in order to check the region where a PS candidate is located. This is due to the undesired effects that masks have if we use them to modify the maps. However, a more sensible use of the masks may be introduced as long as the maps themselves are not modified. For example, we could introduce the use of masks by applying the multiple testing step only to maxima outside the mask.
- Finally, we would like to further develop the code and make it publicly available in a way that may be useful to the scientific community. This could be useful both for researchers interested in the use of needlets and β -map extraction, and for researchers interested in PS identification and the study of the CMB.

Bibliography

- ¹P. Baldi, G. Kerkyacharian, D. Marinucci, and D. Picard, “Asymptotics for spherical needlets”, *The Annals of Statistics* **37**, 1150–1171 (2009).
- ²N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, “Non-Gaussianity from Inflation: Theory and Observations”, *Physics Reports* **402**, 103–266 (2004).
- ³S. Basak and J. Delabrouille, “A needlet ILC analysis of WMAP 9-year polarization data: CMB polarization power spectra”, *Monthly Notices of the Royal Astronomical Society* **435**, 18–29 (2013).
- ⁴E. J. Baxter, C. H. Blake, and B. Jain, “Probing Oort clouds around Milky Way stars with CMB surveys”, (2018).
- ⁵Y. Benjamini and Y. Hochberg, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”, *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).
- ⁶D. Cheng, V. Cammarota, Y. Fantaye, D. Marinucci, and A. Schwartzman, “Multiple testing of local maxima for detection of peaks on the (celestial) sphere”, (2016).
- ⁷D. Cheng and A. Schwartzman, “Distribution of the Height of Local Maxima of Gaussian Random Fields”, *Extremes* **18**, 213–240 (2015).
- ⁸J. Chluba and R. A. Sunyaev, “Induced two-photon decay of the 2s level and the rate of cosmological hydrogen recombination”, *Astronomy & Astrophysics* **446**, 39–42 (2006).
- ⁹T. A. Collaboration, “The Astropy Project: Building an inclusive, open-science project and status of the v2.0 core package”, (2018).
- ¹⁰N. B. Cowan, G. Holder, and N. A. Kaib, “Cosmologists in Search of Planet Nine: The Case for CMB Experiments”, *The Astrophysical Journal Letters* **822**, L2 (2016).
- ¹¹D. Croton, “Damn You, Little h! (or, Real-World Applications Of The Hubble Constant Using Observed And Simulated Data)”, *Publications of the Astronomical Society of Australia* **30** (2013) 10.1017/pasa.2013.31.
- ¹²A. de Graaff, Y.-C. Cai, C. Heymans, and J. A. Peacock, “Missing baryons in the cosmic web revealed by the Sunyaev-Zel’dovich effect”, (2017).

- ¹³J. Delabrouille, M. Betoule, J.-B. Melin, M.-A. Miville-Deschénes, J. Gonzalez-Nuevo, M. L. Jeune, G. Castex, G. de Zotti, S. Basak, M. Ashdown, J. Aumont, C. Baccigalupi, A. Banday, J.-P. Bernard, F. R. Bouchet, D. L. Clements, A. da Silva, C. Dickinson, F. Dodu, K. Dolag, F. Elsner, L. Fauvet, G. Fay, G. Giardino, S. Leach, J. Lesgourgues, M. Liguori, J. F. Macias-Perez, M. Massardi, S. Matarrese, P. Mazzotta, L. Montier, S. Mottet, R. Paladini, B. Partridge, R. Piffaretti, G. Prezeau, S. Prunet, S. Ricciardi, M. Roman, B. Schaefer, and L. Toffolatti, “The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths”, *Astronomy & Astrophysics* **553**, A96 (2013).
- ¹⁴J. Delabrouille, J.-F. Cardoso, and G. Patanchon, “Multidetector multicomponent spectral matching and applications for cosmic microwave background data analysis”, *Monthly Notices of the Royal Astronomical Society* **346**, 1089–1102 (2003).
- ¹⁵R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, “Cosmic Black-Body Radiation.”, *The Astrophysical Journal* **142**, 414–419 (1965).
- ¹⁶C. Dickinson, “CMB foregrounds - A brief review”, (2016).
- ¹⁷A. Einstein, “The Foundation of the General Theory of Relativity”, *Annalen Phys.* **49**, 769–822 (1916).
- ¹⁸H. K. Eriksen, I. J. O’Dwyer, J. B. Jewell, B. D. Wandelt, D. L. Larson, K. M. Górski, S. Levin, A. J. Banday, and P. B. Lilje, “Power Spectrum Estimation from High-Resolution Maps by Gibbs Sampling”, *The Astrophysical Journal Supplement Series* **155**, 227 (2004).
- ¹⁹R. Fernández-Cobos, P. Vielva, R. B. Barreiro, and E. Martínez-González, “Multiresolution internal template cleaning: an application to the Wilkinson Microwave Anisotropy Probe 7-yr polarization data”, *Monthly Notices of the Royal Astronomical Society* **420**, 2162–2169 (2012).
- ²⁰A. Friedman, “On the Curvature of Space”, *General Relativity and Gravitation* **31**, 1991–2000 (1999).
- ²¹A. Friedman, “Über die krümmung des raumes”, *Zeitschrift für Physik* **10**, 377–386 (1922).
- ²²J. Frieman, M. Turner, and D. Huterer, “Dark Energy and the Accelerating Universe”, *Annual Review of Astronomy and Astrophysics* **46**, 385–432 (2008).
- ²³D. Geller and A. Mayeli, “Nearly tight frames and space-frequency analysis on compact manifolds”, *Mathematische Zeitschrift* **263**, 235–264 (2009).
- ²⁴K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman, “HEALPix – a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere”, *The Astrophysical Journal* **622**, 759–771 (2005).
- ²⁵A. H. Guth, “Inflationary universe: A possible solution to the horizon and flatness problems”, *Physical Review D* **23**, 347–356 (1981).
- ²⁶E. Hubble, “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae”, *Proceedings of the National Academy of Science* **15**, 168–173 (1929).
- ²⁷J. D. Hunter, “Matplotlib: A 2D Graphics Environment”, *Computing in Science & Engineering* **9**, 90–95 (2007).

- ²⁸K. Ichiki, “CMB foreground: A concise review”, *Progress of Theoretical and Experimental Physics* **2014** (2014) 10.1093/ptep/ptu065.
- ²⁹E. Jones, T. Oliphant, P. Peterson, et al., *SciPy: Open source scientific tools for Python*, 2001.
- ³⁰T. Kluyver, B. Ragan-Kelley, F. Pérez, B. E. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. B. Hamrick, J. Grout, and S. Corlay, “Jupyter Notebooks-a publishing format for reproducible computational workflows.”, in (2016), pp. 87–90.
- ³¹A. G. Lemaître, “A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-galactic Nebulae”, *Monthly Notices of the Royal Astronomical Society* **91**, 483–490 (1931).
- ³²G. Lemaître, “The Beginning of the World from the Point of View of Quantum Theory.”, *Nature* **127**, 706 (1931).
- ³³G. Lemaître, “Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques”, *Annales de la Société Scientifique de Bruxelles* **47**, 49 (1927).
- ³⁴A. Lewis, “CAMB notes”, (2011).
- ³⁵A. Liddle, *An Introduction to Modern Cosmology* (Wiley, 2015).
- ³⁶A. Loeb, “The Habitable Epoch of the Early Universe”, *International Journal of Astrobiology* **13**, 337–339 (2014).
- ³⁷D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, and N. Vittorio, “Spherical needlets for cosmic microwave background data analysis”, *Monthly Notices of the Royal Astronomical Society* **383**, 539–545 (2008).
- ³⁸J. C. Mather, E. S. Cheng, R. E. Eplee Jr., R. B. Isaacman, S. S. Meyer, R. A. Shafer, R. Weiss, E. L. Wright, C. L. Bennett, N. W. Boggess, E. Dwek, S. Gulkis, M. G. Hauser, M. Janssen, T. Kelsall, P. M. Lubin, S. H. Moseley Jr., T. L. Murdock, R. F. Silverberg, G. F. Smoot, and D. T. Wilkinson, “A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite”, *The Astrophysical Journal Letters* **354**, L37–L40 (1990).
- ³⁹W. McKinney, “Data Structures for Statistical Computing in Python”, in (2010), pp. 51–56.
- ⁴⁰F. Narcowich, P. Petrushev, and J. Ward, “Localized Tight Frames on Spheres”, *SIAM Journal on Mathematical Analysis* **38**, 574–594 (2006).
- ⁴¹T. E. Oliphant, *A guide to NumPy*, Vol. 1 (Trelgol Publishing USA, 2006).
- ⁴²P. Peebles, “Seeing Cosmology Grow”, *Annual Review of Astronomy and Astrophysics* **50**, 1–28 (2011).
- ⁴³A. A. Penzias and R. W. Wilson, “A Measurement of Excess Antenna Temperature at 4080 Mc/s.”, *The Astrophysical Journal* **142**, 419–421 (1965).
- ⁴⁴D. Pietrobon, A. Balbi, P. Cabella, and K. M. Gorski, “NeedATool: A Needlet Analysis Tool for Cosmological Data Processing”, *The Astrophysical Journal* **723**, 1 (2010).
- ⁴⁵Planck Collaboration, “Planck 2015 results. XXVI. The Second Planck Catalogue of Compact Sources”, *Astronomy & Astrophysics* **594**, A26 (2016).

- ⁴⁶Planck Collaboration, “Planck 2018 results. I. Overview and the cosmological legacy of Planck”, [\(2018\)](#).
- ⁴⁷Planck Collaboration, “Planck 2018 results. IV. Diffuse component separation”, [\(2018\)](#).
- ⁴⁸Planck Collaboration, “Planck 2018 results. VI. Cosmological parameters”, [\(2018\)](#).
- ⁴⁹Planck Collaboration, “Planck intermediate results. LV. The Planck Multi-frequency Catalogue of Non-thermal Sources”, [\(2018\)](#).
- ⁵⁰Planck Collaboration, “Planck intermediate results. XXV. The Andromeda Galaxy as seen by Planck”, *Astronomy & Astrophysics* **582**, A28 [\(2015\)](#).
- ⁵¹Planck Collaboration, *Planck PLA Internal Wiki*, (2018) https://wiki.cosmos.esa.int/planck-legacy-archive/index.php/Main_Page (visited on 08/15/2018).
- ⁵²H. P. Robertson, “Kinematics and World-Structure”, *The Astrophysical Journal* **82**, 284 [\(1935\)](#).
- ⁵³A. Schwartzman, Y. Gavrilov, and R. J. Adler, “Multiple testing of local maxima for detections of peaks in 1D”, *Annals of statistics* **39**, 3290–3319 [\(2011\)](#).
- ⁵⁴S. Scodeller, Ø. Rudjord, F. K. Hansen, D. Marinucci, D. Geller, and A. Mayeli, “Introducing Mexican Needlets for CMB Analysis: Issues for Practical Applications and Comparison with Standard Needlets”, *The Astrophysical Journal* **733**, 121 [\(2011\)](#).
- ⁵⁵H. Tanimura, G. Hinshaw, I. G. McCarthy, L. Van Waerbeke, Y.-Z. Ma, A. Mead, A. Hojjati, and T. Tröster, “A Search for Warm/Hot Gas Filaments Between Pairs of SDSS Luminous Red Galaxies”, [\(2017\)](#).
- ⁵⁶M. Tucci, L. Toffolatti, G. De Zotti, and E. Martinez-Gonzalez, “High-frequency predictions for number counts and spectral properties of extragalactic radio sources. New evidences of a break at mm wavelengths in spectra of bright blazar sources”, *Astronomy & Astrophysics* **533**, A57 [\(2011\)](#).
- ⁵⁷N. Vittorio, *Cosmology* (CRC Press, Taylor & Francis Group, 2017), 428 pp.
- ⁵⁸A. G. Walker, “On Milne’s Theory of World-Structure”, *Proceedings of the London Mathematical Society* **2**, 90–127 [\(1937\)](#).

Appendix A

Code repository

The online code repository is currently located at:
<https://github.com/javicarron/MasterThesis>

The structure and main points of the code are described in [Chapter 3](#). There, a simplified version of the main parts of the code can be found. The complete code for the algorithm can be found in the online repository.

We have to mention that, as for now, the code is reported as-is and is only intended for evaluation purposes. This means that the repository contains a (cleaned) version of the Jupyter notebooks used. Although it can be used now, we are working in a simplified interface that will make the use of the programme easier, more flexible, and possibly more efficient.

Apart from the code, the most recent version of this document can also be found there, as well as a high resolution copy of some of the images used in this work. The complete table of results, in pandas format, is also in the repository as ‘`results.pkl`’, and it can be loaded in Python with:

```
import pandas as pd  
results = pd.read_pickle('results.pkl')
```

See [Appendix B](#) for more information about this table and an almost complete copy of it.

Appendix B

Table of results

In this appendix we include the table with the results discussed in [Section 4.2](#). A digital version of this table in `pandas` format can be found in the online repository (see [Appendix A](#)). Each row of the table corresponds to a single execution of the algorithm. Different maps are separated by blocks, since the algorithm runs 12 times per map (4 values of `alpha` and 3 values of `j`). The columns of the table are as follow:

- Map: name of the algorithm used to extract the CMB map.
- Version: version of the map, 2.01 corresponds to the second data release (2015) and 3.00 to the final data release (2018).
- Inpainted: whether or not the map has been inpainted.
- B: value of the parameter B of the needlet.
- j: value of the needlet frequency j .
- nside: $nside$ of the map used.
- alpha: value of the confidence parameter α used for the multiple testing.
- Detections: total number of candidates reported by the algorithm.
- Mask C: number of candidates outside the common confidence mask.
- Mask I: number of candidates outside the inpainting mask.
- Cat: number of candidates that are present in at least one catalogue. A “g” indicates that the PCCS2E catalogue (for sources where the detection confidence could not be determined) has not been used. An “e” indicates that only the PCCS2E has been used. A “C” indicates that only sources outside the confidence mask has been considered.

The complete list of candidates is also stored for each execution of the code, including location and intensity of each candidate, but it is not shown here due to the size of the data. They are present in the full version online.

APPENDIX B. TABLE OF RESULTS

86

Map	version	impainted	B	j	inside	alpha	detections	mask C	mask I	Cat	Cat g	Cat e	Cat C	Cat g C	Cat e C
0	smica	2.01	False	1.2	38	2048	0.05	365	14	25	52	8	48	1	1
1	smica	2.01	False	1.2	38	2048	0.01	294	1	2	42	4	42	0	0
2	smica	2.01	False	1.2	38	2048	0.002	242	0	0	36	4	36	0	0
3	smica	2.01	False	1.2	38	2048	0	0	0	0	0	0	0	0	0
4	smica	2.01	False	1.2	39	2048	0.05	728	22	54	97	20	82	5	4
5	smica	2.01	False	1.2	39	2048	0.01	558	2	10	69	9	64	0	0
6	smica	2.01	False	1.2	39	2048	0.002	482	0	2	58	5	56	0	0
7	smica	2.01	False	1.2	39	2048	0	0	0	0	0	0	0	0	0
8	smica	2.01	False	1.2	40	2048	0.05	1565	39	180	214	49	179	9	8
9	smica	2.01	False	1.2	40	2048	0.01	1197	4	63	153	22	141	2	1
10	smica	2.01	False	1.2	40	2048	0.002	986	1	26	114	14	107	0	0
11	smica	2.01	False	1.2	40	2048	0	0	0	0	0	0	0	0	0
12	nilc	2.01	False	1.2	38	2048	0.05	798	28	39	491	299	338	6	6
13	nilc	2.01	False	1.2	38	2048	0.01	648	7	9	409	233	302	0	0
14	nilc	2.01	False	1.2	38	2048	0.002	520	1	1	337	190	252	0	0
15	nilc	2.01	False	1.2	38	2048	0	0	0	0	0	0	0	0	0
16	nilc	2.01	False	1.2	39	2048	0.05	1579	41	69	908	546	617	11	9
17	nilc	2.01	False	1.2	39	2048	0.01	1222	6	10	753	449	519	1	1
18	nilc	2.01	False	1.2	39	2048	0.002	1050	2	5	658	396	454	0	0
19	nilc	2.01	False	1.2	39	2048	0	0	0	0	0	0	0	0	0
20	nilc	2.01	False	1.2	40	2048	0.05	2959	67	186	1493	838	994	15	12
21	nilc	2.01	False	1.2	40	2048	0.01	2338	8	41	1269	717	865	5	3
22	nilc	2.01	False	1.2	40	2048	0.002	2030	5	16	1126	639	774	3	2
23	nilc	2.01	False	1.2	40	2048	0	0	0	0	0	0	0	0	0
24	commander	2.01	False	1.2	38	2048	0.05	39	0	2	37	37	35	0	0
25	commander	2.01	False	1.2	38	2048	0.01	32	0	1	31	30	0	0	0
26	commander	2.01	False	1.2	38	2048	0.002	24	0	0	23	23	0	0	0
27	commander	2.01	False	1.2	38	2048	0	0	0	0	0	0	0	0	0
28	commander	2.01	False	1.2	39	2048	0.05	64	1	3	61	56	58	0	0
29	commander	2.01	False	1.2	39	2048	0.01	55	0	2	53	52	50	0	0
30	commander	2.01	False	1.2	39	2048	0.002	50	0	2	49	48	47	0	0
31	commander	2.01	False	1.2	39	2048	0	0	0	0	0	0	0	0	0
32	commander	2.01	False	1.2	40	2048	0.05	148	4	13	141	130	108	2	2
33	commander	2.01	False	1.2	40	2048	0.01	108	1	5	106	98	87	1	1
34	commander	2.01	False	1.2	40	2048	0.002	84	0	2	83	78	72	0	0
35	commander	2.01	False	1.2	40	2048	0	0	0	0	0	0	0	0	0
36	seven	2.01	False	1.2	38	2048	0.05	4233	125	432	2282	94	2247	17	16
37	seven	2.01	False	1.2	38	2048	0.01	3506	24	166	2028	72	2009	5	5
38	seven	2.01	False	1.2	38	2048	0.002	3098	4	94	1840	61	1828	0	0
39	seven	2.01	False	1.2	38	2048	0	99	0	0	79	6	78	0	0
40	seven	2.01	False	1.2	39	2048	0.05	8693	256	1526	4654	134	4589	43	36
41	seven	2.01	False	1.2	39	2048	0.01	7064	39	690	4056	90	4029	13	10
42	seven	2.01	False	1.2	39	2048	0.002	6171	6	392	3669	76	3655	4	3
43	seven	2.01	False	1.2	39	2048	0	192	0	165	11	164	0	0	0
44	seven	2.01	False	1.2	40	2048	0.05	16810	490	4633	8104	220	7974	93	72
45	seven	2.01	False	1.2	40	2048	0.01	13676	65	2594	7184	126	7140	21	15
46	seven	2.01	False	1.2	40	2048	0.002	12003	9	1673	6594	104	6568	7	5
47	seven	2.01	False	1.2	40	2048	0	397	0	0	349	18	347	0	0

APPENDIX B. TABLE OF RESULTS

APPENDIX B. TABLE OF RESULTS

Acronyms

Λ CDM Lambda Cold Dark Matter.

AGN Active Galactic Nuclei.

CMB Cosmic Microwave Background.

CO carbon monoxide.

COBE Cosmic Background Explorer.

ESA European Space Agency.

GR General Relativity.

H hydrogen.

He helium.

HEALPix Hierarchical Equal Area isoLatitude Pixelization.

HFI High Frequency Instrument.

LFI Low Frequency Instrument.

NASA National Aeronautics and Space Administration (United States).

NILC Needlet Internal Linear Combination.

PS Point Sources.

SEVEM Spectral Estimation Via Expectation Maximisation.

SMICA Spectral Matching Independent Component Analysis.

STEM Smoothing and TEsting of Maxima.

WMAP Wilkinson Microwave Anisotropy Probe.

Symbols

- B Multipole width of the needlet.
- C_ℓ Angular power spectrum.
- H Hubble parameter.
- N_p Number of pixels in a map.
- T Temperature.
- $Y_{\ell m}$ Spherical harmonic of degree ℓ and order m .
- Λ Cosmological Constant.
- Ω Density parameter of the cosmological fluid.
- α Significance level for the multiple testing.
- β_{jk} Needlet coefficient.
- ℓ Multipole.
- ν Frequency of the light.
- ω Parameter of the equation of state: p/ρ .
- ψ_{jk} Needlet.
- ρ Density of a fluid.
- $a(t)$ Scale factor of the Universe.
- $a_{\ell m}$ Spherical harmonics coefficient.
- b Filter function of the needlet.
- c Speed of light in vacuum.
- $f(x)$ Distribution of maxima in a gaussian map.
- h Dimensionless Hubble constant.
- j Frequency of the needlet.
- k Curvature of the Universe.
- z Redshift.