



HAL
open science

The blind Bayesian approach to Cosmic Microwave Background data analysis

Flavien Vansyngel

► **To cite this version:**

Flavien Vansyngel. The blind Bayesian approach to Cosmic Microwave Background data analysis. Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]. Université Pierre et Marie Curie - Paris VI, 2014. English. NNT : 2014PA066609 . tel-01145843

HAL Id: tel-01145843

<https://theses.hal.science/tel-01145843>

Submitted on 27 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

ED 127 - Astronomie et astrophysique

Institut d'Astrophysique de Paris

**The blind Bayesian approach to
Cosmic Microwave Background data analysis**

Par Flavien Vansyngel
Thèse de doctorat de Cosmologie

Dirigée par Benjamin D. Wandelt

Présentée et soutenue publiquement le mardi 16 décembre 2014
devant un jury composé de :

Bertrand Laforge	Professeur d'université – LPNHE	Président du jury
Jacques Delabrouille	Directeur de recherche CNRS – APC	Rapporteur
Roberto Trotta	Senior lecturer – Imperial College	Rapporteur
Jean-Loup Puget	Directeur de recherche CNRS – IAS	Évaluateur
Guillaume Patanchon	Enseignant-Chercheur – APC	Évaluateur

Abstract

The main topic of this thesis is the analysis of Cosmic Microwave Background (CMB) data. In particular, I present a method, Bayesian Independent component analysis (BICA), that performs both CMB component separation and CMB power spectrum inference.

I begin with an overview of the properties of the CMB. I present the basics of our understanding of this emission and I show the power of CMB to constrain cosmological models. I mainly discuss the CMB power spectrum but I also introduce the study of CMB statistics at the map level. The main interests of this field are the deviation from Gaussianity and isotropy of the CMB map. Some statistical anomalies are present at low level in the CMB map. I therefore highlight the need for a robust error modelling.

Then I present the main source of errors in the CMB products, namely the foregrounds. Foregrounds are astrophysical emissions that prevent us directly observing the primordial temperature fluctuations. I review them; among the most important are the different galactic emissions and the point sources. Secondary anisotropies, which are late time effects on the CMB, can also be considered as separate components to be treated as a foreground. I stress the diversity and the complexity of the maps of the foregrounds. The difficulty to model them is a problem to individually identify them in the data and clean them from the signal.

Component separation is a crucial and delicate step in CMB data analysis. Since the first full sky observation of CMB anisotropies, several approaches have been proposed. I review several methods aiming at cleaning the CMB from foregrounds. The methods may be classified in two categories: blind source separation and physical parametrisation. The methods from the former has a minimal set of assumptions on the physics of the foregrounds whereas the methods of the latter perform the separation based on a physical description of the data. Blind methods have the advantage of providing results and errors that are not constrained by the assumption of too restrictive a model.

Then I present BICA. The method is formulated in a blind Bayesian framework. Thus, from a common blind data description I build a likelihood distribution and I choose prior distributions with minimal assumptions. This process results in a posterior distribution defined on a parameter space that contain all the component maps and power spectra present in the data. The posterior distribution provides an inference of the CMB map and power spectrum from the observation maps. Thus, the errors on the reconstruction include the uncertainties due the presence of foregrounds in the data. I propose two sampling schemes that aim at exploring the posterior. By considering particular choice of prior and sampling scheme, I show how the Bayesian formulation of component separation provide a unifying framework of which previous methods are special cases.

Finally, I present the results of BICA when applied on both simulated data and 2013 Planck data. This method is able to reconstruct the CMB map and power spectrum on a large fraction of the sky. The main contributions of this thesis is to provide: 1) a CMB power spectrum on a large multipole range data whose errors take the the presence of foregrounds into account but without assuming physical models, 2) a CMB map inference together with an error model including both noise and foregrounds residuals.

Contents

General Introduction	5
1 Introduction to Cosmic Microwave Background	8
1.1 Physical context	9
1.2 Our understanding of the CMB	10
1.2.1 Acoustic oscillations	10
1.2.2 Gravitational effects	11
1.2.3 Baryonic effects	13
1.2.4 Damping	13
1.2.5 The transfer function and the growth factor	14
1.2.6 Projection on the sphere	14
1.3 Our understanding from the CMB	15
1.3.1 Parameters of the standard model of cosmology	17
1.3.2 Beyond the power spectrum	18
1.4 Polarisation of the CMB	20
2 The microwave sky	22
2.1 The Planck Sky Model	23
2.2 Galactic emissions	24
2.2.1 Dust emission	24
2.2.2 Synchrotron emission	27
2.2.3 Free-free emission	28
2.2.4 Molecular lines	29
2.2.5 Galactic point sources	32
2.3 Extra-galactic components	33
2.3.1 CMB secondary anisotropies	33
2.3.2 Extra-galactic point sources and the Cosmic Infrared Background	37
3 Basic concepts of CMB component separation	39
3.1 The component separation challenge	40
3.2 Key ideas to solve the problem	41
3.3 Review of component separation methods	42
3.3.1 Data model	42
3.3.2 Internal Linear Combination	43
3.3.3 Independent Component Analysis	45
3.3.4 Sparse blind source separation	46

3.3.5	Template fitting	48
3.3.6	Physical parametrisation	49
4	BICA: a semi-blind Bayesian approach to component separation	53
4.1	Constructing the component separation PDF	54
4.1.1	The blind Bayesian formulation of the problem	54
4.1.2	Likelihood distribution	55
4.1.3	Prior distributions	55
4.1.4	Hierarchical model and power spectrum inference	56
4.1.5	Posterior distribution	57
4.2	Deriving the sampling equations	57
4.2.1	First attempt	58
4.2.2	Marginalisation	58
4.2.3	Sampling scheme	59
4.3	Comparison to previous methods	60
4.3.1	Relevance of the method	60
4.3.2	Comparison to SMICA	61
4.3.3	Comparison to Commander	62
4.3.4	Comparison to ILC	62
4.3.5	Comparison to SEVEM	63
5	Application to simulations	65
5.1	Description of the simulations	65
5.1.1	The components	66
5.1.2	The mixing matrix	68
5.1.3	The noise	68
5.1.4	The data	68
5.2	Model approximations to the simulations	69
5.2.1	Isotropic noise	69
5.2.2	Lack of correlation between components	69
5.2.3	Gaussianity	70
5.3	Results	70
5.3.1	Full Gibbs sampling treatment	70
5.3.2	Self consistent treatments	73
5.3.3	Products of the method	73
5.3.4	CMB power spectrum inference	74
5.3.5	CMB map inference	75
5.3.6	Inference of non-CMB components	81
5.4	Model checking	81
5.4.1	Construction of the mismatch	82
5.4.2	Consistency of the results on simulations	83
5.4.3	Consistency of the results with modified priors	83
5.5	Discussion	84

6	Application to Planck data	87
6.1	The Planck data	87
6.2	Additional modelling	88
6.2.1	Noise	88
6.2.2	Cross-spectra	90
6.2.3	Point sources	90
6.2.4	Beaming	92
6.2.5	Masking, apodising, inpainting	93
6.3	Possible configurations of the data	95
6.3.1	Mask	95
6.3.2	Frequency range	96
6.3.3	Multipole range	97
6.3.4	Number of components	97
6.3.5	Point source model	98
6.3.6	Choice of prior	98
6.3.7	Test cases	99
6.4	Results	101
6.4.1	CMB map and power spectrum inference	101
6.4.2	Inference of the non-CMB components	105
6.4.3	Consistency of the results	106
6.4.4	Comparison with SMICA	109
	General conclusion	113
	Appendices	116
A	Statistical basics	116
A.1	Random variables, distributions and probability density functions	116
A.2	Gaussian distribution and related distributions	117
A.3	Kullback-Leibler divergence	118
A.4	Shannon entropy	119
B	Bayesian inference	120
B.1	Bayes' theorem	120
B.2	Jeffreys priors	121
C	PDF evaluation techniques	123
C.1	Simple approaches	123
C.2	Metropolis-Hastings sampling	123
C.3	Gibbs sampling	124
C.4	Collapsed sampling	125
D	Isotropic Gaussian random field on the sphere	126
D.1	Spherical harmonics	126
D.2	Power spectrum	127
E	HEALPix	128

F	Link between ILC and BICA	129
F.1	Data and notations	129
F.2	ILC and "BICA derived" formulas	129
F.3	Expanding the ILC formula	130
F.4	Relation between the two formulas	130

General Introduction

Elements of modern cosmology

Physical cosmology emerged about a century ago, which makes it one of the youngest scientific fields. Of course, many scientists and philosophers had thought about organising the universe as a whole in a comprehensive way well before the 20th century. However, their reasoning did not stand on solid logical bases. Eventually, cosmology went through the process of mathematisation in the beginning of last century with the then new theories of space-time and gravity, namely the special relativity (SR, 1905) and the general relativity (GR, 1915). SR and GR exploit the purely mathematical theories of space that were developed during the previous century. The results were stupefying, sometimes counter-intuitive. One of the consequences was that the universe in which we were living could, in theory, be dynamic and thus have a history. This was an exotic point of view at that time but it was rather quickly experimentally confirmed with the observation of receding objects, which were interpreted as evidence of the expansion of the universe (1920s).

Since observations show strong evidence for the evolution of the universe, and more specifically expansion, it is worth entering into the details. “What are the parts of the story?” “What are the beginning and the end?” The latter question is a tricky one since it leads to metaphysical considerations. The former is simpler to address. The universe at our time and place is the starting point, the expansion is dynamic, GR the framework to make predictions. “Place” is removed from the equations via the following cosmological principle: the universe is isotropic and homogeneous, there is no privileged point in the universe. Thus, we only need to consider time evolution of the universe.

In order to get to the origins of the universe, or at least to get closer, we have to look backward in time. Reversing the evolution of our universe makes it contract. In addition to GR, thermodynamics and statistical physics are the key theories, here. Indeed, given the cosmological principle, using the same tools for the evolution of the systems “gas in a box” and “matter in our universe” seems relevant. A thermal history of the universe is thus constructed. We learn from the gas analogy that the universe is hotter and hotter as we probe it earlier and earlier. At some point, the pressure and the radiation energy are high enough to break apart any kind of structure, such that the universe is filled

with an infinite fluid of elementary particles. In the ionised universe, the photons are not free anymore because they frequently interact with the electrons. Going forward from then, the time at which the electrons bind to the nuclei is called *recombination*, the first photons to be free to stream form the *Cosmic Microwave Background (CMB)* and the place from which the CMB photons we receive has been emitted is the *Last Scattering Surface (LSS)*.

CMB data analysis

As we will see in the first chapter of this thesis, observations of the Cosmic Microwave Background (CMB) constrain cosmological models. In particular, the CMB fluctuations are very sensitive to the parameters of the so called standard model of cosmology (Jungman et al., 1996). Current and future experiments designed for CMB analysis are signal dominated (Planck Collaboration et al., 2013; Schaffer et al., 2011; Bouchet et al., 2011; Baumann et al., 2009; André et al., 2014). Therefore the remaining issue in deriving cosmological information from CMB is the separation between the CMB signal and any other emission that could hide it or pollute it. These emissions are called *the foregrounds*. Being able to propagate component separation uncertainties to final constraints on fundamental physics is a leading issue in CMB data analysis.

CMB experiments, such as the ongoing Planck mission (Planck Collaboration et al., 2014a), observe in the microwave domain. The CMB is not the only emission that is received when observing from the solar system at these frequencies. Free-free, synchrotron and thermal dust emissions emanating from our galaxy are among the most intense signals in the microwave domain (Sehgal et al., 2010). An observation of the sky at these frequencies is therefore a mixture of the photons from the different sources. Therefore the CMB must be extracted through component separation techniques.

The method I present in this thesis makes no assumption about non-CMB emission. Concerning the CMB, it only assumes that the CMB emission law follows a black body. I make use of Independent Component Analysis (ICA) after assuming the mutual independence of the different signals constituting the data. Blind separation of independent sources (e.g. Cardoso (1998)) is a very general process that finds applications in various fields, from telecommunication to biomedical signals. Blind ICA has previously been applied in cosmology, particularly in CMB analysis (Baccigalupi et al. (2000); Maino et al. (2001) and Cardoso et al. (2002, 2008); Delabrouille et al. (2003)). Other methods (Starck et al., 2004; Starck et al., 2013) exploit sparsity rather than independence to discern between different signals. In this thesis I adopt the first approach and I propose a Bayesian instance of semi-blind ICA.

The different component separation methods are mainly characterised by two im-

portant aspects, the basis in which the data are expressed and the parametrisation of the data. Current methods performs the separation in different bases such as pixel space (Eriksen et al., 2006), spherical harmonic space (Tegmark, 1997; Delabrouille et al., 2003) or needlet space (Delabrouille et al., 2009; Moudden et al., 2005; Fernandez-Cobos et al., 2012). Their description of the data involves either a non-parametric model and exploits the independence between the CMB and the non-CMB component only (Delabrouille et al., 2009; Fernandez-Cobos et al., 2012) or a parametric model that is fitted to the data (Eriksen et al., 2006). Intermediate between this these two, some methods assume (Cardoso et al., 2008) coherence through frequency and complete independence of the components, fitting a non-parametric foreground model to the data *via* likelihood maximisation.

The method I present in this thesis, Bayesian Independent Component Analysis (BICA), goes one step forward. Based on generic assumptions (e.g. statistical independence of the components, spatial coherence between frequencies, spatial or angular scale statistical independence), a generic statistical model of the components is chosen such that it allows a full Bayesian exploration of the posterior density. The introduction of a simple but full generative model that approximates the stochastic model of the component permits propagating the uncertainties within that model. The simplifying assumptions allow either a numerical marginalisation over all nuisance parameters (Gratton, 2008) or, as in the present work, a full exploration of the model and a joint sampling of both the component maps and power spectra. The goal of BICA is to infer both the CMB map and its power spectrum.

Chapter 1

Introduction to Cosmic Microwave Background

Introduction

The Cosmic Microwave Background (CMB) is one of the major observable for modern cosmology and fundamental physics. The study of the CMB has been a very active field of research since its first observation (Penzias and Wilson, 1965) and confirmation (Dicke et al., 1965) and, to a lesser extent since its prediction (Gamow, 1946; Dicke et al., 1946; Alpher et al., 1948) . Since then, many experiments (Crill et al., 2003; Ruhl et al., 2004; Ade et al., 2014; Sievers et al., 2013), including three successful space missions (Smoot et al., 1990; Bennett et al., 1997; Planck Collaboration et al., 2011a), have been dedicated to the observation of the CMB. The reason for this enthusiasm is that the CMB carries the imprint of a vast range of phenomena, from particle physics and inflationary era to galaxy physics and late time universe.

In this chapter I briefly explain the main physical processes that give rise to the features observed in the CMB. I start with perturbation theory in the early universe and treat only primordial perturbation – and thus anisotropies. Secondary anisotropies are treated as if they originated from a different signal (see chapter 2). I link the CMB observables to the primordial properties of the universe. Then I discuss how the change in the cosmological parameters affect the CMB and therefore what we can learn from the study of the CMB.

This chapter is a brief review of Cosmic Microwave Background theory providing a comprehensive reasoning from the basic assumptions and physics needed, to the CMB observations and the cosmological models. The curious reader is encouraged to go further into details with the references therein. Section 1.2 mainly follows the review of Hu (2008) and is enriched by four text books (Dodelson, 2003; Lyth and Liddle, 2009; Gorbunov and Rubakov, 2011; Ryden, 2003).

1.1 Physical context

At early stages, the universe is filled mainly with Cold Dark Matter (CDM) and a fluid in which photons, protons and electrons are tightly coupled. The latter is of particular interest here and it is commonly referred to as the photon-baryon plasma, even if, obviously, electrons are fermions. Within the plasma, a typical photon frequently Compton scatters off free electrons. Its mean free path is then very small compared to the size of the horizon, the universe is said to be opaque. The expansion of the universe cools down the plasma to a point where the number of high energy photons is not large enough to ionise the medium. Protons and electrons bind together to form neutral light elements. The mean free path of the photons suddenly increases such that, from then, a typical photon travels freely through the universe. The photons have very low probability to interact with anything else. The universe is then filled with this remnant light of the primordial plasma that is called the Cosmic Microwave Background.

Two times and one scale are important to remember during the following treatment: the time of recombination, the time of radiation-matter equality and the Hubble horizon.

The time of recombination is simply the time at which recombination happens, under the assumption that the universe goes instantly from the completely ionised state to the completely neutral state.

As universe expands, the energy density of the different species present in the universe dilute at a rate that depends on the equation of state. It turns out that the photons density dilute more rapidly than the matter density. Indeed, in addition to the decay of the number density, the photons undergo further energy dilution due to the expansion of their frequency. Today we observe much less energy contained in the photons than in the matter. Thus, the density of photons and matter were necessarily equal at some point in the past. This time is the time of radiation-matter equality. This is an important turning point because during the matter dominated era, i.e. after the time of matter-radiation equality, the perturbations in the distribution of matter becomes gravitationally unstable. During the radiation dominated era, the photons erase the perturbations.

Because of the finite speed of information transport, a point in space is causally connected to a finite volume of space around it. Because the expansion happens everywhere at the same speed, this region is a filled sphere whose center is the point of interest and the radius the Hubble horizon. From now on, I will refer to the Hubble horizons by the terms horizon or horizon scale (when working in Fourier space). This is an important scale because we are going to extensively use Fourier modes and the modes associated to scales larger and lower than the horizon do not behave the same way.

1.2 Our understanding of the CMB

1.2.1 Acoustic oscillations

In the matter dominated era, the photon-baryon plasma is in thermal equilibrium and experiences acoustic oscillations for scales smaller than the size of the horizon. The ordinary matter tends to fall into the gravitational wells but, because of the tight coupling with the photons, it is repulsed by the radiation pressure. There is a competition between the two phenomena, hence the oscillations.

Because of the frequent interactions between the elements in the plasma, the photon gas is at thermal equilibrium, including the photons that were emitted at the LSS. Therefore the CMB emission law follows a black body. It can be described by a temperature T_{CMB} and the following distribution function (Dodelson, 2003)

$$f(\mathbf{x}, \mathbf{p}, t) = \left[\exp \left(\frac{p}{T_{\text{CMB}}(t) (1 + \Theta(\mathbf{x}, -\hat{\mathbf{n}}, t))} \right) - 1 \right]^{-1} \quad (1.1)$$

where $-\hat{\mathbf{n}} = \mathbf{p}/p$ and

$$\Theta(\mathbf{x}, \hat{\mathbf{n}}, t) \triangleq \frac{\delta T(\mathbf{x}, \hat{\mathbf{n}}, t)}{T_{\text{CMB}}(t)} \quad (1.2)$$

is called the brightness function. Θ is independent of the magnitude p of the photon momentum \mathbf{p} because Compton scattering does not change the magnitude of the momentum. To first order, the perturbed distribution function f keeps the black body form but with a perturbation in temperature that does not depend on p . Thus, $\Theta(\mathbf{x}, \hat{\mathbf{n}}, t)$ is the contrast in temperature that an observer at position \mathbf{x} and time t sees when looking in direction $\hat{\mathbf{n}}$ (i.e. the line of sight, the opposite of the photon direction).

f obeys the Boltzmann equation $\frac{df}{dt} = C[f]$, which is tedious to solve in its most general form. Still, in order to give a hint of the dynamic, we will solve the first moments of the Boltzmann equation. This can be done by integrating the Boltzmann equation and solving for quantities of interests (see e.g. Ryden (2003)). The first two moments are the well known continuity and Euler equations. Compton scattering neither creates nor destroys photons such that changes in the number of photons in a given volume is due flows entering and exiting the volume. Thus, for photons:

$$\frac{\partial n_\gamma}{\partial t} + \nabla \cdot (n_\gamma \mathbf{v}_\gamma) = 0 \quad (1.3)$$

$$\frac{\partial \mathbf{v}_\gamma}{\partial t} + (\mathbf{v}_\gamma \cdot \nabla) \mathbf{v}_\gamma = -\frac{1}{\rho_\gamma} \nabla p_\gamma. \quad (1.4)$$

Since Θ is small we can linearise the continuity equation and the Euler equation. Knowing that the number density evolves as $n_\gamma \propto T^3$, we can relate n_γ to Θ . In addition,

the equation of state of a gas of photon is $p_\gamma = \rho_\gamma/3$. Taking these considerations into account, the continuity equation and the Euler equation become, after a Fourier transform (Hu, 2008),

$$\dot{\Theta} = -\frac{1}{3}i\mathbf{k} \cdot \mathbf{v}_\gamma \quad (1.5)$$

$$\dot{\mathbf{v}}_\gamma = k\Theta, \quad (1.6)$$

where the dot over a quantity means derivation with respect to the conformal time. The combination of these equations gives a harmonic oscillator equation for Θ

$$\ddot{\Theta} + c_s^2 k^2 \Theta = 0, \quad (1.7)$$

where the adiabatic sound speed $c_s^2 \hat{=} \dot{p}_\gamma/\dot{\rho}_\gamma = 1/3$. Inflation predicts adiabatic initial condition (see Kolb and Turner (1990) for a discussion of models). In the adiabatic mode, $\dot{\Theta}(0) = 0$. Thus the solution for a photon gas with an initial state assumed to be adiabatic is

$$\Theta(\eta) = \Theta(0) \cos(kc_s\eta). \quad (1.8)$$

Thus, still in the hypothesis of adiabatic initial condition, this simple treatment implies two important features:

temporal coherence of the oscillations This is possible if all wavelength have been created at the same time. Thanks to inflation, this assumption does not violate causality,

the existence of fundamental scale Indeed, the solution is extremal at $k_n = n\pi/c_s\eta$. In particular, the mode that has oscillated only once ($n = 1$) gives the fundamental scales $k_1 = \pi/c_s\eta^*$, where η^* is the time of recombination, and the corresponding multipole is $\ell_1 \sim 200$. We will see that it is a very good approximation considering the high degree of simplification in this calculation.

The observations of these features in the CMB power spectrum is a strong evidence for the predominance of adiabatic initial conditions, even if other modes of primordial fluctuations are possible (Langlois, 2003).

1.2.2 Gravitational effects

Metric perturbation

The metric perturbation for a flat universe can be written in terms of two potentials Φ and Ψ

$$ds^2 = a^2 \left[-(1 + 2\Psi) d\eta^2 + (1 + 2\Phi) dx^2 \right]. \quad (1.9)$$

In the absence of anisotropic stress and viscosity, Φ and Ψ are the relativistic analogues of the Newtonian potential and they verify $\Psi \sim -\Phi$. Perturbations in the metric, i.e. non zero Φ and Ψ , modify the continuity and the Euler equations as follows (Hu, 2008)

$$\dot{\Theta} = -\frac{1}{3}ikv_\gamma - \dot{\Phi} \quad (1.10)$$

$$\dot{v} = k(\Theta + \Psi). \quad (1.11)$$

Combine again to get the forced harmonic oscillator for Θ

$$\ddot{\Theta} + c_s^2 k^2 \Theta = -\frac{k^2}{3} \Psi - \ddot{\Phi}. \quad (1.12)$$

The effective temperature fluctuation $\Theta + \Psi$ verifies the homogeneous harmonic oscillator equation and the solution is

$$[\Theta + \Psi](\eta) = [\Theta + \Psi](0) \cos(kc_s\eta). \quad (1.13)$$

The competition between gravity and pressure is now explicit in the expression of the effective temperature $\Delta T/T = \Theta + \Psi$. We will return to this phenomenon when addressing the Sachs-Wolfe effect.

Sachs-Wolfe effect

For scales larger than the horizon, physics can not be causal and the anisotropies are due to the initial fluctuations of the gravitational potential. Indeed, during inflation the universe has expanded exponentially with time such that the perturbations on large scales at the time of recombination have preserved the imprint of the primordial perturbations.

At large scales, i.e. small k , the oscillator of equation 1.13 drifts very slowly from its initial condition. This is known as the Sachs-Wolfe effect (Sachs and Wolfe, 1967) and for adiabatic initial conditions the effective temperature is (see White and Hu (1997) for a detailed treatment)

$$\Theta + \Psi = \frac{1}{3}\Psi. \quad (1.14)$$

Ψ is negative in a potential well and positive in underdense regions. Thus, there is a balance between the intrinsic temperature of the photons and the gravitational potential. The effect of the potential prevails such that the photons undergo redshift (resp. blueshift) when escaping the LSS from an overdense (resp. underdense) region.

1.2.3 Baryonic effects

The previous treatments neglect the effect of baryons. Because of their mass, baryons add extra inertia to the baryon-photon fluid. The momentum of the plasma is enhanced by a factor depending on the baryon-to-photon ratio $R = (p_b + v_b)/(p_\gamma + v_\gamma)$. As a result, the sound speed is no longer constant but equal to $c_s^2 = 1/3(1 + R)$ (Hu, 2008). The speed of sound decreases compare to the case of a purely photon fluid because baryons are heavy and reduce the propagation of the perturbations. Then the fluid is no longer relativistic, as baryons raise the effective mass of the fluid but do not contribute to pressure, which remains just the pressure of radiation.

The continuity equation for the photon is unchanged but the Euler equation becomes (Hu, 2008)

$$[(1 + R) v_\gamma]' = k\Theta + (1 + R)k\Psi. \quad (1.15)$$

The oscillator taking gravity and baryons into account is then

$$\left[(1 + R)\dot{\Theta} \right]' + \frac{1}{3}k^2\Theta = -\frac{1}{3}k^2(1 + R)\Psi - \left[(1 + R)\dot{\Phi} \right]. \quad (1.16)$$

The solution is modified as

$$[\Theta + (1 + R)\Psi](\eta) = [\Theta + (1 + R)\Psi](0) \cos(kc_s\eta). \quad (1.17)$$

The introduction of baryons affects the oscillations in three aspects. First, from Sachs-Wolfe effect, we can see that the amplitude of the oscillations are enhanced by a factor $(1 + 3R)$. Second, the perturbations oscillates around a non zero value that is $-R\Psi$. Third, the acoustic scale increases – by a factor $\sqrt{1 + R}$ – since the baryon loading prevent the perturbation from propagate as far as in the case of a zero mass fluid.

1.2.4 Damping

At small scales, the perturbations are smoothed because of free streaming of the photons. The photons have a small but non zero mean free path. Thus, the photons erase the structures smaller than their mean free path because of their diffusion. This effect is known as Silk damping (Silk, 1968). The scale of dissipation λ_D can be well approximated with simple arguments. The mean free-path of a photon is associated to the Compton mean free-path λ_C . Thus, if the photon has scattered N times from the initial time to the time of recombination, the approximate damping distance is given by the square root of the number of scatterings made – because of the stochasticity of its random walk – times the Compton length. Hence (Hu, 2008)

$$\lambda_D = \sqrt{N}\lambda_C = \sqrt{\eta_{\text{rec}}/\lambda_C}\lambda_C = \sqrt{\eta_{\text{rec}}\lambda_C}. \quad (1.18)$$

A similar expression arises from more sophisticated reasoning involving the Navier-Stokes equation in addition to continuity and Euler equations. Silk damping is treated as an anisotropic stress term. A damped oscillator equation results from this consideration and the envelope of the solution is a decaying exponential, $\exp[-(k/k_D)^2]$ with (Hu, 2008)

$$k_D = \lambda_D^{-1} = \left(\frac{2\pi}{\sqrt{6}} \sqrt{\eta\lambda_C} \right)^{-1}. \quad (1.19)$$

1.2.5 The transfer function and the growth factor

During the radiation dominated era, the perturbations at scales smaller than the horizon are suppressed. As time passes, larger and larger scales enter the horizon and are undergoing the damping. The small modes spend more time inside the horizon during the radiation era, such that the smaller the mode is, the more damped it is. The modes entering after matter radiation equality, i.e. modes $k < k_{eq}$ with k_{eq} the mode associated to horizon scale at equality, are much less affected by the damping. Thus the power spectrum of potential P_Φ – and therefore the power spectrum of the brightness function P_θ – has a different shape than the power spectrum of primordial curvature perturbation P_ζ . In linear perturbation theory, these power spectra are related mode by mode via a transfer function $T(k)$ as follows (Lyth and Liddle, 2009)

$$P_\Theta(k, a) = D(a)^2 T(k)^2 P_\zeta(k). \quad (1.20)$$

$D(a)$ is the growth factor that accounts for the wavelength independent growth at late times. By construction, $T(k \ll k_{eq}) = 1$ and $D(a_0) = 1$. $T(k)$ depends on the baryon density $\Omega_b h^2$ and the matter density $\Omega_m h^2$ only.

1.2.6 Projection on the sphere

So far, the quantities of interest were defined on the 3-D space. The CMB photons that we receive here and now from all directions have travelled about the same distance because the recombination was sudden. Thus, they were then emitted from a sphere and reflects the density perturbation on this surface. The goal of this section is to link the 3-D spectrum to our observable, which is a 2-D spectrum. No physics is involved in the calculation since it is a purely geometrical problem.

Let's express the brightness function into the basis of spherical harmonics (see appendix D for more details). We have

$$\Theta(\mathbf{x}, \hat{\mathbf{n}}, t) = \sum_{\ell m} \Theta_{\ell m}(\mathbf{x}, t) Y_{\ell m}(\hat{\mathbf{n}}). \quad (1.21)$$

The $\Theta_{\ell m}$ observed here at \mathbf{x}_0 and now at t_0 are the modes of the CMB anisotropies. Let's define

$$a_{\ell m} \hat{=} \Theta_{\ell m}(\mathbf{x}_0, t_0) \quad (1.22)$$

$$C_\ell \hat{=} \langle |a_{\ell m}|^2 \rangle. \quad (1.23)$$

The C_ℓ are the power spectrum of the CMB. The power spectrum does not depend on m because the statistical isotropy of the emission implies symmetries under rotations of the expected values.

Focusing on the scalar mode of the Fourier transform of the brightness function, we can define multipoles Θ_ℓ by

$$\Theta(\mathbf{k}, \hat{\mathbf{n}}, t) = \sum_{\ell} (-i)^\ell \sqrt{4\pi(2\ell+1)} Y_{\ell 0}(\hat{\mathbf{n}}) \Theta_\ell(\mathbf{k}, t). \quad (1.24)$$

By taking the inverse of this equation, i.e. by expressing Θ_ℓ as an integral involving $\Theta(\mathbf{k})$, and using the orthogonality properties of the spherical harmonics, it can be shown that (Lyth and Liddle, 2009)

$$\Theta_{\ell m}(\mathbf{x}, t) = \frac{4\pi}{(2\pi)^3} i^\ell \int d^3\mathbf{k} \Theta_\ell(\mathbf{k}, t) Y_{\ell m}^*(\hat{\mathbf{k}}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (1.25)$$

hence the following expressions for the observed CMB multipoles and the CMB power spectrum, relating 2-D power spectrum and the 3-D power spectrum of temperature anisotropies (Lyth and Liddle, 2009)

$$a_{\ell m} = \frac{4\pi}{(2\pi)^3} i^\ell \int d^3\mathbf{k} \Theta_\ell(\mathbf{k}, t_0) Y_{\ell m}^*(\hat{\mathbf{k}}) \quad (1.26)$$

$$C_\ell = 4\pi \int_0^\infty dk k^2 P_{\Theta_\ell}(k, t_0) \quad (1.27)$$

$$= 4\pi \int_0^\infty dk k^2 T_\ell(k)^2 P_\zeta(k) \quad (1.28)$$

where $T_\ell(k)$ is the multipole ℓ of the transfer function $T(k)$ defined in equation 1.20. The growth factor today being equal to unity, i.e. $D(a_0) = 1$, it does not appear in the expression for the CMB power spectrum.

1.3 Our understanding from the CMB

Equation 1.1 gives the distribution function of the photons. Other species are significantly present in the universe and coupled to gravity at that time. The distribution function

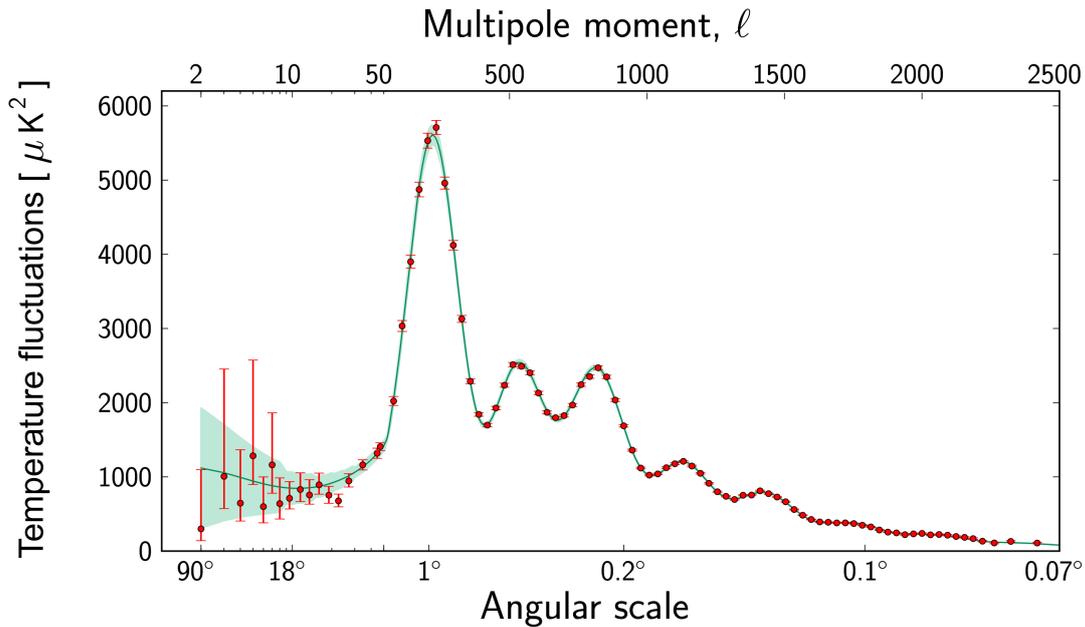


Figure 1.1: Planck CMB power spectrum from Planck Collaboration et al. (2014e). The red dots are the experimental measurements of CMB power spectrum and the green line represents the theoretical power spectrum that best fit data. The power spectrum exhibits three main features: the Sachs-Wolfe plateau for $\ell \lesssim 30$, the three first acoustic peaks and the damping tail for $\ell \gtrsim 1000$.

has a different expression for each of them – baryons, CDM, neutrinos, ... It is possible to solve numerically the full Boltzmann equation to get the evolution of the brightness function Θ . Codes such as CMBFast¹ (Seljak and Zaldarriaga, 1996) or CAMB² (Lewis et al., 2000) or CLASS³ (Blas et al., 2011) take cosmological parameters in input, compute the transfer function, solve numerically the Boltzmann equation, project $\Theta(t_0, \mathbf{x}_0, \hat{\mathbf{n}})$ on the sphere and give the corresponding CMB power spectrum in output. Thus, we have a way to predict the CMB power spectrum. Conversely, if the relation is bijective, i.e. there is no degeneracies, from an observed power spectrum we can learn the values of the parameters of a given model. This is done by exploring the possible values for the set of cosmological parameters and choosing the ones that give the power spectrum that fits the available data the best (Planck Collaboration et al., 2014d), with the use of codes such as PICO⁴ (Fendt and Wandelt, 2007).

Figure 1.1 shows the Planck power spectra, both experimental and theoretical. It is

¹the code is no longer supported by its authors.

²code available at <http://camb.info/>

³code available at <https://lesgourg.web.cern.ch/lesgourg/codes.html>

⁴code available at <https://sites.google.com/a/ucdavis.edu/pico/home>

common to plot $\ell(\ell+1)C_\ell/2\pi$ for a reason that will soon be clear. This figure shows nicely two fundamental scales that separate the different regimes of oscillations. The position of the first peak at $\ell \sim 200$, which correspond to the sound horizon at recombination, separates the oscillating modes – the acoustic peaks – and the modes that never entered the horizon – the Sachs-Wolfe plateau. The Silk scale at $\ell \sim 1000$ separates the damped and undamped modes.

This section presents the current standard model of cosmology. I will show how the change in parameter values affect the shape of the CMB power spectrum.

1.3.1 Parameters of the standard model of cosmology

The standard model is a cosmological model based on a flat FLRW metric taking into account CDM and a cosmological constant, in addition to the ordinary matter and radiation. In its simplest formulation, the standard model has 6 free parameters. Two of them describe the primordial power spectrum as it was after inflation, three of them control the energy content of the universe and one accounts for the modification of the power spectrum when propagating during the epoch of reionisation. Figure 1.2 illustrates the effect of changing the value of the parameters on the shape of the CMB power spectrum.

Initial conditions: A_S and n_S

The leading paradigm to explain the origins of the primordial fluctuations is inflation (Linde, 1982; Guth, 1981). According to the simplest models of this theory, the primordial curvature perturbation ζ would be described by a power spectrum P_ζ such that

$$\frac{k^3 P_\zeta(k)}{2\pi^2} = A_S \left(\frac{k}{k_0} \right)^{n_S-1} \quad (1.29)$$

with k_0 a fiducial normalisation wave number. A_S is the overall amplitude of the spectrum and n_S controls the tilt. The particular case $n_S = 1$ describes a scale invariant spectrum (Harrison, 1970; Zeldovich, 1972).

In the Sachs-Wolfe regime, where the modes are larger than the horizon i.e. small k and ℓ , the transfer function $T(k)$ is by definition equal to unity and therefore its multipoles are simply the square of the spherical Bessel functions, i.e. $T_\ell(k) = j_\ell^2(kD^*)$ with D^* the distance from the last scattering surface. In the case of a primordial scale invariant spectrum, which it has been proved by Planck to be almost true (Planck Collaboration et al., 2014e), the integral in equation 1.28 can be performed and the result is (Gorbunov and Rubakov, 2011)

$$C_\ell \sim \frac{2\pi}{\ell(\ell+1)} A_S. \quad (1.30)$$

Thus, to good approximation at large scales, i.e. $\ell \lesssim 30$, the power spectrum $\ell(\ell+1)C_\ell/2\pi$ is roughly constant (see figure 1.1) and directly probes the variance of the primordial curvature perturbations in the early universe.

The parameter n_S quantifies the deviation from scale invariance. If the matter and baryon transfer effects are determined, then the acoustic spectrum also constrains the tilt of the primordial spectrum (Hu, 2008).

Matter content: Ω_b and Ω_m

Ω_b and Ω_m are the fractions of the total energy content of the universe contained in respectively the baryons and the matter. The baryons and total amount of matter affects the transfer function and, to a lesser extent, the Silk scale.

A change in Ω_b leads to a change in the relative heights of the even and odd peaks of the spectrum, because of the inertia of the baryons in the plasma. Indeed, because of the mass of the baryons and tight coupling with the photons, the plasma spend more time in the potential wells. This is called the baryon loading. To quantitatively see this effect, recall that in the case of non zero baryon-to-photon ratio, the modes oscillate around a non-zero offset proportional to the baryon-to-photon ratio. In this case, the even and odd peaks of the power have different height. Thus, the amount of baryons is probed by evaluating the ratio between consecutive peaks.

Information on Ω_m is obtained by looking at the overall amplitude of the acoustic peaks and the third-to-second peak ratio (Hu, 2008).

Late time effects: Ω_Λ and τ

Ω_Λ is the fraction of the total energy content of the universe contained in the cosmological constant. Since dark energy is completely sub-dominant at epoch of recombination, it can only affect the CMB power spectrum at the largest scales, i.e. the modes entering the horizon at recent time. The lower the multipole is, the more enhanced it is (Hu, 2008).

The universe was reionised at late time, because of the UV radiation of the forming stars. The universe is filled with free electrons and a fraction of the CMB photons interact with them, again via Compton scattering. If τ denotes the optical depth due to travel through the reionised universe until here, the anisotropies are damped by a factor $e^{-\tau}$. All modes are damped by the same factor, except those that were beyond the horizon at recombination (Dodelson, 2003).

1.3.2 Beyond the power spectrum

As we just saw, a lot of information is encoded in the shape of the CMB power spectrum. It means that, even at high resolution, the description of the CMB is compressed in a

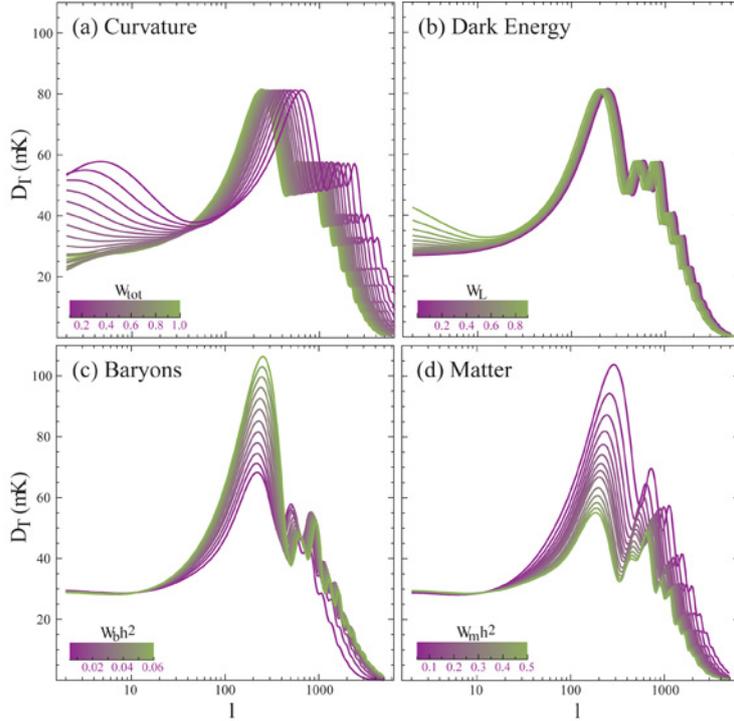


Figure 1.2: This figure from Hu and Dodelson (2002) shows the reshape of the power spectrum under variation of the cosmological parameters. In each panel only one parameter is allowed to vary. The upper left panel shows the effect of a parameter that is not discussed in the text, the curvature. Since the curvature affects the angular distance, the position of the peaks depends strongly on the amount of curvature. Dark energy has an impact on the lowest multipoles only. Conversely, a variation in matter and baryons do not affect the scales that were larger than the horizon at the recombination but change the relative height of the peaks.

couple of thousands numbers and its modelling is reduced to very few parameters. At this level of modelling, everything seems consistent. However, one could naturally question the assumptions of isotropy and Gaussianity of the CMB. This requires an analysis at the map level. Two Planck papers are essentially dedicated to such tests, Planck Collaboration et al. (2014g) and Planck Collaboration et al. (2014f).

Primordial non-Gaussianity

The standard models for generation of primordial perturbations predict an almost Gaussian statistics for the fluctuations. The almost is important here because certain inflationary models have a non-Gaussian signature that would be detectable by Planck. For Planck analysis, several independent statistical tools have been used to study non-Gaussianity such as Minkowski functionals, N-point correlation functions, bispectrum⁵

⁵The bispectrum is the harmonic analogue of the 3-point correlation function.

fitting, estimators based on wavelet statistics. They find very little evidence of non-Gaussianity. The CMB map exhibits some variance, skewness and kurtosis anomalies, but they seem to be related to other statistical anomalies that are discussed in the next paragraph.

Statistical tests on the CMB map

The analysis of the WMAP CMB map shed light on some statistical anomalies. The 2013 Planck analysis confirmed these inconsistencies. Hemispherical asymmetry, extreme hot/cold spots and mode alignment are such anomalies. The fact that two independent experiments, namely WMAP and Planck, point towards the same anomalies shows that they are not artefacts due to data acquirement and processing. The significance of the anomalies is still under debate. But if they turn out to be real deviations from isotropy, no known physical effect could explain them in the current paradigm. Theories beyond the standard model could explain some anomalies, e.g. the Cold Spot would be a natural feature of a Bianchi VII universe (Jaffe et al., 2005), but so far none of these theories has been confirmed.

Limitation due to foregrounds

Some statistical inconsistencies in the CMB are at a very low level. In order to put significant constraints on them, a maximum of information about the CMB is needed and therefore a maximum fraction of the sky of the map must be used. When working with a large sky fraction, the foregrounds are significant contaminants and therefore the CMB must be clean by component separation method before a meaningful analysis can be made. At the map level, the component separation methods do not provide a characterisation of foreground uncertainties. At best, the errors are model dependent and at low resolution (Planck Collaboration et al., 2014c). Unmodelled foreground residuals in the CMB map could be mistaken for deviation from Gaussianity and/or isotropy because no error model that include this foreground residuals is available. Late time effect, like the so called iSW effect, could also be a source of mode alignment and low variance (Francis and Peacock, 2010).

1.4 Polarisation of the CMB

My main interest during my three years Ph.D. thesis was the temperature of the CMB. An other important property of the CMB photons is their polarisation. About 10% of the CMB photons are polarised, which makes a signal of the order of a few μK . Zaldarriaga and Seljak (1997) proposed a decomposition of the polarisation into two modes,

E-modes and B-modes. This is in analogy to electrostatics, in which the electric field (E-field) has a vanishing curl and the magnetic field (B-field) has a vanishing divergence. Primordial B-modes are of particular interest since they are produced by tensor modes of the perturbation only. A detection of primordial B-modes would be a strong evidence for gravitational waves arising from the epoch of inflation. Detecting the B-modes is difficult, partly because the degree of foreground contamination is unknown, and detecting primordial B-modes is even more difficult because of the leakage of power from E-modes to B-modes, due to gravitational lensing. This effect has been measured by the South Pole Telescope in 2013 (Hanson et al., 2013).

Conclusion

I have briefly presented the Cosmic Microwave Background, from its necessary existence in an expanding universe to its role as a probe in modern cosmology. I have outlined the physical processes that give the CMB power spectrum its main features: the Sachs-Wolfe plateau, the acoustic peaks and the damping tail. Different values for the set of cosmological parameters lead to different shapes of CMB power spectrum. Boltzmann codes provide a way to quickly predict the shape of the CMB power spectrum and thus allow testing possible cosmological models. I also presented studies that show that the CMB contains little but crucial information at the map level. I have therefore introduced the need for separation between the primordial CMB fluctuation and any other contribution in the data. I present a list of the brightest components to take into account in chapter 2 and address the problem of removing them in chapter 3.

Chapter 2

The microwave sky

Introduction

We saw that the CMB is a powerful probe that allows testing and challenging models of fundamental physics. Beside the perturbation theory described in the previous chapter, one can use atomic and statistical physics to infer the thermodynamic properties of the CMB. In particular, nuclear physics predicts that the recombination between baryons and electrons occurred at a temperature of the order of $3 \times 10^3 \text{K}$ (Hu, 2008). Since then, the universe has expanded by a factor 10^3 which had the consequence of dropping the temperature of the CMB by roughly the same factor. The emission of the CMB forms a black body radiation of about 3K. The CMB signal is therefore the brightest in the microwave frequency domain, between 10GHz and 500GHz.

Experiments dedicated to the study of the CMB naturally observe in the microwave domain. Many other emissions than the CMB radiation, from cosmological as well as from astrophysical source, light up the microwave sky. In CMB analysis, any signal that is observed together with the CMB is referred to as a foreground.

In order to get a clean CMB signal, foregrounds must be removed, and to be removed they must be understood, at least in the frequency domain of interest. In this chapter I will give a list of the foregrounds that need to be taken into account during the separation step. For each of them I will overview the physical processes from which they emerge and describe their expected signals, sometimes empirically. This list is meant to be exhaustive but only in the scope of the Planck mission, i.e. for full sky data between 30GHz and 857GHz at a maximum resolution of 5 arcmin. Also, this overview reflects our current knowledge of the microwave sky, which is of course not perfect.

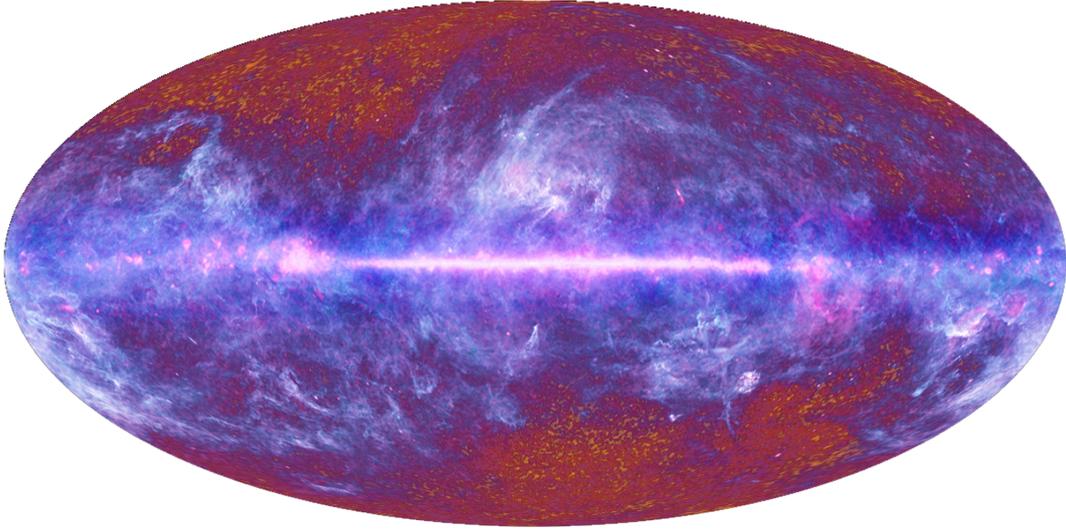


Figure 2.1: The Planck all sky. This image was released with the early results of the Planck collaboration for communication purpose only, but shows how hard is the task of CMB cleaning is. In the highest Planck frequencies, the galaxy signal in blue is much higher than that of CMB in red-orange. In addition, Planck has a high enough resolution to measure the fine and complex structures of the galaxy, which are another challenge to overcome for CMB cleaning. Credits: ESA / LFI & HFI Consortia.

2.1 The Planck Sky Model

I will often refer to the Planck Sky Model (PSM) throughout this chapter. The PSM is a project that aims to simulate a multi-component microwave sky, from a few GHz to above the THz. It was first developed within the Planck collaboration, especially for this mission. Thus the software simulates the sky as it can be seen by experiments such as WMAP or Planck. The goal is to produce realisations of the data as close to the real data as possible. The software does not propose just one simulation of the sky since the components can be simulated with different values for their physical parameters. The simulations are very useful to visualise the data and its expected components and to test component separation methods.

The components included in the PSM and their simulation process are described in Delabrouille et al. (2012). For some realisation of certain components, the PSM extensively uses previous attempts of mapping them. In this chapter, original works will be cited when appropriate, along with the PSM. The code and an exhaustive list of collaborators can be found on the web page of the project¹.

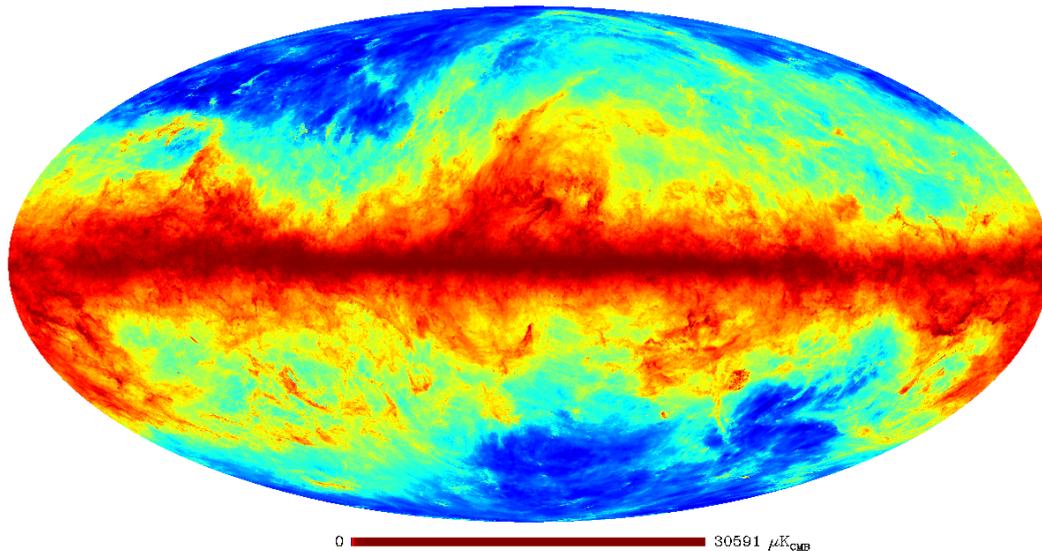


Figure 2.2: Thermal dust spatial distribution as simulated by the PSM, based on the work of Finkbeiner et al. (1999) and Finkbeiner (2003). The high sensitivity of Planck and the inclusion of detectors observing up to 857GHz makes the dust the brightest component in the data. Spinning dust is highly correlated with thermal dust and therefore has almost the same spatial distribution.

2.2 Galactic emissions

The most important foreground that needs to be taken into account during CMB cleaning is the galaxy. Its signal is several order of magnitudes higher than that of the CMB on some part of the sky. The galaxy itself is a mixture of different components and each of them can be simply described by a single physical process. The diffuse emissions are produced by the interstellar medium (ISM).

Figure 2.1 shows the extent of the problem. This figure was released with the Planck early results and shows a full sky map made up from a mixture of the Planck observation maps. It emphasize the angular extent of galactic dust emission. The galaxy is not only overwhelming all other components in the galactic plane but is also present over almost all the sky, mostly at high frequencies. The CMB fluctuations are barely distinguishable in some parts of the sky around the poles.

2.2.1 Dust emission

The ISM is principally composed of dust, which is an aggregation of grains. The electric and magnetic molecular dipoles within the grains have two modes of excitation: oscillation

¹<http://www.apc.univ-paris7.fr/~delabrou/PSM/psm.html>

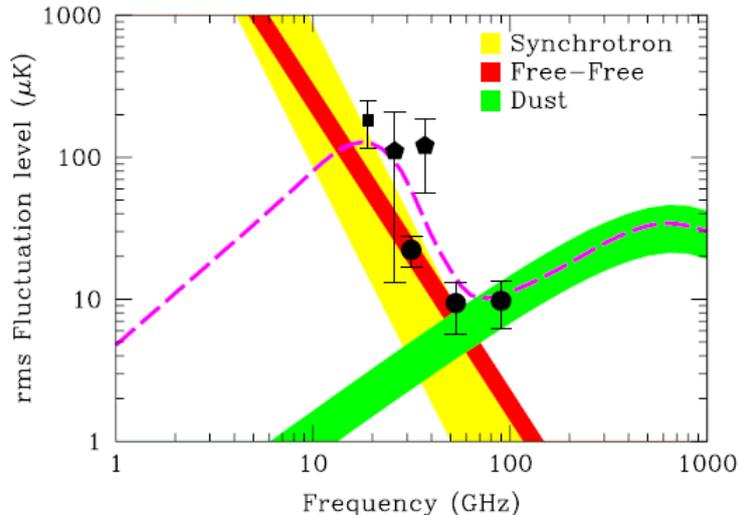


Figure 2.3: Electromagnetic spectrum of thermal and spinning dust combined (de Oliveira-Costa et al., 1998). Both components peak outside of the range of observed frequency and are approximated by power law emissions. Spinning dust peaks around 20GHz while thermal dust, hotter, peaks further in the infrared, around 1000GHz.

(or vibration) and rotation. Each mode is source of a specific radiation. Dust emissions are the brightest signals of the infrared band.

Thermal dust emission

The thermal dust emission is the dominant component at frequencies above about 80GHz. Around star formation regions, dust is heated up by interstellar radiation field. The grains absorb UV and optical photons that leave them in an excited electronic state. Most of the electronic energy is converted into vibrational energy. The dipoles then emit in infrared. At Planck frequencies, the dominant source of photons is the big ($1\mu\text{m}$) grains that radiate at thermal equilibrium. They have a thermal emission with a frequency-dependent emissivity. The thermal dust emission law peaks in the infrared, between 100 and $200\mu\text{m}$, i.e. about 1000GHz. Therefore, within the Planck frequencies, it can be approximated by a power law with a positive spectral index varying between 1.5 and 2 across the sky.

Spinning dust

The spinning dust component was not identified as a dust component at the moment of its discovery Kogut et al. (1996); Leitch et al. (1997); Watson et al. (2005). It was first named anomalous microwave emission. This emission was significantly correlated to the thermal dust emission, but the two of them could not be explained by one simple dust model. The emission is actually due to the spinning of the small dust particles that

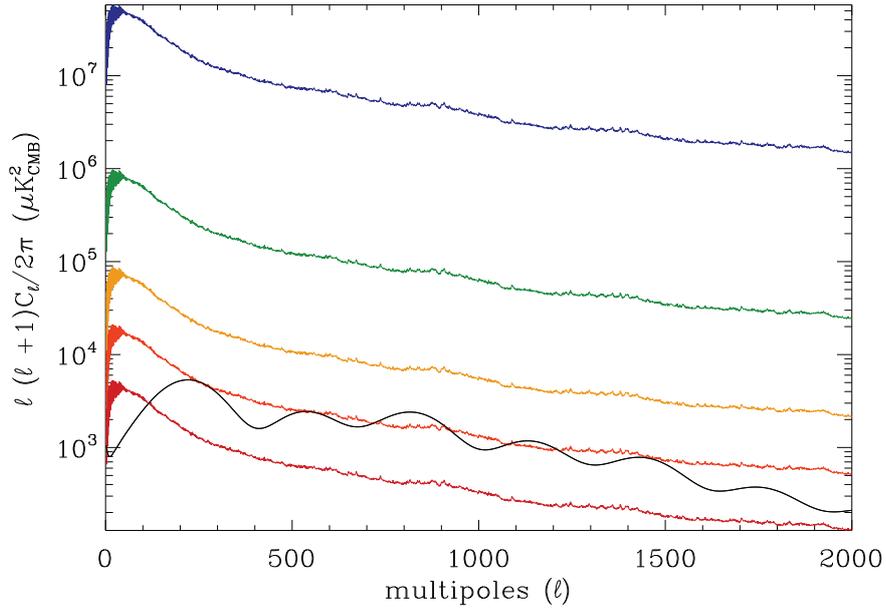


Figure 2.4: Thermal dust power spectrum through frequencies from the simulated full sky maps of the PSM. The black body law of thermal dust peaks at a higher frequency than the highest observed frequency of Planck. Dust emission can therefore be approximated as a rising power law emission. Colours from red to blue respectively represents the amount of dust in the frequency bands 70GHz, 100GHz, 143GHz, 217GHz and 353GHz. The CMB anisotropies, whose power spectrum is represented by the black solid line, is at most as powerful as dust considering the entire sky at low frequencies.

induces rotational electric dipole emission. The intensity of the emission peaks around 20GHz, such that its frequency spectrum can be approximated by a power law emission, even if there is a high-frequency cut off due to the limited speed at which a grain can spin. Figure 2.3 shows the frequency dependence of both the spinning dust and the thermal dust.

Visualisation

Figure 2.2 shows the spatial distribution of thermal and spinning dust over the sky, as predicted by the PSM from the works of Finkbeiner et al. (1999) and Finkbeiner (2003). The signal is very high in the galactic plane, as expected, and is also present at high latitudes which make the dust a component that extends over almost the whole sky. Figure 2.4 shows the power spectrum of thermal dust at several frequencies together with the power spectrum of the CMB that best fits the current available data. For any component separation method, it is very challenging to clean the CMB map from dust of the highest frequencies.

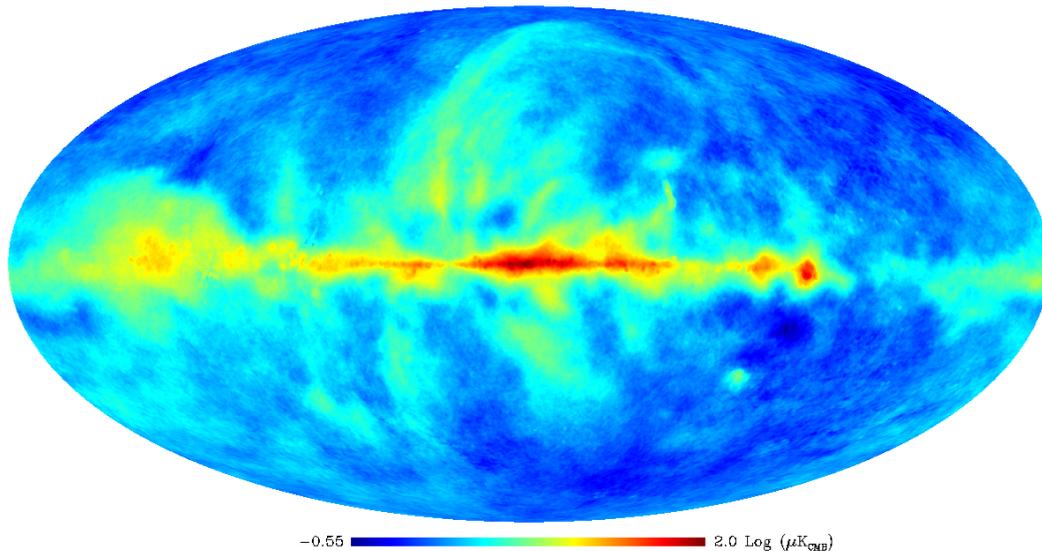


Figure 2.5: Synchrotron spatial distribution as simulated by the PSM, based on the work of Haslam et al. (1982) and Giardino et al. (2002). This map shows the presence of spurs above and below the galactic plane. These features are due to the field lines of the magnetic field of the galaxy, in which the electrons are trapped. Since the magnetic field is present all around the galaxy, emission from electrons comes from a large fraction of the sky.

2.2.2 Synchrotron emission

Synchrotron emission is due to free electrons spiralling in the magnetic field of the galaxy (Rybicki and Lightman, 1979). Because of their circular acceleration, the electrons radiate. The intensity of the synchrotron emission depends on the electron density and the strength of the magnetic field perpendicular to the line of sight. The galactic magnetic fields being weak (1-5 μG), the electrons need to be relativistic in order to produce detectable light. One possible source of relativistic electrons is the ejection and the acceleration by supernovae. Thus, cosmic ray electrons come from future star forming regions and there is a correlation between dust and synchrotron emissions. The intensity of synchrotron emission peaks in the radio band, such that it can be approximated by a negative slope emission law (cf. Davies and Wilkinson (1998) for a review). Synchrotron has a steep spectral index, varying over the sky between -3.3 and -2.7. The spatial distribution of synchrotron emission will change substantially with frequency.

Figure 2.5 shows the spatial distribution of the synchrotron emission over the whole sky, as predicted by the PSM from the work of Haslam et al. (1982) and Giardino et al. (2002). As for dust, the signal is the highest in the galactic plane. The broader regions on both sides of the galactic center are sliced views of the spiral arms of the galaxy. This

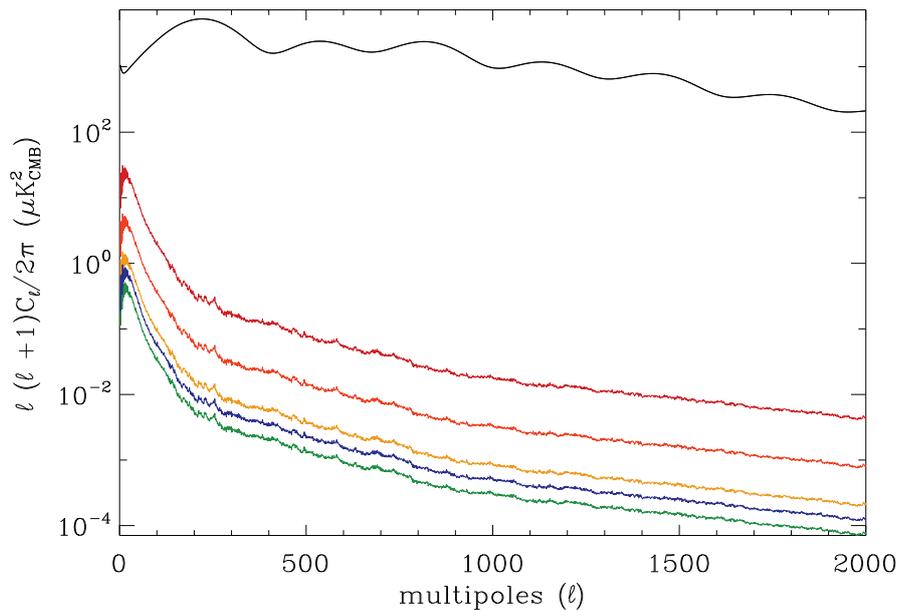


Figure 2.6: Power spectrum of the synchrotron through frequencies from PSM simulations based on the work by Haslam et al. (1982) and Giardino et al. (2002) on full sky. Colours from red to blue respectively represents the amount of dust in the frequency bands 70GHz, 100GHz, 143GHz, 217GHz and 353GHz.

map gives a hint on the structures of the galactic magnetic field, for instance with the bright curved region over the galactic center, called the north galactic spur. Figure 2.6 shows the power spectrum of the synchrotron at different frequencies, together with the CMB power spectrum that best fits the current data.

2.2.3 Free-free emission

As synchrotron, the free-free emission (or Bremsstrahlung) is a component from the radio band. Because of the dust radiation coming from the infrared band, the steep index of synchrotron emission in the radio band and the CMB in between, free-free is the least well understood galactic component. It is due to electron-ion scattering in the warm ionised ISM, near star forming regions. The UV radiation from the very young hot stars ionises the surrounding gas. Then the electrons are decelerated in the electric fields of the ions and emit photons during the interaction with the nucleus. The emission reaches its highest intensity in the radio domain. At Planck frequencies, free-free emission can therefore be modelled by a power law with negative slope. The spectral index varies little, between -2.15 and -2.14 , such that free-free has almost the same spatial distribution at all frequencies. The velocity distribution of the electrons is well described by a Maxwellian. This spectral index depends slightly on the electronic temperature of the

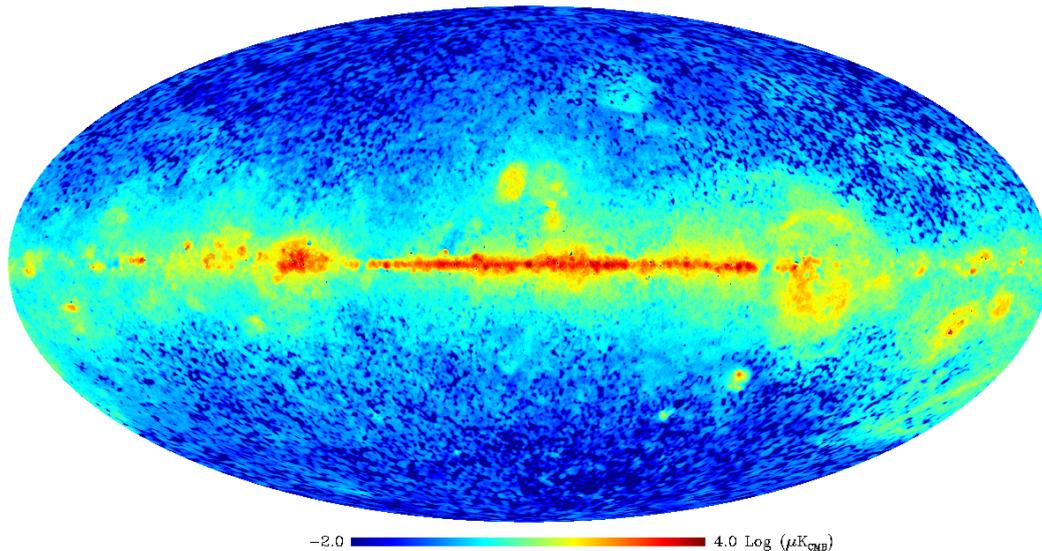


Figure 2.7: Spatial distribution of free-free emission as simulated by the PSM, based on the work of Dickinson et al, MNRAS, 341, 369 (2003) and Bennett et al. The emission is the brightest in the galactic plane. It becomes quickly very low compare to the other components as the latitude decreases or increases. Note the noise in the map at high galactic latitude.

velocity dispersion. Radio frequency $\text{H}\alpha$ maps provides template for free-free emission, in regions of low dust optical depth where the maps can be corrected. $\text{H}\alpha$ maps trace the α line of the Balmer series that probes the surface brightness of the ionised emission from the entire Galaxy. The largest source of uncertainty comes from the dust absorption of the α line.

Figure 2.7 shows the spatial distribution of the free-free component over the sky, as predicted by the PSM and works by Miville-Deschênes et al. (2008). The emission is more locally distributed than that of dust and free-free but is still very diffuse. Figure 2.8 shows the power spectrum of the free-free component at different frequencies, together with the CMB power spectrum that best fits the current data.

2.2.4 Molecular lines

Part of the ISM is composed of dense molecular clouds. Photons from the stars excite these molecules. Then, the molecules emit through de-excitations. The strongest emission lines in the frequency range of the Planck mission are those of ^{12}CO . Although highly correlated, each transition line has a different spatial distribution. The most probable transitions, and therefore the brightest ones, are $(J=1\rightarrow 0)$, $(J=2\rightarrow 1)$ and $(J=3\rightarrow 2)$,

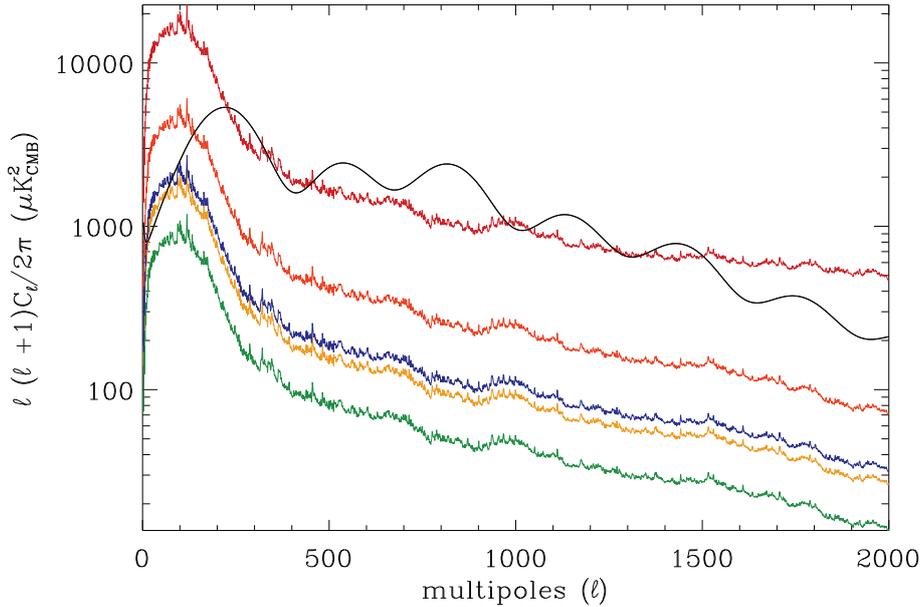


Figure 2.8: Power spectrum of the free-free emission through frequencies of PSM simulations based on the work of Miville-Deschênes et al. (2008). Colours from red to blue respectively represents the amount of dust in the frequency bands 70GHz, 100GHz, 143GHz, 217GHz and 353GHz. Thermodynamics unit curves the power law of free-free such that in units of K_{CMB} free-free looks like having more power at high frequencies. Free-free emission is particularly powerful at large scales and is therefore a major component in the range of frequencies of interest.

where J is the total angular momentum of the molecule. Other molecules than ^{12}CO are present in the ISM – ^{13}CO , HCN and HCO^+ among the most abundant (Naylor et al., 2010) – but they have fainter contributions to the total signal. CO-molecule only is significant during the component separation step in Planck analysis. PSM maps of CO spatial distribution are based on the works of Dame et al. (2001).

As a by-product of their 2013 cosmological results, the Planck collaboration released the first full sky mapping of the spatial distribution of the CO-molecule present in our galaxy. Figure 2.9 shows the spatial distribution of the CO-molecule component over the sky, as extracted from 2013 Planck data (Planck Collaboration et al., 2014b). This map and maps of other transition lines are discussed in the Planck paper on CO emission (Planck Collaboration et al., 2014b). Figure 2.10 shows the principal emission lines in the Planck frequency response. The brightest components, ($J=1\rightarrow 0$) and ($J=2\rightarrow 1$), are present in two of the cleanest channels. Emissions from CO molecules were not expected components to be present in the data. The response of the instruments were chosen without considering the frequency of emission lines, which could explain the lines right in the middle of some frequency bands. The Planck collaboration turned that disadvantage

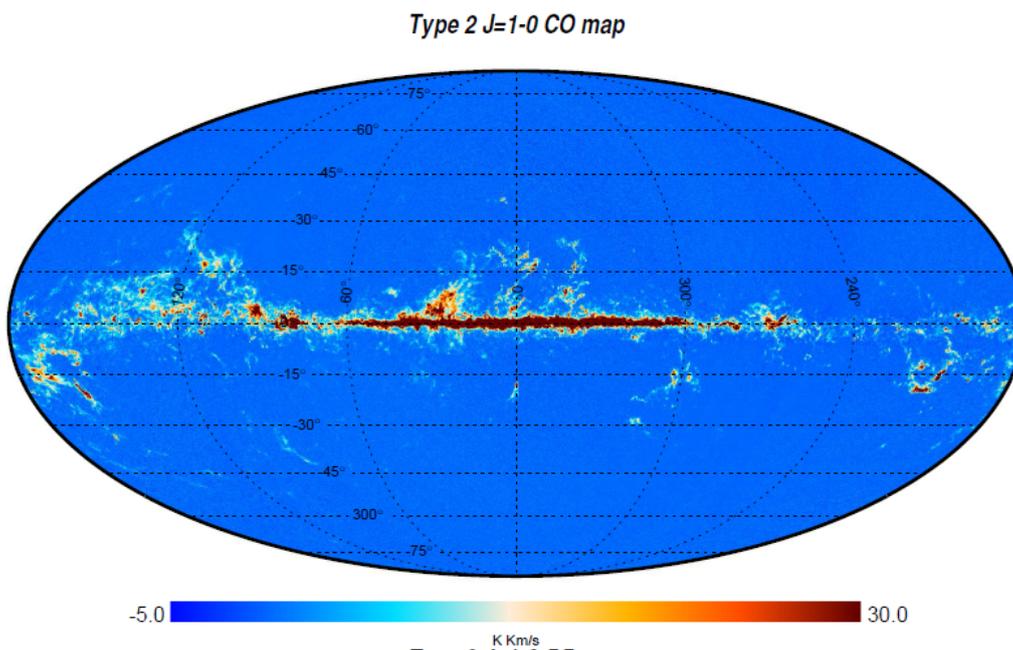


Figure 2.9: Spatial distribution of CO ($J=1 \rightarrow 0$) transition line. This map is a result of component separation methods applied to real 2013 Planck data (Planck Collaboration et al., 2014b). The emission is concentrated in the denser parts of the galaxy.

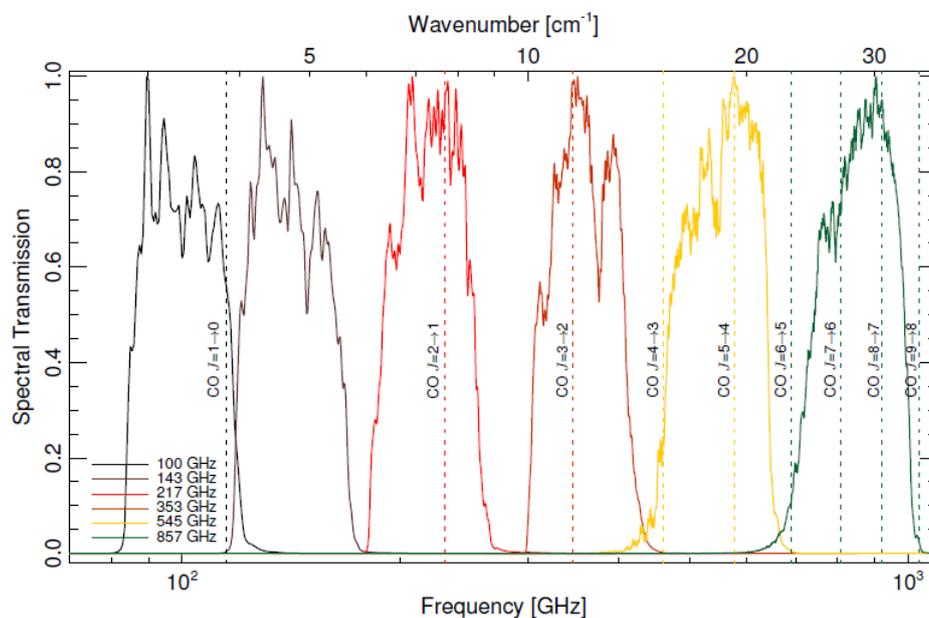


Figure 2.10: The averaged spectral response of Planck/HFI instrument (Planck Collaboration et al., 2014b). The vertical dashed bars represent the CO emission lines. The three most probable and therefore the brightest transitions happens right in the middle of the 100, 217 and 353GHz bands.

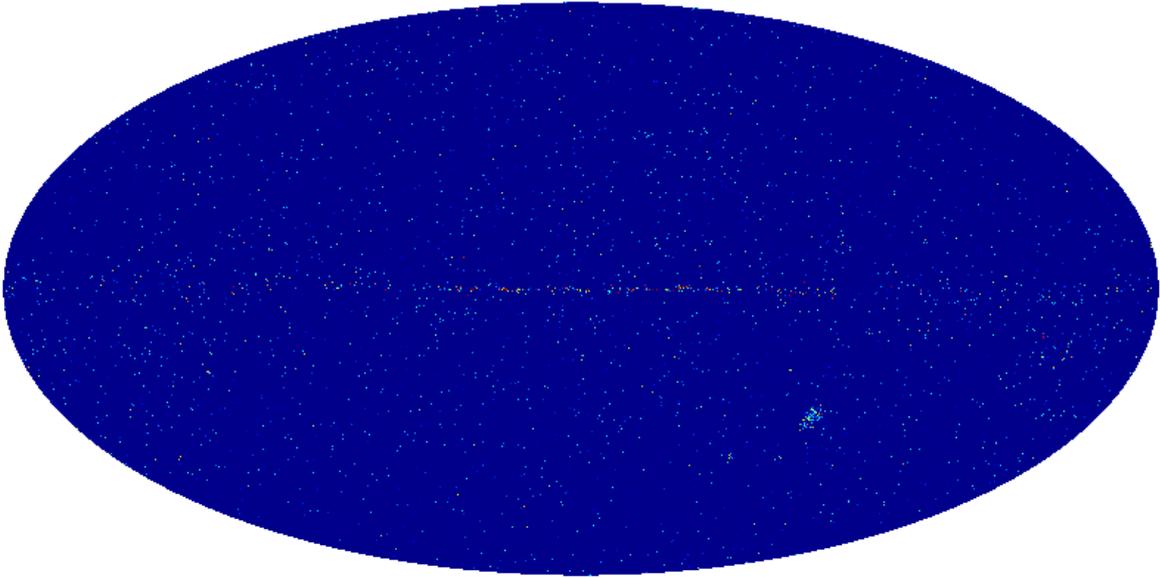


Figure 2.11: Point source catalogue as generated by the PSM, partly based on Wood and Churchwell (1989); Kurtz et al. (1994). As expected, this map looks like a full sky starry night. Point sources are concentrated in the galactic plane but a non-negligible amount of them are distributed all across the sky. The clusters of point sources in the bottom right of the map are the Magellanic clouds.

into a world première result: the first full sky mapping of CO line emission.

2.2.5 Galactic point sources

A compact source is a small source that spreads over a very few pixels. It is spatially resolved and is very bright compared to the sky around it. The main object causing the galactic point sources are stars and Ultra Compact HII (UCHII) regions. UCHII regions are dense HII regions whose size does not exceed the parsec.

Figure 2.11 shows a catalogue of galactic point source based on Wood and Churchwell (1989); Kurtz et al. (1994). There are many of them in the microwave domain. Diffuse components and point sources have very different generative models, close to Poisson process. Since a combined treatment of Poisson and Gaussian processes is difficult, point sources are treated separately from the diffuse emissions. There are several methods to remove the point sources from the observation maps prior to their use for component separation. A common way to remove them is to set to zero the pixels that contain a point source and then fill with fake data the holes thus made in the maps. These processes are respectively called masking and inpainting. Alternatively, the point source function (PSF) of the instrument can be fitted to the brightest point sources. The fit is then subtracted off the data.

2.3 Extra-galactic components

2.3.1 CMB secondary anisotropies

CMB photons were emitted at a very early time and those that we receive have travelled through the late universe. During their journey, they crossed the structures of the universe and they interact with them and their content. The properties of the CMB radiation are found to be slightly altered as photons undergo physical processes in the structures. The difference between the CMB before the modifications due to late time effects and after the modifications are called CMB secondary anisotropies and are considered as components on their own.

Sunyaev-Zel'dovich effects

Galaxy clusters contain hot electron gas. Photons of the CMB scatter via inverse-Compton scattering on these electrons. The electrons that interact with the CMB transfer part of their energy to the photons. Thus, after crossing a galaxy cluster, the CMB has a modified black body with slightly more energetic photons. This is known as the Sunyaev-Zel'dovich effect (SZ effect) (Sunyaev and Zeldovich (1972), see Carlstrom et al. (2002) for a review). The change in temperature is not the same for all lines of sight, since photons coming from different directions have crossed different structures. Thus, the SZ effect induces secondary CMB anisotropies.

The SZ distortions due to the thermal energy of the electrons are called the thermal SZ (tSZ) effect. A second order effect, coined the kinetic SZ (kSZ), is due to the bulk velocity of the electrons with respect to the rest frame of the CMB. The interaction of the CMB photons with a moving gas of electrons results in a redshift or blueshift of the photons, depending on the direction of the cluster velocity with respect to the observer. Thus, it is a priori undistinguishable from primary anisotropies. However, in a typical galaxy cluster, the kSZ effect is small, inducing a change in temperature of the order of 10^{-5} .

Figure 2.13 shows the 3K black body, i.e. the CMB electromagnetic spectrum as it has been emitted at the last scattering surface, the CMB electromagnetic spectrum of the anisotropies due to the tSZ effect and the kSZ effect. The tSZ spectrum is negative then positive because the whole CMB spectrum is blueshifted because of the kick of the hot electrons. It crosses zero at about 217GHz, one of the frequency observed by HFI. The kSZ spectrum is negative because of the broadening of the CMB spectrum, which is due to both blueshifting and redshifting of the CMB photons. Figure 2.12 shows a map of the CMB secondary anisotropies due to the SZ effect. Since the SZ effect is due to galaxy clusters and is very faint, this component looks like a point source map and therefore is highly non-Gaussian.

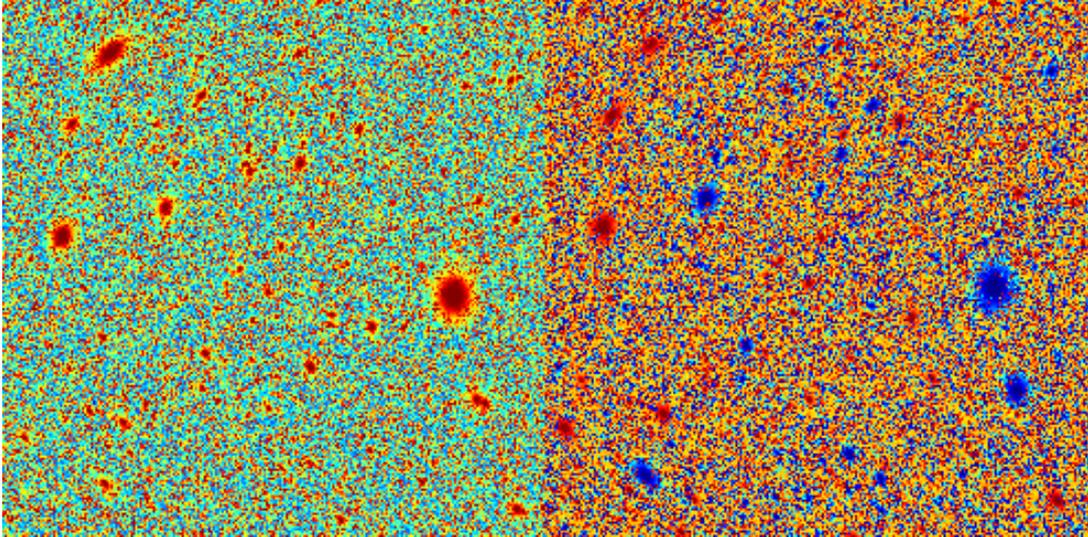


Figure 2.12: Detail of tSZ (left) and kSZ (right) maps as simulated by the PSM at 353GHz. Blue means lower temperature and red means higher temperature. The size of the squares is 30 degrees on each side. CMB photons that undergo tSZ effect in spherical galaxy clusters gain energy. Thus the tSZ map at higher (resp. lower) frequency than 217GHz is made of wide red (resp. blue) dots in direction of the clusters. Since kSZ blueshifts or redshifts the CMB photons depending on the bulk motion of the cluster, kSZ map is a repartition of blue and red wide dots.

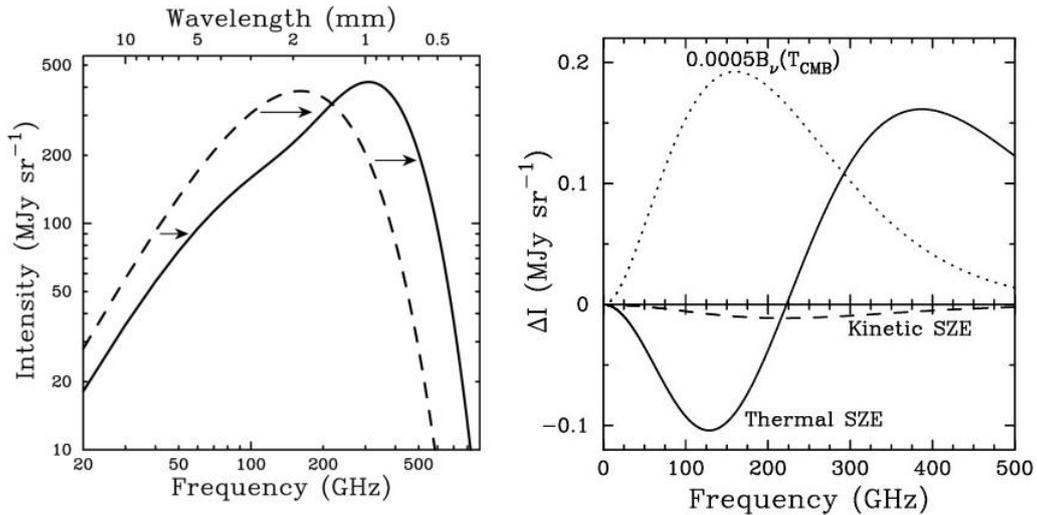


Figure 2.13: CMB spectral distortions due to thermal SZ effect (left) and the spectrum of tSZ and kSZ (right) (Sunyaev and Zeldovich, 1980). The dashed curve on the left panel represent the perfect 3K black body and the solid line represent the same distribution of photons but that have undergone large y -type distortions. The solid line is a first order approximation in y . The solid line on the right panel represent the electromagnetic spectrum of the tSZ difference map in the direction of a galaxy cluster, the dashed line represents the electromagnetic spectrum of the kSZ difference map, still in the direction of a cluster where the effect occurs. For comparison, the 3K black body spectrum is also plotted, after being scaled to tSZ and kSZ spectra.

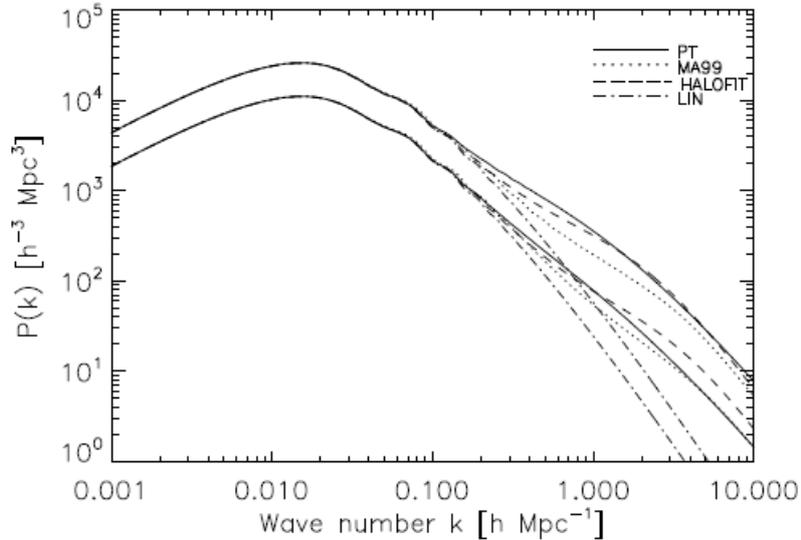


Figure 2.14: 3D matter power spectrum including the effects of the structures. The dashed-dotted line shows the power spectrum of the iSW effect, which treats the gravitational effect in linear perturbation. The Rees-Sciama effect is the full resolution, i.e. including higher orders in the perturbation, of the integrated gravitational effect of the structures. The figure shows the result of three codes, represented here by the solid, dotted and dashed lines. Figure from Junk and Komatsu (2012).

Integrated Sachs-Wolfe effect

During its propagation, a CMB photon crosses inhomogeneities. The photons undergo relativistic effects due to the change of the gravitational potential along their path. For instance, if a photon crosses an overdensity, it gains energy when falling into the potential well and loses exactly the same amount of energy when escaping the overdensity. Now, if the overdensity grows between the moments the photon enters and exits, then the photon has to escape a deeper potential well than the one it entered in. Therefore, between before and after the crossing of the overdensity, the photon loses energy. The integrated Sachs-Wolfe (iSW) effect (Sachs and Wolfe, 1967; Rees and Sciama, 1968) is the overall effect on the photon energy over its whole path, from the moment it was emitted at the surface of last scattering to the moment it was detected.

In the CMB power spectrum, the typical scale of the effect of the structures is the horizon size because smaller structures tend to cancel out. The effect is hard to detect because of the weak effect on the energy of the photons and because it has the same electromagnetic spectrum than that of the CMB. In addition, the effect is the biggest just where cosmic variance is the largest. Figure 2.14 shows a matter power spectrum including the effect of structures.

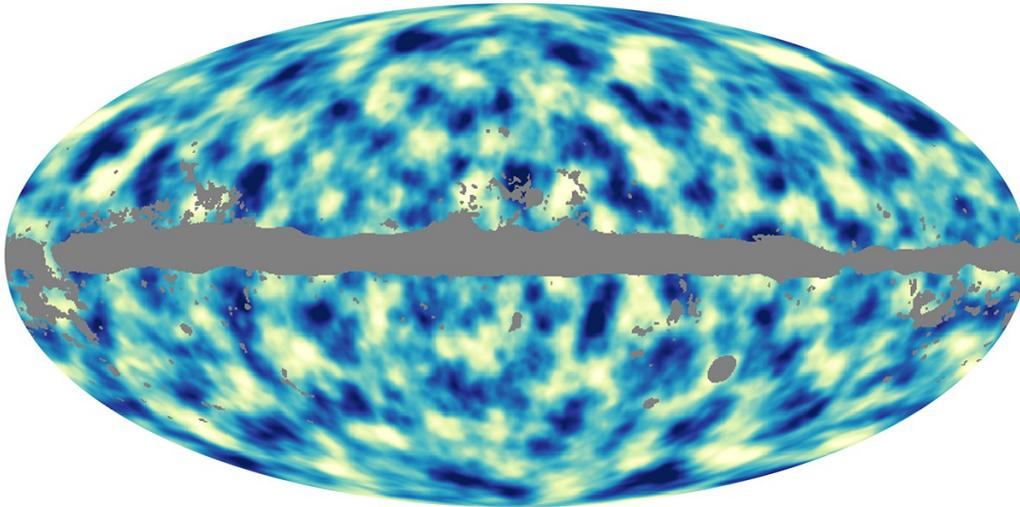


Figure 2.15: Map of integrated gravitational potential (Planck Collaboration et al., 2014c), from the last scattering surface to our galaxy. In other words, this map traces all the matter of the observable universe along our light cone.

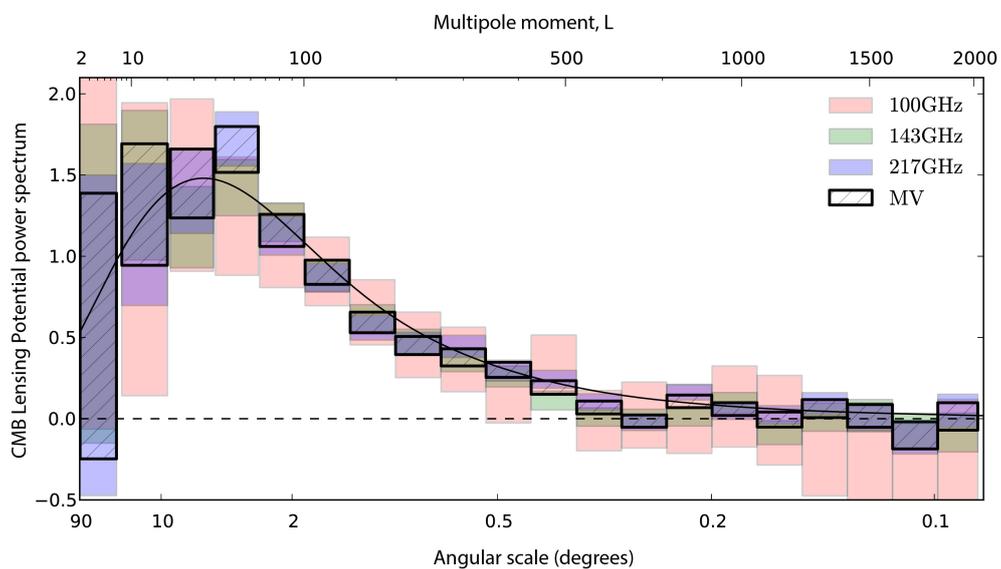


Figure 2.16: Power spectrum of maps such as that of figure 2.15 (Planck Collaboration et al., 2014c).

Lensing

Locally, the universe is not homogeneous. Thus, when the geodesic of a CMB photon passes near an overdensity, the direction of propagation of the photon changes, due to the gradient of the gravitational potential. The whole CMB map is distorted due to the weak lensing. The order of magnitude of the deflection is $2'$. Thus, the deflection angle has a lower value than the resolution of the Planck instruments. But since the deflection is correlated over several degrees, the effect is still noticeable in the data.

Gravitational lensing is achromatic. Thus, lensing does not change the temperature of the CMB. The anisotropies have the same frequency dependence than the CMB. As explained later, the assumption that the CMB follow a black body law emission is crucial for CMB cleaning. Component separation can not unlens the CMB map. But it must be treated because weak lensing alters the statistical properties of the primordial CMB (Lewis and Challinor, 2006) – for example, it introduces non-Gaussianity. Lensing also provides information about the late universe. For these two reasons, the output CMB of component separation must be treated in post-processing.

Figure 2.15 shows a full sky gravitational potential map, integrated over the whole path of the photons. It was reconstructed from a the deflection angle map computed from the lensed Planck CMB map. Figure 2.16 shows the power spectrum of that map.

2.3.2 Extra-galactic point sources and the Cosmic Infrared Background

There are two kinds of extra-galactic point sources, one coming from the radio band, one from the infrared band. The radio sources correspond to galaxies hosting a radio loud active galactic nuclei (AGN). The infrared sources correspond to dusty galaxies in the process of forming stars.

As for the galactic point sources, the sources that are resolved need a special treatment, like fitting and removing or masking and inpainting. The unresolved infrared sources produce a diffuse anisotropic background known as the Cosmic Infrared Background (CIB). This component is nearly a random Gaussian field and is handled as a diffuse component in the separation process.

Figure 2.17 shows the relative contribution of the galaxies at different redshift to the total CIB signal.

Conclusion

I have presented an account of both foreground emissions and CMB secondary anisotropies. These components are present in the data of any CMB experiment and therefore mix with

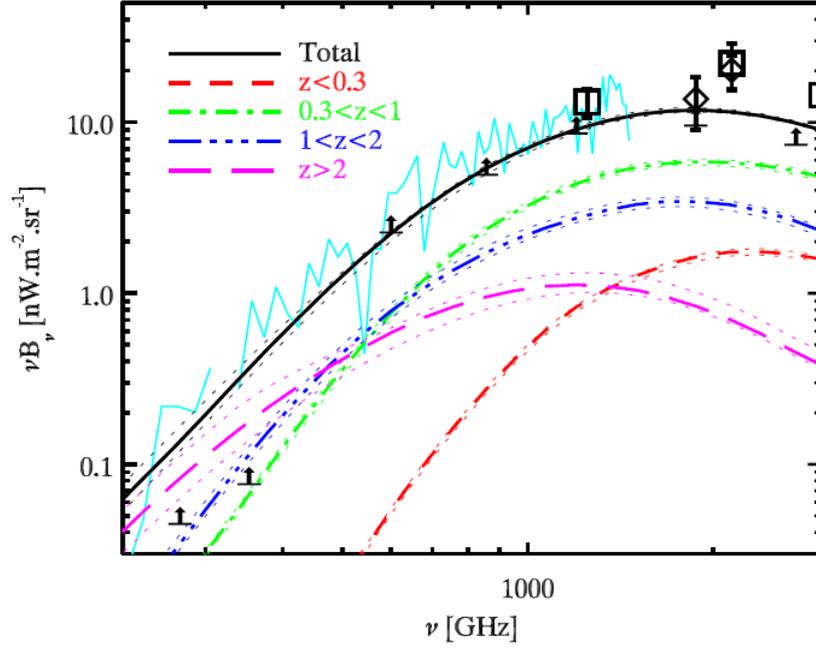


Figure 2.17: Dusty galaxies contribution to CIB per bin of redshift (Planck Collaboration et al., 2011b). Each color represent the mean galaxy spectrum within a particular redshift bin. The galaxies that contribute the most to the total CIB signal are that of intermediate redshifts, around $z = 1$.

the CMB. The diversity and the complexity of the foregrounds make them difficult to model and therefore challenging to separate. In the next chapter, I present the main approaches to CMB component separation.

Chapter 3

Basic concepts of CMB component separation

Introduction

All photons emitted through the various physical processes presented in chapter 2 mix with the CMB and all the components are observed as a single mixed signal per frequency. After the pre-processing of the data, the operation consisting in recovering the individual components that compose the signal is called component separation.

The main cosmological goal of missions such as Planck is to extract the map of the CMB anisotropies because of its power to constrain cosmological models and fundamental physics, as we glimpsed it in chapter 1 and section 2.3.1. In this case the step is just called CMB foreground cleaning. The data of an experiment that observes the sky in the microwave domain is full of information on the late universe and local astrophysics. It would then be a waste to use the data to just produce the CMB map. Recovering either the CMB only or a whole set of various physical maps requires different approaches. Thus, several classes of component separation methods exist. Some of them are more CMB cleaning oriented and others are more ambitious and aim to recover as many physical component maps as possible.

In this chapter I first point out the difficulties that component separation has to overcome, then I sketch the fundamental assumptions that, in principle, allow source separation. Finally I give an overview of existing component separation methods. Beyond recovering clean individual maps, getting meaningful errors on them is equally as important. I will especially highlight the difference between blind and parametric methods because both approaches introduce different error estimates.

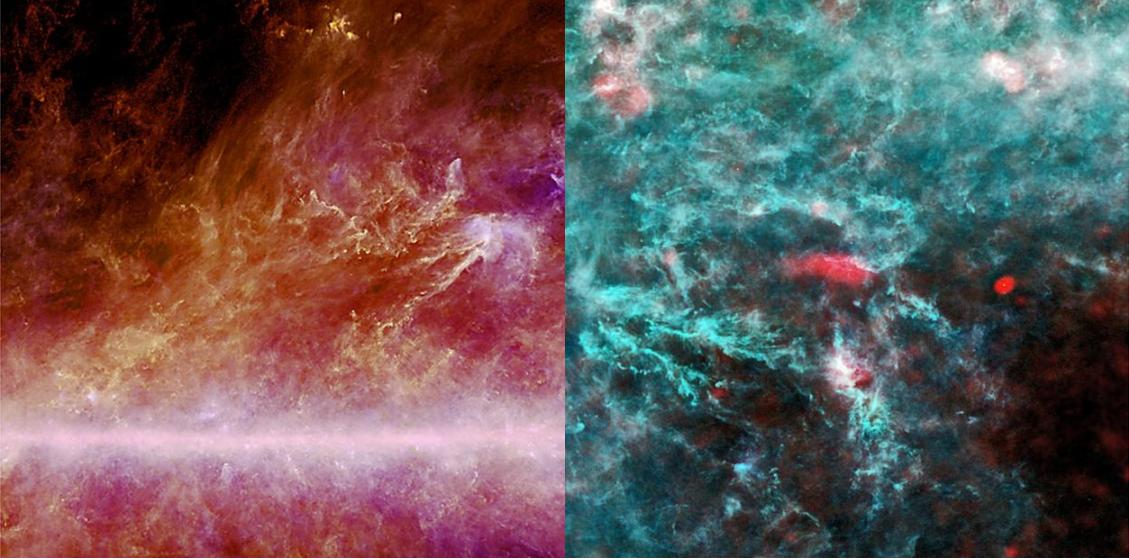


Figure 3.1: Details of Planck all sky (figure 2.1, ESA/ LFI & HFI Consortia). Left: Detail of the galactic bulge. Right: region in the constellation Perseus. If the galaxy were easy to model the high intensity would not be a problem. But the galaxy exhibits fine and complex structures in the Planck data that are impossible to model right without a good physical model. In case of full sky analysis and poor modelling, a bright component leaves too much residual contamination in the cleaned CMB map.

3.1 The component separation challenge

The principal difficulty in separating the CMB from other components occurs in regions where the components are much brighter than the CMB or in cases where the components are similar to the CMB in several respects.

As shown by the Planck all sky in figure 2.1 and by the figures of section 2.2, the galaxy is brighter than the CMB by several order of magnitudes. That would not be a problem if the generative model of the galactic components were well known (Cardoso, 1998). But for all sky observations, complexity of foregrounds increases with the sensitivity and the resolution of the instruments. Thus, very fine structures of the galaxy can be seen, as in figure 3.1, and the galactic components can only be accurately described by physical or very complex statistical models. The necessity of approximating the galactic model makes the identification of the components uncertain in regions where the galaxy is the brightest, e.g. the galactic plane. In those cases, foregrounds residuals leak into the estimate of the CMB map.

Some components, like the diffuse emission of the unresolved point sources, share several properties with the CMB such as coherence through frequencies, isotropy and same level of power at multipole $\ell \gtrsim 2000$. The presence of noise then blurs the remaining difference between the components and degeneracies therefore occur between them.

Each component has its own peculiarities that need to be modelled either physically

or in terms of phenomenological model. Each component can be well described, or well approximated in some basis, but there is no unique basis that models all components properly. Thus, the diversity of the components can be a problem in writing a model that would allow a simple description of all components. Approximations are necessary and lead to non-accuracy in the results. However, as it shown in the following diversity is also exploited to make the difference between the components and therefore separate them.

3.2 Key ideas to solve the problem

As a general rule, to separate different components mixed together, we have to make use of what makes them different from one another. In CMB experiments the feature that differentiates the most the components is their different frequency behaviour. The electromagnetic spectrum of the CMB is accurately known, since it has been measured with high precision by previous missions (Mather et al., 1994). All component separation methods assume that the CMB is a perfect black body since primordial CMB spectral distortions are too small to be detected (Chluba, 2014).

Although our knowledge on spectral emissions of most of the individual components in the data is rather poor, we know that their electromagnetic spectra are distinct. Furthermore, the data can be interpreted as a superposition of the components, i.e. each frequency contain a specific amount of each component. Thus a multi-frequency data set is the starting point of many component separation methods.

As an example, figure 3.2 shows the frequency spectra of the main galactic components in the frequency domain covered by the WMAP satellite. As the components mix linearly, each band of observation (K, Ka, Q, V, W) is a linear combination of the component maps, the linear coefficients depending on the emission laws of the components. The total contribution of the galaxy is also shown. But choosing just one map that scales through frequencies for the whole galaxy is not enough because the spatial distribution of the entire galaxy vary from one frequency to another, unlike the individual components that are approximately coherent through frequency.

Performing blind source separation requires at least one statistical assumption about the source, either independence (Cardoso, 1998; Cardoso et al., 2008) or sparsity (Zibulevsky and Pearlmutter, 2000; Bobin et al., 2007). Both are a measure of diversity of the sources. Independence force the separation of sources that have been independently emitted whereas sparsity use clear morphological diversity to disentangle between several signals.

The goal of component separation is to recover the individual components by combining a linear mixture of them. In other words, it is an inverse problem, from the frequency

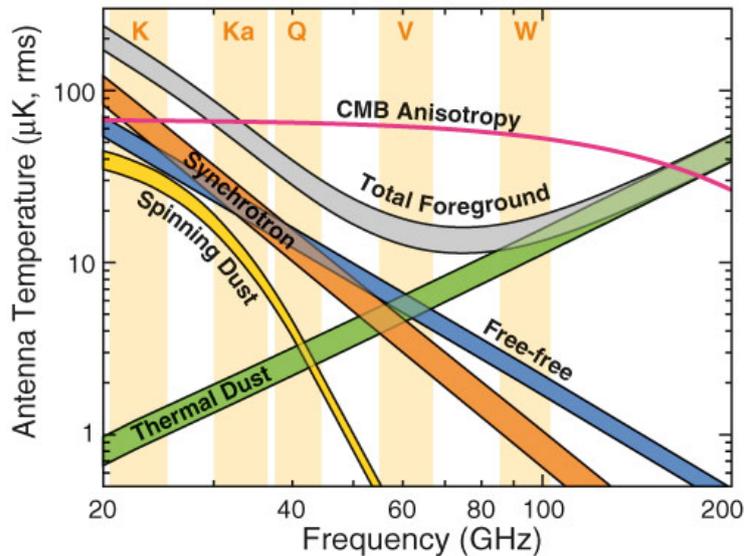


Figure 3.2: The emission laws of the foregrounds taken into account for the WMAP nine year release (Bennett et al., 2013). The amount of total foreground at each frequency band of observation (K, Ka, Q, V, W) is a specific superposition of the same foregrounds. Multi-frequency observations are a powerful way of removing foregrounds. For instance dust dominates the W band, which makes it easier to remove in other bands. Note that in units of antenna temperature, the CMB response is not constant through frequencies.

map space to the component map space. The next section discusses methods to solve this problem.

3.3 Review of component separation methods

The optimal component separation method for CMB data does not exist yet. Source separation specifically dedicated to the analysis of the CMB signal has been an active field for a couple of decades now and several fundamental ideas have been introduced. This section gives an overview of the principal approaches to component separation together with their respective advantages and disadvantages. I will particularly stress the differences between the methods that are blind source separation methods and those that makes use of physical parametric modelling.

3.3.1 Data model

I will first describe the usual description of the data in the space of the spherical harmonics, without loss of generality. The available data is a collection of maps, each of them is an observation of the full sky at a given frequency and each of them is a specific mixture

of all the components presented in chapter 2. To good approximation, the flux of the physical emissions has no influence on each others. This assumption allows us to model the mixing as a linear combination. Hence the following decomposition for the piece of data $d_{i\ell m}$ contained in the spherical harmonic coefficient (ℓ, m) of the observation maps of frequency i (out of N_f frequency bands)

$$d_{i\ell m} = \sum_{k=1}^{N_c} A_{ik} s_{k\ell m} + n_{i\ell m}, \quad (3.1)$$

where the sum runs over the assumed number of components N_c ,

$s_k = \{s_{k\ell m}; \ell = \llbracket \ell_{\min}, \ell_{\max} \rrbracket, m \in \llbracket -\ell, \ell \rrbracket\}$ is the *spatial distribution*, or *map*, of the k th component, A_{ik} is the amount of component k in frequency band i and $n_{i\ell m}$ is the instrumental noise present in $d_{i\ell m}$. Equation 3.1 reads in matrix form

$$d_{\ell m} = \mathbf{A} s_{\ell m} + n_{\ell m}. \quad (3.2)$$

The matrix \mathbf{A} , which gathers all the coefficients A_{ik} , is called the *mixing matrix*.

3.3.2 Internal Linear Combination

The idea of the Internal Linear Combination (ILC) method (Bennett et al., 1992; Tegmark, 1997; Delabrouille et al., 2009) is encapsulated in its name: taking a linear combination of the multi-frequency data maps in order to cancel all components but one. When all components are random Gaussian fields, the optimal choice of coefficients is the one that minimises the variance of the linear combination.

The method can be applied in any basis that can describe the data. The resulting map is then expressed in the same basis. Historically it has been developed to work in pixel space and spherical harmonic space. A version of ILC using a needlet decomposition of the data, named NILC, was one of the four component separation methods that were retained by the Planck collaboration for the 2013 results and data release. The needlet basis is a basis for functions defined on the sphere. Its basis vectors look like localised waves on the sphere, as in figure 3.3. Needlets are then well defined to describe the complex structures of the galaxy while still having some of the properties of spherical harmonic basis.

For a recovering of non-CMB component, the method is parametric since the emission law of the component to be recovered is needed to perform the separation. ILC is therefore a powerful method for CMB cleaning because its electromagnetic spectrum is accurately known but is not suitable for blind separation of any other component whose frequency behaviour is uncertain.

One drawback of ILC is that the errors on the CMB map are only a propagation of

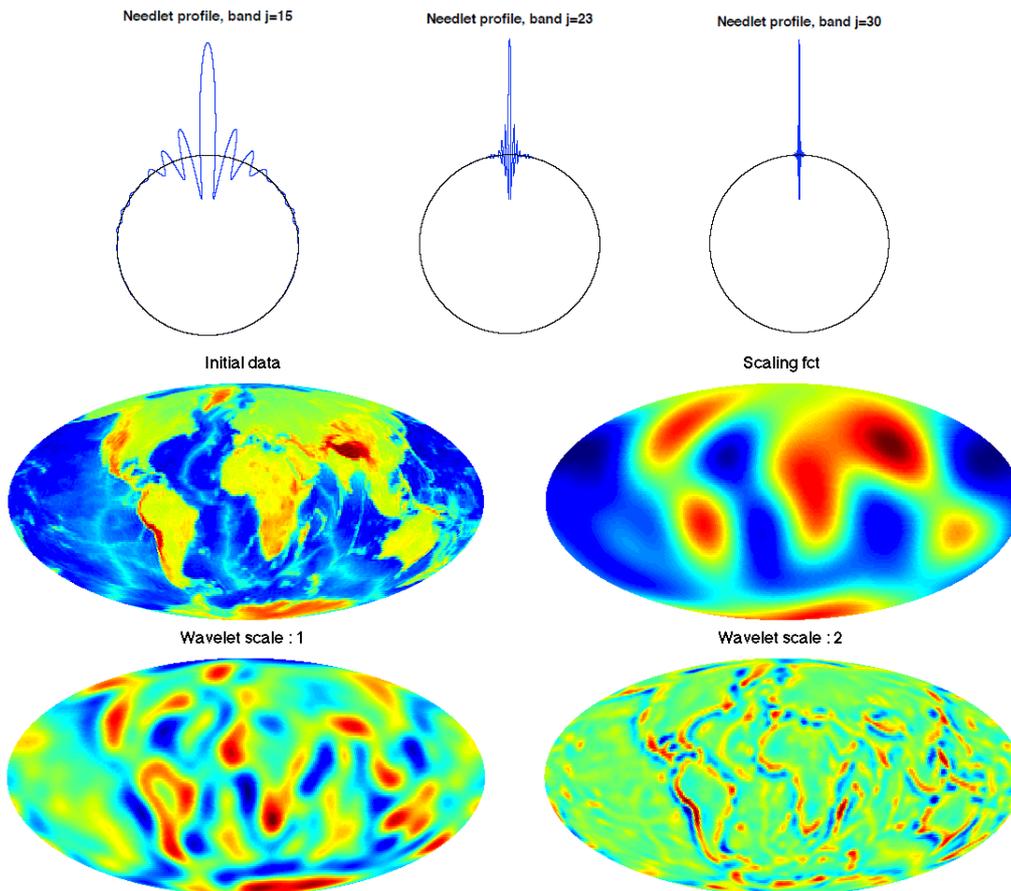


Figure 3.3: Needlets (or wavelets on the sphere) in pixel space (Leistedt et al., 2013; Faÿ and Guilloux, 2011). The top figure shows three basis vector of the needlet basis. They are more or less localized, depending on the scale of the vector. Also, they have properties of the spherical harmonics since the needlet set is constructed from this basis – they are localised wavelet on the sphere. Bottom figures shows such an example with the the function on the sphere "altitude on Earth". The signal is made of needlets of different scales.

the instrumental noise. In other words, it supposes that the separation is perfect. The ILC method is optimal if the mixed signals are Gaussian but some of the components, e.g. the galactic components, are strongly non-Gaussian. Thus, the separation is not perfect and some unmodelled residual contamination from foregrounds is expected in the ILC map. Figure 3.6 shows the NILC CMB map together with the CMB map of the three other official Planck component separation methods (Planck Collaboration et al., 2014c).

The method involves a weighted average of the observation maps in order to cancel all components but the one of interest. Usually, ILC weights are derived by minimizing the variance of a linear combination of the observation maps.

The ILC CMB map is derived as follows. Since the CMB has constant response

through frequencies, the data can be written, for each frequency band i ,

$$d_i = s_{\text{CMB}} + s_{\text{fg},i},$$

where $s_{\text{fg},i}$ contains the mixture all the components but the CMB and the noise. Now construct an estimator depending on a set of weights $\{w_i; i \in \llbracket 1, N_f \rrbracket\}$, such that $\sum_i w_i = 1$

$$\hat{s}_{\text{CMB}}(w) = \sum_i w_i d_i = s_{\text{CMB}} + \sum_i w_i s_{\text{fg},i}.$$

Assuming the CMB and the foregrounds to be uncorrelated, one can see that the weights $\{w_i; i \in \llbracket 1, N_f \rrbracket\}$ that minimise the variance of $\hat{s}_{\text{CMB}}(w)$ are the ones that maximally reduce the spread of the other components in the estimator. The technique of Lagrange multipliers is used for the minimisation of the variance, since it has to be done under the constraint $\sum_i w_i = 1$. At minimum, the estimator becomes:

$$\begin{aligned} \forall (\ell, m), \quad \hat{s}_{\text{CMB},\ell m} &= w^T d_{\ell m} \\ &= \frac{e^T C_d^{-1}}{e^T C_d^{-1} e} d_{\ell m} \end{aligned} \quad (3.3)$$

where C_d is an estimate of the data covariance and $e = (1 \dots \dots)^T$, i.e. e is the frequency response of the CMB.

3.3.3 Independent Component Analysis

In blind source separation, the simple assumption of mutual independence between the sources allows the separation of a mixture of non-Gaussian and at most one Gaussian signals. In CMB component separation, independence can be assumed between each component or just between the CMB and the foregrounds. In practice, the separation is performed by minimising of mutual information between the sources, making use of statistical tools such as the Kullback-Leibler (KL) divergence or Shannon entropy (see appendix A for more details on KL divergence and Shannon entropy).

The ICA instance of CMB component separation that was chosen for the presentation of the 2013 Planck results is called Spectral Matching ICA (SMICA) (Cardoso et al., 2008, 2002; Delabrouille et al., 2002). The method works at the power spectrum level. Therefore, it processes the data covariance. SMICA proceeds by minimising the divergence between the data and a model where the CMB is independent with all the other components.

Depending on the way that the component covariance and the mixing matrix of the model are parametrised, SMICA is either a blind method or a parametric method. Apart from independence and coherence, the model assumptions are usually minimal. It is

usually used as blind as possible, assuming only the CMB anisotropies to be the derivative of a black body. Thus the products of the minimisation are the elements of the mixing matrix and the power spectra of the components. The maps of the components can be estimated by Wiener-filtering the data, using the component spectra as prior covariance. Figure 3.6 shows the SMICA CMB map together with the CMB map of the three other official Planck component separation methods.

Since SMICA works at the power spectrum level, the quantities of interest are the covariances of equation 3.1

$$R_\ell = \mathbf{A}C_\ell\mathbf{A}^T + N_\ell \quad (3.4)$$

where R_ℓ , C_ℓ and N_ℓ are the spectral covariances of the data, the components and the noise. Independence implies that C_ℓ and N_ℓ are diagonal matrices. As explained above, the aim is to fit the model covariance of equation 3.4 to the covariances of the data $\hat{R}_\ell = \frac{1}{2\ell+1} \sum_m d_{\ell m} d_{\ell m}^T$ by minimising with respect to \mathbf{A} and all C_ℓ the function

$$\phi(\mathbf{A}, \{C_\ell\}) = \sum_\ell (2\ell + 1) D(\hat{R}_\ell, R_\ell(\mathbf{A}, C_\ell)) \quad (3.5)$$

where D is a divergence between \hat{R}_ℓ and R_ℓ . In SMICA, the divergence used is the KL divergence¹ between two Gaussians of zero mean and covariance R_1 and R_2

$$D_{KL}(R_1, R_2) = \frac{1}{2} (\text{tr}(R_1 R_2^{-1}) - \log |R_1 R_2^{-1}| - N_f) . \quad (3.6)$$

Minimising the SMICA divergence is equivalent to maximising the SMICA likelihood P_{SMICA} which verifies

$$-2 \log P_{\text{SMICA}}(d | \mathbf{A}, C) = \sum_{\ell, m} \log |2\pi R_\ell| + d_{\ell m}^T R_\ell^{-1} d_{\ell m} . \quad (3.7)$$

Thus the errors on the parameters at maximum are estimated by inverting their Fisher Information matrix (see appendix B for more details on Fisher information), which is then the covariance of the Gaussian distribution that approximates the peak. Since the likelihood takes into account the presence of foregrounds, the uncertainties of the component separation process are incorporated in the errors. But the Fisher information is an average over possible data and does not depend on the actual data.

3.3.4 Sparse blind source separation

The concept of sparsity can be used to solve the blind source separation problem. A signal is said to be sparse in a basis if a small number of elements of the basis is sufficient

¹see appendix A for more details on Kullback-Leibler divergence.

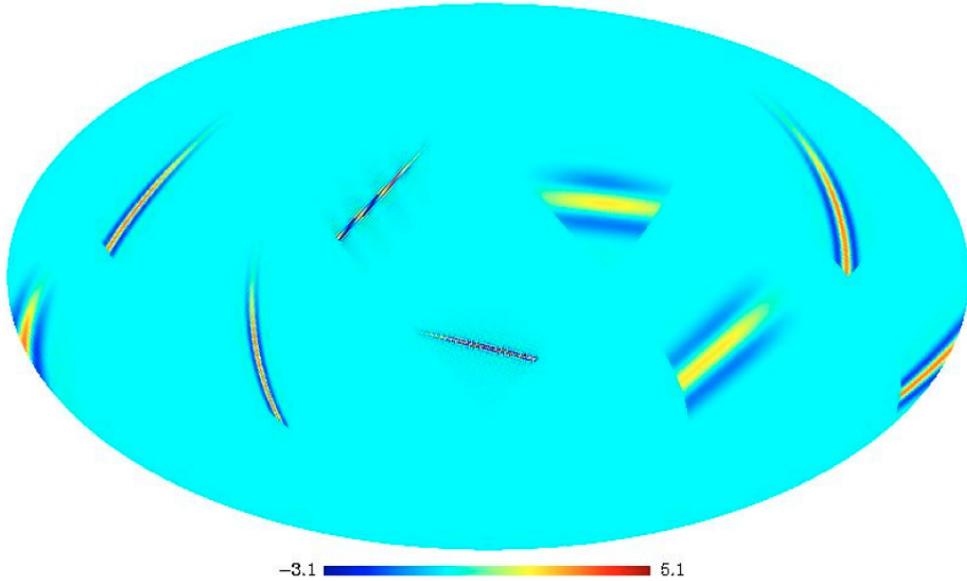


Figure 3.4: Ridgelets basis vectors in spatial space. This basis is used within sparsity methods because just one vector can model a long part of a contour. Thus, a component can be expressed with very few coefficients and is sparse in this basis.

to (almost) completely represent the signal. Thus, if all the differences between several signals are compressed in a few basis coefficients then the components are clearly distinguishable. In CMB component separation, the basis is often chosen over-complete, i.e. is often a collection of several bases, since one basis is not able to catch the diversity of all components.

Morphological Component Analysis (MCA) (Starck et al., 2013; Starck et al., 2004; Bobin et al., 2007) is a component separation method that uses the sparse representation of the components in different bases. Because of the complexity of the spatial structure of the galaxy, the bases are especially constructed to represent specific shapes, such as needlets, ridgelets (cf. figure 3.4), curvelets among others. Apart from sparsity, MCA relies on the incoherence of the bases to represent the components, i.e. a component must be sparse in one basis only. In practice, MCA minimises the distance between the data model and the data together with the coherence of the representations of the components. Figure 3.5 shows the CMB map produced by a generalisation of MCA, from WMAP and Planck data.

Mathematically, a signal d is said to be sparse in a basis \mathcal{B} if most of its coefficients α in \mathcal{B} are zero or almost zero. If \mathcal{B} is over-complete, then α is not unique. The sparsest α is found by solving the following minimisation problem

$$\min_{\alpha} \|\alpha\|_0 \quad \text{such that} \quad d = \alpha\mathcal{B}. \quad (3.8)$$

This is an NP-hard optimisation problem. This is overcome by substituting the ℓ_0 norm by

the ℓ_1 norm in equation 3.8. The minimisation is then performed much more efficiently and is still correct. If the signal d is a composite of N_c components, i.e. if

$$d = \sum_k s_k = \sum_k \alpha_k \mathcal{B}_k, \quad (3.9)$$

where \mathcal{B}_k are orthonormal basis and $\mathcal{B} = \cup_k \mathcal{B}_k$, then the full minimisation problem becomes

$$\{s_k\} = \operatorname{argmin}_{\{s_k\}} \left[\sum_k \|s_k \mathcal{B}_k^T\|_1 + \kappa \left\| d - \sum_k s_k \right\|_2^2 \right]. \quad (3.10)$$

In equation 3.10, the first term of the sum assures the sparsity and the second the consistency with the data, weighted by a factor κ . GMCA is a straightforward generalisation to mixing of components and multichannel observation and it aims to solve the following minimisation problem

$$(\mathbf{A}, \{s_k\}) = \operatorname{argmin}_{\mathbf{A}, \{s_k\}} \left[\sum_k \|s_k \mathcal{B}_k^T\|_1 + \kappa \|d - \mathbf{A}s\|_2^2 \right]. \quad (3.11)$$

GMCA is a blind method. It is difficult to write a parametric version of this method. Indeed, the components can be well phenomenologically described in various basis but a physical parameter can not model simply a component in different basis. For example, it is complicated to associate a different spectral index per pixel of a component and use the same indices to parametrise the modes of the same components.

3.3.5 Template fitting

One of the four Planck component separation methods is more specific to CMB component separation. The method, named Spectral Estimation Via Expectation-Maximisation (SEVEM) (Fernandez-Cobos et al., 2012), gives an estimation of the mixing coefficients of the components, i.e. their frequency spectrum, given a set of templates for the foreground maps. The statistical criteria that the mixing coefficient must satisfy is simple. A set of foreground templates $\{t_{k\ell m}\}$ is chosen and they are linearly combined. The linear coefficients are the mixing coefficients A_{ik} to be estimated. The sum of these templates is subtracted from the data and the coefficients are chosen such that the residuals look like CMB and noise, i.e.

$$\mathbf{A} = \operatorname{argmin}_{\mathbf{A}} \sum_{\ell m} (d_{\ell m} - \mathbf{A}t_{\ell m})^T (C_{\ell}^{cmb} e e^T + N_{\ell})^{-1} (d_{\ell m} - \mathbf{A}t_{\ell m}), \quad (3.12)$$

where \mathbf{A} has no column dedicated to the CMB and e is N_f -vector filled with ones. The CMB map is then the difference between the data and the mixing of templates.

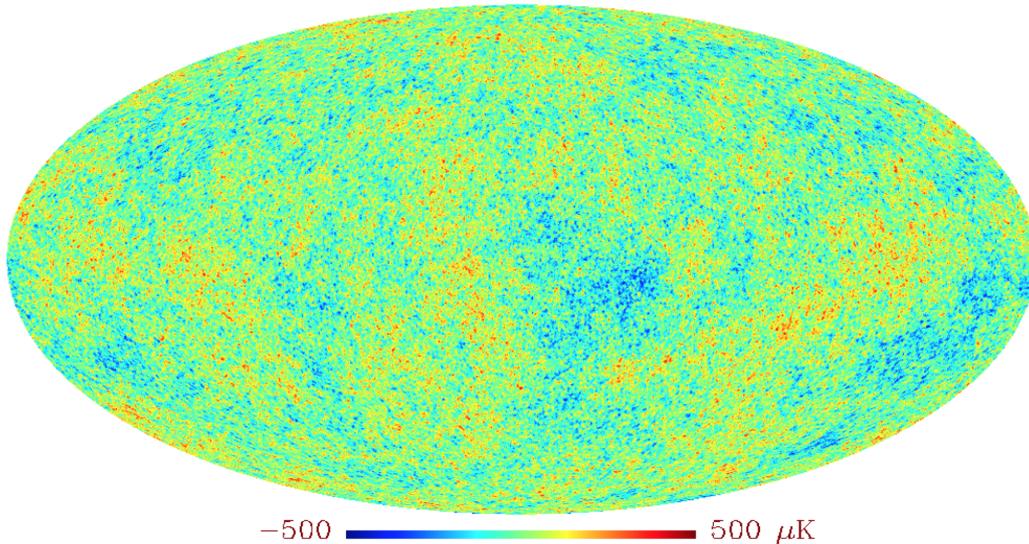


Figure 3.5: CMB map produced by Localised Generalised MCA (LGMCA) (Bobin et al., 2014), a generalisation of MCA. The analysis took both WMAP nine year data and Planck 2013 data as input, and used an external dust template to clean the galactic plane.

Figure 3.6 shows the SEVEM CMB map together with the CMB map of the three other official Planck component separation methods. A template is not necessarily a model of just one foreground, it could be a linear combination of several of them. Thus, for Planck 2013 results, the templates are constructed by taking the difference between observation maps close in frequency. Thus, since the components are not constant through frequency but coherent, the templates contain all relevant components and they are naturally consistent with the data. Also, since the CMB signal is constant through frequency, the CMB component is absent from the template to be sure to not fit for a CMB frequency spectrum.

Since the template are chosen without satisfying any statistical criteria, the choice is then arbitrary. The method is highly parametric because the whole spatial distributions of the foregrounds are supposed to be exactly known. In the case where the template were not reflecting the physical components or a mixture of them, the errors would be underestimated because the approximation would not be taken into account.

3.3.6 Physical parametrisation

The output of blind source separation methods are free of almost any assumption. They are the start of the physical analysis and they must be post-processed to give meaningful information. In contrast, physical component separation proposes to perform source separation and extraction of physical knowledge in one go. The principle is to write the

data model just like in any method but to make the mixing coefficients and the component maps and power spectra depends on physical parameters. Any estimation process is then directly applied to the physical quantities.

Commander (Eriksen et al., 2006, 2008; Stompor et al., 2009) is a method that infers the CMB map together with foreground physical parameters in a Bayesian framework (see appendix B for more details on Bayesian analysis). It is one of the four methods that were retained by the Planck collaboration for the 2013 results. Commander makes use of Markov Chain Monte Carlo (MCMC) to explore a likelihood or a posterior distribution written as function of the physical parameters. The method works at the map level. The data model is the same for every pixel but the parameters of the model can have different values at different positions on the sky. The model is a sum of components that are more or less physically parametrised. A common parametrisation of mixing coefficients arises from the assumption that some components follow power law emissions. Then all the mixing coefficients of one component depend on its spectral index only.

Let's denote by θ the set of parameters on which the mixing matrix depends. The aim of Commander is to explore the following posterior distribution (see appendix A for more details on notations)

$$P(\theta, s | d) \propto \prod_p \mathcal{N}(d_p | \mathbf{A}(\theta)s_p, N_p)P(\theta, s) \quad (3.13)$$

where $P(\theta, s)$ is the prior distribution on the parameters θ and the component maps s . Since Commander works at the pixel level, all index (ℓ, m) have been replaced by an index p . The exploration is done by sampling according to the following two-step scheme

$$\begin{aligned} s &\leftarrow P(s | \theta, d) \\ \theta &\leftarrow P(\theta | s, d). \end{aligned}$$

If the model allows spatial variation of the physical parameters then the parameter space to explore is huge. For a realistic data modelling, Commander is achievable at low resolution only and a algorithm called Ruler computes the high-resolution maps from the Commander maps. The high-resolution maps are estimated by fixing the spectral dependence of the component pixels, that then allow solving a least square system for the pixel of the high-resolution maps. Formally, the high resolution piece of component s_p is

$$s_p = \underset{s_p}{\operatorname{argmin}} (d_p - \mathbf{A}(\theta)s_p)^T N_p^{-1} (d_p - \mathbf{A}(\theta)s_p) \quad (3.14)$$

$$= (\mathbf{A}(\theta)^T N_p^{-1} \mathbf{A}(\theta))^{-1} \mathbf{A}(\theta)^T N_p^{-1} d_p. \quad (3.15)$$

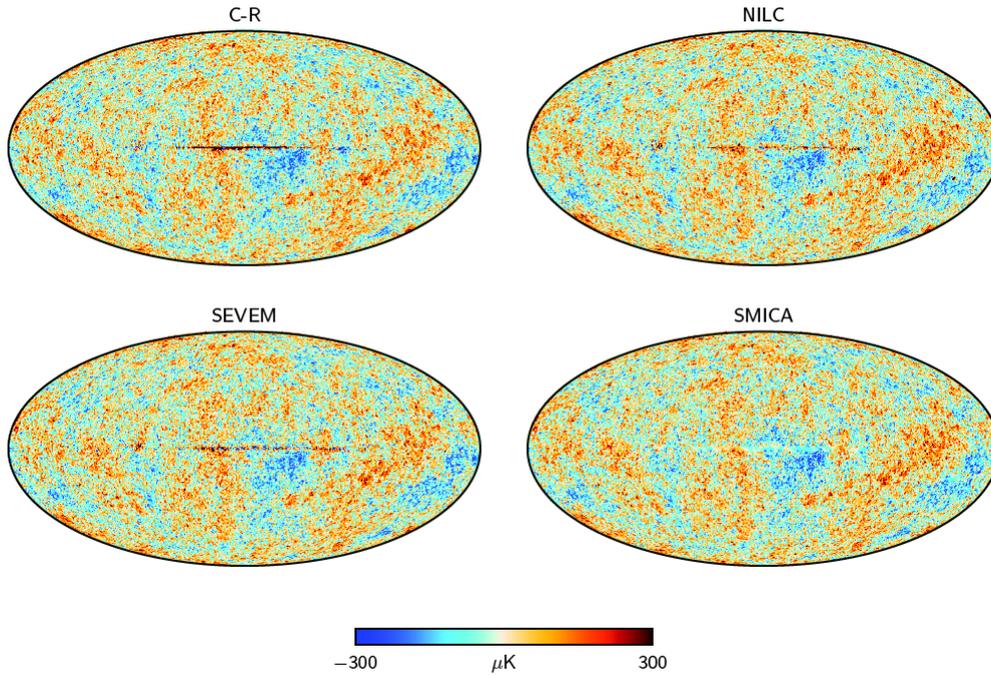


Figure 3.6: The CMB maps resulting from the four official component separation methods used by the Planck collaboration, Commander-Ruler, NILC, SEVEM and SMICA (Planck Collaboration et al., 2014c). All of them are in very good agreement for both the map and the power spectrum. All the method produce a map which is either still polluted by galactic residuals or masked and filled in.

Figure 3.6 shows the Commander-Ruler CMB map together with the CMB map of the three other official Planck component separation methods. Commander-Ruler is a parametric method. On one hand, the joint inference of physical parameters and CMB map allows a propagation of the uncertainty due to component separation to the CMB and physical models. On the other hand, the results are forced to be in accordance with the physical model, which can introduce biased estimations if the physical assumptions are wrong or approximate.

Also, the errors depend on how much the physical model is constraining. On one hand, if the data model is very specific then there is little space for error exploration. In this case the error are underestimated if the model is wrong. On an another hand, Commander can infer parameters of a very general model by allowing independent frequency spectra at every low resolution pixel. This move is motivated by the will of modelling any sky signal but the physical model might be too flexible and introduce much freedom. In this case, in addition to needlessly increase the dimension of the parameter space, the error bars are overestimated.

Conclusion

I have presented the main CMB component separation methods, especially the four methods that were retained by the Planck collaboration. The main different approaches are blind source separation and physical parametrisation, i.e. each of them brings more or less information about the components and the way they mix. Also, each method has its own error model. The Bayesian formalism used by the parametric method Commander allows an exploration of the physical parameter space and therefore provides errors that take the presence of foregrounds into account. The drawbacks of Commander are that it works at low resolution only and the errors depend on the assumed physical model. In the next chapter, I present the method I developed during my thesis. It also works within a Bayesian framework but makes use of a phenomenological model and works at high resolution. This method performs both CMB cleaning and CMB power spectrum inference that includes component separation uncertainties.

Chapter 4

BICA: a semi-blind Bayesian approach to component separation

Introduction

In this chapter I will describe the principles of Bayesian Independent Component Analysis (BICA), the blind and Bayesian formulation of CMB component separation that I have developed during my Ph.D.. The method arises from an ICA-like formulation of the problem and solves it in a Bayesian framework. ICA is suitable for blind source separation because of the general statistical assumption on the components. Bayesian analysis allows the introduction of priors with which the degree of information can be controlled.

Blind separation aim to recover the components of a mixture of signals without assuming any physical model in the analysis. They generally make assumption about their statistical properties. In contrast, physical parametrisation methods base their separation on physical models. BICA lies in between these two approaches and is then semi-blind.

BICA exploits the full power of Bayesian analysis since it evaluates the full posterior PDF rather than just some of its properties. From the data, the method provides in one go an inference of the CMB map and its power spectrum. In other words, without pre-processing of the observation maps, BICA performs CMB cleaning and CMB analysis at the same time. By exploring the joint posterior PDF on the component maps, BICA infers the component maps, their power spectra and their respective amount in the data. The uncertainty on the parameters is encoded in the shape of the posterior. Thus, since the foregrounds and the clean CMB map and its power spectrum are inferred from the same posterior PDF, the uncertainties due to foreground cleaning residuals naturally propagate to the CMB products.

BICA has a general formulation of the component separation problem. The possibility of providing prior information can specify the method. Thus, BICA can be seen as a root formulation for many different methods. It unifies the frameworks in which previous

methods have been formulated and allows an easier comparison between them on an analytic level.

The description of the method and the results of this chapter are gathered into the paper Vansyngel et al. (2014).

4.1 Constructing the component separation PDF

This section details the construction of BICA. From a simple data model, I build the likelihood and choose the priors. I will show how turning on the stochasticity of the prior parameters leads to power spectrum inference in addition to CMB cleaning.

4.1.1 The blind Bayesian formulation of the problem

The starting point is the description of the data in equation 3.1, with N_c is the assumed number of components and N_f the number of frequency channels. The model in this equation is said to be blind because it makes use of a phenomenological description of the data. No physics is involved in the model since the observation maps d_i are just linear combinations of some component maps, which are simply described by the collection of their harmonic coefficients. Therefore a blind method aims at recovering the spatial distribution of the components, and the amount of each component in each frequency band. The parameters of interest to be sampled are the N_c component maps s_k and the mixing matrix \mathbf{A} . The first step of the method is to write Bayes' theorem for these parameters. Following the basics of Bayesian analysis described in appendix B, the probabilistic equation that relates d , \mathbf{A} and s is

$$P(\mathbf{A}, s | d) \propto \mathcal{L}(d | \mathbf{A}, s) P(\mathbf{A}, s) . \quad (4.1)$$

Although in this work component separation is addressed with a blind analysis, the flexibility of the Bayesian formalism allows the introduction of a physical parametrisation of the problem. The current understanding of physical galactic and extra-galactic phenomena can be progressively introduced by a more detailed data model and by assigning a parametrized prior PDF on the foregrounds. The CMB power spectrum is also highly parametrizable since its shape depends on a small number of cosmological parameters (Planck Collaboration et al., 2014e). Thus a joint inference of cleaned CMB map, CMB power spectrum and cosmological parameters is conceivable, thanks to fast and accurate Boltzmann code emulators like PICO (Fendt and Wandelt, 2007). A less parametric approach would be to exploit the smoothness of the power spectra through binning or representation in terms of smooth basis function, such as splines.

Also, as in other component separation methods, the Bayesian formulation presented in this thesis can be extended to infer the CMB polarisation power spectrum.

In the following, I will keep the blind Bayesian approach. In the next subsections, I derive an expression for the likelihood, choose priors for the parameters and enlarge the scope of the analysis.

4.1.2 Likelihood distribution

The form of the likelihood depends on the data model and the instrument properties. Suppose that the noise n in 3.1 is Gaussian with zero mean, anisotropic and independent from one spherical harmonic coefficient to another and from one frequency band to another. The noise is therefore entirely determined by a diagonal covariance N . This is a good approximation for the WMAP and Planck data (Planck Collaboration et al., 2013; Jarosik et al., 2011). If \mathbf{A} and s are fixed, the stochasticity of d relies on the noise only. Therefore the likelihood in accordance with the data modelling in equation 3.1 and Gaussian noise is:

$$\begin{aligned} \mathcal{L}(d | \mathbf{A}, s) &= \prod_{\ell, m} \mathcal{N}(d_{\ell m} | \mathbf{A}s_{\ell m}, N_{\ell}) \\ &\propto \exp -\frac{1}{2} \sum_{\ell, m} (d_{\ell m} - \mathbf{A}s_{\ell m})^T N_{\ell}^{-1} (d_{\ell m} - \mathbf{A}s_{\ell m}) . \end{aligned} \quad (4.2)$$

4.1.3 Prior distributions

The prior does not depend on the data, it is a PDF that is chosen prior to the data acquisition. Multiplying the likelihood by a prior PDF shapes the posterior PDF. Thus, the prior PDF brings a priori information into the inference of the parameters and must be chosen with care. The spatial distribution of the component should not depend on the way the component mix together. Thus, the prior I put on the mixing matrix and the component maps are independent, i.e. $P(\mathbf{A}, s) = P(\mathbf{A})P(s)$.

Prior on the component maps The simplest inflationary theories predict that CMB fluctuations are very nearly a Gaussian random field (e.g. Mukhanov (2013)). Therefore, choosing a Gaussian distribution as a prior for the CMB is physically well motivated. The generative model of the other components violates Gaussianity. However, a Gaussian distribution is, in a sense, the least informative prior for the component maps. Indeed, if one constrains a random variable to be described by its variance σ^2 only, the Gaussian distribution whose PDF has zero mean and variance σ^2 is the one that maximises among all distributions the Shannon entropy of that random variable. Thus, if the components are constrained to be described by their power spectra C only, the least informative prior

is:

$$P(s | C) = \prod_{\ell, m} \mathcal{N}(s_{\ell m} | 0, C_{\ell}) \propto \prod_{\ell, m} \frac{1}{\sqrt{|C_{\ell}|}} \exp -\frac{1}{2} (s_{\ell m} - \bar{s}_{\ell m})^T C_{\ell}^{-1} (s_{\ell m} - \bar{s}_{\ell m}) . \quad (4.3)$$

The maps $\{\bar{s}_{\ell m}\}$ represent a step forward in the physical modelling of the component maps. Indeed, choosing non-zero mean, for example templates of galactic emissions, amounts to include physical information. Concerning the C_{ℓ} 's, they are chosen diagonal. In other words, the model forces the Gaussian components to be independent. Later implementations, including the one used for the application of BICA to the Planck data in chapter 6, include the cross-spectra of the non-CMB components in the inference. Also, since the prior is the same for all modes m within the same multipole ℓ , the components are assumed to be statistically isotropic. This is a good assumption for the CMB, but it is an approximation for the galactic components since the power is concentrated in the galactic plane. The Gaussian prior is suitable for inclusion of physical information. For example, I could choose a flat power spectrum (i.e. all multipoles are restrained to the same value) in order to model a point source like component. This is investigated in chapter 6 when dealing with diffuse point sources.

Prior on the mixing matrix Some physical information on the mixing matrix is input in the inference *via* its prior PDF because it is necessary to recover the CMB map. Since no CMB spectral distortion has been detected yet, I suppose that the CMB anisotropies have the emission law of a derivative with respect to temperature of a black body, at CMB temperature. In thermodynamic units, the response is then constant through frequencies. Thus the prior distribution for the column of the mixing matrix that is related to the CMB is a Dirac delta function. These parameters are therefore perfectly known and not stochastic.

On the remaining elements, I put a prior from the class of priors called Jeffreys prior¹. These priors are said to be non-informative because they conserve the Fisher information under a change of variable. Driven by the will of having an as blind as possible method, a Jeffreys prior is put on the mixing matrix. As the likelihood is Gaussian in the coefficients of \mathbf{A} , its Jeffreys prior is a flat distribution:

$$P(\mathbf{A}) \propto 1 . \quad (4.4)$$

4.1.4 Hierarchical model and power spectrum inference

The data can be further described in a hierarchical model, in which the prior parameter C of equation 4.3 becomes a *hyperparameter* of the model. The additional parameter is then

¹see appendix B for more details on Jeffreys priors.

stochastic and must also be sampled. This move has two motivations. First it introduces more flexibility in the modelling of the components since all shapes of power spectrum are potentially investigated. Second and foremost, the posterior on the covariance C provides an inference of the component power spectra. C must be included in the parameters of the posterior, and one must chose a prior for it. Equation 4.1 then reads:

$$P(\mathbf{A}, s, C | d) \propto \mathcal{L}(d | \mathbf{A}, s) P(\mathbf{A}) P(s | C) P(C). \quad (4.5)$$

Prior on the power spectra For the same reasons than for the mixing matrix, I choose a Jeffreys prior for C . The Jeffreys prior for a diagonal covariance Σ is proportional to $1/|\Sigma|$. Hence:

$$P(C) \propto \prod_{\ell} 1/|C_{\ell}| = \prod_{\ell, k} 1/C_{\ell k}. \quad (4.6)$$

4.1.5 Posterior distribution

Injecting equations 4.2, 4.3, 4.4 and 4.6 into equation 4.5 leads to the explicit expression of the posterior:

$$\begin{aligned} P(\mathbf{A}, s, C | d) \propto & \prod_{\ell, m} \exp -\frac{1}{2} (d_{\ell m} - \mathbf{A} s_{\ell m})^T N_{\ell}^{-1} (d_{\ell m} - \mathbf{A} s_{\ell m}) \\ & \prod_{\ell, m} |C_{\ell}|^{-\frac{1}{2}} \exp -\frac{1}{2} (s_{\ell m} - \bar{s}_{\ell m})^T C_{\ell}^{-1} (s_{\ell m} - \bar{s}_{\ell m}) \\ & \prod_l |C_{\ell}|^{-1}. \end{aligned} \quad (4.7)$$

The results of chapters 5 and 6 do not include a physical model for the spatial distribution of the component, i.e. $\bar{s} = 0$.

Using sampling techniques described in appendix C, samples from the posterior will provide a PDF of the component maps, their power spectra and the mixing matrix. Next section discusses approaches to sample from the posterior.

4.2 Deriving the sampling equations

Since the set $\{\mathbf{A}, s, C\}$ has a huge number of parameters (of the order of 10^7), the brute force grid evaluation of the posterior is out of the question. Also, no algorithm exists to directly sample all the parameters at once from the posterior. Indirect sampling is needed in order to evaluate the posterior (see appendix C for an overview of sampling techniques).

4.2.1 First attempt

The posterior has a complex dependence on the whole vector $\{\mathbf{A}, s, C\}$. If one partitions $\{\mathbf{A}, s, C\}$ in the subsets mixing matrix \mathbf{A} , component maps s and power spectra at each multipole ℓ C_ℓ , then the conditionals $P(X | \{\mathbf{A}, s, C\} \setminus X, d)$, $X \in \{\mathbf{A}, s, C_{\ell_{\min}}, \dots, C_{\ell_{\max}}\}$, are easy to sample from. Indeed, by looking at equation 4.7, one can see that the conditional for \mathbf{A} and the conditional for s are Gaussian distributions, and the conditional for each C_ℓ is an inverse-Wishart distribution that does not depend on the value of C at multipoles different from ℓ . Thus, to construct a chain of I samples from the posterior, a straightforward Gibbs sampler would be:

Sampler 1

1. begin with some initial value of \mathbf{A} , s and C , $A^{(0)}$, $s^{(0)}$ and $C^{(0)}$,
2. for $i \in \llbracket 1, I \rrbracket$,
 - 2.1 sample $A^{(i)}$ from the Gaussian distribution $P(A | s^{(i-1)}, C^{(i-1)}, d)$,
 - 2.2 sample $s^{(i)}$ from the Gaussian distribution $P(s | A^{(i)}, C^{(i-1)}, d)$,
 - 2.3 for $\ell \in \llbracket \ell_{\min}, \ell_{\max} \rrbracket$,
sample $C_\ell^{(i)}$ from the inverse-Wishart distribution $P(C_\ell | A^{(i)}, s^{(i)}, d)$,
3. a histogram of the chain $(A^{(i)} \cup s^{(i)} \cup C^{(i)}; i \in \llbracket 1, I \rrbracket)$ provides an estimate of the full posterior $P(\mathbf{A}, s, C | d)$.

However, this sampling scheme has a major drawback. Since there is a large number of spherical harmonics coefficients, the dimensionalities of \mathbf{A} ($\mathcal{O}(1-10)$) and C ($\mathcal{O}(10^3-10^4)$) are much smaller than that of s ($\mathcal{O}(10^7)$ for high resolution data). Thus fixing s while sampling the other parameters confines the sampled parameters to a tiny region of the overall posterior volume. At each drawing, the parameters move very little away from their current state, which leads to a slowly converging sampler and highly correlated successive samples. Besides, the sampling of the maps is much more time-consuming than the sampling of the mixing matrix or of the power spectra.

The next two subsections present an alternative and more appropriate sampler to solve the problem.

4.2.2 Marginalisation

One solution to speed up the sampling is to marginalise over as many parameters as possible. Marginalisation is an operation on a PDF that allows to discard a subset of parameters while taking into account all its possible values. For example, consider a

PDF of two variables x and y , $P(x, y)$. Then the *marginal distribution* of x is obtained by integrating the PDF over y :

$$P(x) = \int dy P(x, y), \quad (4.8)$$

and the following relation holds:

$$P(x, y) = P(y | x)P(x). \quad (4.9)$$

Thus, being able to sample x from $P(x)$ then y from $P(y | x)$ amounts to sample $\{x, y\}$ from $P(x, y)$.

Integrating out the component maps s considerably reduces the number of parameters to be sampled as well as the correlation length in the chain. After marginalising over the component maps, the dependence of the marginal posterior on the remaining parameters is more complex than in the full posterior:

$$\begin{aligned} P(\mathbf{A}, C | d) &= \int ds P(\mathbf{A}, s, C | d) \\ &\propto \prod_{\ell} |N_{\ell} + \mathbf{A}C_{\ell}\mathbf{A}^T|^{-\frac{2\ell+1}{2}} \exp -\frac{1}{2} \text{tr} \left\{ (N_{\ell} + \mathbf{A}C_{\ell}\mathbf{A}^T)^{-1} V_{\ell} \right\} \\ &\quad \prod_{\ell} |C_{\ell}|^{-1}, \end{aligned} \quad (4.10)$$

where $V_{\ell} = \sum_m d_{\ell m} d_{\ell m}^T$, and

$$P(\mathbf{A}, s, C | d) = P(s | \mathbf{A}, C, d)P(\mathbf{A}, C | d). \quad (4.11)$$

4.2.3 Sampling scheme

The marginal posterior can not be sampled from directly and the set of parameters $\{\mathbf{A}, C\}$ can not be partitioned in order to obtain simple conditional distributions. Besides, $P(s | \mathbf{A}, C, d)$ is Gaussian. From these considerations arises a sampling scheme involving a Metropolis-Hastings-Gibbs hybrid sampler. This sampling scheme for the complete set of parameter $\{\mathbf{A}, s, C\}$ is:

Sampler 2

1. begin with initial values of \mathbf{A} , C and s , $A^{(0)}$, $C^{(0)}$ and $s^{(0)}$,
2. build a chain $(\{\mathbf{A}, C\}^{(i)}, i \in \llbracket 1, I \rrbracket)$. For $i \in \llbracket 1, I \rrbracket$,
 - 2.1 propose a candidate \mathbf{A}^* and accept/reject it according to $P(\mathbf{A}, C | d)$,

- 2.2 for $\ell \in \llbracket \ell_{\min}, \ell_{\max} \rrbracket$,
propose a candidate C_ℓ^* and accept/reject it according to $P(\mathbf{A}, C | d)$,
3. build a chain $(s^{(i)}; i \in \llbracket 1, I \rrbracket)$. For $i \in \llbracket 1, I \rrbracket$,
sample $s^{(i)}$ from the Gaussian distribution $P(s | \{\mathbf{A}, C\}^{(i)}, d)$,
4. the chain $(\{\mathbf{A}, s, C\}^{(i)}; i \in \llbracket 1, I \rrbracket)$ allows us to calculate any moments of the full posterior $P(\mathbf{A}, s, C | d)$ that we wish to examine.

The sampling of s being much slower than that of \mathbf{A} and C , in practice the sampling of the maps makes use of a subset of the chain $(\{\mathbf{A}, C\}^{(i)}, i \in \llbracket 1, I \rrbracket)$ only.

4.3 Comparison to previous methods

As mentioned in section 3.3, various approaches and methods for CMB estimation have been developed prior to the method presented in this thesis. In this section I stress the particularities of BICA and I link it to the methods that have been used for the Planck 2013 release.

4.3.1 Relevance of the method

BICA provides a joint inference of the component maps and of their power spectra, together with the mixing matrix. Thus, the errors due to multiple components in the data are encoded in the shape of the posterior, and uncertainties due to component separation are fully explored.

The number of physical components in the data is larger than the number of observed frequencies. The advantage of BICA (and of some other methods like SMICA, see next subsection) is that blindness implies that the assumed components are effective components that may combine several physical components. Thus, BICA aim at recovering the dominant modes of the foregrounds, which contain the largest part of the signal.

Apart from the assumption that the CMB has a black body emission law, which has been shown to be the case with high accuracy, the data model is a phenomenological model. Thus, errors do not rely on any strong physical assumptions. If a physical model were to be used to parametrise the data, the PDF should be exploited with care because it then would be valid within that particular model only.

Phenomenological models prevent biases that could arise from assuming a constraining physical context. However, if the data is described via physical laws instead, the physical quantities are the parameters of the posterior PDF. In this case, the method provides a direct inference of the parameters of the theory. In BICA there is the possibility of introducing progressively a parametric model. For instance, instead of describing the

CMB power spectrum by its value at each multipole, one can sample the few parameters of the standard model from which the whole shape of the CMB power spectrum can be derived. Then the uncertainties of the separation step propagate to the inference of the cosmological parameters. The foregrounds also could be constrained by priors. For example the mixing matrix can be parametrised with a few spectral indices or include a prior mean and use the Gaussian to parametrise the deviation from the mean. Since any value can be chosen for such prior parameters, the information can be controlled and tuned as wanted.

4.3.2 Comparison to SMICA

Spectral Matching Independent Component Analysis (SMICA) (Cardoso et al., 2008) is a method that shares several similar aspects with BICA. The parametrisation is the same, both methods are blind. The parameters are estimated by reducing the divergence between two Gaussian distributions, one whose covariance is constructed from the model, one whose covariance is the empirical data covariance. In the model, the cross power spectra between the CMB and the foregrounds are null in order to force uncorrelation between the component to be separated. Depending on the binning of the power spectra, and following the original formulation of SMICA, the method is equivalent to a maximisation with respect to \mathbf{A} and C of the likelihood $P(d | \mathbf{A}, C)$ of equation 3.7.

We can understand SMICA analysis from a Bayesian perspective as follows:

1. Begin with the Bayesian formulation, equation 4.5,
2. Construct the same likelihood as that of BICA and choose Gaussian priors for the component maps, flat priors for the mixing matrix and the power spectra,
3. Marginalise the posterior over all component maps,
4. Maximise the obtained marginal distribution with respect to \mathbf{A}, C .

As in BICA, data are post-processed with the obtained value of \mathbf{A}, C and N to produce the maps of the components.

In BICA formalism, the maximisation with respect to $\{\mathbf{A}, C\}$ of the marginal PDF $P(\mathbf{A}, C | d)$ in equation 4.10 is equivalent to a SMICA estimation of the parameters. Thus instead of just finding the peak of the distribution like SMICA does, the Bayesian sampler explores the whole distribution over \mathbf{A} and C . The shape of the distribution and in particular the spread around the peak is better appreciated.

4.3.3 Comparison to Commander

As BICA, Commander (Eriksen et al., 2006, 2008) is a Bayesian formulation of the joint component separation and CMB power spectrum inference problem. The main difference between the two approaches is the parametrisation of the problem.

Commander makes intensive use of parametric models to describe the physical emissions while BICA adopts a phenomenological description of the different components. Thus Commander infers component maps and power spectra within a constraining physical model, and therefore the most probable values of the parameters and their errors depend on this model. The results of BICA do not rely on any physical assumption, except for the constant response of the CMB signal across frequencies. In Commander, since a lot of information is given to recover the individual components, the output components are identified as physical emissions. In BICA, apart from the CMB, no component has a physical meaning but the non-trivial difference is that our results do not depend on any complex physical model.

In addition, Commander works at the map level whereas BICA works mainly at the power spectrum level. Thus, at equal resolution, BICA is faster.

4.3.4 Comparison to ILC

The Internal Linear Combination (ILC) method (Bennett et al., 1992; Tegmark, 1997; Delabrouille et al., 2009) provides a map of a component, given its frequency spectrum. Therefore, except for the CMB, it is not a blind method. Recall the ILC the estimator:

$$\forall(\ell, m), \quad \hat{s}_{\text{CMB},\ell m} = \frac{e^T C_d^{-1}}{e^T C_d^{-1} e} d_{\ell m}, \quad (4.12)$$

where C_d is an estimate of the true data covariance:

$$C_{d_{\ell m}} = C_{\text{CMB},\ell} e e^T + \mathbf{A}_{\text{true}} C_{\text{true},\ell m} \mathbf{A}_{\text{true}}^T + N_{\ell m}. \quad (4.13)$$

Here the mixing matrix \mathbf{A} has no column dedicated to the CMB, $C_{\text{true},\ell m}$ is the true underlying covariance of all components but the CMB at multipole numbers (ℓ, m) and e a vector filled with 1's, i.e. the frequency response of the CMB.

From the Bayesian perspective, one can understand ILC analysis as follows:

1. Begin with the Bayesian formulation, equation 4.5,
2. Construct the same likelihood as that of BICA and choose Gaussian priors for the component maps, flat priors for the mixing matrix and the power spectra,
3. Marginalise the posterior over all component maps but the CMB map,

4. Maximise the obtained marginal distribution with respect to the CMB map.

The solution is:

$$\hat{s}_{\text{CMB},\ell m} = \frac{e^T (N_\ell + \mathbf{A}C_\ell\mathbf{A}^T)^{-1}}{C_{\text{CMB},\ell}^{-1} + e^T (N_\ell + \mathbf{A}C_\ell\mathbf{A}^T)^{-1} e} d_{\ell m}. \quad (4.14)$$

If the prior variance of the CMB is infinite (i.e. flat prior) then the two approaches (equations 4.12 and equation 4.14) are equivalent². By comparing the usual derivation of ILC and its Bayesian interpretation, we can see that the true mixing matrix and true power spectra are approximated by the data covariance. That is why, despite its very simple formulation, the ILC method gives satisfactory results.

The ILC method avoids degeneracies since it produces a single CMB map. It has a unique solution because for a fixed mixing matrix and component covariance there is only one peak in the component subspace. In the Bayesian analysis, since the components are not individually identifiable, there is a degeneracy in the mixing matrix subspace. Thus, the BICA posterior distribution is multimodal.

4.3.5 Comparison to SEVEM

Spectral Estimation Via Expectation-Maximisation (SEVEM) (Fernandez-Cobos et al., 2012) is a template fitting method. This method does the opposite of what ILC does. Providing a set of template for the component maps, SEVEM finds the electromagnetic spectra of the templates that fit the data the best. In the Bayesian formulation, templates can be seen as infinitely narrow priors for the component maps. Thus, in a similar fashion than for the other methods SEVEM can be understood from a Bayesian perspective as follows

1. begin with the Bayesian formulation, equation 4.5,
2. choose a Gaussian prior for the mixing matrix, a Gaussian prior for the CMB map and delta functions centred around the templates for the foregrounds,
3. marginalise over all the component maps,
4. maximise the marginal with respect to the mixing matrix,
5. recover the CMB map as the difference between the data and the mixing of the templates.

Step 4 is equivalent to a chi-squared minimisation like it is usually carried in template fitting.

²see appendix F for more details.

In SEVEM, the model for mixing matrix is the same than that of BICA but the model for the component maps is much more constraining. The differences between the templates and the actual components are not evaluated and end up as residuals in the CMB map. Since the foreground model is supposed to be perfect, only the instrumental noise is propagated to the CMB map and the errors are therefore underestimated. A compromise between the strong constraints of SEVEM and the very general blind formulation of BICA would be to use the foregrounds as mean of the Gaussian priors for the component maps in BICA.

Conclusion

I have presented a new formulation for the CMB foreground cleaning. In my analysis, I avoid physical assumptions, except that the CMB behaves like a black body and I model the components as Gaussian random fields. The CMB is then cleaned by jointly inferring CMB, galactic residuals, and point source power spectra and frequency spectra. This Bayesian method provides an evaluation of a posterior PDF for the CMB power spectrum which thus takes into account uncertainties due to the removal of foreground contamination. I also showed that previous component separation methods can be derived as special cases of our Bayesian formulation, which thus provides a unified approach for semi-blind foreground cleaning from multi-frequency CMB data. In the next chapter, I present an application of the method to a set of simulated data.

Chapter 5

Application to simulations

Introduction

In order to test the BICA method presented in the previous chapter, I applied it on a set of simulated data. The simulations of the observation maps contain a minimal foreground model. This simple data set approaches the model assumed by BICA and is not meant to faithfully represent the real sky. The aim here is to prove the efficiency of the method in an optimistic case. An adaptation of BICA for more complex data and its application to real data can be found in the next chapter.

In this chapter I will first describe the simulations, then qualitatively compare them to the data model of chapter 4 and explain in which respect they are appropriate for the test. In the previous chapter I have presented in section 4.2 two samplers that are supposed to solve the problem. I present the results of both approaches. Sampler 1 does not provides very conclusive results because of convergence issues but the errors on the reconstruction still has some interesting properties. Concerning the results of Sampler 2, I will spend more time describing them since they are very satisfactory. They consist of a joint blind inference of components maps and power spectra, in particular the CMB. They are then confronted to a consistency check.

The description of the method and the results of this chapter are gathered into the paper Vansyngel et al. (2014).

5.1 Description of the simulations

The primary goal of the method is to clean the CMB from foreground contaminations. It therefore must be expressed in a basis where the CMB anisotropies are modelled as accurately as possible. The simplest basis to fully describe the CMB is the set of the spherical harmonics. That is why the method has been built to work in the harmonic space and that the multi-frequency data is described in this basis.

In the case of real data, noise is anisotropic and pixels are correlated because of the noise. Thus, in principle, it would be better to work in pixel space. But then a full covariance would be needed in order to model the CMB, and this is not achievable. Since the signal-to-noise ratio is high and the deviation from isotropy and the correlations are rather low, I work in harmonic space where diagonal covariances are sufficient to almost perfectly model the CMB and the noise. Also, since the covariances of the component prior are also inferred, applying the method in harmonic space provides a CMB power spectrum inference within the component separation step.

Without loss of generality, I chose to use $N_f = 4$ observation frequencies of a mixing of $N_c = 3$ components, including the CMB. In the following, I use a simulation of the first four HFI frequencies, i.e. the 100GHz, 143GHz, 217GHz and 353GHz channels. For each frequency i , the coefficients of the simulated data d_{ilm} is constructed following the data model of equation 3.1, i.e. the set of simulations is a noisy composite of CMB and two galactic emissions.

Figure 5.1 shows the four simulated observation maps. The plot shows the CMB power spectrum in black and the power spectra of all the foregrounds at each frequency channel in color, from red to purple being from 100GHz to 353GHz.

5.1.1 The components

Beside the CMB radiation, the two components chosen for the analysis are the thermal dust emission and the free-free emission. This is motivated by the fact that they are the two most intense galactic components above 80GHz. In addition, one has a power law emission with a negative spectral index (free-free) and one a power law emission with a positive spectral index (dust) such that foregrounds are dominant at high and low frequencies in the simulated data and alike real data. The data model in equation 3.1 implies that the components are coherent through frequencies. Therefore one map of each component is simulated.

The power spectrum of the CMB is simulated using the CAMB software (Lewis et al., 2000) in a standard cosmology model. CAMB is a Boltzmann emulating code that takes cosmological parameters in input and computes the expected CMB power spectrum based on perturbation theory. From the resulting power spectrum, a CMB map is simulated using the HEALPix software¹ (see appendix E for more details on HEALPix) (Górski et al., 2005).

The spatial distributions of the thermal dust and free-free components are simulated using the publicly available version of the Planck Sky Model (PSM) (Delabrouille et al., 2012). The free-free map from the PSM has an electron temperature of 7000K, a power law with spectral index close to -0.15 and is a composite of maps from Dickinson et al.

¹code available at <http://healpix.sourceforge.net>

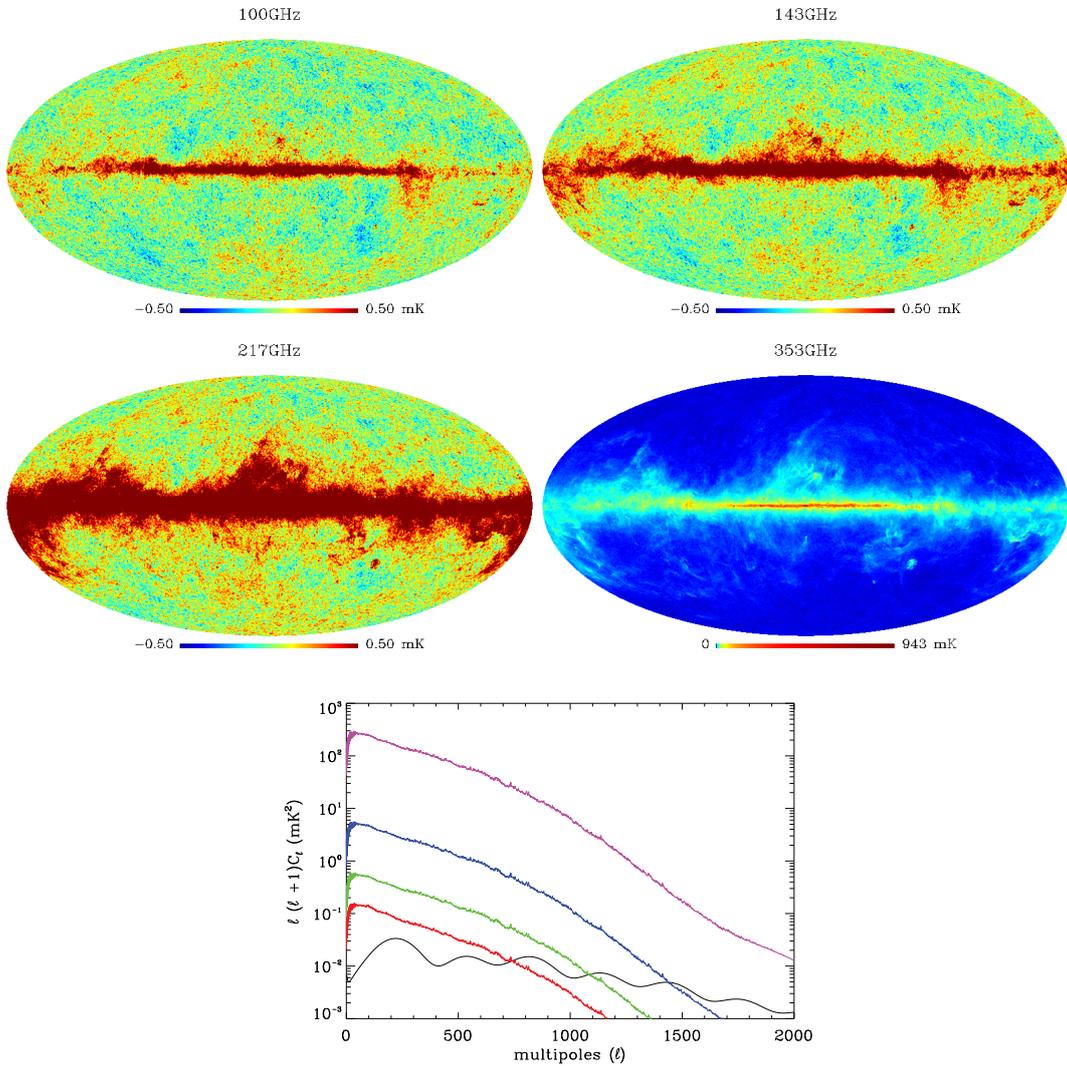


Figure 5.1: The simulated data maps at four of the Planck HFI frequencies, from 100GHz to 353GHz, using realistic spatial distributions of free-free and thermal dust emissions from the PSM. In these simulated data maps the templates of the component maps are scaled through frequency according to the mixing matrix. We chose to work with this set of channels because the CMB is the least contaminated by foregrounds and noise in this frequency range. The plot shows the power spectrum of the CMB (black line) and the level of foregrounds at each frequency channel in color (red to purple is 100GHz to 353GHz).

(2003) and from the WMAP MEM map (Bennett et al., 2013). The thermal dust map from the PSM is simulated from the Schlegel-Finkbeiner-Davies map, to which the ultra-compact HII regions are subtracted. More details can be found in Delabrouille et al. (2012).

5.1.2 The mixing matrix

The mixing matrix scales each component map according to its frequency spectrum. It is a $N_f \times N_c$ matrix where each column represents the frequency behaviour of each component. As stated above, the CMB is assumed to have constant response through the range of observed frequencies. In each of the columns dedicated to the non-CMB components, one entry is fixed in order to break the scale degeneracy between the intensity of a component map and its amount present in each channel. The simulated mixing matrix obeys these assumptions. The remaining elements of the mixing matrix to be inferred are chosen such that they agree with the PSM. A typical mixing matrix is

$$A_{\text{true}} = \begin{pmatrix} 1 & 1 & t \\ 1 & a & u \\ 1 & b & v \\ 1 & c & 1 \end{pmatrix}.$$

Here, the lowest frequency element is fixed for the free-free component and conversely for dust. The fixed mixing coefficient of free-free and dust are fixed at the frequency channel where they are the brightest and therefore where data are supposed to have the most of information about them. As a consequence, all elements a, b, c, t, u, v take their value between 0 and 1, such that free-free and dust emissions are at their maximum of intensity at respectively 100GHz and 353GHz in the simulations.

5.1.3 The noise

The noise is simulated at the pixel level and is uncorrelated pixel to pixel. The noise standard deviation maps are designed to be consistent with the scanning strategy chosen for the Planck spacecraft. Therefore it is anisotropic. In the harmonic domain the noise is characterized by one white power spectrum per frequency, derived from the noise standard deviation maps. Our inference approximates the noise as isotropic. The impact of this approximation will be assessed in section 5.4.

5.1.4 The data

The component spatial maps and the noise maps are simulated and combined at the pixel level. Each observation map is then transformed into a collection of spherical harmonic

coefficients.

5.2 Model approximations to the simulations

The data is not simulated exactly according to the model upon which BICA is based on. There are three features in the simulations that are not supported by the model: cross correlation between two different components, non-Gaussianity of the components and anisotropic noise.

5.2.1 Isotropic noise

A power spectrum completely describes a Gaussian and isotropic random field on the sphere. The contrapositive is also true: a non-Gaussian or non-isotropic random field is not completely describes by its power spectrum. The simulated noise is a Gaussian process but it is anisotropic. Since BICA is working at the power spectrum level, only a power spectrum of the noise is used. All harmonic coefficients $a_{\ell m}$ of an anisotropic random field have different variance $C_{\ell m}$. An estimate of the noise power spectrum is then $C_\ell = \frac{1}{2\ell+1} \sum_m C_{\ell m}$. It is a good approximation when at each ℓ the dispersion of noise power in the $C_{\ell m}$ for $m = -\ell, \dots, \ell$ is small. In addition, at large scales, where the noise is the most anisotropic, the signal of the components is dominant by several order of magnitude in the data. Thus error in estimating the noise is insignificant in the total power.

5.2.2 Lack of correlation between components

The model for the components is a collection of random Gaussian fields that are uncorrelated between them. It is an approximation because some realistic components are spatially correlated. For example, the galactic components are all very bright in the galactic plane and get fainter as latitude increases or decreases. The cross correlation between the maps gives a spectrum that is far from being null, i.e. the off-diagonal terms of the covariances C_ℓ are in principle non negligible, at least on large scales. In the simple formulation presented in chapter 4 BICA will find the uncorrelated components that fit the data the best. The resulting components are then non-physical emissions. Concerning the CMB anisotropies, no physical processes link them to the galactic emissions. Thus, their spatial distribution is uncorrelated with the galactic component maps. Forcing uncorrelation between the CMB map and the foregrounds is physically well motivated, which leads to a recovery of the true CMB map. This is sufficient if we do not aim at reconstructing physical foreground maps.

5.2.3 Gaussianity

Apart from the CMB, the underlying distribution of the component maps is clearly non-Gaussian. However, in the *a priori* model, they are represented by isotropic random Gaussian fields with zero mean. As we will see in the next chapter, it is too much of an approximation to reconstruct the whole sky component maps. Indeed, in the galactic plane, where the non-CMB components violate Gaussianity the most, the Gaussian prior does not fit the component signal. However, at higher latitudes, the components are less bright and look more Gaussian and isotropic than in the full sky. The Gaussian prior is then adapted to the sampling problem. One way to reduce non-Gaussianity would be to include non-zero prior means for the components. In this case, the Gaussian prior model the fluctuation around that mean and not the entire spatial distributions of the components. This is still to be investigated.

5.3 Results

This section presents the results of the BICA method, when it is applied to the simulations presented in the previous section. Since the semi-blind property of the method prevents from distinguishing between the non-CMB components, this section presents first the results concerning the CMB followed by the results concerning all the other components.

5.3.1 Full Gibbs sampling treatment

Sampler 1 has several advantages. First, the chain moves at each step. Indeed Gibbs sampling can be derived from Metropolis-Hastings by choosing the conditional PDF as the proposal PDF and by construction the acceptance ratio is always 1. Second, there is no need to test different proposal PDF and of adapting the proposal along the chain. Third, there is no need of post-processing the samples, the whole set of parameters is sampled in one chain. Fourth, which comes with second and third, Sampler 1 is quick and is easy to implement.

Sampler 1 has also two major drawbacks. First, the successive samples in the chain are very correlated to one another. If the correlation between two samples drops quickly with their separation in the chain, the chain is said to be mixing well and uncorrelated samples from the distribution are quickly drawn. Because the dimensionality of the problem, as explained in section 4.2.1, the mixing matrix and power spectra part are moving very little each step of the chain, i.e. the chain mixes poorly. The mixing matrix correlation function of figure 5.2 illustrates this sampling issue. Second, the sampling of one sample of high resolution component maps takes about 2 seconds on one CPU, which prevents from constructing the long chains needed to overcome the high correlation length. At

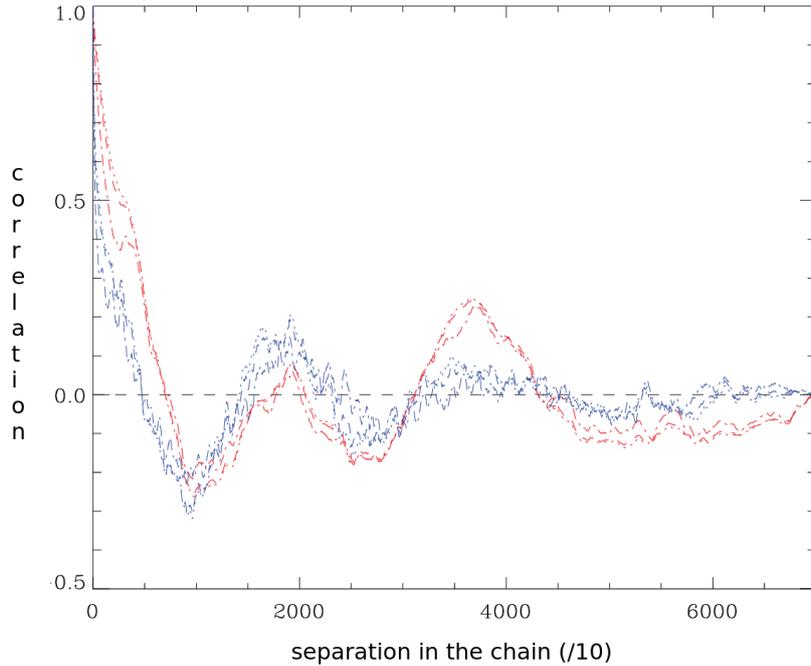


Figure 5.2: Correlation function of the mixing matrix chain. The six lines correspond to the six free parameters of the mixing matrix. Each colour blue and red correspond to a column of the mixing matrix. The sampling of the maps being computationally demanding, the sampling of this chain took one week. Thus, this figure shows that the typical time scale between two uncorrelated samples of mixing matrix is of the order of one day.

each step the thousands of covariances have to be recomputed and the millions of pixels have to be sampled. thus, if the chain does not start close to an area where the posterior has support, the parameters will never get to their most probable values. It is a heavy burn-in problem. Despite all the advantages, these two drawbacks prevent us from using Sampler 1.

Before realising how to overcome the convergence issue exposed in section 4.2.1, I tested Sampler 1 described in the same section. The first runs were conducted in pixel space and with fixed prior variance for the components. Prior covariances then became hyperparameters of the problem since I noticed that any error on the component modelling induced bias into the CMB reconstruction, even if the prior is right for the CMB. I also moved to the spherical harmonic domain in order to be able to have a complete model for the CMB anisotropies on the full sky. The results of this subsection are the products of such a configuration, i.e. a low resolution version of the data and the components are expressed in terms of spherical harmonic coefficients and the covariances are additional parameters to be sampled.

In order to show some results from adopting this approach, I set the initial values of the component maps to their true value, and I let the chain evolve. Figure 5.3 shows the

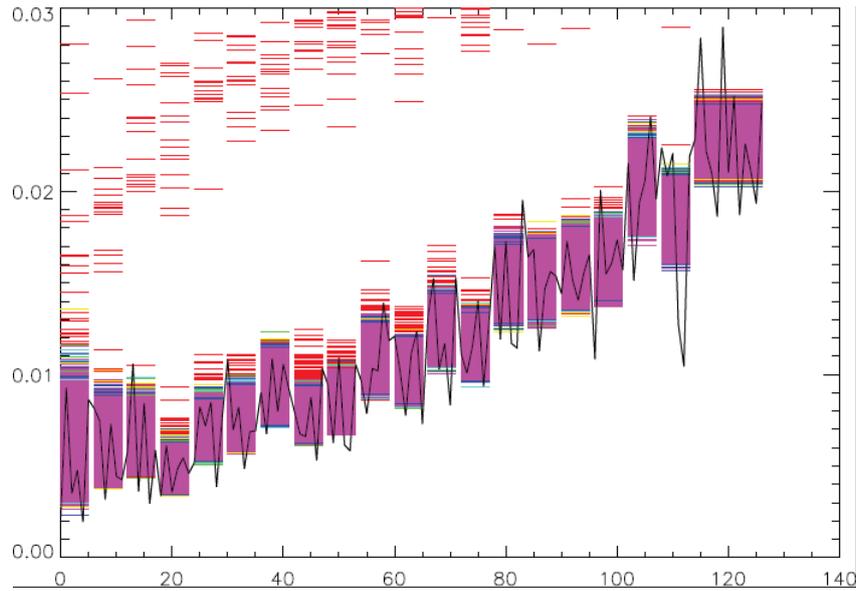


Figure 5.3: Low resolution binned CMB power spectrum inference. The full Gibbs sampling approach still gives good results. Black solid line is the input scattered C_ℓ 's and the coloured horizontal lines show the evolution of the chain, from red to purple being from beginning to the end. The length of one horizontal bar corresponds to the size of the bin.

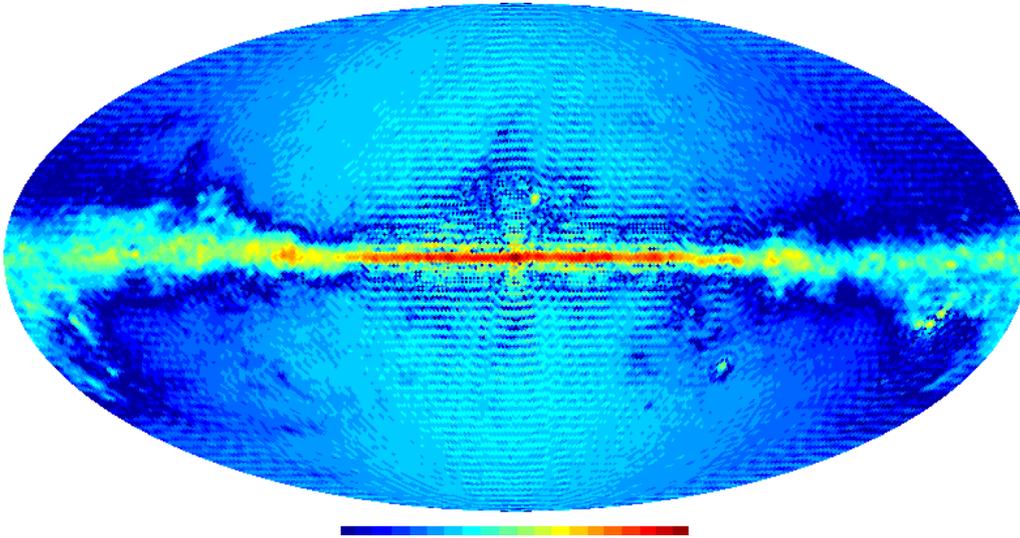


Figure 5.4: Low resolution sample variance map of CMB samples. This map does not have exactly the same structure than that of Sampler 2 case (see figure 5.8). This one has more structures and variance decreases less quickly along the latitudes.

inference of the first 126 multipoles above the dipole of the CMB. Each of the coloured horizontal bars represents one sample of a bin and the black line represents the input power spectrum. If the sampling issues are avoided by starting the chain close to the peak of the posterior, then the CMB power spectrum is well reconstructed. Figure 5.4 shows the sample variance of the CMB map. As expected, the errors are larger in the galactic plane.

5.3.2 Self consistent treatments

In order to test the algorithm, I applied the method on simulated data that have been drawn from the priors and add isotropic noise. The simulated data are then fully consistent with the model. In this case, the resulting mixing matrix maps and power spectra match their input values. Also, the mismatch (as defined later in equation 5.5) is completely flat and has the appropriate number of degrees of freedom if a set of simulations completely consistent with the data model is used. Such test involves Gaussian foregrounds and isotropic noise. data composed of foregrounds drawn from an isotropic Gaussian random field and isotropic noise is the ideal configuration in which BICA can fully solve the component separation problem, but in this case the foregrounds have no physical meaning and the noise model is far from that of real experiments.

I also test the impact of anisotropic noise on the results. I drew a set of data from the priors and this time added anisotropic noise. I observed no significant deviation from the results of the fully consistent case.

I will now move on to more physically interesting cases.

5.3.3 Products of the method

The output of the BICA sampler presented in the previous chapter is a collection of samples from the posterior distribution. Each sample, containing one realisation of the component maps and the power spectra and the mixing matrix, has dimension N_{para} . A normalized N_{para} -dimension histogram would give an estimate of the posterior over all parameters, but the number of dimension is too large to compute higher moments of the full posterior distribution with accuracy. Instead, a 1-dimension histogram is computed for each parameter. It can be shown that this histogram is proportional to an estimate of the posterior marginalised over all parameters but the one of interest. Then this histogram is used to compute various quantities (mean, variance, mode, the number of degrees of freedom, ...) relative to the marginal distribution.

All elements of one column of the mixing matrix are fixed to the same arbitrary constant (I chose 1). This prior information leads the sampler to distribute the information contained in the data. Any emission that is constant through frequency is transferred

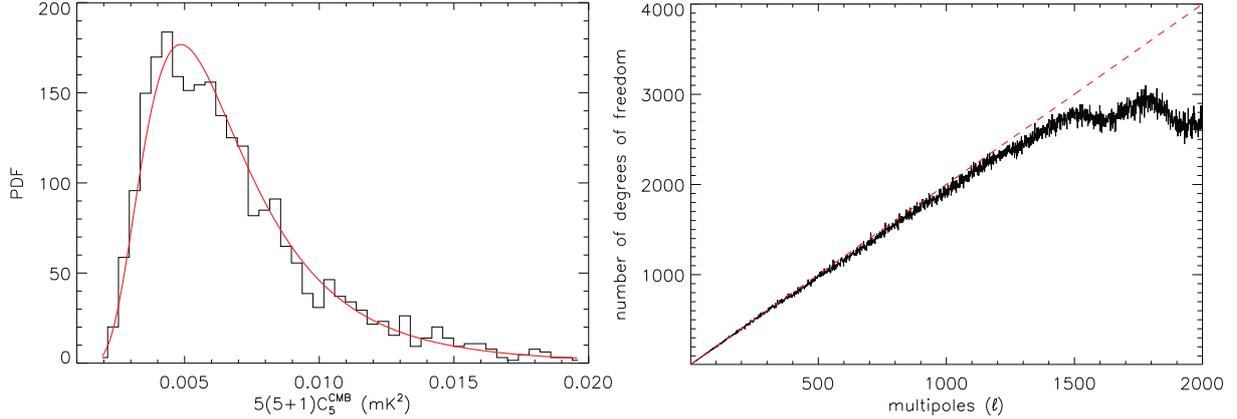


Figure 5.5: Left: Posterior PDF marginalised over all parameters but the CMB power spectrum at multipole $\ell = 5$. The histogram is an estimation of the PDF and the solid red curve is the best fit of an inverse-gamma function to the histogram. Right: number of degrees of freedom of the CMB C_ℓ 's individual marginal distributions (black line). The $2\ell + 1$ degrees of freedom expected for noiseless all sky is showed by the red dashed line. Deviation from the ideal case comes from the presence of noise that enlarge uncertainties on the power spectrum at large multipoles.

into the power spectrum corresponding to the constant mixing matrix column, and any other emission is transferred into the other power spectra. Since the CMB is the only coherent signal with constant response across all frequencies, our analysis amounts to infer a CMB power spectrum in the presence of foregrounds systematics.

Regarding the other components, the unmixing is not unique. Since there is no physical information on either the power spectra or the frequency spectra, the outcome of these parameters are mixtures of the data maps that obtains plausible configurations. *A priori*, the individual input power spectra of the non-CMB components are not expected to be identifiable as dust and free-free because I force no correlation between the two spatial distributions². The degeneracies come from the permutation of two column of the mixing matrix. Thus, there is a finite number of degeneracies introduced by using a phenomenological model. Since each solution are afar from each other, the degeneracies do not entail the sampling efficiency.

The run took 20 CPU hours to produce the chain from which the posterior is evaluated.

5.3.4 CMB power spectrum inference

The value of a power spectrum at multipole ℓ , C_ℓ , is the variance for each of the $2\ell + 1$ harmonic coefficients $a_{\ell m}$ of the corresponding isotropic map. Thus, for a full sky treatment, the expected PDF for the C_ℓ 's is an inverse-gamma distribution with $2\ell + 1$

²Physically, dust and free-free are spatially correlated since both of them are more prevalent in the galactic disk than at high latitudes.

number of degrees of freedom. Indeed an inverse-gamma distribution fits the individual marginal distributions well, see figure 5.5, as expected (see e.g. Wandelt et al. (2004)). Figure 5.5 shows the number of degrees of freedom of the inverse-gamma distributions that fit the marginal distribution of the C_ℓ 's. From $\ell \sim 500$, the number of degrees of freedom is lower than that of the optimal case of noiseless full sky treatment. This is due to the noise at large ℓ 's that enlarges the variance of the marginal distribution.

The inverse-gamma distribution is a non-symmetric PDF. To visualise the inference of the full CMB power spectrum I choose to show the mode of each distribution rather than the mean. To give an idea of the shape around the peak I compute the variances of the distribution below and above the mode, as if the distribution were a two-sided Gaussian. Figure 5.6 shows such a visualisation after a run of BICA. On the upper panel the black dots show the modes of the marginal distribution of each C_ℓ . The grey region represents the shape as if the distribution were a two-sided Gaussian distribution: upper error is one upper standard deviation and the lower error is one lower standard deviation. The solid red line shows the input power spectrum. The lower panel of each plot in the figure shows the relative error to the input power spectrum. The input power spectrum of the CMB lies within the error bars and the recovery is accurate at better than the percent level.

5.3.5 CMB map inference

In section 4.2.3, I described the sampling used to explore the posterior distribution. In practice I let the Metropolis-Hastings sampler of the marginal evolve until it converges and draws N_{sam} uncorrelated samples of $\{\mathbf{A}, C\}$. Then, for each sample $\{\mathbf{A}_n, C_n\}$, I draw a sample of the component maps s_n from the conditional $P(s | \mathbf{A}=\mathbf{A}_n, C=C_n, d)$. The chain $\{\{\mathbf{A}_n, s_n, C_n\}; n = 1 \dots N_{\text{sam}}\}$ provides a samples from the full posterior $P(\mathbf{A}, s, C | d)$.

The sampling of the s_n is straightforward since the conditional $P(s | \mathbf{A}, C, d)$ is a Gaussian distribution, independent from one harmonic coefficient to another. For each piece of maps $s_{\ell m}$ the mean and covariance are

$$\mu_{\ell m} = \Sigma_\ell \mathbf{A}^T N_\ell^{-1} d_{\ell m} \quad (5.1)$$

$$\Sigma_\ell = (\mathbf{A}^T N_\ell^{-1} \mathbf{A} + C_\ell^{-1})^{-1} . \quad (5.2)$$

Thus, the $\mu_{\ell m}$'s are obtained by Wiener-filtering the data. There is a loss of power at high multipole in the mean due to the filter. Sampling the maps corrects this and the samples of the maps have the correct covariance. Following Wandelt et al. (2004), the sampling of the maps is done by solving the system

$$\Sigma_\ell^{-1} s_{\ell m} = \mathbf{A}^T N_\ell^{-1} d_{\ell m} + \xi_{\ell m}$$

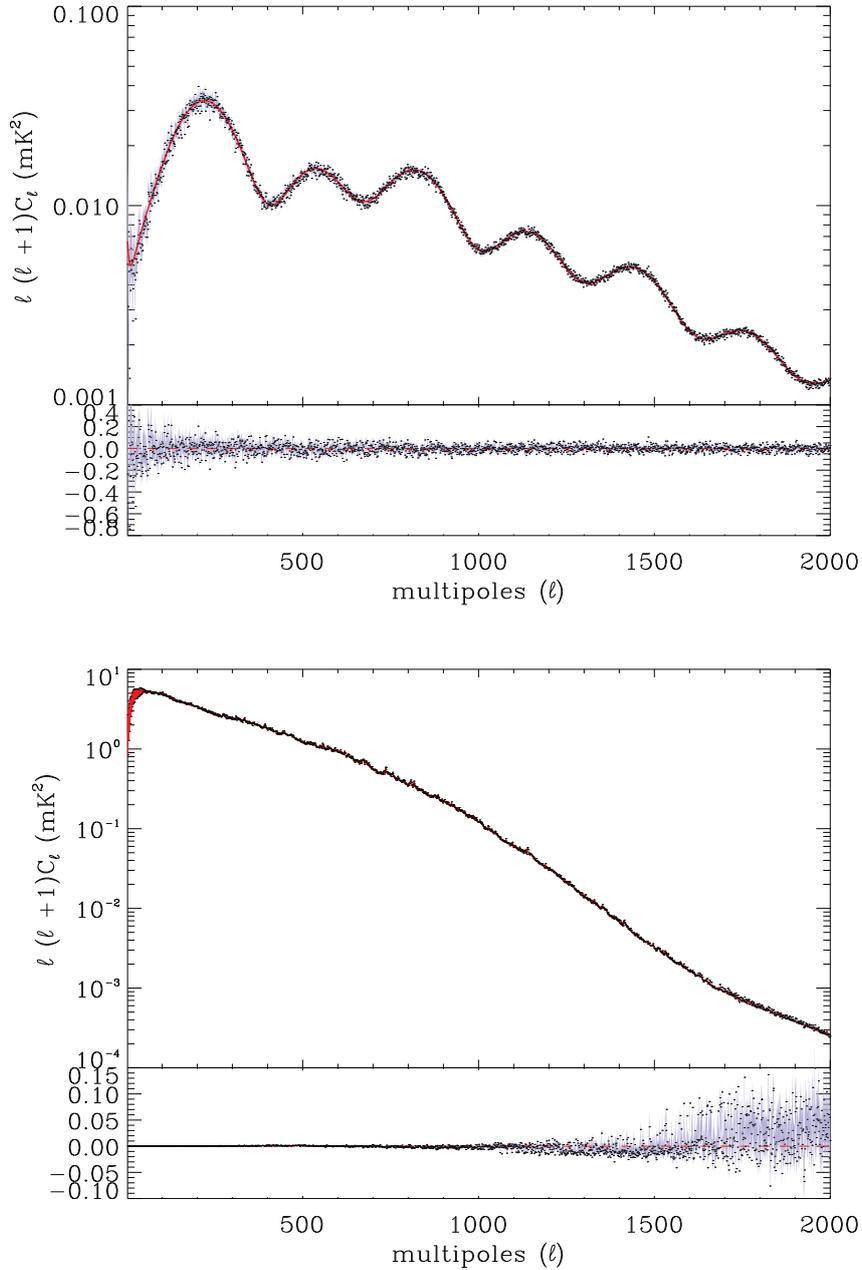


Figure 5.6: Input and inferred power spectra of the CMB (top) and the sum of the non-CMB components (i.e. dust and free-free) at 353GHz (bottom). In the upper panel of each figure, the black dots at each multipoles represent the peaks of the marginal posterior, the grey region shows the asymmetric $\pm 1\sigma$ -error bar derived from the marginal posterior, the red line is the input power spectrum. The lower panel represents the relative error to the input power spectrum. The sampler accurately recovers the power spectrum of the CMB.

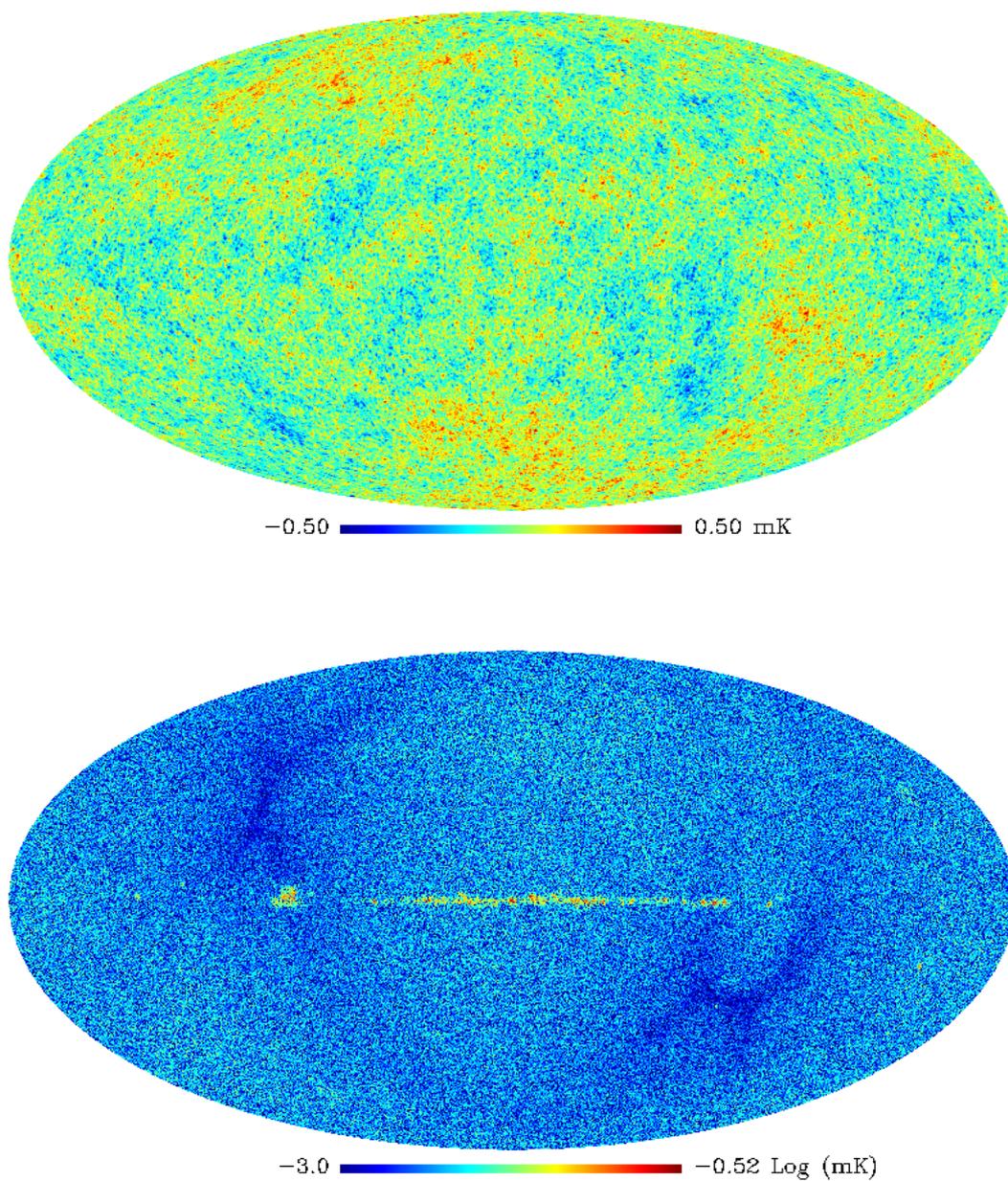


Figure 5.7: Input and residual CMB map. This residual map represents the absolute value of the difference between the sample mean and the input map. The errors are wider in the galactic plane but the uncertainties in this region of the sky are also larger, as shown on Fig. 5.8. To show the noise in the residual map, it is shown on a decimal logarithm scale.

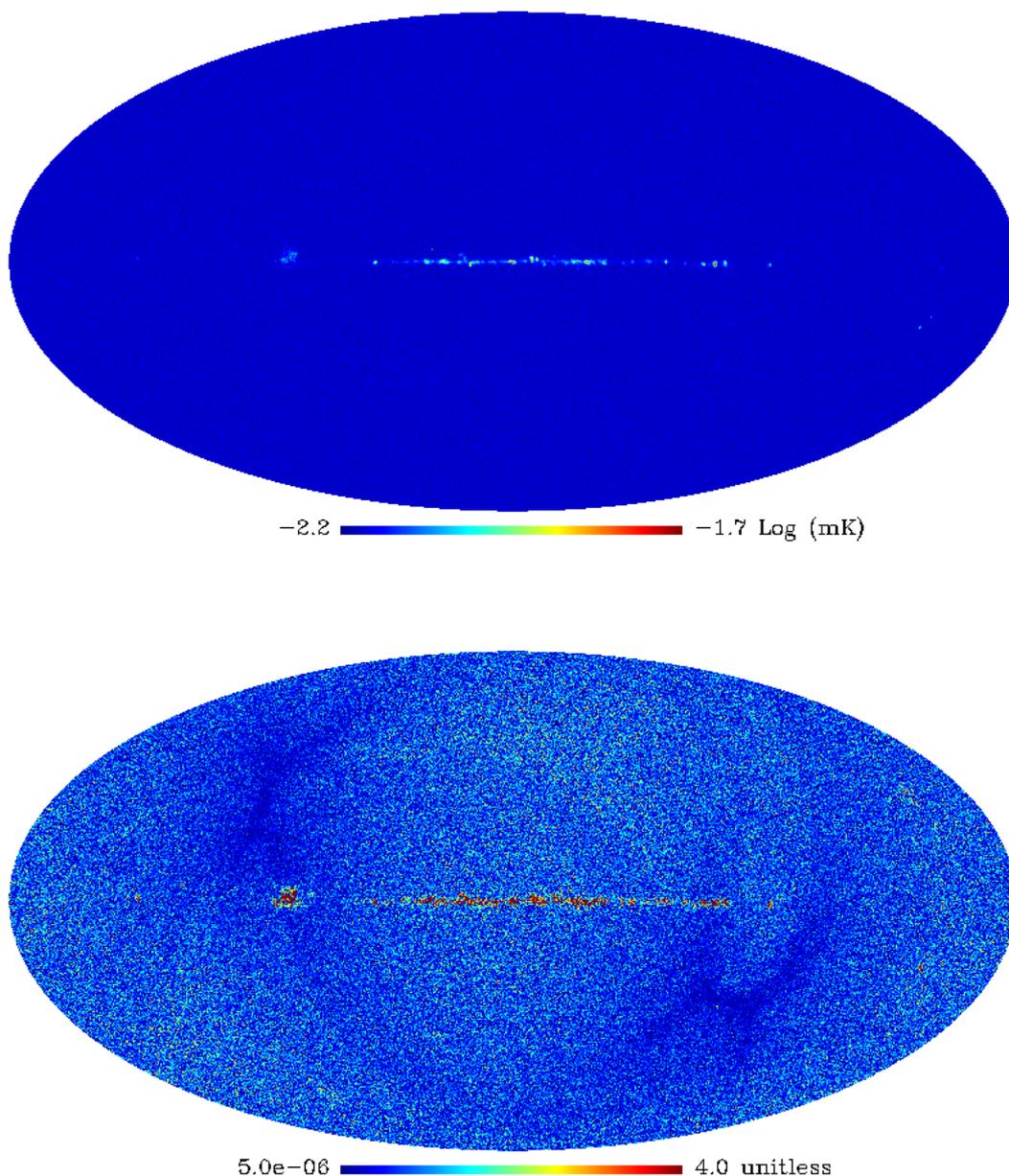


Figure 5.8: Top: Standard deviation map of the CMB map samples. The posterior distribution is wider in the region of the galactic plane. Bottom: Standardized error map, all red pixel have value 4 or more. This map is the ratio between the residual map of the CMB and its standard deviation map (top). Note that the posterior standard deviation map only represents the part of the uncertainty that is uncorrelated from pixel to pixel while the Bayesian analysis returns a fully correlated error model for the recovered map shown in Fig. 5.7. Standardizing with the uncorrelated errors reveals two things: the isotropic noise approximation leads to overestimated uncertainties in low noise regions; and an uncorrelated error model does not capture the uncertainties in regions where foregrounds dominate. See Fig. 5.9 for a visualization of correlated uncertainties.

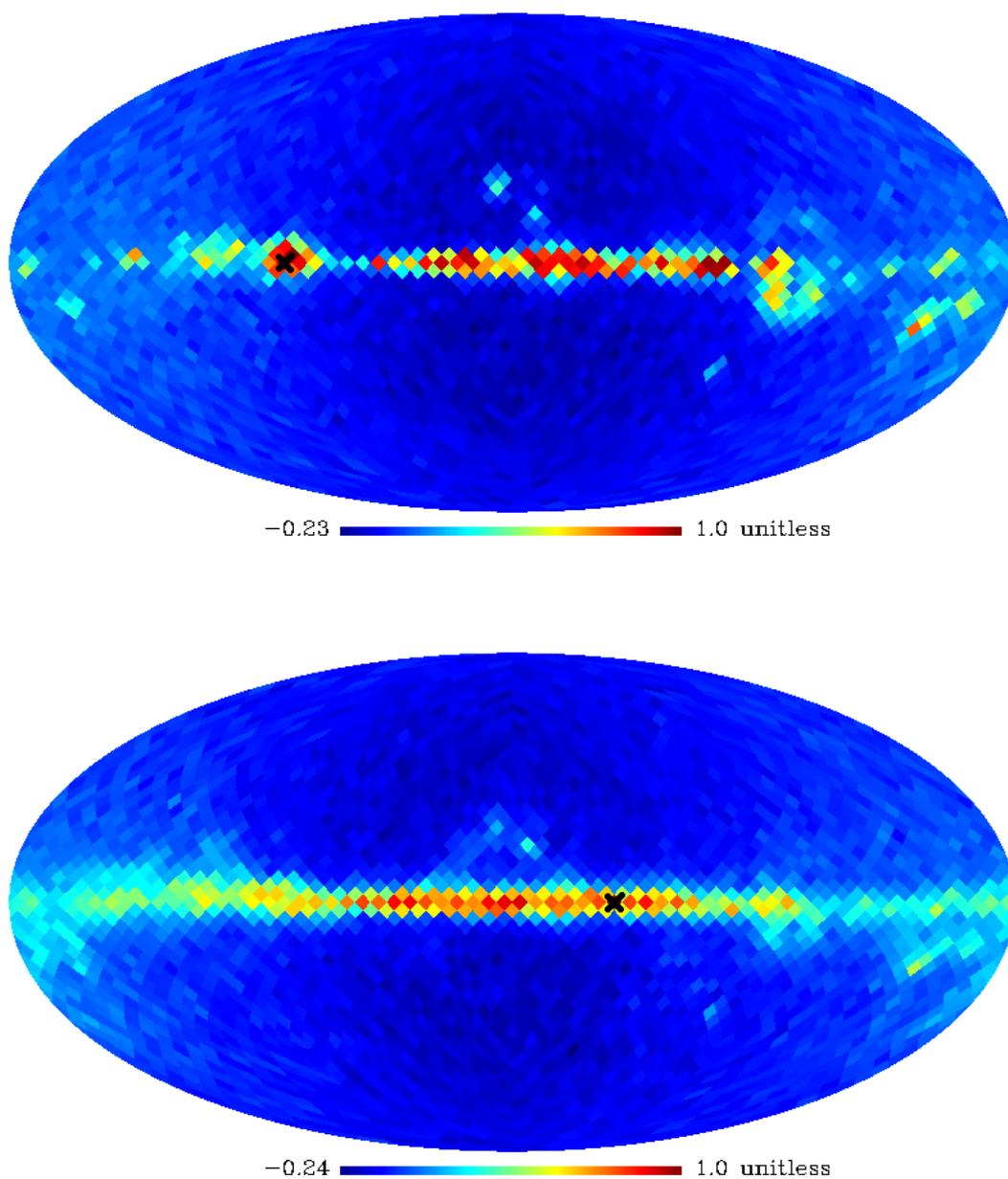


Figure 5.9: The figure shows two rows of the posterior correlation matrix for the 2 pixels marked by a black cross in each map at HEALPix resolution parameter $N_{\text{side}} = 16$. The inferred uncertainties due to foreground removal are highly correlated in the galactic plane and must be taken into account in a meaningful statistical interpretation of the recovered CMB map.

where $\xi_{\ell m}$ is i.i.d. for each $s_{\ell m}$ from a Gaussian distribution with zero mean and covariance Σ_{ℓ}^{-1} .

Marginalising the posterior $P(\mathbf{A}, s, C | d)$ over all parameters but one pixel of one component map leads to a distribution which is consistent with a Gaussian distribution. I therefore consider the sample mean of the map samples, which is an estimate of the mean posterior CMB map, as a reference for a recovered CMB map. Fig. 5.7 shows the input CMB map and the absolute value of the residual map. There is more residual error in the galactic plane because of foreground contamination. There are pixel to pixel correlation in the posterior but qualitative errors on the recovered map are given by the sample variance of each pixel.

Fig. 5.8 shows a map containing the sample standard deviation of each pixel of the CMB map. As stated above, the errors on the CMB map include the uncertainty due to the presence of galactic emission. I also plot in Fig. 5.8 the standardized error map, i.e. the ratio between the residual map and the standard deviation map. If the model were a perfect description of the data, and if the pixels were not correlated in the posterior, then this map should be a realisation of an isotropic random Gaussian field, with zero mean and unit covariance. However, because of the approximations in the modelling presented in section 5.2, this is not the case. The isotropic noise approximation leads to an overestimation of the error bars in the regions of low noise. The per-pixel error is underestimated in highly contaminated regions. The residuals have strong spatial correlations, see Fig. 5.9. Another explanation could be that the Gaussian model is too coarse an approximation in regions where the foregrounds are the most intense and highly non-Gaussian. If it is the case, these results could be used to construct masks from the sample variance map to mask the observation maps where necessary, since these regions are the regions of high variances in the posterior variance map (top of Fig. 5.8).

Since the uncertainty in pixels are correlated in the posterior, the variance map gives only a qualitative idea of the errors on the map, since it is only the diagonal of the posterior covariance. Pixels in the posterior are correlated and the sample variance map only is not sufficient to fully describe the error on the reconstructed CMB map. To show the correlation, I compute the correlation matrix of the CMB map samples on a lower resolution map. The correlation matrix has a row for every pixel showing the correlation of this pixel to all other pixels. Figure 5.9 shows the correlation maps of two pixels in the galaxy plane. The pixels in the galactic plane are highly correlated, which explains at least part of the excess of standardized error in the galactic plane of Fig. 5.8. An eigenvalue analysis of the correlation matrix shows that, in addition to noise uncertainties on small scales, the foreground subtraction uncertainties are dominated by a few, highly correlated modes, see Fig. 5.10.

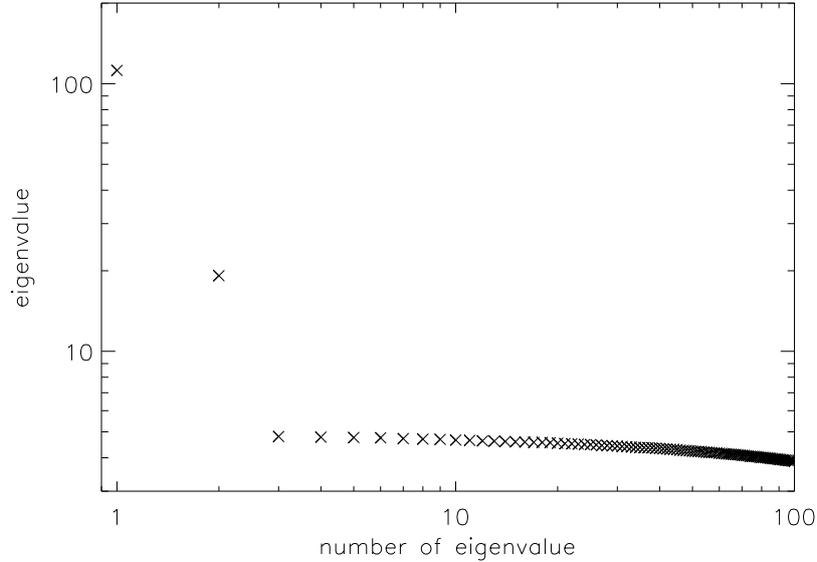


Figure 5.10: The hundred largest eigenvalues of the posterior correlation matrix of the low resolution CMB map. Two modes dominate.

5.3.6 Inference of non-CMB components

Without any physical information and assuming lack of correlation, the inferred non-CMB components are not identifiable as free-free and dust. However, the sum of all non-CMB components is recovered. Figure 5.6 shows the power spectrum of the simulated whole galactic emission and its inference. Since the CMB is cleaned of all foregrounds, the sum of all non-CMB components is also well recovered but small biases appear from $\ell = 1000$. The biases are due to the fact that the correlation between the components are not taken into account, as explained in section 5.4. These biases are small compared to the CMB power and therefore have no significant effect on the CMB inference.

5.4 Model checking

The uncertainties on the reconstruction of power spectra directly rely on the shape of the posterior. Therefore the errors are correctly estimated if the *a priori* model correctly describes the data. It is therefore important to assess the quality of the fit achieved by the model through model checking (Gelman and Meng, 1996). In order to check for biases in the reconstruction due to assumptions on the statistical properties of the components, a consistency check is needed. The measure of some mismatch between the empirical covariance of the data and the covariance of the data model will provide a test of the model at the power spectrum level. In this section, I construct such a check, starting from basic statistical tools.

5.4.1 Construction of the mismatch

At the power spectrum level, only covariances are considered. The empirical data covariances at each multipole ℓ , $\hat{R}_\ell = \frac{1}{2\ell+1} \sum_m d_{\ell m} d_{\ell m}^T$, will be used as references. $R_\ell = \mathbf{A}C_\ell\mathbf{A}^T + N$ is the data covariance from the model of equation 3.1. If the model correctly represents the data, \hat{R}_ℓ and R_ℓ at the peak should be close to each other. In order to compare them, I consider the following generative model for the data at each multipole ℓ :

$$p_\ell(d | \Sigma) = \prod_{m=-\ell}^{\ell} \mathcal{N}(d_{\ell m} | 0, \Sigma), \quad (5.3)$$

and compute the KL divergence for $p_\ell(\cdot | \hat{R}_\ell)$ and $p_\ell(\cdot | R_\ell)$. Since the KL divergence between two Gaussian distribution \mathcal{N}_1 and \mathcal{N}_2 with zero mean and respective $n \times n$ covariances Σ_1 and Σ_2 is

$$D_{KL}(\mathcal{N}_1, \mathcal{N}_2) = \frac{1}{2} [\text{tr}(\Sigma_1 \Sigma_2^{-1}) - \log |\Sigma_1 \Sigma_2^{-1}| - n] \quad (5.4)$$

and thanks to the additivity property we have for each ℓ

$$\begin{aligned} D_{KL}(p_\ell(\cdot | \hat{R}_\ell), p_\ell(\cdot | R_\ell)) &= \frac{2\ell+1}{2} \left[\text{tr}(\hat{R}_\ell R_\ell^{-1}) - \log |\hat{R}_\ell R_\ell^{-1}| - N_f \right] \\ &\hat{=} D_\ell/2. \end{aligned} \quad (5.5)$$

Following the analysis of Delabrouille et al. (2003), I will consider D_ℓ , i.e. half the KL divergence, because it can be shown that $\exp -\frac{1}{2}D_\ell$ is proportional to the likelihood. The posterior PDF on the CMB power spectrum approaches a Gaussian as ℓ increases, such that D_ℓ has the properties of the chi-squared distribution for sufficiently large ℓ . For example the number of degrees of freedom is obtained by subtracting the number of stochastic parameters per multipole from the number of degrees of freedom of a symmetric matrix $N_f \times N_f$. N_c spectra are sampled per multipole and one mixing matrix is sampled for all multipoles. Thus, if the number of degrees of freedom within a mixing matrix are distributed over all multipoles, the number of degrees of freedom of the pseudo chi-squared distribution followed by each D_ℓ is

$$\begin{aligned} N_{\text{dof}} &= N_f(N_f + 1)/2 - \left[N_c + \frac{(N_f - 1)(N_c - 1)}{(\ell_{\text{max}} - \ell_{\text{min}} + 1)} \right] \\ &\simeq N_f(N_f + 1)/2 - N_c. \end{aligned}$$

In particular $\mathbb{E}[D_\ell] = N_{\text{dof}}$, which does not depend on ℓ . In the case of the application on the simulations, $N_{\text{dof}} = 7$.

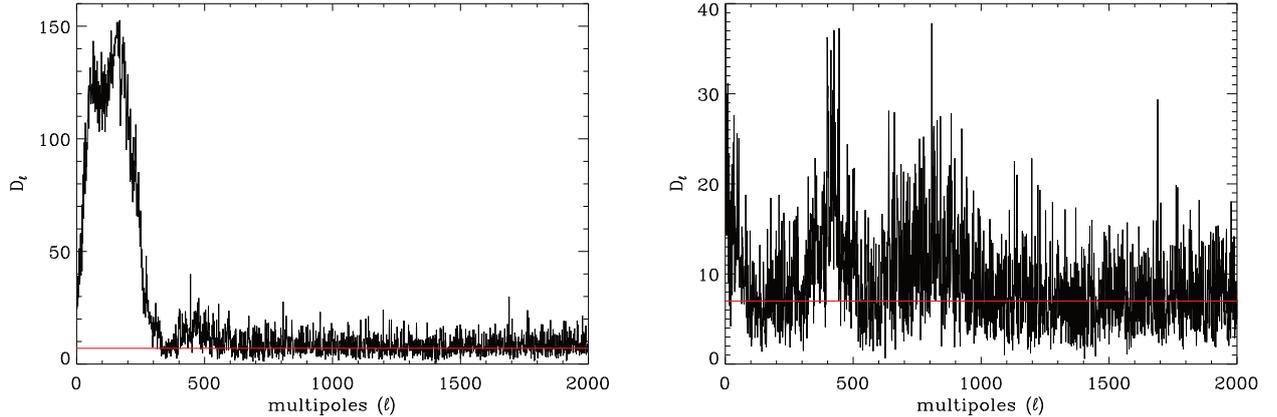


Figure 5.11: Mismatch between the data and the data modelling. Left: Divergence between the data covariance and input parameters. The large mismatch at low ℓ is due to correlations between the input component maps. Right: Divergence between the data covariance and the recovered parameters. During sampling, I impose no cross-correlation. Thus the sampler converges towards components that are uncorrelated but whose power spectra are almost capable of capturing the covariance of the input component maps. The red line is the mean of the expected chi-squared distribution that the D_ℓ 's should approximate.

5.4.2 Consistency of the results on simulations

In Fig 5.11 I plot the D_ℓ 's for R_ℓ containing the inferred value of $\{\mathbf{A}, C\}$. For comparison, I also plot the D_ℓ 's in the case where \mathbf{A} and C are set to their input values. Although the input parameters are the true parameters to be recovered, the inferred values have lower mismatch because the components are correlated in the data but not in the model and the sampler finds uncorrelated components that fit the data better. If the foreground modelling matches the statistical properties of the input foregrounds, the D_ℓ 's should follow a chi-squared distribution with a number of degrees of freedom N_{dof} , whose mean N_{dof} is represented by the horizontal red line on Fig. 5.11.

5.4.3 Consistency of the results with modified priors

I performed a separation where the cross power spectra of the component are taken into account during the sampling. I do not sample the cross power spectra but I use the covariance of the input components instead and I keep them fixed during the sampling, i.e. each C_ℓ is a non-diagonal matrix but only the diagonal is stochastic. In Fig 5.12 I plot the D_ℓ 's with the output values of $\{\mathbf{A}, C\}$ of such a run. Taking the correlation of the input component maps into account erases the discrepancy at low ℓ . A chi-squared distribution fits the histogram of the D_ℓ for ℓ large enough ($\ell > 700$ for this plot). In the figure, the fit is shown by three blue solid lines, the solution of the fit and the $\pm 1\sigma$ error

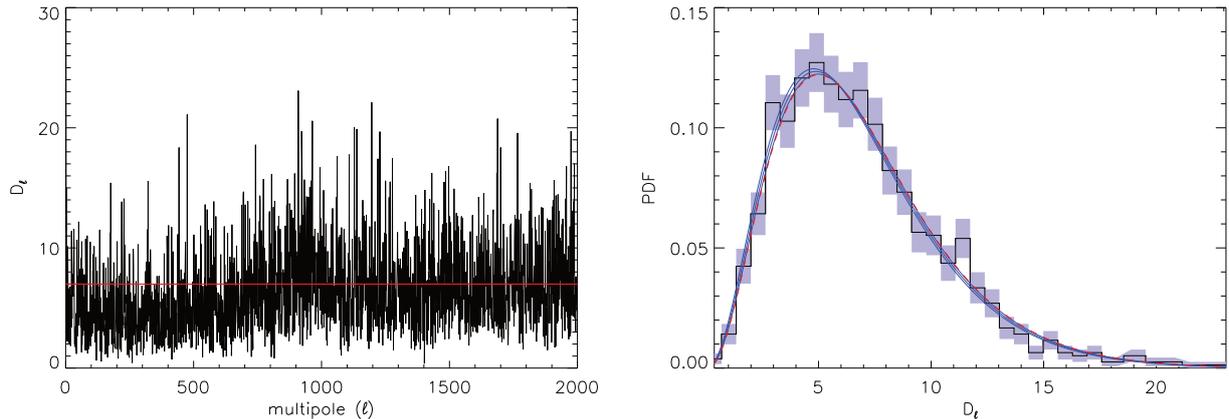


Figure 5.12: Mismatch between the data and the result of a sampler that includes the input correlations between the components during the sampling process (left) and its PDF (right). In the left panel the red line represents the mean of the expected chi-squared distribution that the D_ℓ 's should follow. In the right panel, the blue lines are the best χ^2 -fit to a chi-squared curve and the ± 1 standard deviation error from the fit. The red dashed line represents the chi-squared distribution with the expected number of degrees of freedom $N_{\text{dof}} = 7$. The introduction of the correlation between the input component erases the large discrepancy at low multipole.

on the fit. The red dashed line represents the chi-squared distribution with the expected number of degrees of freedom N_{dof} and it lies within the error bars. The remaining deviations from an ℓ -independent distribution are due to the differences between the data model and the data actually used, as explained in section 5.3.2.

Considering a Gaussian model for the components and neglecting the correlations between the CMB and the non-CMB components do not affect the reconstruction of the CMB power spectrum since the CMB is not correlated to the foregrounds.

5.5 Discussion

The results are very encouraging. The CMB power spectrum is well reconstructed even with a simple Gaussian model for the non-Gaussian foregrounds. The production of variance maps and correlation maps taking the presence of the foregrounds blindly into account is a major step forward in CMB data analysis.

The fact that the sampler can not distinguish between non-CMB emissions is due to an additional degree of freedom. Without any information about the physical emissions, all the recovered components others than CMB are a mixture of the true signals. Putting a prior either on the mixing matrix, i.e. import knowledge about the frequency spectra, or on the shapes of the component power spectra would break the degeneracies. Although the constraints of the priors can be controlled, the blindness of the method would be

lost with this introduction of *a priori* information. Furthermore, in this work I impose decorrelation between the component maps. The foregrounds could be modelled in more detail to get a full component separation method, but the focus here is on CMB reconstruction. I showed that including the input cross spectra of the components improves the mismatch. In the next chapter, I include the cross spectra as additional parameters to infer.

Unresolved point sources appear as extra power at small angular scales of the inferred CMB power spectrum. Masking the listed point sources and inpainting in the mask would be a way to address the point sources issue. In addition allowing the components to mix differently in different regions of the sky or in different angular scale ranges by allocating different mixing matrices in each range would reduce mismatch due to lack of coherence. Also, foregrounds that are not fully coherent from frequency to frequency may be modelled by increasing the number of components in our model.

In this chapter I reconstruct the CMB maps on the full sky. It remains to be seen if this is achievable on realistic data and next chapter address this problem. Since our approach is similar to the one by SMICA, which provides a clean map on a large fraction of the sky in Planck analysis (Planck Collaboration et al., 2014c), treatment of almost full sky data should be feasible. Because of the assumption of diagonal covariances in ℓ -space the effect of a small mask needs to be tested. If a mask has non negligible effect, masking may be avoided with inpainting. If necessary a Wiener filter method such as Elsner and Wandelt (2013) could be used to implement a full sampling approach.

The next step is to apply the method to real data (see chapter 6). The main problem is the instrumental noise. In the work presented in this chapter I assume a simple noise model. Dealing with real data noise requires developing a more realistic noise model. A full model of correlated noise involves very large covariance matrices. Therefore alternative ways to deal with noise like half-ring half difference maps or noise simulations should be considered.

If joint modelling of foregrounds allows working with a large part of the sky, mode coupling effects due to the mask may be ignored with high accuracy. This needs to be tested and it is also addressed in the next chapter.

Conclusion

I have presented the results of BICA on a set of simulations. Both the CMB and galaxy posser spectra are correctly inferred at the percent level. The results match the expected statistical properties such as the shape of the marginal posterior for the CMB multipoles or the number of degrees of freedom of the data mismatch. Full maps of CMB anisotropies are recovered with their own PDF which reveals that the dominant foreground residuals

are captured in terms of a small number of error modes. In the next chapter, I apply the BICA method to the Planck data.

Chapter 6

Application to Planck data

Introduction

Now that I showed that BICA has robust results when it is applied to simulations, the next step is to apply it to real data. I worked for a short time with the WMAP data, but I set this analysis aside when the Planck collaboration released its 2013 data.

The sky simulation of the previous chapter was a simplistic model but the purpose of the application of BICA to simulations was to validate the concepts of the method. The Planck data are much more complex than the simulations. Thus, after giving a brief overview of the Planck data, I build a refined data model that takes into account additional instrumental and physical modelling. Then I present the first results on the inference of the parameters of this model and compare them to the results of SMICA.

6.1 The Planck data

The satellite of the Planck mission (Planck Collaboration et al., 2011a) was launched in May 2009 and reached its orbital position at L2 Lagrange point in July of the same year. It has since observed the full sky twice. Planck embarked two instruments aboard: the Low Frequency Instrument (LFI) and the High Frequency instrument (HFI). The two instruments are designed with different technologies. LFI is composed of transistors whereas HFI makes use of bolometers cooled down to 0.1K. Both instruments map temperature as well as polarisation of the CMB.

In this work I worked with the Planck public release of 2013 (Planck Collaboration et al., 2014a). The temperature maps as they were released is an estimate of the sky signal at each frequency band centred around 30GHz, 44GHz, 70GHz for LFI and 100GHz, 143GHz, 217GHz, 353GHz, 545GHz and 837GHz for HFI. The maps are pre-processed in order to remove known systematic effects as much as possible. The dipole induced by the non zero velocity of the solar system in the rest frame of the CMB (Corey and Wilkinson,

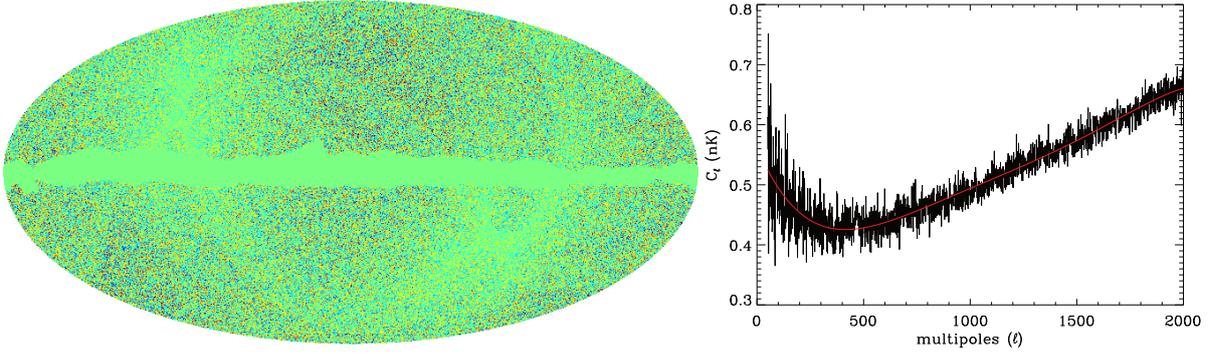


Figure 6.1: Masked half-ring half-difference map and its power spectrum. This map was obtained by taking the half difference of the two half-ring maps of the 100GHz channel. Then it was masked with the galactic mask only because the point source mask is inpainted and filled with noise. The black line on the plot below is the power spectrum of the map. The power spectrum thus computed is just an estimator of the variance of the noise in each mode. The red line is a fit to the estimate of the power spectrum that gets closer to the true noise power spectrum. This operation is repeated for all channels and the noise model is the collection of the fits.

1976) is also removed. I took the maps from the nominal mission as the observation maps. The nominal maps are built from all the valid detectors of a frequency channel.

Separately are also provided two maps per frequency band called half-ring maps. One is made of the first half of the scanning ring, one is made of the second half of the scanning ring only and both cover the full sky. They are very important in order to characterise some properties of the instrumental noise, as detailed in the next section.

6.2 Additional modelling

The simulations on which I have tested BICA in chapter 5 are an approximation of reality. Real data are more complex and need a more detailed modelling, of the physics as well as of the instrument. Through out this section, I will improve parametrisation of the problem starting from the model I validated against simulations in chapter 5.

6.2.1 Noise

The noise in the nominal maps is anisotropic and correlated pixel to pixel. Thus, a full $N_{\text{pix}} \times N_{\text{pix}}$ covariance is necessary to model it accurately in pixel space. For high resolution maps, such an array is too large to be contained in an ordinary computer RAM, and other ways of representation are needed. Correlation in pixel space induces colouring of the noise in harmonic space, i.e. the power spectrum of the noise is not flat. Conversely, a Gaussian random field with a non-constant power spectrum has correlated

directions on the sphere but the field is isotropic. Using a power spectrum only does not account for statistical anisotropies in the spatial distribution of the noise. Thus, choosing to model the noise by a power spectrum amounts to making the approximation of isotropy of the noise but allows including noise correlations in this isotropic approximation. The approximation of isotropy do not affect the results by much because the power spectrum is little biased by the anisotropy (Wandelt et al., 2001).

As it was done for Planck analysis (e.g Planck Collaboration et al. (2014g), Planck Collaboration et al. (2014d), Planck Collaboration et al. (2014c)), I use the half-ring maps in order to find the power spectrum that approximates the properties of anisotropy and correlation the best. Half-ring maps are built using only the first or the second half of the scanning ring of the nominal mission. For each frequency band, the half-ring maps contain the same physical signal but with a different realisation of the noise in half the full data. Thus taking the half difference of the maps cancels out the physical signal and gives a possible realisation of the noise in the nominal map. Computing the power spectrum of that map leads to an estimate of the noise variance per multipole in the nominal map. The higher the multipole is, the largest number of modes it contains, the better the estimate is. Even if the estimate is good for small scales, the variance estimate can not be used as it is because of the scatter. I fit a slow varying function to the estimate to get the noise power spectrum that I use as a model of the instrumental noise.

Due to specifics in the pre-processing of the data, the estimate is a fraction of a percent lower than the actual power spectrum (Planck Collaboration, 2013). In order to model this inaccuracy, I add a noise power spectrum normalisation factor f_i per frequency band i to the parameters to sample. Thus, the model becomes

$$d_{i\ell m} = \sum_k A_{ik} s_{k\ell m} + \sqrt{f_i} n_{i\ell m}, \quad (6.1)$$

which reads in terms of covariances

$$R_\ell = \mathbf{A} C_\ell \mathbf{A}^T + f N_\ell, \quad (6.2)$$

where $f = \text{diag}(f_i; i \in \llbracket 1, N_f \rrbracket)$.

The collection of the f_i are added in the set of parameters to infer. Including these parameters adds N_f more dimensions. It is a low number compare to already existing number of dimensions, such that the inclusion of the normalisation factors does not disrupt the convergence of the sampler. Since f is actually close to identity, I put a prior on the f_i 's that constrain them to stay around 1 for the LFI channels and 1.005 for the HFI channels. I chose an inverse-gamma distribution as prior for each f_i because f acts like a diagonal covariance.

6.2.2 Cross-spectra

The spatial distributions of the components are correlated. For instance the different components of the galactic signal are all the brightest in the galactic plane. If two maps a and b are spatially correlated, i.e. $\langle a(\vec{n}_1)b(\vec{n}_2) \rangle \neq 0$ for some pixels, then their cross-power spectrum C_ℓ^{ab} is non zero:

$$C_\ell^{ab} \hat{=} \langle a_{\ell m} b_{\ell m}^* \rangle \propto \int d\vec{n}_1 d\vec{n}_2 \langle a(\vec{n}_1)b(\vec{n}_2) \rangle Y_{\ell m}^*(\vec{n}_1) Y_{\ell m}(\vec{n}_2) \neq 0. \quad (6.3)$$

In order to reduce the mismatch between the data and the data model, I add the cross-spectra of the foregrounds in the set of parameters to be sampled. This move breaks the assumption of independence between the components. To keep the separation between the CMB and the foregrounds, I keep the cross-spectra between them equal zero. For now on, for each multipole ℓ , the non-diagonal covariance C_ℓ of the components in the prior model and equation 6.2 looks like

$$C_\ell = \left(\begin{array}{c|c} C_\ell^{\text{cmb}} & \mathbf{0}^T \\ \hline \mathbf{0} & C_\ell^{\text{fg}} \end{array} \right), \quad (6.4)$$

where C_ℓ^{cmb} is the CMB power spectrum at multipole ℓ and C_ℓ^{fg} is the non-diagonal covariance of the foregrounds at multipole ℓ . The CMB power spectrum C_ℓ^{cmb} and the free parameters of the symmetric matrix C_ℓ^{fg} are sampled for all multipoles ℓ . The Jeffreys prior for $N_c \times N_c$ matrices like that of equation 6.4 is $P(C_\ell) \propto \left(C_\ell^{\text{cmb}} | C_\ell^{\text{fg}} |^{\frac{N_c+1}{2}} \right)^{-1}$.

6.2.3 Point sources

In addition to the galactic signal, a important component to take into account is the unresolved point sources. Since this component mainly behaves like an isotropic random Gaussian field, its power spectrum is flat, i.e. it has the same power at all scales (Planck Collaboration et al., 2014d). This is a major issue because all the other kinds of components, CMB among them, have decreasing power spectra such that point sources dominates the signal in the multipole range $\ell \gtrsim 2000$.

It is compulsory to input information through prior distributions. Indeed, if I increase the number of components, I open flat directions because of degeneracies and therefore adding point source component requires specifying information about the expected power spectrum. The power spectrum of the point sources is flat because the point sources are uncorrelated on the map to good approximation. Since point sources are a component that looks like a realisation of a random Gaussian white noise, a straightforward way

to model them is to add a component p_{ilm} to the right hand side of equation 6.1 that satisfies $\langle p_{ilm}p_{j'l'm'} \rangle = \delta_{\ell\ell'}\delta_{mm'}\Phi_{ij}$. Note that Φ , the point sources covariance in harmonic space, does not depend on ℓ or m . Hence the further modelling of the covariance

$$R_\ell = \mathbf{A}C_\ell\mathbf{A}^T + \Phi + fN_\ell. \quad (6.5)$$

Point sources are spatially uncorrelated. In contrast, the Planck likelihood analysis showed that the point sources are strongly correlated frequency to frequency (Planck Collaboration et al., 2014d). This makes the $N_f \times N_f$ covariance matrix Φ a low rank matrix that then can be represented as follows

$$\Phi = \beta BB^T,$$

with B a $N_f \times N_{\text{pt}}$ matrix and β a scalar. B and β can be seen as a mixing matrix of N_{pt} component that are coherent through frequencies, uncorrelated from one another, and all of them have a power spectrum with constant value β .

The covariance Φ is not mathematically low rank but has some low eigenvalues. So we can see point sources as a mixing of very few component maps that are scaled through frequency by the matrix B plus one map q_{ilm} per frequency band i in the right hand side of equation 6.1 that will account for lack of coherence. The coefficients q_{ilm} are completely independent and $\langle q_{ilm}q_{j'l'm'} \rangle = \delta_{\ell\ell'}\delta_{mm'}\delta_{ij}\phi_i$. Hence the further modelling of data and its covariance

$$d_{ilm} = \sum_k A_{ik}S_{k\ell m} + \sum_h B_{ih}P_{h\ell m} + q_{ilm} + \sqrt{f_i}n_{ilm} \quad (6.6)$$

$$R_\ell = \mathbf{A}C_\ell\mathbf{A}^T + \beta BB^T + \phi + fN_\ell, \quad (6.7)$$

where $\phi = \text{diag}(\phi_i; i \in \llbracket 1, N_f \rrbracket)$. When $N_{\text{pt}} > 0$, ϕ_i is just a correction in frequency band i , so there is no need to have a non-zero ϕ_i in each band. Let's denote the number of non-zero ϕ_i by N'_{pt} . We have then

$$N_{\text{pt}} + N'_{\text{pt}} \leq N_f. \quad (6.8)$$

Splitting the point sources into a correlated part (βBB^T) and an uncorrelated part (ϕ) is very convenient because it allows to model the point sources with some flexibility. For examples, the model considering just one point source completely coherent through frequencies is $(N_{\text{pt}}, N'_{\text{pt}}) = (1, 0)$, the model considering only one uncorrelated point sources map in each frequency band will be $(N_{\text{pt}}, N'_{\text{pt}}) = (0, N_f), \dots$

There are constraints on N_{pt} and N'_{pt} . Indeed, if we suppose that Φ in equation 6.5 and $\beta BB^T + \phi$ in equation 6.7 have the same physical meaning the maximum number of

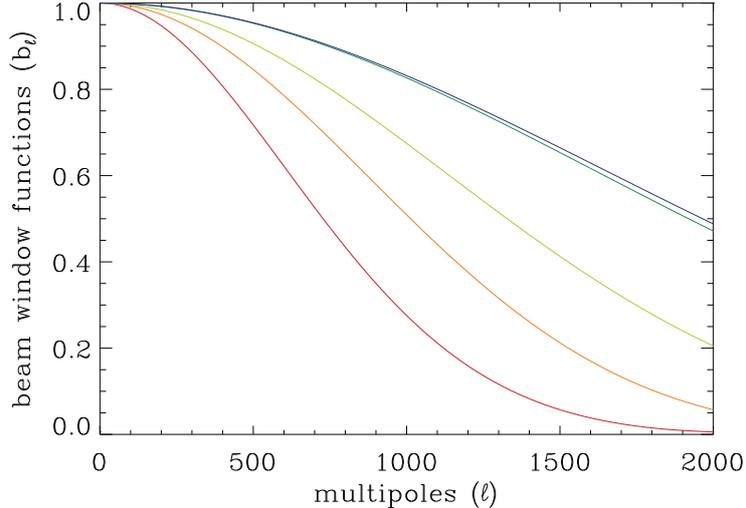


Figure 6.2: Beam window functions of the frequency bands from 70GHz to 353 GHz, respectively from red to blue. The data of an instrument is the result of a weighting by its own beam window function of the true signal on the sky. The highest the frequency, the highest the resolution. Thus, almost all information is suppressed at small scales in the lowest frequencies.

free parameters of $\beta BB^T + \phi$ is the number of free parameters in Φ . Thus, the covariance Φ being a $N_f \times N_f$ symmetric matrix, we have

$$N_{\text{pt}}N_f + N'_{\text{pt}} \leq \frac{N_f(N_f + 1)}{2}, \quad (6.9)$$

which becomes

$$N_{\text{pt}} \leq \frac{N_f}{2} \quad (6.10)$$

when equality in equation 6.8 is reached.

The parameters to sample are all the elements of the mixing matrix B and the non-zero entries of ϕ . The normalisation factor for the power spectra β is fixed. As for the mixing matrix \mathbf{A} , the Jeffreys prior for B is a flat prior. The Jeffreys prior for ϕ is $P(\phi) \propto \prod_i \phi_i^{-1}$, the product dismissing the vanishing ϕ_i .

6.2.4 Beaming

The instruments have a finite resolution. The signal seen in each direction is actually a weighted average of the underlying signal of neighbouring directions. Since the resolution is the same in each direction, the observed signal can be seen as a convolution on the sphere of the true signal with a function, called beam function. Each instrument has a specific resolution and therefore has a specific beam function.

The beams of the Planck instruments are very close to Gaussian functions. A convolu-

tion on the real sphere is a simple multiplication in harmonic space, just like between the real Euclidean space and the Fourier space. In addition, Gaussian functions on the sphere transform into Gaussian functions with inverse width in harmonic space, with the same value for all the modes of the same multipole. The transform of the real beam function is called the beam window function because just a range of scales is selected. This can be understood as follows: the lower the resolution of the instrument is, the narrower the beam window function is because the smaller scales of the true signal are smoothed. The beamed signal is thus less powerful at large multipoles, i.e. at small scales. Figure 6.2 shows the beams of 5 frequency bands of Planck.

The noise is not affected by the beams since it is not part of the sky signal. Let's denote $b_{i\ell}$ the beam factor at frequency i and scale ℓ . The signal only undergoes the beams such that equation 6.7 and 6.7 become

$$d_{i\ell m} = b_{i\ell} \sum_k A_{ik} s_{k\ell m} + b_{i\ell} \sum_h B_{ih} p_{h\ell m} + b_{i\ell} q_{i\ell m} + \sqrt{f_i} n_{i\ell m} \quad (6.11)$$

$$R_\ell = b_\ell \mathbf{A} C_\ell \mathbf{A}^T b_\ell + \beta b_\ell B B^T b_\ell + b_\ell \phi b_\ell + f N_\ell, \quad (6.12)$$

with $b_\ell = \text{diag}(b_{i\ell}; i \in \llbracket 1, N_f \rrbracket)$.

The beams are assumed to be perfectly characterised and are not sampled.

6.2.5 Masking, apodising, inpainting

The galaxy is very bright in the galactic plane and no model is good enough to remove it without leaving uncharacterised residuals. As I described in the previous chapter, simulations show that the contaminations in the recovered CMB map are the largest where the galactic signal is the largest. When the data are not under control, the best way of treating them is to not use them at all. Thus the usual way to get around the galaxy issue is to mask it. The Planck collaboration produced several masks that mask more or less aggressively the galaxy. A similar problem arises with the resolved point sources and they also need to be masked. Figure 6.3 shows the galactic mask and the point source mask.

After masking the data maps, no additional modelling enters in equations 6.11 and 6.12. However, masks introduce mode coupling. To correct for this bias, either the masked observation maps need to be pre-processed or the results need to be post-processed, or both, depending on the masks applied to the data.

Let's consider a random isotropic field on the sphere with power spectrum C_ℓ and a mask. Then the power spectrum \tilde{C}_ℓ of the masked field is related to the original power spectrum C_ℓ as follows

$$\tilde{C}_\ell = \sum_{\ell'} M_{\ell\ell'} C_{\ell'}. \quad (6.13)$$

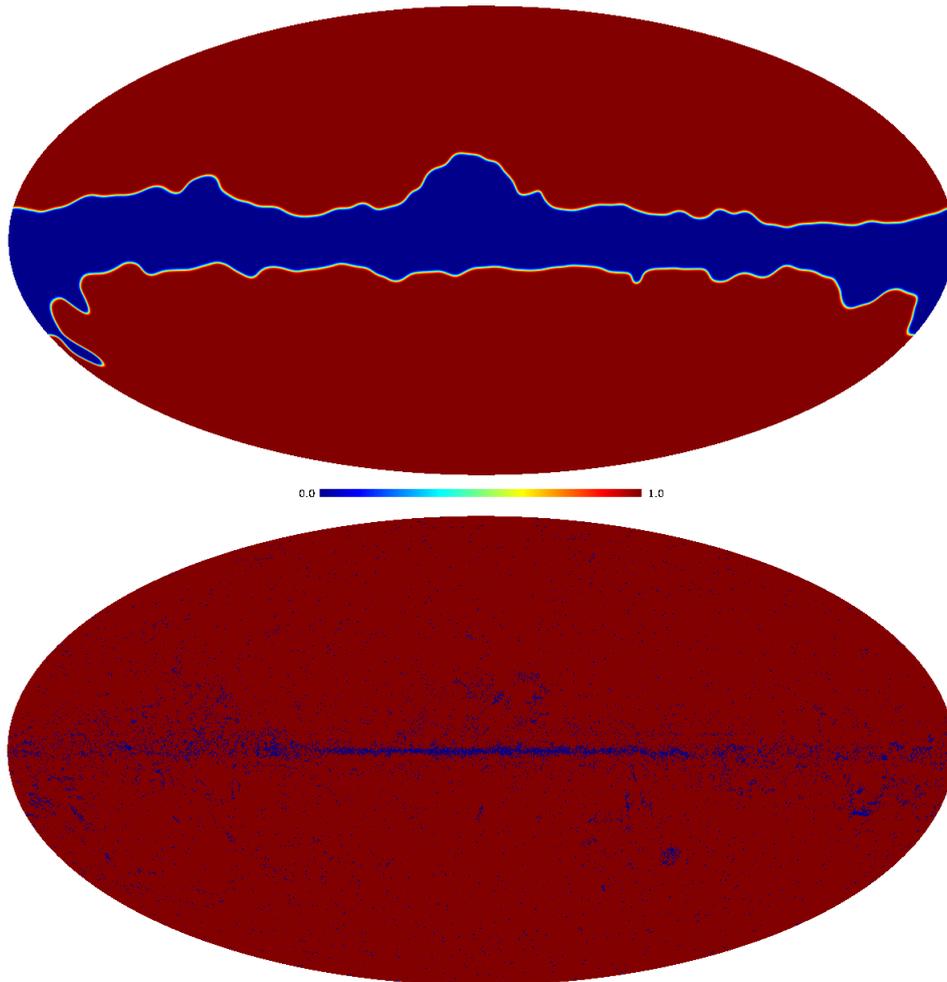


Figure 6.3: Galactic mask (top) and point source mask (bottom). The galactic mask leaves 80% of the sky. It was apodised with a Gaussian beam with full width at half maximum of 1 degree. That is why it takes continuously values between 0 and 1 at its edges. The point source mask masks about 1% of the sky outside the 80% galactic mask.

The coupling matrix M depends on the mask power spectrum only (e.g. Hivon et al. (2002)). Since masking is a simple multiplication in pixel space, equation 6.13 can be seen as a convolution in harmonic space. Thus, after having performed a CMB inference, correctly deconvolving the resulting power spectrum via the inversion of equation 6.13 is necessary to produce a rigorous final power spectrum. This step is deterministic and does not introduce further error.

Apodising is an operation that smooths the effect of masking. It consists in smoothing the edges of the mask in order to reduce the mode coupling, i.e. to make the coupling matrix M of equation 6.13 more diagonal. The galactic masks in figure 6.3 are apodised by smoothing them with a Gaussian beam that has a full width at half maximum of 1 degree. Thus, the pixels of a mask $\{m_p\}_p$ can take any value between 0 and 1. For the

following, let's define the fraction of the sky used

$$f_{sky} = \frac{1}{N_{\text{pix}}} \sum_p m_p^2. \quad (6.14)$$

The point source mask is not apodised because its effects are treated in a different manner.

Imputation, or inpainting, is a technique that fills the holes that a mask produces in an observation map with fake data. Inpainting is a general technique for image and video treatment so very sophisticated algorithms have been developed (see e.g. He and Sun (2012) for a review). However, I use one of the simplest methods, called diffusive inpainting, because despite its simplicity it is adapted to the problem of point source masking. Diffusive inpainting consists in filling in the holes by diffusing the observation map from the edges to the inside of the holes. Since the point source mask contains only small holes whose sizes do not exceed ten pixels or so, the spread of the smoothing due to the diffusion is weak. Moreover, diffusive inpainting solve the Laplace's equation, which leads to a $1/\ell^2$ spectrum, which is similar to the CMB spectrum on large scales. In order to correct for the small loss of power that the masking still induces at small scales, I add the noise realisation built from the half-ring half-difference maps in the inpainted region.

6.3 Possible configurations of the data

After pre-processing we have to make choices in the modelling because our model now allows many different configuration in terms of number of components, in particular for the restrictive point source components. In this section I list the possible data selections and model choices, and discuss the impact of these choices on the results.

6.3.1 Mask

Since just a fraction of the available data is used, the reconstruction is less constrained because the likelihood has a wider shape than in the case of a full sky analysis. Thus, the smaller the fraction of the sky used is, the larger the errors on the parameters are. I have chosen to work mostly with rather high sky fractions, that is f_{sky} between 80% and 97%, but I also used lesser sky coverage for consistency checks.

As discussed above, masking reshapes the power spectrum. This geometric effect needs to be treated for rigorous comparisons with the results of other methods. For these first test, I have just used the approximation that the coupling matrix M of equation 6.13 is proportional to the identity matrix (Hivon et al., 2002), with

$$M_{ij} = \delta_{ij} f_{sky}. \quad (6.15)$$

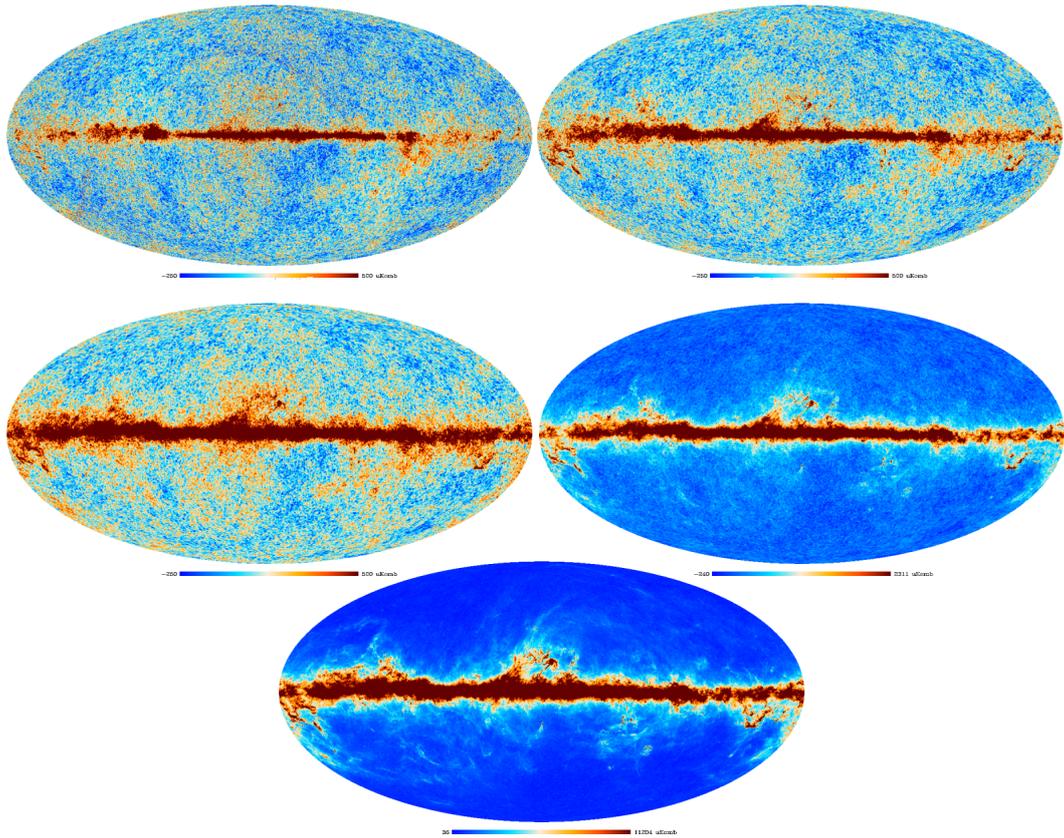


Figure 6.4: The five nominal maps 70GHz, 100GHz, 143GHz, 217GHz and 353GHz. This is the set I mostly used. The first three frequency bands have the same scale in order to show the CMB. The two other maps have a larger dynamic range to keep the galaxy from spreading all over the sky. Far outside the galactic plane even the low frequencies are dominated by point sources on small scales.

This is motivated by the fact that I work on a large fraction of the sky, i.e. with small masks that have low level of expected coupling.

6.3.2 Frequency range

The Planck instruments observe nine frequency bands. The bands have to be chosen with care because there is a trade off. The middle bands are the cleanest. The highest and the lowest bands are very contaminated by dust in the infrared band and synchrotron in the radio band, respectively, but also contain more information on the foregrounds to clean the CMB.

The number of assumed components in the data can not exceed the number of frequency bands chosen for the analysis without opening up flat directions in the parameter space. If a large fraction of the sky is used, then the physical emissions are complex, including deviation from coherence. In this case, multiple components are necessary to fit

one foreground. So when using small masks, a maximum of dimensions in the component space is necessary, and therefore many frequency bands are needed.

Since the current version of the method works with the same multipole range for each frequency band, using low resolution maps imposes limitations. Some of the observation maps, particularly that from LFI, are almost unusable for an analysis that uses multipoles above $\ell \sim 2000$. I have worked mostly with the frequency bands ranging from 70 to 353 GHz, i.e. with a number of frequency bands $N_f = 5$. More recently, I have been using the $N_{\text{side}} = 2048$ version of the 70GHz observation map.

6.3.3 Multipole range

There are several reasons for not taking the full range of multipole available. At large scales, i.e. low multipoles (below $\ell \sim 150$), the galactic foregrounds dominate, even after masking. If the separation is performed on a large range of multipoles, then the constant mixing matrix through frequency is constrained by very large scales as well as low scales, which can reduce the efficiency of the sampling. Foregrounds are more non-Gaussian on large scales than on small scales, they are also more dominant and more anisotropic. In addition, they do not necessarily mix the same way at small and large scales, like it is assumed in the model *via* the constant mixing matrix. Therefore there are several special problems at low multipoles that need to be treated separately. Furthermore, due to the cosmic variance, rejecting the first hundred multipoles does not affect the constraints on the cosmological parameters by much. At small scales, i.e. large multipoles (above $\ell \sim 2000$), noise and point sources dominate and become sources of high contamination and uncertainty. For all these reasons, I mostly use multipoles ranging from $\ell_{\text{min}} = 50$ to $\ell_{\text{max}} = 2000$.

6.3.4 Number of components

Some methods, like Remazeilles et al. (2011), find the number of dimensions of the component subspace from the data themselves. In the general case, the number of components is a limitation of many component separation models. Indeed, it has to be chosen prior to the analysis and is fixed during the separation. Furthermore, in BICA the recovered foregrounds are not physical which means that the true emissions distribute over the available dimensions in order to minimise the mismatch as much as possible. Thus, the optimal number of components that will balance chi-squared reduction, convergence and economy of parameters can only be chosen after several trials. For the frequency range and the multipole range I chose to work with, the assumed number of components N_c in the separations I have ran are maximal or almost, i.e. $N_c = 4$ or 5 .

In principle, one could explore a Bayesian model selection approach which chooses

the number of components necessary to model the data. Here, I take a model checking approach (see section 5.4) where I use the mismatch as a quality indicator for the analysis.

6.3.5 Point source model

The choice of exact point source model has important consequences for the CMB reconstruction on the small scales. There are numerous different parametrisations of the point source components *via* the numbers of uncorrelated and correlated point source components. The choice of a mask, of a frequency range and of a multipole range is just a data selection. The number of components is not a critical parameter of the model. Too few of them will increase the mismatch between the model and the data. Too many of them will just create convergence problems or degeneracy problems. Physically, the number of assumed components in the data is never too large because the number of physical components in the data is always large (cf. chapter 2). The situation with the point sources is different because there are degeneracies between them and the CMB in the posterior. Thus, a model that overfits the point sources will remove power from the CMB at small scales and a model that is not sufficient to capture all the diffuse point sources will lead to leakage into the CMB power spectrum at small scales. I mostly have tried $N_p = 1, 2$ correlated point source components and N'_p uncorrelated point sources such that $0 \leq N'_p \leq N_f - N_p$.

I mainly used $N_p = 1$ and $N'_p = N_f - N_p$. This configuration is motivated by the fact that this choice amounts to choose one point source map that is coherent through frequencies and to correct for non coherence with one uncorrelated point source map per channel. The Jeffreys prior applied to the uncorrelated point sources will act as a sparsity prior by driving to zero the correction that are not needed.

6.3.6 Choice of prior

The prior on the parameters is the same as that in the case of application to simulations, except for the component auto and cross spectra. Because of burn-in and convergence issues, the Jeffreys prior is replaced by a flat prior. Using the Jeffreys prior increases the correlation between adjacent samples in the Markov chain. I therefore assess in this section the difference between using a flat prior and a Jeffreys prior. I performed importance sampling (e.g. Arouna (2003)) on the samples drawn from the posterior with flat prior in order to see the impact of the change of priors on the sample mean.

Let's $\{C_\ell^i; i \in \llbracket 1, N_{\text{sam}} \rrbracket\}$ be a collection of samples of the CMB power spectrum at multipole ℓ . Then an estimator of the mean would be the sample arithmetic mean

$$A = \frac{1}{N_{\text{sam}}} \sum_i C_\ell^i. \quad (6.16)$$

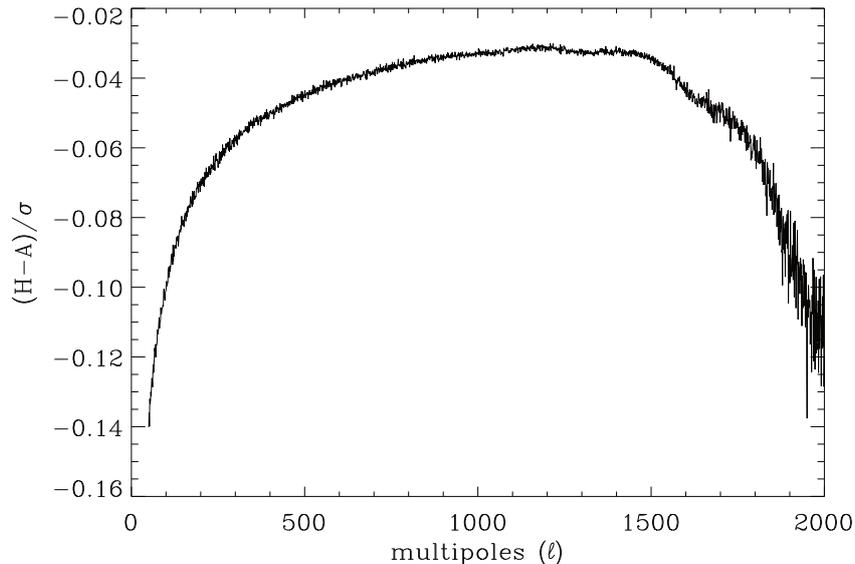


Figure 6.5: Difference between the sample harmonic mean and the sample arithmetic mean, divided by the sample standard deviation. The estimator of the mean the posterior containing the Jeffreys prior is lower than that of the posterior without a Jeffreys prior. The results deviates by a few percent of the error bars.

Now, if the samples are weighted by the Jeffreys prior, i.e. by weights $w_i = 1/C_\ell^i$ because the CMB is independent from the other component in the Jeffreys prior, then the estimator becomes

$$H = \left(\sum_i w_i \right)^{-1} \sum_i w_i C_\ell^i = N_{\text{sam}} \left(\sum_i \frac{1}{C_\ell^i} \right)^{-1}, \quad (6.17)$$

i.e. the sample harmonic mean. Figure 6.5 shows the difference between the sample harmonic mean and the sample arithmetic mean, divided by the sample standard deviation, for all multipoles, for case#1 (see section 6.3.7). In the worst cases, the mean deviates of a few percent of the error bars. Thus, the results on the CMB power spectrum are not altered by much, but an improved sampling scheme that takes the Jeffreys prior into account is necessary to recover the full properties of the posterior containing the Jeffreys prior.

6.3.7 Test cases

This section presents four data parametrisations on which BICA has been applied. The results are presented in the next section. In each case the number of assumed components and/or the fraction of the sky used and/or the range of multipole is different. All cases use the same frequency bands and the same point source model, whose parameters are

Frequency bands	$N_f = 5$	70GHz to 353GHz
Point source model	$(N_p, N'_p) = (1, 4)$	reference channel: 143GHz

Table 6.1: Table resuming the specifics of the common parameters to the four cases of table 6.2. Since $N'_p + N_p = N_f$, the reference channel of the point source model refers to the frequency band in which there is no uncorrelated point sources, i.e. the correction in all other bands are due to the non-coherence of the correlated point source map at this frequency.

	f_{sky}	N_c	ℓ_{min}
case #1	80%	4	50
case #2	70%	4	50
case #3	80%	5	2
case #4	90%	4	50

Table 6.2: Table summarising the differences between the four cases treated in this chapter. An additional component is needed in case #3 because a multipole range larger in the large scales is taken into account.

displayed in table 6.1. The frequency band range from 70GHz to 353GHz, i.e. $N_f = 5$, and the point source model is a mixture of one correlated point source component and four uncorrelated point source components, i.e. $N_p = 1$ and $N'_p = 4 = N_f - N_p$.

The cases differ in the number of components, which is either $N_c = 4$ or $N_c = 5$, the fraction of the sky used, which is either $f_{sky} = 70\%$ or $f_{sky} = 80\%$ or $f_{sky} = 90\%$, the range of multipoles, which is either from $\ell_{min} = 50$ or $\ell_{min} = 2$ to $\ell_{max} = 2000$. These choices are arranged in four cases gathered in table 6.2. Each case was ran took 3 days of CPU time.

As discussed above, the correlated part is a component that scales through frequencies and the uncorrelated part provides a necessary correction due to the slight non-coherence of the point sources. Since $N'_p + N_p = N_f$, there are N_f possible choices for a channel of reference for the correlated point sources. This choice should not change the results but changes the rate of convergence. Indeed, at low frequency the noise dominates the large scales and at high frequencies the galactic emissions dominates, even the small scales. As for the CMB, most of the information on the point sources is contained in the intermediate frequency bands. Changing the frequency of reference changes the shape of the posterior because then this PDF is defined on different parameters. If the reference channel is either the highest or the lowest frequency band, then the posterior is wider on the phenomenological parameters but the inference on the physical parameters is the same. Choosing the 143GHz band as the reference channel for the point sources pins down the correlated point source component such that the uncertainty on the uncorrelated point source components is minimal¹.

¹This is analogous to choosing to pivot of $P(k)$ for cosmological parameters analysis

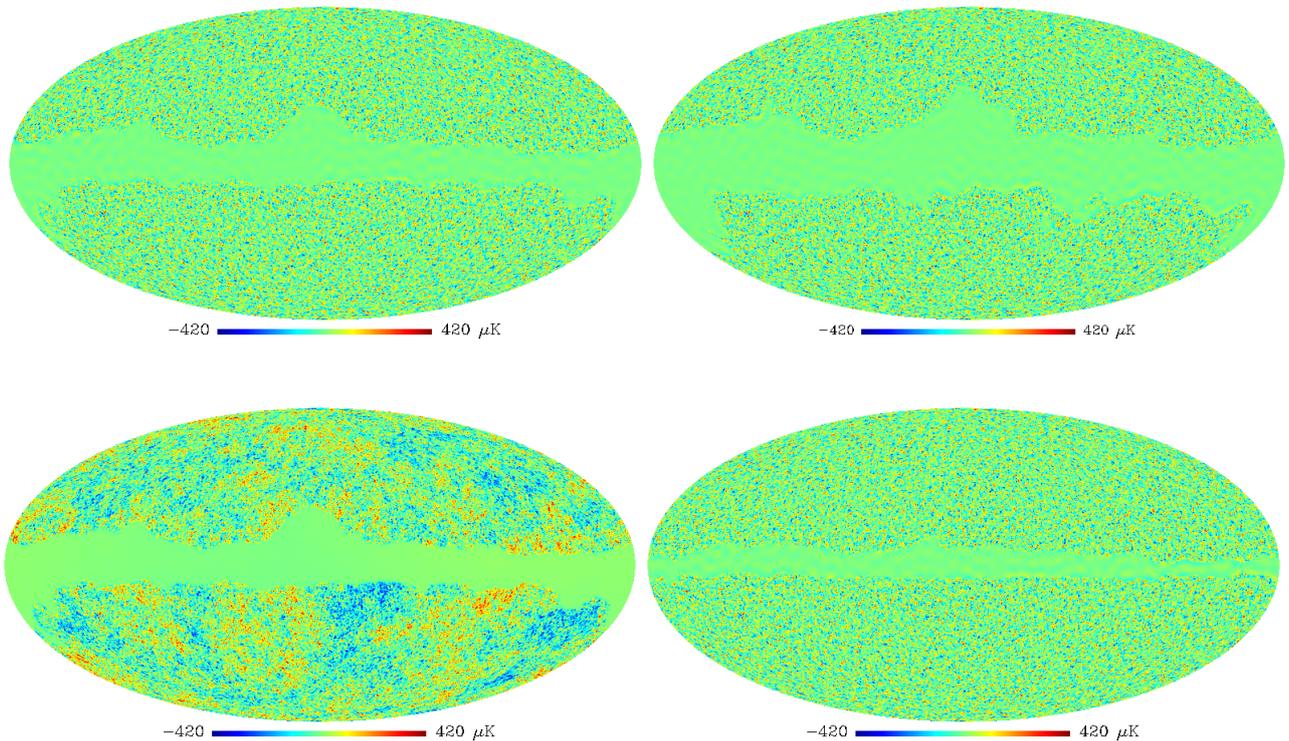


Figure 6.6: CMB sample mean map of the four cases. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. Large structures of the CMB can be seen only in case #3 because the other maps are constructed as if there were no power at multipoles lower than 50. One can notice on the maps stripped from the Sachs-Wolfe plateau the power of the first peak of the CMB that causes the most visible fluctuations at scale $\sim 1^\circ$. While the eye is drawn to the familiar large scale pattern in the CMB temperature map, case #3 is misleading because at the power spectrum level this map has discrepancies with the 2013 Planck power spectrum at small scales.

6.4 Results

This section presents the first results of applying BICA to the Planck 2013 data in the four cases gathered in table 6.2. Case #1 is the case using a maximum of data. Case #2 is a consistency check, the difference with case #1 being just a reduction of the fraction of the sky used. Cases #3 and #4 are two extensions of case #1. The first has a multipole range that goes down to $\ell_{\min} = 2$ and, the second has a larger f_{sky} .

6.4.1 CMB map and power spectrum inference

Figure 6.6 shows the sample mean maps of the four test cases. A future work will be to explore the difference maps and compare them to the error maps returned by the Bayesian inference, in order to check for consistency of the method.

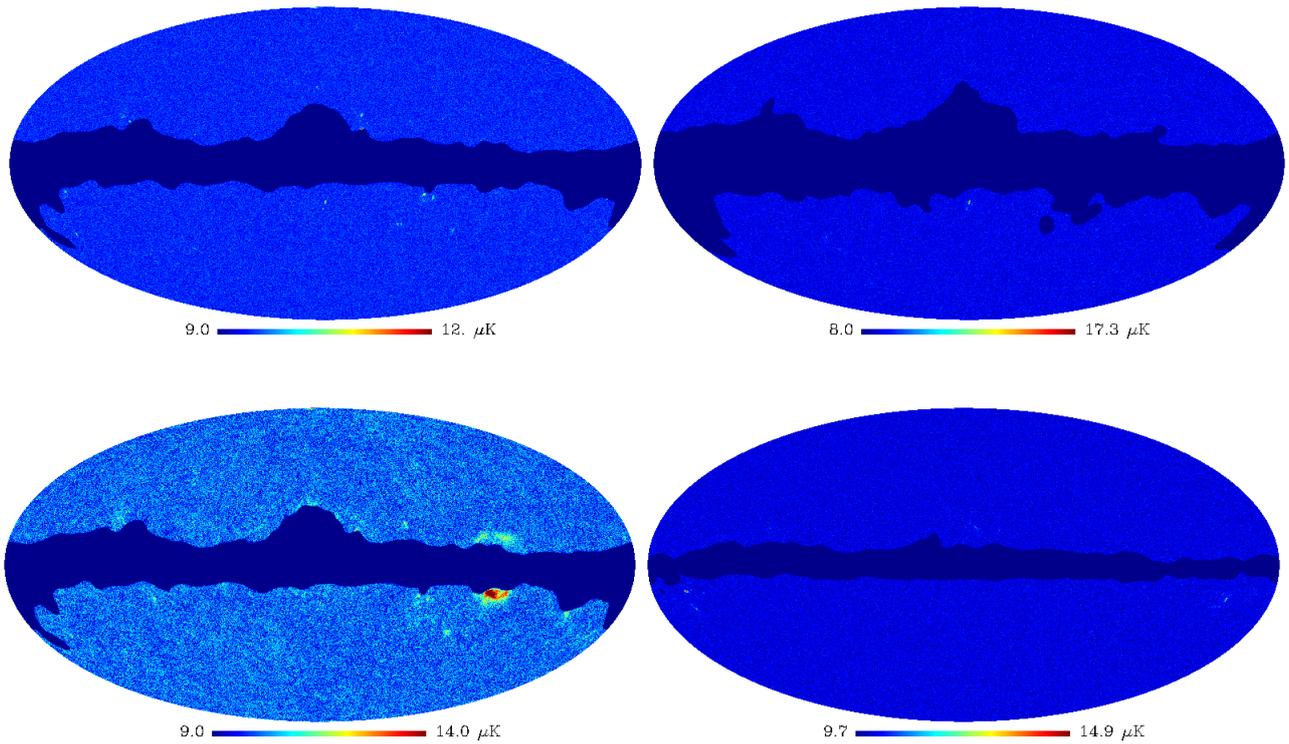


Figure 6.7: CMB sample standard deviation maps of the four cases. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. The errors are larger in regions where high foreground signals are not masked. In case #3, only a few hundreds samples of the CMB map are used to construct this error model. Thus, convergence towards the true value of the variance is not reached and some features do not stand out from the background.

Figure 6.7 shows the sample standard deviation maps of the four cases. Each map highlights the regions with highest uncertainties. As expected, the errors are larger in the regions of high foreground signal. The same features appears on all the maps. The large region of uncertainty around the galactic in the map of the case with increased multipole range (case #3) do not show up in the other maps. The galactic masks provided by the Planck collaboration do not mask these regions. Thus, the standard deviation map would be a useful tool to construct masks, which eventually are different for the range of multipole used. These maps do not reflect all the uncertainty encoded into the posterior. They only shows the variance in each pixel. However, it is clear that because the uncertainties are included in the analysis, the cases using a large fraction of the sky like case #1 can be used for cosmological parameter estimation or recover a clean CMB map. As it was done for simulations, correlation will have to be computed in order to better characterise the errors on the reconstructed CMB.

Figure 6.9 shows the inference of the CMB power spectrum. As for the simulation,

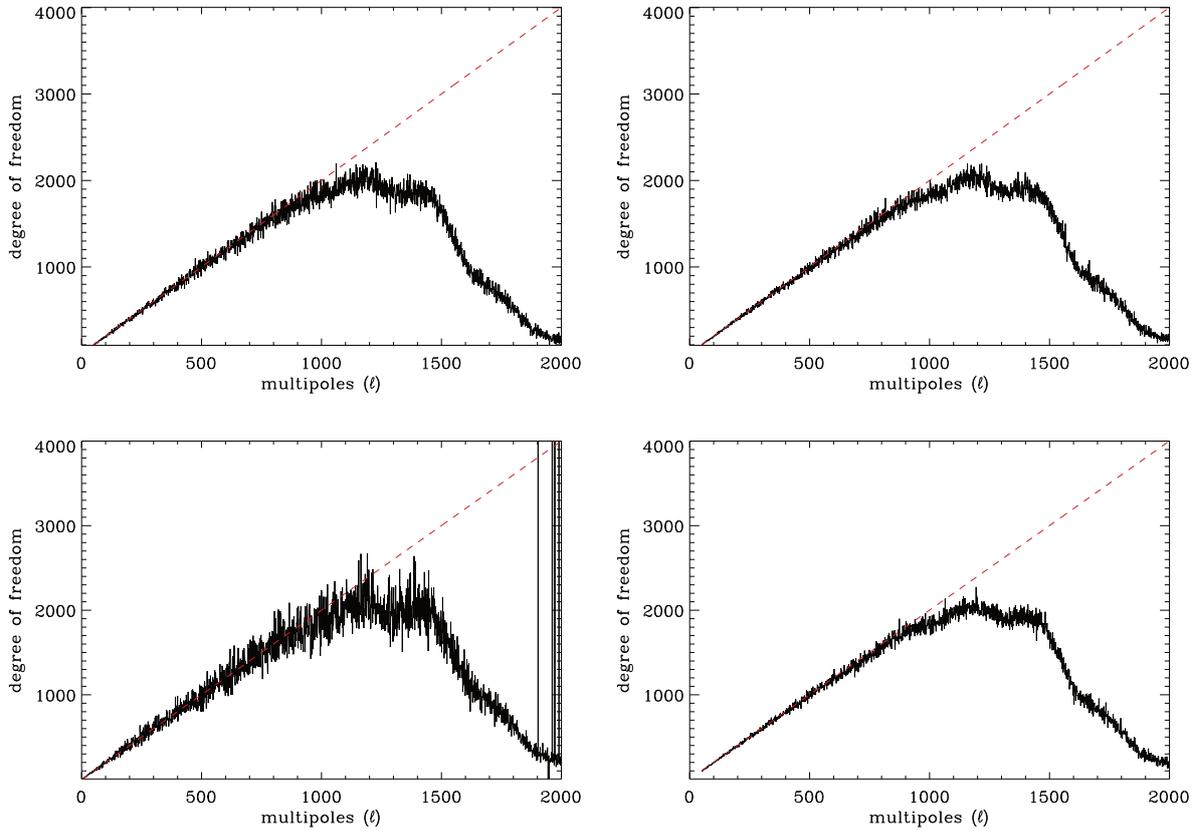


Figure 6.8: Number of degrees of freedom of the marginal distributions. The black line is the estimated number of degrees of freedom of the marginal distribution over all parameters but one multipole of the CMB power spectrum. The dashed red line represents the expected number of degrees of freedom in the case of a noiseless full sky CMB, i.e. $2\ell + 1$. Thus, this red line represents the maximum information on the CMB power spectrum that can be reached with one CMB sky only. The sampler of case #3 has not fully converged such that a few multipoles at small scales have non physical values. Using a larger range of multipole or a larger fraction of the sky seems to bring more information but since the inferred power spectrum has larger values at high multipoles in these cases, the relative error decreases.

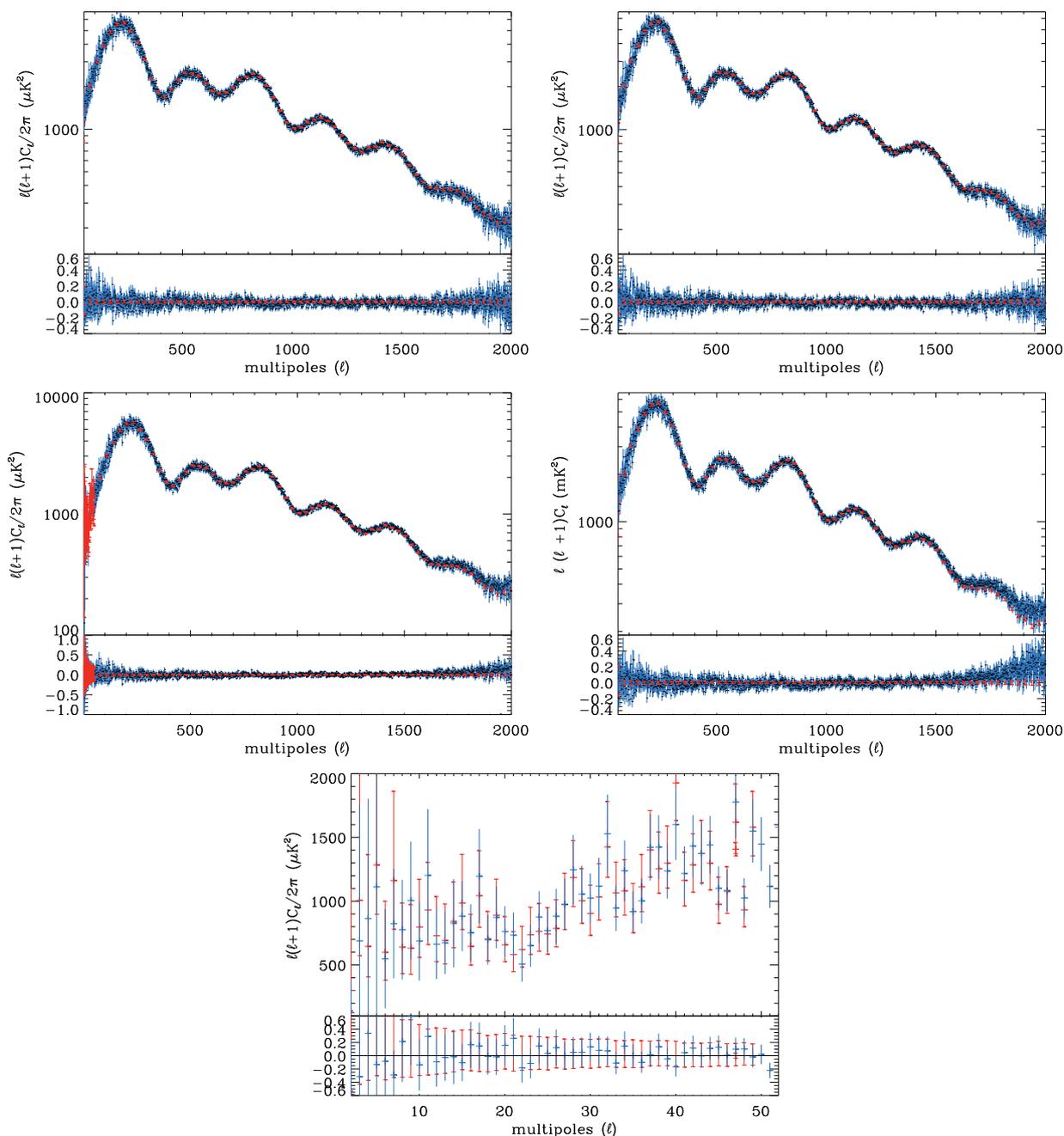


Figure 6.9: CMB power spectrum for the four cases. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. The fifth panel at the bottom shows the inference of the lowest multipole of case #3. In all figures, the black dots and the blue lines show respectively the mode and the variance of the marginal distribution. The red power spectrum is the 2013 Planck power spectrum. Cases #3 and #4 are consistent with the 2013 Planck power spectrum up to $\ell \sim 1500$, especially at very low multipoles for case #3, cases #1 and #2 are consistent with the 2013 Planck power spectrum on the whole multipole range considered for the analysis.

the black dots represent the mode of the marginal distribution of each multipole and the blue line around each black dots are the 1σ errors. For comparison the 2013 Planck CMB power spectrum is also plotted (red points). In each plot, the bottom panel contains the relative difference between the 2013 Planck power spectrum and the BICA power spectrum. The first two cases are consistent with the 2013 Planck spectrum, and the error bars are slightly larger in the case where $f_{sky}=70\%$. For the two other cases, a deviation appears at small scales. The inferred CMB has a larger value than expected because it is contaminated by point sources. The introduction of a fifth component in case #3 reduces the convergence rate such that some multipoles of the CMB power spectrum have not converged yet. It remains to be investigated what the result would be after convergence. In principle, since point source signal is significant at small scales only, there should not be any difference between cases #1 and #3. Since the diffuse point sources arise from a similar physical process than that of galactic emissions, part of the galaxy signal could be captured by the point source components. When a larger fraction of the sky is used, like in case #4, the regular components and the point source components are not able to compensate for the increase of power in the galactic plane. Thus, part of the signal leaks to the CMB power spectrum.

An inverse-gamma distribution fits the individual marginal distribution of each multipole of the CMB. The black line in figure 6.8 is the estimated number of degrees of freedom of the marginal distribution. The dashed red line represents the expected number of degrees of freedom in the case of a noiseless full sky CMB, i.e. the number of modes per multipole. Thus, this red line represents the maximum available information about the underlying CMB power spectrum with one CMB sky. Information is close to maximal at multipole $\ell \leq 1000$ but drops at small scales because of noise, beaming and point sources. Beaming and point sources were not involved in the test on simulations and the information about the CMB power spectrum was limited by the noise only. Using a larger range of multipole or a larger fraction of the sky, cases #3 and #4, slightly increases the information. But this might be due to the larger statistical power at small scales that the inference of the CMB power spectrum gives in these cases, and therefore the error bars are relatively lower.

6.4.2 Inference of the non-CMB components

Except the CMB, the recovered individual components are not physical. Since the CMB is extracted, the rest of the data is clean of CMB anisotropies. This section presents the inference of the total of the non-CMB and non-point source like components.

Figure 6.10 shows the sample mean map of the foregrounds at high frequency. Supposing that the CMB is clean of all foregrounds, these maps contain all the non-CMB components, except point sources, at 353GHz, and particularly the galactic emissions.

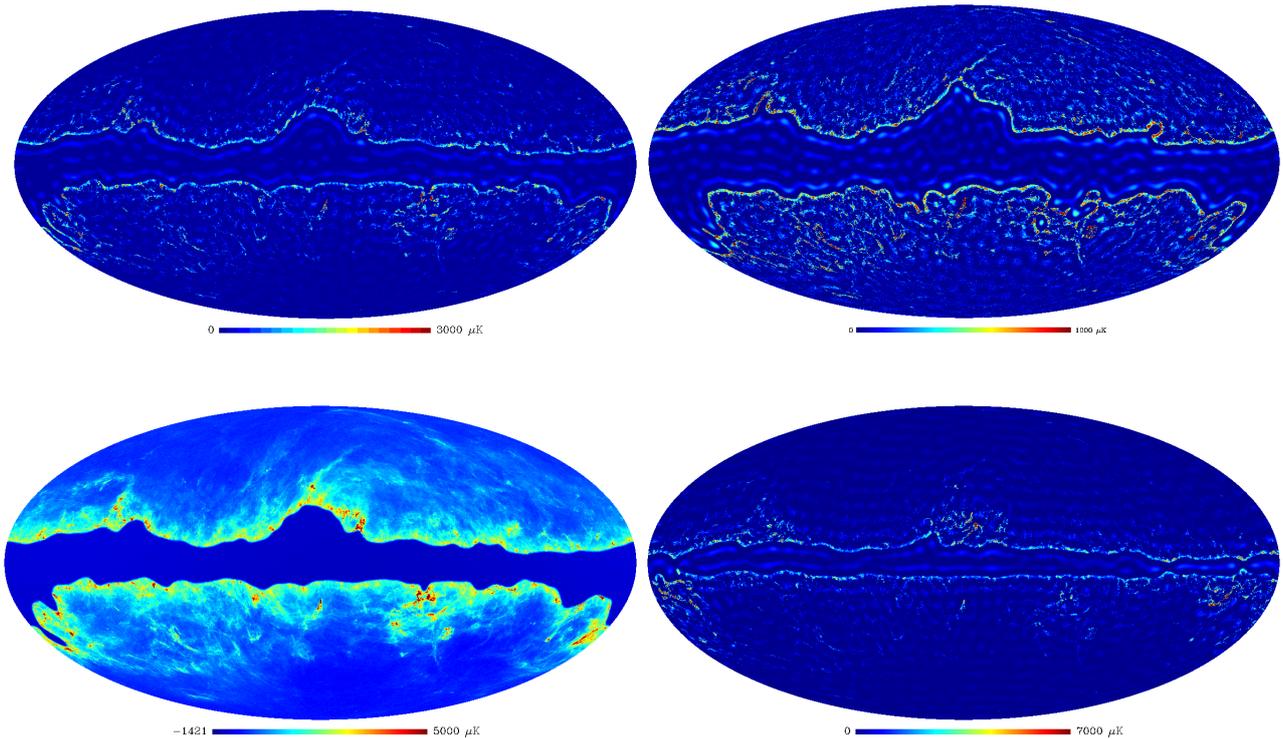


Figure 6.10: Foreground sample mean for the four cases. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. This is a map of the sum of all the non-CMB and non-point source components at the highest frequency band, i.e. 353GHz. Dust is the brightest component at high frequencies and is the component that spreads the most over the sky. That is why it is the main component in these maps. The ringing structure in maps of cases #1, #2 and #4 comes from cutting the map at $\ell_{\min} = 50$.

Dust has a spatial distribution that covers more sky than the other components. In addition, it is the brightest component in the frequency bands I used. Thus, the main visible component in these maps is the emission from dust of the ISM. Since the power of the galaxy lies principally in the lowest multipoles, the galaxy is more visible in case #3, where the minimum multipole included in the analysis is the lowest, $\ell_{\min} = 2$.

Figure 6.11 shows the inferred power spectra of the galaxy. Since the galaxy is not isotropic, the shape of its power spectrum changes if a different fraction of the sky is used for the analysis.

6.4.3 Consistency of the results

Figure 6.12 shows the mismatch D_ℓ of equation 5.5 between the data covariances of each multipole and the model covariance R_ℓ of equation 6.12. Case #2 is the case where the mismatch is the flattest because in this case the galaxy is heavily masked and less

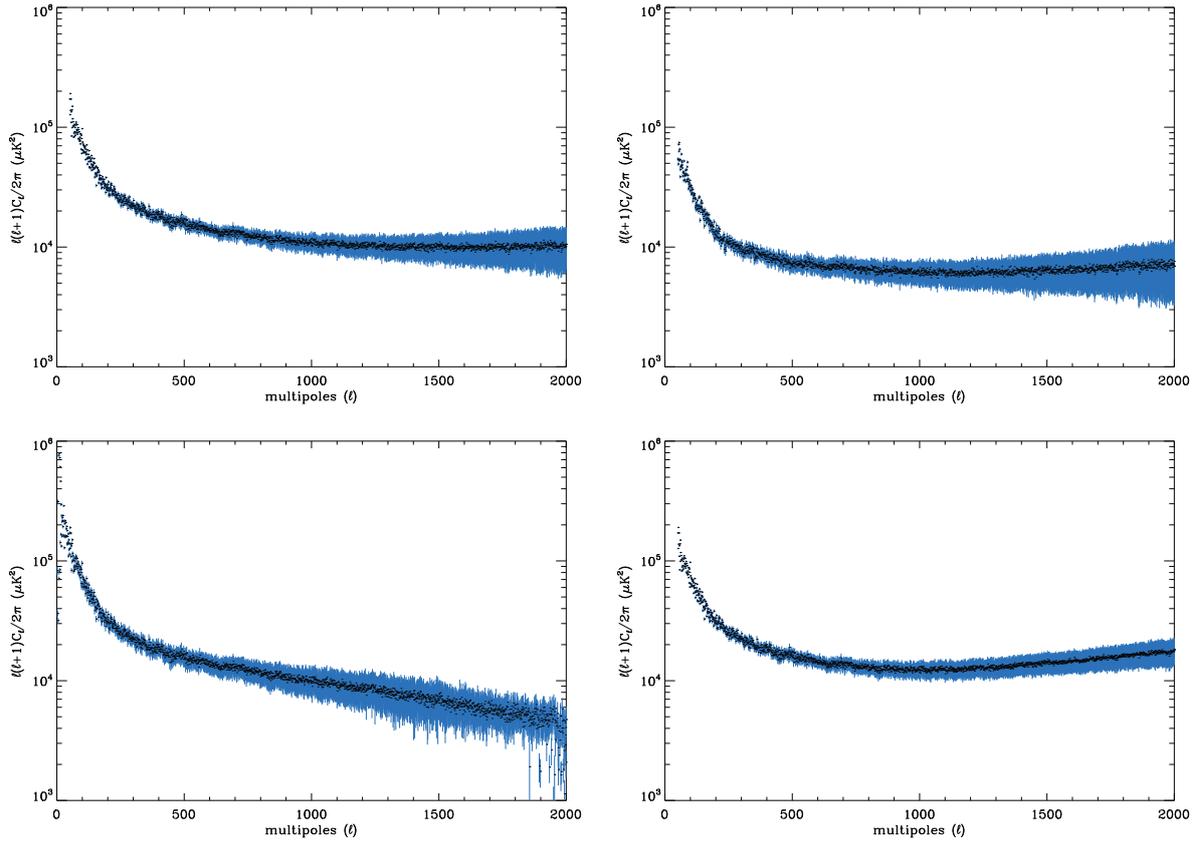


Figure 6.11: Non-CMB non-point source like component power spectrum for the four cases. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. As for the CMB, the black dots represents the peak of the individual marginal distributions at each multipole, and the blue bar around each dot represents the one-sigma error around the peak.

components are needed to describe it. In all other cases the mismatch rises at the largest scales, where the galaxy has more power, in a lesser extent in case #3 because more the assumed number of components is larger ($N_c = 5$ against $N_c = 4$ in cases #1 and #4).

As detailed in section 5.4.1, the number of degrees of freedom is found by subtracting the number of sampled parameters per multipole to the degrees of freedom of the data covariance of that multipole. However, for example, if two parameters were completely correlated in the posterior, then only one parameter, i.e. only one number of degrees of freedom, would be enough to explain both of them. Figure 6.13 shows the eigenvalues of the posterior correlation matrix of the component power spectra and cross spectra, at two multipoles. Only case #1 is shown here, but the same properties are found in all the other three cases. There are 7 eigenvalues because there are 4 components, including the CMB, and 3 cross-spectra because cross-spectra between the CMB and the other components are fixed to zero. In the case of uncorrelated parameters, the two lines would be flat and all eigenvalues would be 1. This is clearly not the case here and the effective

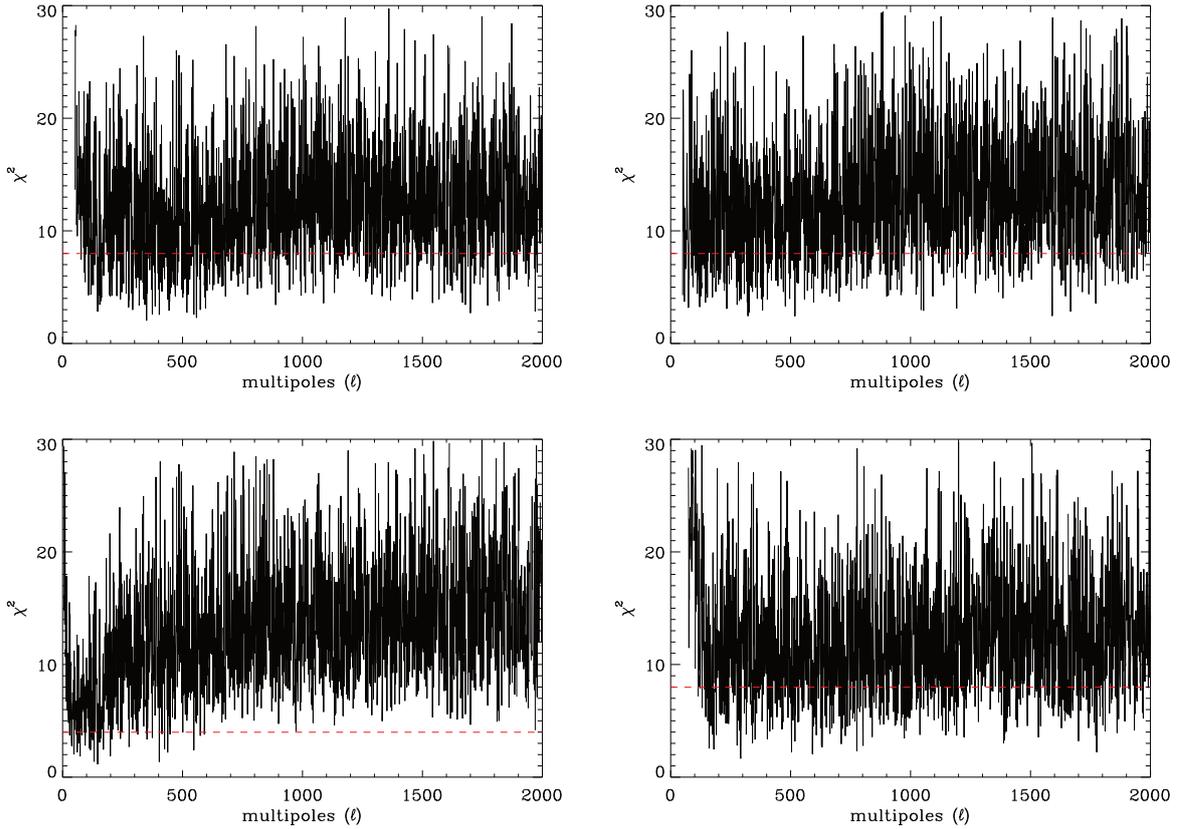


Figure 6.12: Mismatch between the data covariance and the model covariance at each multipole, for the four cases. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. Black line is the mismatch at each multipole, red dashed line represents the expected number of degree of freedom of the distribution of the mismatch at each multipole. The expected number of degrees of freedom is lower in case #3 because more parameters per multipole are fitted to the data. The number of degrees of freedom is higher than expected, maybe because of correlation between parameters (see text).

number of degrees of freedom is larger than shown in figure 6.12 which explain why the CMB spectrum is a good fit, consistent between cases #1 and #2, in spite of the offset in mismatch seen in figure 6.12. Furthermore, the blue and red crosses shows the eigenvalues of a correlation matrix at respectively low ($\ell \sim 100$) and high ($\ell \sim 2000$) multipoles. Thus, the eigenvalues of high multipoles are decreasing more rapidly than that of low multipoles, which could explain the difference of heights in the mismatch between low and high multipoles.

As another test for consistency, I directly compare the values of the data covariance and the model covariance R_ℓ at each multipole ℓ . Figure 6.16 shows the relative difference of these two quantities. The model slightly overpredicts the data power spectra, which is another explanation for a higher than expected mismatch.

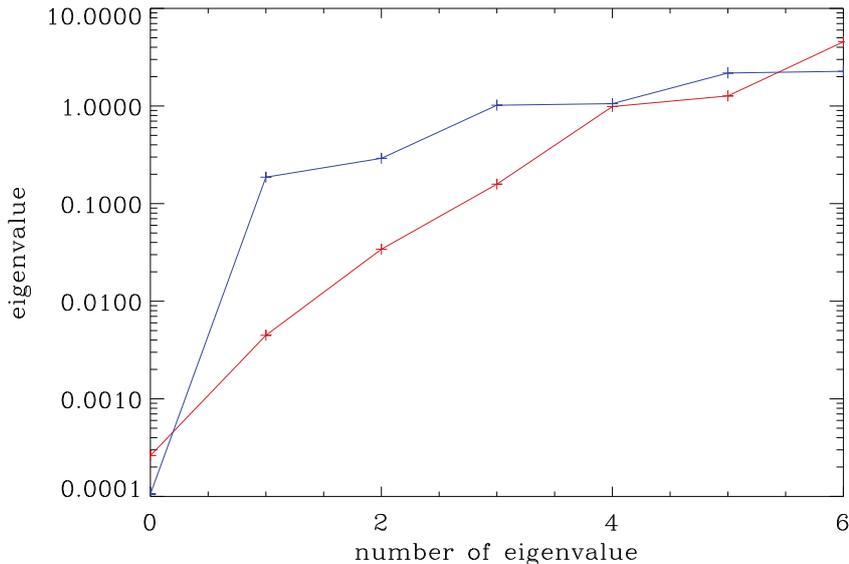


Figure 6.13: Two set of eigenvalues of the posterior correlation matrices of the parameters of two multipoles, one at large scale ($\ell \sim 100$, blue line) one at high multipole ($\ell \sim 2000$, red line) (see text).

6.4.4 Comparison with SMICA

Among all methods presented in chapter 3.3, SMICA is the one to which BICA is the most closely related. In this subsection I compare the SMICA map and the BICA map of the four cases. It is important to note that the procedures of BICA and SMICA for the Planck analysis are different in several aspects. The BICA model is more complete than the SMICA one because it includes a description of the point sources. SMICA uses a very large fraction of the sky ($f_{sky} = 97\%$) and assumes more components ($N_c = 6$) than BICA ($N_c = 4$ or 5). SMICA performs the separation on a large multipole range, $\ell = 2 - 4000$ and the power spectra are binned. SMICA takes the calibration errors between the observation channels into account, i.e. the CMB frequency response is a vector filled with ones, corrected by a factor close to 1 (at the subpercent level or less) in each channel.

Figure 6.14 shows the difference between the BICA map and the SMICA map. Red shows an excess in the BICA map, blue an excess in the SMICA map. The maps have been masked by the mask that I used for the analysis and the SMICA confidence mask. In cases #1, #2 and #4, i.e. in cases of minimum multipole for analysis $\ell_{\min} = 50$, the maps agree on most of the sky. Case #3 indicates that large scales may need a special treatment. Figure 6.15 shows the power spectra of the maps of figure 6.14.

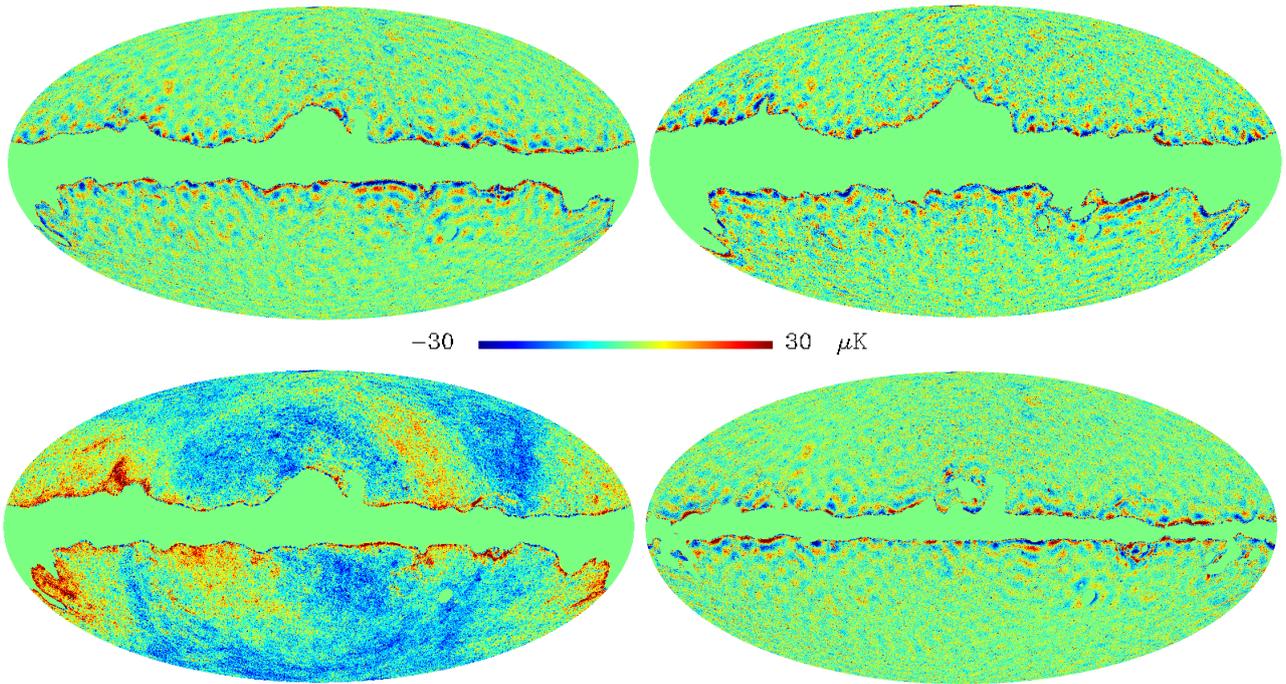


Figure 6.14: The difference map between the BICA map and the SMICA map for the four cases. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. Red and blue means excess in respectively the BICA map and SMICA map. Case #3 seems to indicate that galactic residuals are present in this map.

Conclusion

I have presented the first application of BICA to the 2013 Planck data. Although preliminary, the results are promising. BICA is able to recover the CMB power spectrum using a large fraction of the sky on large multipole range. The Bayesian framework and the phenomenological model of the components allows to construct an error map of the recovered CMB map that takes the presence of the foregrounds into account and that is free of any physical assumption – apart from the point source model. The results are consistent with the 2013 Planck results at the power spectrum level. BICA produces a rich data product ready for further and deeper analysis.

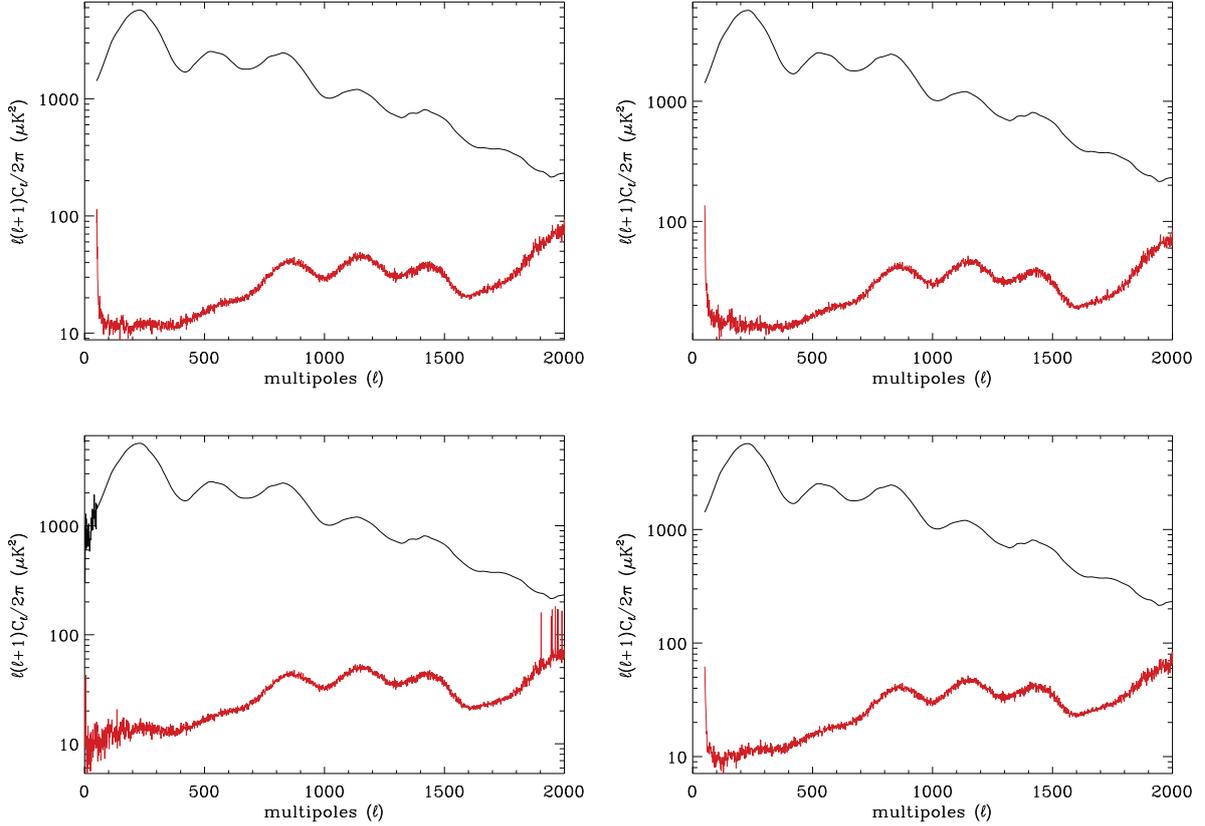


Figure 6.15: Power spectrum of the difference map between the BICA map and the SMICA map for the four cases, i.e. the power spectra of the maps of figure 6.14. Top left: case #1. Top right: case #2. Bottom left: case #3. Bottom right: case #4. The red line represents the power spectrum of the difference. For comparison, the 2013 Planck power spectrum is plotted in black. Case #3 suffers of foreground contamination at low ℓ the least because the model assumes 5 components. The oscillatory pattern may be due to masking, since the maps are masked with an unapodised point sources mask (the SMICA confidence mask and the point source mask used in BICA analysis).

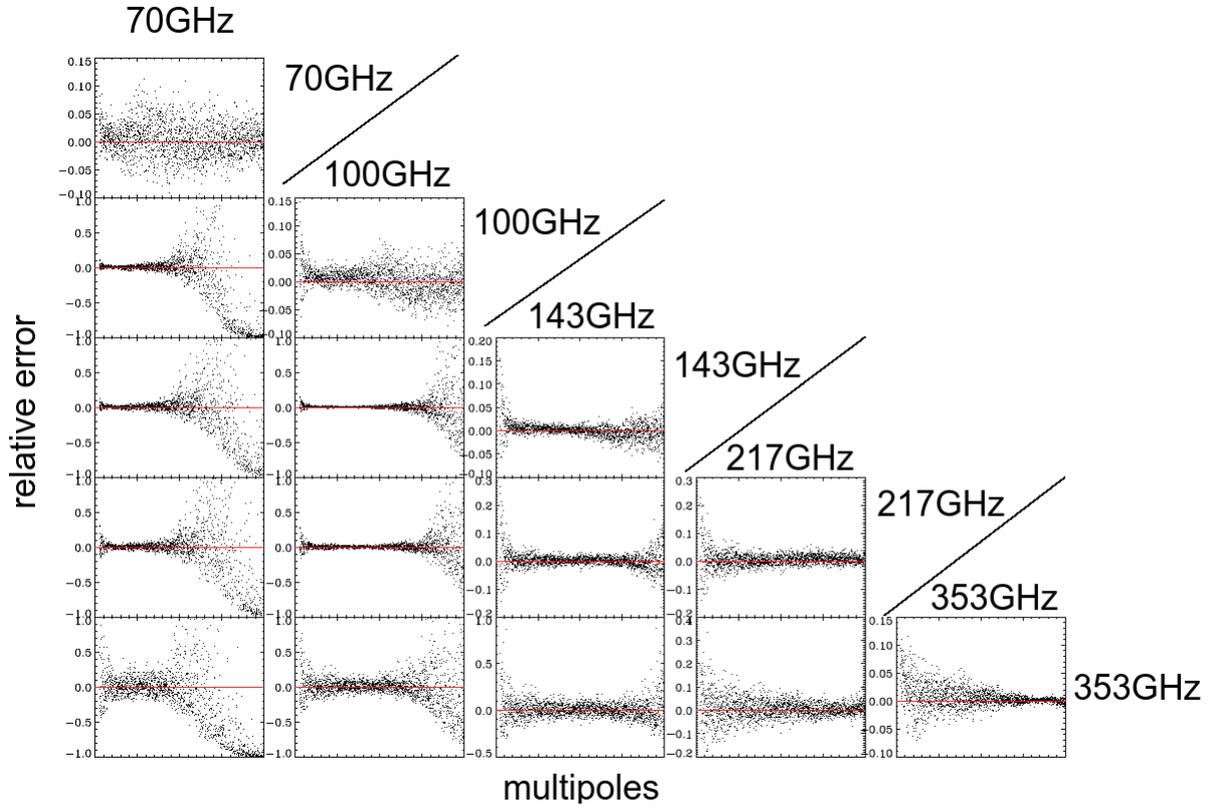


Figure 6.16: Comparison between the data covariance and the model covariance. Only case #1 is shown here. Each window shows the relative difference between an element of the data covariance and the model covariance, for the whole range of multipoles, i.e. between $\ell_{\min} = 50$ and $\ell_{\max} = 2000$. Each panel is a frequency cross spectrum $A \times B$ with $A, B = 70\text{GHz}, 100\text{GHz}, 143\text{GHz}, 217\text{GHz}, 353\text{GHz}$. The black dots represent the relative difference, red line is zero. There is an overall good fit. the discrepancies in the 70GHz channel comes from the fact that the data are very noisy above $\ell \sim 1000$ at this frequency band. An offset between the data and the model is present in the auto spectra in frequency bands between 100GHz and 217GHz, which could explain the excess of mismatch in figure 6.12.

General conclusion

Accomplishments

In this thesis I have presented a new blind formulation of the CMB inference problem, named BICA. For all-sky and high resolution data, it unifies parts of the analysis that have previously been treated separately, such as component separation and power spectrum inference. I have also presented previous component separation methods and we saw that two approaches are exploited: blind source separation, where the assumptions about the sources mixed in the signal are minimal, and physical parametrisation, where the sources are modelled as faithfully as possible. The most attractive advantage of blind source separation is to provide results without a priori knowledge that could bias the results if the physical assumptions are approximate. Thus, the approach of BICA relies on a phenomenological model of the multi-frequency microwave sky without the need for physical models of the individual components.

BICA fully makes use of Bayesian statistics. In this framework, I construct a posterior PDF free of almost any physical assumption and I describe an efficient sampling scheme that fully explores the component separation uncertainties on the inferred CMB products such as maps and/or power spectra. The method constructs a PDF over the CMB map and power spectrum directly from the observation maps. The shape of the PDF, and therefore the error model, accounts for the presence of foregrounds in the data.

External information about individual components can be incorporated as a prior giving a flexible way to progressively and continuously introduce physical component separation from a maximally blind approach. I showed that from the general formulation of BICA and by ad hoc prior choices, it is possible to connect the Bayesian formalism to existing component separation methods such as Commander, SMICA, ILC and SEVEM.

I have presented the results of the application of the method to both simulations and data.

In the case of the application to a simple set of simulations, BICA infers correctly the CMB power spectrum. The statistical properties of the recovered CMB power spectrum are close to what is expected if a clean full sky map of the CMB were available. The map is also well recovered, although the residuals point to the need of masking a small part

of the galactic plane.

Additional modelling was necessary before the application to the 2013 Planck data. The main differences with the simulations are the approximate estimation of the noise power spectra and the inclusions of beams for the instrumental modelling, the masking of the observation maps, the inclusion of cross-spectra between the non-CMB components and the special treatment of diffuse point sources for the physical modelling. The results reflects only works in progress but I have shown that the inference of the CMB power spectrum seems robust, since different treatments of the data lead to the same results. Also, I show consistency with the 2013 Planck power spectrum (Planck Collaboration et al., 2014d) which is obtained with a different methodology.

The results provide a point of departure for more in-depth analysis, for example cosmological implications and the impact of the foreground removal uncertainties on cosmological parameters. Given the promising results of this first application, additional choices for data selection and model choice remain to be explored.

Future works

First, although the robust inference of the CMB power spectrum, all the results of the application of BICA to Planck data need further studies. The masking effects even though small need to be addressed for example using MASTER (Hivon et al., 2002). Full inpainting of the maps, involving Gaussian realisations under constraints, would also be a way of avoiding the effects of masking. Due to slow convergence, few map samples were available to present the results in this thesis. More statistics will help to get a deeper understanding of the results.

The impact of the choice of priors needs to be addressed. BICA is capable of incorporating various priors and a continuous range of values for prior parameters. Thus, many cases have to be tested. In particular, adding a non-zero mean in the Gaussian prior of the components would reduce the non-Gaussianity and the anisotropy of the galactic residuals. In this case the model would be closer to the case in which BICA is optimal.

Thanks to its phenomenological method and the possibility of introducing information with various degrees of confidence, BICA is a very flexible method and it will take time to explore all of its possibilities. In particular, the posterior PDF on the CMB can be used as input for a likelihood analysis of the cosmological parameters. Alternatively, a parametrisation of the CMB power spectrum while keeping a blind description of the foregrounds would lead to a direct inference of the cosmological parameters within the CMB cleaning step. The errors would then incorporate the uncertainties due to all the intermediate steps, from the observation maps to the cosmological parameters, that are so far needed. Bayesian model provide a way to compare cosmological models. This

method was applied on WMAP data without considering component separation (Trotta, 2007). In the case of a parametric formulation of BICA, Bayesian model selection would compare cosmological model directly from the observation maps.

Application to polarisation is a natural extension of the method. The 2014 release of the Planck data, expected for December of this year, will include polarisation data but tests on simulation would be necessary before the application to the data.

General covariance learning

The previous paragraph discusses applications to CMB analysis only. However, BICA can be seen as a particular case of a general method for covariance learning that has applications beyond the scope of CMB analysis.

Thanks to phenomenological description of the data, it can be applied to any kind of inverse problem. For example it could be applied to detection of the 21cm line emission from the reionisation epoch (e.g. Madau et al. (1997)). The 21cm signal is dominated by foregrounds by several orders of magnitude. Thus, a blind source separation is necessary because the slightest error in the the foreground modelling induce a huge leakage in the cosmological component.

The CMB data analysis method presented in this thesis is a particular case of a more general covariance learning method. The work of thesis shows the power of the Bayesian approach. Writing the component separation problem by a blind description and solving it in a Bayesian framework opened new perspectives in CMB data analysis.

Appendix A

Statistical basics

For any data analysis, statistics are a necessary tool. Sophisticated and powerful methods have been developed over the past 25 years especially for CMB analysis since COBE, the first mission to observe the CMB anisotropies (Smoot). Statistics and probabilities are well suited for CMB analysis because the CMB *is* a statistical object. Indeed, our CMB anisotropy map is a picture of quantum fluctuations from inflation that have barely evolved.

This appendix gives informal but practical definitions of the mathematical objects that I use to build the BICA method (see chapter 4) and analyse its products.

A.1 Random variables, distributions and probability density functions

A **random variable** is a variable that can take a whole set of values. Each value is more less possible, the likeliness is quantified by a probability. The **probability distribution** – or simply distribution – is a set that gathers all the probabilities. In the scope of this thesis, four quantities associated to random variables and their distributions are important¹

- the **mean** or **expected value** of a random variable, which is the average value that the random variable takes after having change its value according to its probability distribution an infinite number of time,
- the **mode** of a distribution, which is the most probable value of the random variable,
- the **standard deviation** of a random variable, which quantifies the average deviation from the mean after an infinite number of changes in the value of the random variable, noted

¹For some particular probability distributions, part of these quantities do not exist.

- the **correlation** between two random variables, which is a quantification among others of the influence of one variable on the other one (and conversely), it is a dimensionless quantity and can take values in $[-1, 1]$.

Equivalently, we can define the **variance** of a random variable which is the squared standard deviation and the **covariance** between two random variables which is the correlation times the standard deviations of the variables of interest.

Let X be a continuous random variable than can take any real value. Then it has a probability distribution and the function $f : \mathbb{R} \rightarrow \mathbb{R}_+$ is the **probability density function (PDF)** of X if

$$\forall x \in \mathbb{R}, \quad P(X \leq x) = \int_{-\infty}^x du f(u) \quad (\text{A.1})$$

where $P(X \leq x)$ is the probability of the event "The value of X is lower or equal to x ". Since $P(X \leq +\infty) = 1$ for any distribution, a PDF f is normalised to 1. Unlike probabilities, a PDF can take values greater than 1. It gives the relative probabilities between events. For example, if $f(x) \geq f(y)$ then the event "The value of X is equal to x " is more probable than the event "The value of X is equal to y ".

The variable X is not necessarily defined on \mathbb{R} and can be, for example, a vector of length n or an $n \times n$ square matrix. In these cases, f still is real-valued but is defined on \mathbb{R}^n or the set of matrices $\mathcal{M}_n(\mathbb{R})$.

A.2 Gaussian distribution and related distributions

The most common PDFs that come across in probability are the PDFs of Gaussian processes. A random vector that has a **Gaussian distribution** can take any value in \mathbb{R}^n and its PDF \mathcal{N} is

$$\forall x \in \mathbb{R}^n, \quad \mathcal{N}(x | \mu, \Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) . \quad (\text{A.2})$$

The two parameters μ and Σ are respectively the mean and the covariance matrix of the Gaussian variable. If an infinite number of realisations of the Gaussian random variable were available then μ would be the averaged value of the realisations, i.e. the mean. The diagonal elements of Σ are the variances of the entries of the random vector and the off-diagonal of Σ are the covariances between the entries of the random vector.

In equation A.2, if Σ is the random variable and if x and μ are fixed, then we get a PDF for the covariance Σ , up to a normalisation factor. A generalisation of this distribution

is the **inverse-Wishart distribution**, whose PDF is

$$\forall \Sigma \in \mathcal{M}_n^+(\mathbb{R}), \quad \mathcal{W}^{-1}(\Sigma | \Psi, \nu) = \frac{|\Psi|^{\frac{\nu}{2}}}{2^{\frac{\nu n}{2}} \Gamma_n(\frac{\nu}{2})} |\Sigma|^{-\frac{\nu+n+1}{2}} \exp -\frac{1}{2} \text{tr}(\Psi \Sigma^{-1}). \quad (\text{A.3})$$

where $\mathcal{M}_n^+(\mathbb{R})$ is the set of the positive-definite matrices and Γ_n is the multivariate Gamma function. The parameter ν is the number of degrees of freedom of the distribution. From a likelihood point of view, the more available data there is, the highest the number of degrees of freedom is and the sharpest the distribution for Σ is. The univariate specialisation of the inverse-Wishart distribution is called the **inverse-gamma distribution**.

In equation A.2, the exponent $Y = (X - \mu)^T \Sigma^{-1} (X - \mu)$ has its own distribution. It is closely related to the Gaussian distribution and its PDF can be derived from the n -variate Gaussian PDF. The distribution for the random variable Y is called the **chi-squared distribution** with n degree of freedom and its PDF is

$$\forall x \in \mathbb{R}_+, \quad \chi^2(x | n) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{(n/2-1)} \exp(-x/2). \quad (\text{A.4})$$

It is defined on \mathbb{R}_+ because Σ^{-1} is positive-definite.

A.3 Kullback-Leibler divergence

The Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) is a statistical tool that measures the dissimilarity between two distributions. If p and q are the PDFs of two distribution P and Q , then the KL divergence of P and Q is defined as

$$D_{KL}(P, Q) = \int_{\mathbb{R}} dx \log \left(\frac{p(x)}{q(x)} \right) p(x). \quad (\text{A.5})$$

In this thesis, the bijection between a zero mean Gaussian distribution P and its covariance Σ_P will lead me to write $D_{KL}(\Sigma_P, \Sigma_Q)$ rather than $D_{KL}(P, Q)$. Note that the KL divergence is not symmetric in P and Q and that it is additive for independent distributions, e.g. if $p(x, y) = p_1(x)p_2(y)$ and $q(x, y) = q_1(x)q_2(y)$ then

$$D_{KL}(P, Q) = D_{KL}(P_1, Q_1) + D_{KL}(P_2, Q_2). \quad (\text{A.6})$$

A.4 Shannon entropy

Shannon entropy measures the information content of a random variable. It is defined as follows. Let X be a random variable with PDF f_X . Then its entropy H is

$$H(X) = \int dx f_X(x) I(x) = - \int dx f_X(x) \log f_X(x). \quad (\text{A.7})$$

The less probable the outcome of a random variable is, the more information $I(X) = \log\left(\frac{1}{f_X(X)}\right)$ it brings. Thus, the Shannon entropy is the sum of the information of all values weighted by the probability of their appearance.

Appendix B

Bayesian inference

Our universe – and *a fortiori* the CMB – is the only realisation of universes that is at our disposition. Bayesian statistics is a probabilistic framework in which the data of a one-event experiment can be treated. Bayesian inference is often opposed to frequentist estimation, a statistical framework in which the accuracy of the results depends on the number of times the experiment has been conducted. For example, frequentist statistics are widely used in particle physics because experiments can be repeated as many times as wanted, thanks to particle accelerators. The component separation that I have been developing during my thesis is based on Bayesian statistics. This appendix focuses on the basics of Bayesian analysis that are needed to understand the method and the interpretation of its products.

B.1 Bayes' theorem

In data analysis, Bayesian inference is a particular interpretation of the Bayes' theorem. Formally this theorem gives a relation between the probabilities of two events A and B and their conditional probabilities:

$$P(A|B)P(B) = P(B|A)P(A). \quad (\text{B.1})$$

Now, let say I have some data and a model that might explain the data and, in the Bayes' theorem, let A be the event "Each parameter of the model has a certain value" and B the event "The data has a certain value". For that specific model, I want to know the probability of the event A to be true, with an analysis based on the data. In other words, I ask the question "What is the probability of the parameters to take certain values, knowing the value of the available data". By definition, $P(A|B)$ is the answer to this question.

The Bayesian way of thinking data analysis is to use the Bayes' theorem as follows

$$P(\text{hypotheses} \mid \text{data}) \propto \mathcal{L}(\text{data} \mid \text{hypotheses})P(\text{hypotheses}). \quad (\text{B.2})$$

$P(\text{hypotheses} \mid \text{data})$ is called the **posterior** probability, $\mathcal{L}(\text{data} \mid \text{hypotheses})$ the **likelihood** probability and $P(\text{hypotheses})$ the **prior** probability. As soon as the data is acquired, $P(\text{data})$ has taken some fixed value that will never change, thus we can disregard this normalisation quantity. The likelihood measures how likely the available data is, given the assumed model. The prior is the probability of a particular event to occur, even before the acquisition of the data. Therefore the prior does not depend on the data, it depends on the analyst's choice.

A Bayesian analysis should provide the probability of any event. It is much more common to make use of probability density functions rather than just probabilities. Doing a Bayesian inference of some model's parameters is by definition evaluating the posterior PDF over these parameters with the use of Bayes' theorem.

B.2 Jeffreys priors

As it can be seen in equation B.2, the prior PDF shapes the posterior PDF. Thus, the information that the prior brings to the inference forces the results towards a particular direction. Jeffreys priors are a class of prior PDF that tends to minimise the information carried by the prior PDF with respect to the likelihood, they are uninformative priors.

Let \mathcal{L} be likelihood, x a set of data and $\theta = \{\theta_i; i \in \llbracket 1, N \rrbracket\}$ the set of parameters that we want to infer. The score function $V(\theta_i, x)$ measure how sensitive the likelihood is under a variation in θ_i :

$$V(\theta_i, x) = \frac{1}{\mathcal{L}(x \mid \theta)} \frac{\partial \mathcal{L}(x \mid \theta)}{\partial \theta_i} = \frac{\partial \log \mathcal{L}(x \mid \theta)}{\partial \theta_i}, \quad (\text{B.3})$$

and the Fisher information $\mathcal{I}(\theta_i)$, which quantifies how much information on θ_i is contained in the likelihood, is the variance¹ of the score function:

$$\mathcal{I}(\theta_i) = \mathbb{E} [V(\theta_i, x)^2 \mid \theta] = -\mathbb{E} \left[\frac{\partial^2 \log \mathcal{L}(x \mid \theta)}{\partial \theta_i^2} \mid \theta \right]. \quad (\text{B.4})$$

The mutual information between two entries of θ is the covariance of the score function

$$\mathcal{I}(\theta_i, \theta_j) = \mathbb{E} [V(\theta_i, x)V(\theta_j, x) \mid \theta] = -\mathbb{E} \left[\frac{\partial^2 \log \mathcal{L}(x \mid \theta)}{\partial \theta_j \partial \theta_i} \mid \theta \right]. \quad (\text{B.5})$$

¹because the mean is zero.

The matrix $\mathcal{I}(\theta)_{ij} = \mathcal{I}(\theta_i, \theta_j)$ is called the Fisher information matrix.

If π_J is an uninformative prior then it should be invariant under a change of variable. The answer is the Jeffreys prior:

$$\pi_J(\theta) = |\mathcal{I}(\theta)|^{1/2} . \tag{B.6}$$

Appendix C

PDF evaluation techniques

In Bayesian analysis, once the data is available and the analytical form of the posterior PDF is derived, quantitative informations are extracted from evaluations of the posterior PDF. This appendix proposes strategies to evaluate a PDF. I first present two simple intuitive approaches to give but these are impracticable, in the general case complex sampling techniques has to be used to evaluate a PDF. This kind of algorithms constructs chains containing a series of samples from the PDF. I describe two algorithms that perform indirect sampling: Metropolis-Hastings sampling and Gibbs sampling.

C.1 Simple approaches

Grid evaluation The brute force way to evaluate a PDF is to discretise the parameter space and to compute the value of the PDF at each node. This solution is rarely achievable since the time of a full PDF evaluation grows exponentially with the number of parameters. For example, if one wants to evaluate the posterior over a 10×10 pixel image with two bins only per pixel, and if one evaluation of the posterior takes $1\mu s$, it would require $2^{100} \times 1\mu s$, i.e. three million times the age of the universe.

Direct sampling The usual way of evaluating a PDF is by means of random generators. In some particular cases, the PDF is a common distribution from which it is easy to sample from, e.g. a Gaussian distribution whose mean and covariance do not depend on the parameters. The method to evaluate the PDF is then to draw samples from the distribution and construct a histogram with the samples.

C.2 Metropolis-Hastings sampling

Metropolis-Hastings sampling (Metropolis et al., 1953; Hastings, 1970) estimates the distribution by building a chain of samples $(\Theta^{(i)}; i \in \llbracket 0, I \rrbracket)$. It proceeds as follows:

Metropolis-Hastings sampling

1. begin with some initial value $\Theta^{(0)}$ of Θ ,
2. for $i \in \llbracket 1, I \rrbracket$,
 - 2.1 propose a random candidate Θ^* depending on the current position of the chain $\Theta^{(i-1)}$,
 - 2.2 compute the acceptance ratio $a = P(\Theta^*)/P(\Theta^{(i-1)})$,
 - 2.3 accept/reject Θ^* , i.e. $\Theta^{(i)} = \Theta^*$ with probability $\min(1, a)$, $\Theta^{(i)} = \Theta^{(i-1)}$ with probability $1 - \min(1, a)$.

It can be shown that the histogram of the resulting chain $(\Theta^{(i)}; i \in \llbracket 0, I \rrbracket)$ converges towards the target distribution $P(\Theta)$ as $I \rightarrow \infty$.

C.3 Gibbs sampling

Gibbs sampling (Gelman and Rubin, 1992) is a particular case of Metropolis-Hastings sampling where the proposal distribution is the marginal distribution. Then it can be shown (Gelman and Rubin, 1992) that, by construction, the acceptance ratio is always one and the samples drawn from the marginal are always accepted. The idea of Gibbs sampling is to split one difficult sampling problem into several easy sampling problems. Say that we have a set of parameters $\Theta = \{\theta_n; n \in \llbracket 1, N \rrbracket\}$ and that we want to sample Θ from a given distribution $P(\Theta)$, which has not a common form. The principle of Gibbs sampling is to build a chain of samples $(\Theta^{(i)}; i \in \llbracket 0, I \rrbracket)$ following the scheme:

Gibbs sampling

1. partition $\Theta = \bigcup_{k=1}^K \Theta_k$ such that for all k , the conditional distribution $P(\Theta_k | \Theta \setminus \Theta_k)$ is easy to sample from,
2. begin with some initial value $\Theta^{(0)}$ of Θ ,
3. for $i \in \llbracket 1, I \rrbracket$, for $k \in \llbracket 1, K \rrbracket$,
sample the set of parameters $\Theta_k^{(i)}$ from $P(\Theta_k | \Theta \setminus \Theta_k)$ where $\Theta \setminus \Theta_k$ is fixed to its latest value, i.e. $\Theta_1^{(i)} \cup \dots \cup \Theta_{k-1}^{(i)} \cup \Theta_{k+1}^{(i-1)} \cup \dots \cup \Theta_K^{(i-1)}$

As for Metropolis-Hastings sampling, it can be shown that the histogram of the resulting chain $(\Theta^{(i)}; i \in \llbracket 0, I \rrbracket)$ converges towards the target distribution $P(\Theta)$ as $I \rightarrow \infty$. Gibbs sampling is achievable only if the set of parameters Θ and its distribution $P(\Theta)$ verify the condition in step 1. If they do not, one should rely on Metropolis-Hastings algorithm.

C.4 Collapsed sampling

Collapsed sampling is a sampling technique that reduces the dimension of the parameter space in order to speed up the sampler. The dimension reduction is achieved via marginalisation of the posterior. For example, say that we want to sample a set of three parameters $\{\Theta_1, \Theta_2, \Gamma\}$ from the distribution $P(\Theta_1, \Theta_2, \Gamma)$ and that the marginal $P(\Theta_1, \Theta_2)$ is known. In addition, each conditional distribution of both the joint distribution and its marginal can be sampled from directly. A collapsed Gibbs sampler build a chain of samples $(\{\Theta_1^{(i)}, \Theta_2^{(i)}, \Gamma^{(i)}\}; i \in \llbracket 0, I \rrbracket)$ according to the following scheme

Collapsed Gibbs sampling

1. begin with some initial value $\{\Theta_1^{(0)}, \Theta_2^{(0)}, \Gamma^{(0)}\}$,
2. for $i \in \llbracket 1, I \rrbracket$,
 - 2.1 sample $\Theta_1^{(i)}$ from $P(\Theta_1 | \Theta_2^{(i-1)})$,
 - 2.2 sample $\Theta_2^{(i)}$ from $P(\Theta_2 | \Theta_1^{(i)})$,
 - 2.3 sample $\Gamma^{(i)}$ from $P(\Gamma | \Theta_1^{(i)}, \Theta_2^{(i)})$.

The implementation of BICA in chapter 4 makes use of such a sampler except that the sampling from the marginal distribution (steps 2.1 and 2.2 in the example) is replaced by a "propose and accept/reject" step as in Metropolis-Hastings.

Appendix D

Isotropic Gaussian random field on the sphere

D.1 Spherical harmonics

The spherical harmonics are a set of functions that form a basis that are convenient for representation of a function on the sphere. They are defined as follows

$$\forall \ell \geq 0, m \in \llbracket -\ell, \ell \rrbracket, \quad Y_{\ell m}(\theta, \varphi) = \sqrt{\frac{(2\ell + 1)(\ell - m)!}{4\pi(\ell + m)!}} P_{\ell}^m(\cos \theta) e^{im\varphi}, \quad (\text{D.1})$$

where P_{ℓ}^m is the associated Legendre polynomials of degree ℓ and order m and they verify the following orthonormal relation

$$\int_0^{\pi} \int_0^{2\pi} \sin \theta d\theta d\varphi Y_{\ell m}(\theta, \varphi) Y_{\ell' m'}^*(\theta, \varphi) = \delta_{\ell \ell'} \delta_{m m'}. \quad (\text{D.2})$$

A square-integrable function on the sphere $f : (\theta, \varphi) \in \llbracket 0, \pi \rrbracket \times \llbracket 0, 2\pi \rrbracket \rightarrow f(\theta, \varphi) \in \mathbb{R}$ can be expressed on the basis of spherical harmonics as a linear combination of the $Y_{\ell m}$ as follows

$$\forall (\theta, \varphi) \in \llbracket 0, \pi \rrbracket \times \llbracket 0, 2\pi \rrbracket, \quad f(\theta, \varphi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \varphi), \quad (\text{D.3})$$

where the $a_{\ell m}$ are the modes of the function f . The inverse relation can be found thanks to orthonormal relation and

$$\forall (\ell, m) \in \mathbb{N} \times \llbracket -\ell, \ell \rrbracket, \quad a_{\ell m} = \int_0^{\pi} \int_0^{2\pi} \sin \theta d\theta d\varphi f(\theta, \varphi) Y_{\ell m}^*(\theta, \varphi). \quad (\text{D.4})$$

The mode a_{00} is called the **monopole** and the set of modes such that $\ell = 1$ constitute

the **dipole**.

D.2 Power spectrum

The function is equivalently represented by either $\{f(\theta, \varphi); (\theta, \varphi) \in [0, \pi] \times [0, 2\pi]\}$ or $\{a_{\ell m}; (\ell, m) \in \mathbb{N} \times [-\ell, \ell]\}$. Thus, if $f(\theta, \varphi)$ is stochastic so are the $a_{\ell m}$. For convenience, I now change the notation for the coordinate on the sphere: $(\theta, \varphi) \rightarrow \vec{n}$ where \vec{n} is the unit vector that points toward the point of coordinate (θ, φ) .

Let f be an isotropic Gaussian random field on the sphere, i.e. the mean is the same over the sphere and the covariance between two points \vec{n}_1 and \vec{n}_2 depends on the angle between them z_{12} only:

$$\langle f(\vec{n}) \rangle = \bar{f}, \quad (\text{D.5})$$

$$\langle (f(\vec{n}_1) - \bar{f})(f(\vec{n}_2) - \bar{f}) \rangle = \xi(z_{12}). \quad (\text{D.6})$$

The function ξ is called the **two-point correlation function**. Since the monopole is the average of f over the sphere i.e.

$$a_{00} = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} \sin \theta d\theta d\varphi f(\theta, \varphi) \hat{=} \bar{f}, \quad (\text{D.7})$$

we can redefine f and the $a_{\ell m}$ in order to work with zero mean quantities

$$a_{\ell m} \rightarrow a_{\ell m} - a_{00} \quad (\text{D.8})$$

$$f(\theta, \varphi) \rightarrow f(\theta, \varphi) - \bar{f}. \quad (\text{D.9})$$

The isotropy of f implies that the covariance of the modes is diagonal, i.e.

$$\langle a_{\ell m} a_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{m m'}. \quad (\text{D.10})$$

The collection of the C_ℓ is called the **power spectrum** of the Gaussian random field and we have the following relation between the two-point correlation function and the power spectrum:

$$\xi(z_{12}) = \sum_{\ell} C_\ell \frac{2\ell + 1}{4\pi} P_\ell(z_{12}) \quad (\text{D.11})$$

where P_ℓ is the Legendre polynomial of degree ℓ , i.e. the power spectrum are the multi-poles of the two-point correlation function.

Appendix E

HEALPix

Hierarchical Equal Area isoLatitude Pixelization¹ (HEALPix) (Górski et al., 2005) is a partition of the 2-D sphere. The pixelisation scheme has three important properties:

- at a given resolution, the pixels have the same areas
- the sphere is hierarchically tessellated, meaning that at a given resolution, each pixel divides up into four pixels to give the resolution above. The lowest resolution is a 12 piece tessellation of the sphere
- the centres of the pixels occur on a minimal number of latitude rings.

The last property is particularly important to allow fast analyses on the sphere, like transformation from real to harmonic space or Gaussian random realisation of a map from a power spectrum.

Since a map in HEALPix is hierarchically divided from a 12 piece partition, the number of pixels of a given resolution is $12N_{\text{side}}^2$ with $N_{\text{side}} \in \mathbb{N}^*$. N_{side} is called the HEALPix resolution parameter of the map.

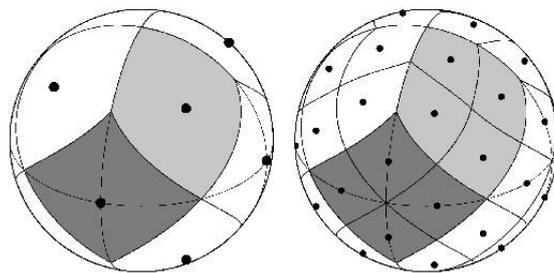


Figure E.1: HEALPix pixelisation scheme. The figure on the left shows the lowest resolution of the pixelisation possible with its 12 pixels. The figure on the right shows the hierarchical pixelisation.

¹code available at <http://healpix.jpl.nasa.gov/index.shtml>

Appendix F

Link between ILC and BICA

F.1 Data and notations

For each mode (ℓ, m) , the data model is

$$d_{\ell m} = s_{\ell m}^{cmb} e + A s_{\ell m}^f + n_{\ell m} \quad (\text{F.1})$$

$$R_{\ell} = C_{\ell}^{cmb} e e^T + A \Sigma_{\ell} A^T + N_{\ell} \quad (\text{F.2})$$

where $e = (1 \dots 1)^T$, $s_{\ell m}^{cmb}$ represents the CMB map, $s_{\ell m}^f$ is a vector containing all the foreground maps, A has no column dedicated to the CMB, R_{ℓ} is the covariance of $d_{\ell m}$. From now on, for simplicity I define

$$R \equiv R_{\ell} \quad (\text{F.3})$$

$$C \equiv C_{\ell}^{cmb} \text{ (is a scalar)} \quad (\text{F.4})$$

$$F \equiv A \Sigma_{\ell} A^T + N_{\ell}. \quad (\text{F.5})$$

These quantities are related as follows

$$R = C e e^T + F. \quad (\text{F.6})$$

F.2 ILC and "BICA derived" formulas

The ILC weights are (Tegmark et al., 2003)

$$w = \frac{e^T R^{-1}}{e^T R^{-1} e}. \quad (\text{F.7})$$

The weights derived following the bayesian approach are (equation 4.14)

$$w' = \frac{e^T F^{-1}}{C^{-1} + e^T F^{-1} e}. \quad (\text{F.8})$$

Next sections shows the relation between those two formulas.

F.3 Expanding the ILC formula

Let's expand the ILC formula of equation F.7 following the decomposition of equation F.6. Let's start by the inverse of R , using Woodbury matrix identity¹. Cee^T is not invertible so the only way to use the Woodbury formula is as follows:

$$R^{-1} = F^{-1} - \frac{1}{C^{-1} + e^T F^{-1} e} F^{-1} e e^T F^{-1}. \quad (\text{F.9})$$

Then the numerator of ILC formula (equation F.7):

$$e^T R^{-1} = e^T F^{-1} - \frac{e^T F^{-1} e}{C^{-1} + e^T F^{-1} e} e^T F^{-1} \quad (\text{F.10})$$

$$= \left(1 - \frac{e^T F^{-1} e}{C^{-1} + e^T F^{-1} e} \right) e^T F^{-1} \quad (\text{F.11})$$

Then the denominator of ILC formula (equation F.7): multiplying equation F.10 on the right by e and factorising by $e^T F^{-1} e$ on the right we get

$$e^T R^{-1} e = \left(1 - \frac{e^T F^{-1} e}{C^{-1} + e^T F^{-1} e} \right) e^T F^{-1} e. \quad (\text{F.12})$$

Dividing equation F.11 by equation F.12 we get the following relation

$$w = \frac{e^T R^{-1}}{e^T R^{-1} e} = \frac{e^T (Cee^T + F)^{-1}}{e^T (Cee^T + F)^{-1} e} = \frac{e^T F^{-1}}{e^T F^{-1} e}. \quad (\text{F.13})$$

F.4 Relation between the two formulas

Finally, we have $w = w'$ if $C \rightarrow \infty$, i.e. flat prior for CMB map. The limit is well defined since (with the notation $x \equiv \frac{C^{-1}}{e^T F^{-1} e}$)

$$1 - \frac{1}{1+x} \underset{x \rightarrow 0}{=} x + \mathcal{O}(x^2) \quad \text{and} \quad \frac{x + \mathcal{O}(x^2)}{x + \mathcal{O}(x^2)} = \frac{1 + \mathcal{O}(x)}{1 + \mathcal{O}(x)} \underset{x \rightarrow 0}{\rightarrow} 1. \quad (\text{F.14})$$

¹The Woodbury matrix identity is $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$, where A is n -by- n and invertible, U is n -by- k , C is k -by- k and invertible and V is k -by- n (Woodbury, 1950).

Bibliography

- Ade, P. et al. (2014). BICEP2 II: Experiment and Three-Year Data Set. *Astrophys.J.*, 792:62.
- Alpher, R. A., Bethe, H., and Gamow, G. (1948). The Origin of Chemical Elements. *Physical Review*, 73:803–804.
- André, P. et al. (2014). PRISM (Polarized Radiation Imaging and Spectroscopy Mission): An Extended White Paper. *JCAP*, 1402:006.
- Arouna, B. (2003). Adaptive monte carlo method, a variance reduction technique. *Monte Carlo Methods and Applications*, 10:1–24.
- Baccigalupi, C., Bedini, L., Burigana, C., De Zotti, G., Farusi, A., et al. (2000). Neural networks and separation of cosmic microwave background and astrophysical signals in sky maps. *Mon. Not. Roy. Astron. Soc.*, 318:769.
- Baumann, D. et al. (2009). CMBPol Mission Concept Study: Probing Inflation with CMB Polarization. *AIP Conf. Proc.*, 1141:10–120.
- Bennett, C., Smoot, G., Hinshaw, G., Wright, E., Kogut, A., and de Amici, G. (1992). Preliminary separation of galactic and cosmic microwave emission for the COBE differential microwave radiometer. *Astrophys. J.*, 396:L7–L12.
- Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Limon, M., Mather, J., Meyer, S. S., Page, L., Spergel, D. N., Tucker, G., Wilkinson, D. T., Wollack, E., and Wright, E. L. (1997). The Microwave Anisotropy Probe (MAP) Mission. In *American Astronomical Society Meeting Abstracts*, volume 29 of *Bulletin of the American Astronomical Society*, page 1353.
- Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., Smith, K. M., Hill, R. S., Gold, B., Halpern, M., Komatsu, E., Nolte, M. R., Page, L., Spergel, D. N., Wollack, E., Dunkley, J., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., and Wright, E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *ApJS*, 208:20.
- Blas, D., Lesgourgues, J., and Tram, T. (2011). The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *J. Cosmology Astropart. Phys.*, 7:34.

- Bobin, J., Starck, J.-L., Fadili, J., and Moudden, Y. (2007). Sparsity and morphological diversity in blind source separation. *Image Processing, IEEE Transactions on*, 16(11):2662–2674.
- Bobin, J., Sureau, F., Starck, J.-L., Rassat, A., and Paykari, P. (2014). Joint Planck and WMAP CMB map reconstruction. *A&A*, 563:A105.
- Bouchet, F. R. et al. (2011). CORE (Cosmic Origins Explorer) A White Paper. arXiv:1102.2181.
- Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proceedings of the IEEE*, 86:2009–2025.
- Cardoso, J.-F., Martin, M., Delabrouille, J., Betoule, M., and Patanchon, G. (2008). Component separation with flexible models. Application to the separation of astrophysical emissions. arXiv:0803.1814.
- Cardoso, J.-F., Snoussi, H., Delabrouille, J., and Patanchon, G. (2002). Blind separation of noisy Gaussian stationary sources. Application to cosmic microwave background imaging. arXiv:astro-ph/0209466.
- Carlstrom, J. E., Holder, G. P., and Reese, E. D. (2002). Cosmology with the Sunyaev-Zel’dovich Effect. *ARA&A*, 40:643–680.
- Chluba, J. (2014). Science with CMB spectral distortions. arXiv:1405.6938.
- Corey, B. E. and Wilkinson, D. T. (1976). A Measurement of the Cosmic Microwave Background Anisotropy at 19 GHz. In *Bulletin of the American Astronomical Society*, volume 8 of *Bulletin of the American Astronomical Society*, page 351.
- Crill, B. P., Ade, P. A. R., Artusa, D. R., Bhatia, R. S., Bock, J. J., Boscaleri, A., Cardoni, P., Church, S. E., Coble, K., de Bernardis, P., de Troia, G., Farese, P., Ganga, K. M., Giacometti, M., Haynes, C. V., Hivon, E., Hristov, V. V., Iacoangeli, A., Jones, W. C., Lange, A. E., Martinis, L., Masi, S., Mason, P. V., Mauskopf, P. D., Miglio, L., Montroy, T., Netterfield, C. B., Paine, C. G., Pascale, E., Piacentini, F., Polenta, G., Pongetti, F., Romeo, G., Ruhl, J. E., Scaramuzzi, F., Sforna, D., and Turner, A. D. (2003). BOOMERANG: A Balloon-borne Millimeter-Wave Telescope and Total Power Receiver for Mapping Anisotropy in the Cosmic Microwave Background. *ApJS*, 148:527–541.
- Dame, T. M., Hartmann, D., and Thaddeus, P. (2001). The Milky Way in Molecular Clouds: A New Complete CO Survey. *ApJ*, 547:792–813.
- Davies, R. D. and Wilkinson, A. (1998). Synchrotron Emission from the Galaxy. *ArXiv Astrophysics e-prints*. arXiv:astro-ph/9804208.
- de Oliveira-Costa, A., Tegmark, M., Page, L., and Boughn, S. (1998). Galactic emission at 19 ghz. arXiv:astro-ph/9807329.

- Delabrouille, J., Betoule, M., Melin, J.-B., Miville-Deschenes, M.-A., Gonzalez-Nuevo, J., et al. (2012). The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths. *arXiv:1207.3675*.
- Delabrouille, J., Cardoso, J.-F., Le Jeune, M., Betoule, M., Fay, G., and Guilloux, F. (2009). A full sky, low foreground, high resolution CMB map from WMAP. *A&A*, 493:835–857.
- Delabrouille, J., Cardoso, J.-F., and Patanchon, G. (2003). Multi-detector multi-component spectral matching and applications for CMB data analysis. *Mon. Not. Roy. Astron. Soc.*, 346:1089.
- Delabrouille, J., Patanchon, G., and Audit, E. (2002). Separation of instrumental and astrophysical foregrounds for mapping cosmic microwave background anisotropies. *MNRAS*, 330:807–816.
- Dicke, R. H., Beringer, R., Kyhl, R. L., and Vane, A. B. (1946). Atmospheric absorption measurements with a microwave radiometer. *Phys. Rev.*, 70:340–348.
- Dicke, R. H., Peebles, P. J. E., Roll, P. G., and Wilkinson, D. T. (1965). Cosmic Black-Body Radiation. *ApJ*, 142:414–419.
- Dickinson, C., Davies, R. D., and Davis, R. J. (2003). Towards a free-free template for CMB foregrounds. *MNRAS*, 341:369–384.
- Dodelson, S. (2003). *Modern Cosmology*. Academic Press. Academic Press.
- Elsner, F. and Wandelt, B. D. (2013). Efficient Wiener filtering without preconditioning. *Astron. Astrophys.*, 549:A111.
- Eriksen, H., Jewell, J., Dickinson, C., Banday, A., Gorski, K., et al. (2008). Joint Bayesian component separation and CMB power spectrum estimation. *Astrophys. J.*, 676:10–32.
- Eriksen, H. K., Dickinson, C., Lawrence, C., Baccigalupi, C., Banday, A., et al. (2006). CMB component separation by parameter estimation. *Astrophys. J.*, 641:665–682.
- Faÿ, G. and Guilloux, F. (2011). Spectral estimation on the sphere with needlets: high frequency asymptotics. *Statistical Inference for Stochastic Processes*, 14(1):1–25.
- Fendt, W. A. and Wandelt, B. D. (2007). Pico: Parameters for the Impatient Cosmologist. *ApJ*, 654:2–11.
- Fernandez-Cobos, R., Vielva, P., Barreiro, R., and Martinez-Gonzalez, E. (2012). Multi-resolution internal template cleaning: An application to the Wilkinson Microwave Anisotropy Probe 7-yr polarization data. *MNRAS*, 3:2162–2169.
- Finkbeiner, D. P. (2003). A Full-Sky $H\alpha$ Template for Microwave Foreground Prediction. *ApJS*, 146:407–415.
- Finkbeiner, D. P., Davis, M., and Schlegel, D. J. (1999). Extrapolation of Galactic Dust Emission at 100 Microns to Cosmic Microwave Background Radiation Frequencies Using FIRAS. *ApJ*, 524:867–886.

- Francis, C. L. and Peacock, J. A. (2010). An estimate of the local integrated Sachs-Wolfe signal and its impact on cosmic microwave background anomalies. *MNRAS*, 406:14–21.
- Gamow, G. (1946). Expanding universe and the origin of elements. *Phys. Rev.*, 70:572–573.
- Gelman, A. and Meng, X.-L. (1996). *Model checking and model improvement*. Springer.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.*, 7(4):457–472.
- Giardino, G., Banday, A. J., Górski, K. M., Bennett, K., Jonas, J. L., and Tauber, J. (2002). Towards a model of full-sky Galactic synchrotron intensity and linear polarization: A re-analysis of the Parkes data. *A&A*, 387:82–97.
- Gorbunov, D. and Rubakov, V. (2011). *Introduction to the Theory of the Early Universe: Hot Big Bang Theory*. World Scientific.
- Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622:759–771.
- Gratton, S. (2008). Prescription for Cosmic Information Extraction from Multiple Sky Maps. arXiv:0805.0093.
- Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356.
- Hanson, D., Hoover, S., Crites, A., Ade, P. A. R., Aird, K. A., Austermann, J. E., Beall, J. A., Bender, A. N., Benson, B. A., Bleem, L. E., Bock, J. J., Carlstrom, J. E., Chang, C. L., Chiang, H. C., Cho, H.-M., Conley, A., Crawford, T. M., de Haan, T., Dobbs, M. A., Everett, W., Gallicchio, J., Gao, J., George, E. M., Halverson, N. W., Harrington, N., Henning, J. W., Hilton, G. C., Holder, G. P., Holzappel, W. L., Hrubes, J. D., Huang, N., Hubmayr, J., Irwin, K. D., Keisler, R., Knox, L., Lee, A. T., Leitch, E., Li, D., Liang, C., Luong-Van, D., Marsden, G., McMahan, J. J., Mehl, J., Meyer, S. S., Mocanu, L., Montroy, T. E., Natoli, T., Nibarger, J. P., Novosad, V., Padin, S., Pryke, C., Reichardt, C. L., Ruhl, J. E., Saliwanchik, B. R., Sayre, J. T., Schaffer, K. K., Schulz, B., Smecher, G., Stark, A. A., Story, K. T., Tucker, C., Vanderlinde, K., Vieira, J. D., Viero, M. P., Wang, G., Yefremenko, V., Zahn, O., and Zemcov, M. (2013). Detection of B-Mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope. *Physical Review Letters*, 111(14):141301.
- Harrison, E. R. (1970). Fluctuations at the threshold of classical cosmology. *Phys. Rev. D*, 1:2726–2730.
- Haslam, C. G. T., Salter, C. J., Stoffel, H., and Wilson, W. E. (1982). A 408 MHz all-sky continuum survey. II - The atlas of contour maps. *A&AS*, 47:1.
- Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.

- He, K. and Sun, J. (2012). Statistical image completion. *IEEE*. http://research.microsoft.com/pubs/192345/stat_completion.pdf.
- Hivon, E., Gorski, K., Netterfield, C., Crill, B., Prunet, S., et al. (2002). Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets. *Astrophys.J.*, 567:2.
- Hu, W. (2008). Lecture Notes on CMB Theory: From Nucleosynthesis to Recombination. arXiv:0802.3688.
- Hu, W. and Dodelson, S. (2002). Cosmic microwave background anisotropies. *Ann.Rev.Astron.Astrophys.*, 40:171–216.
- Jaffe, T. R., Banday, A. J., Eriksen, H. K., Górski, K. M., and Hansen, F. K. (2005). Evidence of vorticity and shear at large angular scales in the wmap data: A violation of cosmological isotropy? *The Astrophysical Journal Letters*, 629(1):L1.
- Jarosik, N., Bennett, C. L., Dunkley, J., Gold, B., Greason, M. R., Halpern, M., Hill, R. S., Hinshaw, G., Kogut, A., Komatsu, E., Larson, D., Limon, M., Meyer, S. S., Nolte, M. R., Odegard, N., Page, L., Smith, K. M., Spergel, D. N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. *ApJS*, 192:14.
- Jungman, G., Kamionkowski, M., Kosowsky, A., and Spergel, D. N. (1996). Cosmological parameter determination with microwave background maps. *Phys. Rev.*, D54:1332–1344.
- Junk, V. and Komatsu, E. (2012). Cosmic microwave background bispectrum from the lensing-Rees-Sciama correlation reexamined: Effects of nonlinear matter clustering. *Phys. Rev. D*, 85(12):123524.
- Kogut, A., Banday, A. J., Bennett, C. L., Gorski, K. M., Hinshaw, G., Smoot, G. F., and Wright, E. I. (1996). Microwave Emission at High Galactic Latitudes in the Four-Year DMR Sky Maps. *ApJ*, 464:L5.
- Kolb, E. W. and Turner, M. S. (1990). *The early universe*. Front. Phys., Vol. 69.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.
- Kurtz, S., Churchwell, E., and Wood, D. O. S. (1994). Ultracompact H II regions. 2: New high-resolution radio images. *ApJS*, 91:659–712.
- Langlois, D. (2003). Isocurvature cosmological perturbations and the CMB. *Comptes Rendus Physique*, 4(8):953 – 959. Dossier: The Cosmic Microwave Background.
- Leistedt, B., McEwen, J. D., Vandergheynst, P., and Wiaux, Y. (2013). S2LET: A code to perform fast wavelet analysis on the sphere. *A&A*, 558:A128.

- Leitch, E. M., Readhead, A. C. S., Pearson, T. J., and Myers, S. T. (1997). An Anomalous Component of Galactic Emission. *ApJ*, 486:L23–L26.
- Lewis, A. and Challinor, A. (2006). Weak gravitational lensing of the CMB. *Phys. Rep.*, 429:1–65.
- Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476.
- Linde, A. D. (1982). A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems. *Phys.Lett.*, B108:389–393.
- Lyth, D. and Liddle, A. (2009). *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*. Cambridge University Press.
- Madau, P., Meiksin, A., and Rees, M. J. (1997). 21-CM tomography of the intergalactic medium at high redshift. *Astrophys.J.*, 475:429.
- Maino, D., Farusi, A., Baccigalupi, C., Perrotta, F., Banday, A., et al. (2001). All-sky astrophysical component separation with fast independent component analysis (fastica). arXiv:astro-ph/0108362.
- Mather, J. C., Cheng, E. S., Cottingham, D. A., Eplee, Jr., R. E., Fixsen, D. J., Hewagama, T., Isaacman, R. B., Jensen, K. A., Meyer, S. S., Noerdlinger, P. D., Read, S. M., Rosen, L. P., Shafer, R. A., Wright, E. L., Bennett, C. L., Boggess, N. W., Hauser, M. G., Kelsall, T., Moseley, Jr., S. H., Silverberg, R. F., Smoot, G. F., Weiss, R., and Wilkinson, D. T. (1994). Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *ApJ*, 420:439–444.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087–1092.
- Miville-Deschênes, M.-A., Ysard, N., Lavabre, A., Ponthieu, N., Macías-Pérez, J. F., Aumont, J., and Bernard, J. P. (2008). Separation of anomalous and synchrotron emissions using WMAP polarization data. *A&A*, 490:1093–1102.
- Moudden, Y., Cardoso, J.-F., Starck, J.-L., and Delabrouille, J. (2005). Blind component separation in wavelet space. Application to CMB analysis. *EURASIP J. Appl. Signal Process.*, 15:2437–2454.
- Mukhanov, V. (2013). Quantum Cosmological Perturbations: Predictions and Observations. arXiv:1303.3925.
- Naylor, B. J., Bradford, C. M., Aguirre, J. E., Bock, J. J., Earle, L., Glenn, J., Inami, H., Kamenetzky, J., Maloney, P. R., Matsuhara, H., Nguyen, H. T., and Zmuidzinas, J. (2010). A Census of the High-density Molecular Gas in M82. *ApJ*, 722:668–681.
- Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421.

- Planck Collaboration (2013). PLA explanatory supplement. http://wiki.cosmos.esa.int/planckpla/index.php/Frequency_Maps#Types_of_maps.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Alves, M. I. R., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Aussel, H., and et al. (2014a). Planck 2013 results. I. Overview of products and scientific results. *A&A*, 571:A1.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Alves, M. I. R., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., and et al. (2014b). Planck 2013 results. XIII. Galactic CO emission. *A&A*, 571:A13.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014c). Planck 2013 results. XII. Diffuse component separation. *A&A*, 571:A12.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014d). Planck 2013 results. XV. CMB power spectra and likelihood. *A&A*, 571:A15.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014e). Planck 2013 results. XVI. Cosmological parameters. *A&A*, 571:A16.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014f). Planck 2013 results. XXIII. Isotropy and statistics of the CMB. *A&A*, 571:A23.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014g). Planck 2013 results. XXIV. Constraints on primordial non-Gaussianity. *A&A*, 571:A24.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Baker, M., Balbi, A., Banday, A. J., and et al. (2011a). Planck early results. I. The Planck mission. *A&A*, 536:A1.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Balbi, A., Banday, A. J., Barreiro, R. B., and et al. (2011b). Planck early results. XVIII. The power spectrum of cosmic infrared background anisotropies. *A&A*, 536:A18.
- Planck Collaboration, Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., and et al. (2013). Planck 2013 results. III. LFI systematic uncertainties. arXiv:1303.5064.
- Rees, M. J. and Sciama, D. W. (1968). Large-scale Density Inhomogeneities in the Universe. *Nature*, 217:511–516.

- Remazeilles, M., Delabrouille, J., and Cardoso, J.-F. (2011). Foreground component separation with generalised ILC. *ArXiv e-prints*. arXiv:1103.1166.
- Ruhl, J. E. et al. (2004). The South Pole Telescope. *Proc.SPIE Int.Soc.Opt.Eng.*, 5498:11.
- Rybicki, G. B. and Lightman, A. P. (1979). *Radiative processes in astrophysics*.
- Ryden, B. (2003). *Introduction to Cosmology*. Addison-Wesley.
- Sachs, R. K. and Wolfe, A. M. (1967). Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. *ApJ*, 147:73.
- Schaffer, K., Crawford, T., Aird, K., Benson, B., Bleem, L., et al. (2011). The First Public Release of South Pole Telescope Data: Maps of a 95-square-degree Field from 2008 Observations. *Astrophys. J.*, 743:90.
- Sehgal, N., Bode, P., Das, S., Hernandez-Monteagudo, C., Huffenberger, K., et al. (2010). Simulations of the Microwave Sky. *Astrophys. J.*, 709:920–936.
- Seljak, U. and Zaldarriaga, M. (1996). A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469:437.
- Sievers, J. L. et al. (2013). The Atacama Cosmology Telescope: Cosmological parameters from three seasons of data. *JCAP*, 1310:060.
- Silk, J. (1968). Cosmic Black-Body Radiation and Galaxy Formation. *ApJ*, 151:459.
- Smoot, G. Observation of the cosmic microwave background anisotropy. *IAU XXI Highlights of Astronomy*, 9:281–286.
- Smoot, G., Bennett, C., Weber, R., Maruschak, J., Ratliff, R., Janssen, M., Chitwood, J., Hilliard, L., Lecha, M., Mills, R., Patschke, R., Richards, C., Backus, C., Mather, J., Hauser, M., Weiss, R., Wilkinson, D., Gulkis, S., Boggess, N., Cheng, E., Kelsall, T., Lubin, P., Meyer, S., Moseley, H., Murdock, T., Shafer, R., Silverberg, R., and Wright, E. (1990). COBE Differential Microwave Radiometers - Instrument design and implementation. *ApJ*, 360:685–695.
- Starck, J.-L., Donoho, D. L., Fadili, M. J., and Rassat, A. (2013). Sparsity and the Bayesian perspective. *A&A*, 552:A133.
- Starck, J.-L., Elad, M., and Donoho, D. L. (2004). Redundant multiscale transforms and their application for morphological component separation. *Adv. Imag. Elect. Phys.*, 132(DAPNIA-2004-88):287–348.
- Stompor, R., Leach, S., Stivoli, F., and Baccigalupi, C. (2009). Maximum likelihood algorithm for parametric component separation in cosmic microwave background experiments. *MNRAS*, 392:216–232.
- Sunyaev, R. A. and Zeldovich, I. B. (1980). Microwave background radiation as a probe of the contemporary structure and history of the universe. *ARA&A*, 18:537–560.

- Sunyaev, R. A. and Zeldovich, Y. B. (1972). The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173.
- Tegmark, M. (1997). Removing real world foregrounds from cmb maps. arXiv:astro-ph/9712038.
- Tegmark, M., de Oliveira-Costa, A., and Hamilton, A. J. (2003). High resolution foreground cleaned CMB map from WMAP. *Phys. Rev. D*, 68(12):123523.
- Trotta, R. (2007). Applications of Bayesian model selection to cosmological parameters. *MNRAS*, 378:72–82.
- Vansyngel, F., Wandelt, B. D., Cardoso, J.-F., and Benabed, K. (2014). Semi-blind Bayesian inference of CMB map and power spectrum. arXiv:1409.0858.
- Wandelt, B. D., Hivon, E., and Gorski, K. M. (2001). The pseudo- C_ℓ method: cosmic microwave background anisotropy power spectrum statistics for high precision cosmology. *Phys. Rev.*, D64:083003.
- Wandelt, B. D., Larson, D. L., and Lakshminarayanan, A. (2004). Global, exact cosmic microwave background data analysis using gibbs sampling. *Phys. Rev. D*, 70:083511.
- Watson, R. A., Rebolo, R., Rubiño-Martín, J. A., Hildebrandt, S., Gutiérrez, C. M., Fernández-Cerezo, S., Hoyland, R. J., and Battistelli, E. S. (2005). Detection of Anomalous Microwave Emission in the Perseus Molecular Cloud with the COSMO-SOMAS Experiment. *ApJ*, 624:L89–L92.
- White, M. and Hu, W. (1997). The Sachs-Wolfe effect. *A&A*, 321:8–9.
- Wood, D. O. S. and Churchwell, E. (1989). The morphologies and physical properties of ultracompact H II regions. *ApJS*, 69:831–895.
- Woodbury, M. A. (1950). Inverting modified matrices. *Statistical Research Group, Memo Rep. no. 42*. Princeton University, Princeton, N. J.
- Zaldarriaga, M. and Seljak, U. (1997). All-sky analysis of polarization in the microwave background. *Phys. Rev. D*, 55:1830–1840.
- Zeldovich, Y. (1972). A Hypothesis, unifying the structure and the entropy of the universe. *Mon.Not.Roy.Astron.Soc.*, 160:1P–3P.
- Zibulevsky, M. and Pearlmutter, B. A. (2000). Blind source separation by sparse decomposition. In *AeroSense 2000*, pages 165–174. International Society for Optics and Photonics.